Faculty of Science
Utrecht Institute for Pharmaceutical Sciences
Department of Pharmaceutics
Daily Supervisor: Danny Wilbie, MSc
Examiner: Prof. Dr. Enrico Mastrobattista

Cas9-Protein Bioconjugation via Click Chemistry to Enhance HDR

Research Report

Asimina Kogkalidou

Abstract

The CRISPR/Cas9 system is a newly-discovered method in gene therapy, showing great potential. However, after the introduction of a double strand break by the Cas9 protein to the genomic DNA, the error-prone non-homologous end-joining DNA repair pathway predominates over the precise homology-directed repair (HDR) one. For the CRISPR/Cas9 system to be more efficient in accurate genome editing and thus be utilized for the correction of disease-causing mutations, the efficiency of the HDR response should be increased. In this study, an effort has been made to conjugate the Cas9 protein with other protein domains, namely the dominant-negative domain of 53BP1, the human phosphatase inhibitor-2 and the N-terminal region of the human geminin, in order to enhance the frequency of the HDR response either directly or in a time-controlled manner by taking advantage of the characteristics of those three protein domains. An unnatural amino acid was incorporated in the Cas9 sequence, for the purpose of the bioconjugation, which was done utilizing copper-free click chemistry. Our results show that the incorporation of the unnatural amino acid is possible in different positions and can be used for conjugation of the protein with molecules, such as PEG or fluorescent dye. Based on the conjugation efficiency and the relative activity that the SpCas9 variants demonstrated, a preferred position for the amino acid substition was highlighted. The corresponding SpCas9 variant is highly active, even when conjugated with the fluorescent dye. The results of the protein-protein bioconjugation remain inconclusive.

Layman’s Summary

Genetic disorders are a common type of disease, and are attributable to changes in the genome of the organism. These changes in the DNA are called mutations, and in some cases in which they are the cause of diseases, they can be found only in one place in the genome. These diseases are called monogenic and can potentially be cured. A lot of research has been dealing with the correction of these specific mutations, and gene therapy is the field of medicine that regards this correction using different techniques. One tool that seems promising in repairing mutations is the CRISPR/Cas9 system. Basically, the Cas9 is a protein that can act as a "scissor" and induce cuts in the DNA at specific points. After the cutting caused by Cas9, the human organism has two different main ways of repairing the break of the DNA, one being extremely accurate and called homology-directed repair (HDR) and the other one being flawed and rough, and called non-homologous end-joining (NHEJ). The HDR response fixes the cut by precisely copying a template, but unfortunately occurs less frequently than NHEJ and only at predetermined times. As one can easily understand, in order to cure a monogenic disease, Cas9 has to cut the DNA where the problematic mutation is, and then HDR can mend the mutation using the correct, exogenously provided, template. In hopes of using the CRISPR/Cas9 system as a gene therapy tool, it is evident the HDR efficiency should be increased.

This project aims at exactly that, in enhancing the frequency of HDR after the cut caused by Cas9. In order to do that, some molecules will be connected with the Cas9 by conjugation, and the final product would have the ability to function as a "scissor" and also augment HDR. These molecules are three proteins that can be found (or derived from other proteins) that are naturally existent in the organism. They act by directly boosting HDR or helping the Cas9 accumulate where it needs to be, only when HDR is happening.

In order to connect the two molecules, a foreign building block able to click the modifications together (think Lego pieces) has to be introduced inside the Cas9 sequence, in a site-specific manner, in order not to disturb the ability of the Cas9 to function properly and cut the DNA. Then, this block can react with an artificial linker that will conjugate the Cas9 with the other protein part. This linker has been tailor-made to react at the one end only with the introduced in the Cas9 block, and at the other end with the other protein. The final product would retain the activities of both proteins.

This study aimed at producing the conjugate of the Cas9 with the other protein domain. It is not clear to what extent it succeed in that goal, but it has been proved that the block can be introduced at distinct positions of the Cas9 sequence and can be used to conjugate the protein with other molecules, such as a fluorescent coloring agent.

Acknowledgements

Sincere gratitude to my daily supervisor, Danny Wilbie, for his constant guidance and outstanding patience. Many thanks also to my examiners, prof Dr. Enrico Mastrobattista and Dr. Olivier de Jong, for their support and suggestions throughout the project. I would also like to thank all the people belonging to the group of Pharmaceutics, for always making me feel welcome and providing their knowledge and encouragement.

Table of Contents

Chapter 1

Introduction

The technique of gene editing has a profound impact in many fields, ranging from medicine to agriculture (Adli, 2018). Gene editing involves site-specific modifications in the DNA, which lead to precise changes in genomes (Adli, 2018; Bak et al., 2018). Introducing Double Strand Breaks (DSBs) at the targeted sequence of the genome and then modulating their repair is an effective way of gene editing (Rouet et al., 1994; Rudin et al., 1989), and to this end various engineered nucleases have been produced (Bibikova et al., 2003; Miller et al., 2010; F. Zhang et al., 2011). The most recent groundbreaking gene editing platform of nucleases is the Clustered regularly interspaced short palindromic repeats/CRISPR-associated nuclease 9 (CRISPR/Cas9) (Barrangou, 2014): the CRISPR system is an intrinsic defense mechanism of some bacteria and archaea against foreing DNA, while the Cas9 is the protein involved in a specific type of this mechanism. Cas9 is an RNA-guided DNA endonuclease responsible for the precise cleavage of the DNA (Jinek et al., 2012). It functions by binding to the guide RNA, and then recognizing a complementary between the guide RNA and the target DNA 20-nucleotide sequence with the involvement of other necessary elements such as the protospacer-adjacent motif (PAM) in the DNA sequence. The 20-nucleotide DNA sequence allows the binding of the Cas9 to the targeted DNA, and Cas9 introduces a DSB in the targeted site (Gasiunas et al., 2012). Specific characteristics of the CRISPR/Cas9 system, such as cost-effectiveness and easily achieved specificity, have underlined its potential as a therapeutic option for genetic diseases (Bak et al., 2018; Karimian et al., 2019).

The introduction of a DSB by Cas9 to genomic DNA leads to the activation of the DNA repair mechanisms in mammalian cells (Doudna and Charpentier, 2014), the main ones being Non-Homologous End Joining (NHEJ) and Homology-Directed Repair (HDR). NHEJ is mostly preferred and is error-prone, resulting in insertions or deletions at the cleaved site of the gene and thus usually in gene knockouts. On the other hand, HDR occurs less (Ran et al., 2013; Salsman et al., 2018), is mainly active during the S/G2 phase of the cell cycle (Salsman et al., 2018) and is in need of a DNA repair template, but leads to precise gene modifications (Ran et al., 2013; Salsman et al., 2018), and therefore can be utilized to correct disease-causing mutations.

Increasing the efficiency of HDR would allow the implementation of the CRISPR/Cas9 system as a gene therapy tool (Memi et al., 2018), and this defines the aim of the current project: to produce Cas9 conjugates with protein domains that enhance HDR, either directly or in a time-controlled manner. The later can be accomplished by causing specific degradation or specific accumulation of the Cas9 inside the nucleus of the cell, so Cas9 can cleave the DNA when HDR is most active. In order to achieve elevated levels of HDR, three protein domains were chosen to be bioconjugated with Cas9, utilizing click chemistry reaction tools (Kolb et al., 2001).

Regarding the characteristics of the proteins used in this project, the Cas9 is the one derived from *Streptococcus pyogenes*, called SpCas9 (Adli, 2018), which has a molecular weight of about 160 kDa (Jinek et al., 2014), and is the most commonly used by researchers. It has a bilobed structure, the two lobes being the N-terminal recognition one (REC) and the C-terminal nuclease one (NUC). The NUC lobe contains the two different nuclease domains (Jinek et al., 2014), namely HNH and RuvC, which cut the complementary to the guide RNA and the non-complementary to the guide RNA DNA strand respectively (Chen et al., 2014; Gasiunas et al., 2012; Jinek et al., 2012), and also the PAM-interacting domain. The REC lobe has three separate domains, the Bridge Helix, REC1, and REC2. Keeping in mind the complex structure of SpCas9, in combination with the fact that the two terminals of the protein are proximal in the protein's conformation and orient inward (Oakes et al., 2016), and also both interact with the ends of the target DNA when the latter is bound to SpCas9 (Oakes et al., 2014), protein engineering options beside N- and C-terminal fusions may seem appealing. As for the protein domains, the first one is the dominant negative domain of tumor suppressor p53-binding protein 1 (53BP1), called DN1S. Normally, 53BP1 promotes NHEJ and blocks HDR, while DN1S acts by (locally, when fused with SpCas9) blocking the activity of 53BP1, thus increasing the frequency of HDR and decreasing the one of NHEJ. Its molecular weight is 46 kDa (Jayavaradhan et al., 2019). The conjugation of this domain with SpCas9 would ideally lead to a direct increase of the HDR efficiency after the DSB introduced by SpCas9. Another protein domain

consists of the first 110 amino acids of the human geminin protein (Gem110), which has a molecular weight of approximately 15 kDa. The role of this domain concerning our goal is its action as a substrate for the APC/Cdh1 complex, which is active during the late M and G1 phase of the cell cycle and leads to the ubiquitination of Gem110 among other proteins. Fused Gem110 with Cas9 has proven to degrade during the M and G1 phase of the cycle and result in enhanced HDR (Gutschner et al., 2016), and this is the reason that Gem110 was chosen to be conjugated with SpCas9. The final protein domain is the human phosphatase inhibitor-2 (IPP2). In this case, the whole protein is used and not just a domain of it, but for simplicity reasons it is being referred to as a protein domain. It is a 23 kDa protein, and its relevant role is its localization inside the nucleus only during the S phase of the cell cycle (Kakinoki et al., 1997), thus hopefully facilitating, when conjugated with SpCas9, the specific accumulation of SpCas9 in the nucleus when HDR is active (Gutschner et al., 2016; Salsman et al., 2018).

In order for the domains to be conjugated with SpCas9 using click chemistry tools, an unnatural amino acid (UAA) containing a biorthogonal functional group should be incorporated in the protein sequence (Nwe and Brechbiel, 2009). The chosen UAA is 4-Azido-L-phenylalanine (pAzF), for its advantages such as high incorporation rate (Chin et al., 2002). Using polymerase chain reaction (PCR)-mediated site-directed mutagenesis (SDM) (Carey et al., 2013) to genetically incorporate the TAG nonsense codon (Chin et al., 2002) in distinct positions in the Cas9 sequence and then providing the necessary machinery of aminoacyl tRNA synthetase and tRNA pair for the *in vivo* incorporation of the pAzF in *Escherichia coli* (*E. coli*), the pAzF could be incorporated at specific positions in SpCas9 where the UAG codon would be translated (Chin et al., 2002). The azide group is then utilized in a copper-free reaction to avoid the toxic effects of copper (Baskin et al., 2007), namely the Strain-Promoted Azide Alkyne Cycloaddition (Agard et al., 2004) and more precisely the dibenzocyclooctyne-based strain-promoted azide–alkyne [3+2] cycloaddition (Ning et al., 2008). The other molecule taking part in the reaction is either a polyethylene glycol (PEG) of 30 kDa containing a DBCO functional group, or a PEG linker of 5 kDa containing a DBCO functional group and another reactive group, the latter being either an N-hydroxysuccinimide activated ester (NHS ester) to react with the lysines (Anderson et al., 1963) in Gem110, or a maleimide group to react with the cysteine in DN1S and IPP2 (Ravasco et al., 2018). The purpose of the PEG conjugation experiment is to mimic and optimize the reaction conditions of the protein-protein bioconjugation, as the DBCO-mPEG MW 30,000 could resemble the protein domains regarding the functional for click chemistry group and the (mean) molecular weight.

The goal of this study was initially to produce ten SpCas9 variants with pAzF incorporated. The positions where pAzF would be introduced were chosen based on their minimal interference with the protein's structure and function (Oakes et al., 2016), and also on their accessibility to solvents, were amino acids containing a phenol group [tyrosine (Y) and phenylalanine (F)] in order to correlate with the pAzF regarding chemical characteristics, and were the following: F196, F256, F446, F539, F688, Y943, Y1010, Y1036, Y1237 and F1258. The 446 position orients into the DNA binding pocket of SpCas9 (Nishimasu et al., 2014), thus acting as a negative control as modifications to that amino acid would not easily succeed. Then, the optimal variant regarding activity and reactivity would be selected, conjugated with the domains, the conjugates would be purified and finally tested regarding their activity in cells. As fusion proteins of SpCas9 with DN1S (Jayavaradhan et al., 2019) or Gem110 (Gutschner et al., 2016) have already been tried, a comparison between the fusion and the conjugated protein would also be made, regarding activity and stability. Afterwards, the continuation of the project would involve a fusion protein at the C-terminal region of the SpCas9 with each of the domains, and based on the most promising conjugate, pAzF would also be introduced inside the SpCas9 sequence and then bioconjugated with the ideal domain. The end product would be a triple-activity fusion and conjugate protein.

The SpCas9 variants are mentioned throughout this report either by F/YxAzF or F/Yx* SpCas9 or SpCas9-xAzF, where x is the position of the altered amino acid and * indicates the incorporation of pAzF.

Chapter 2

Materials and Methods

Most of the chemicals and reagents used were acquired from Sigma-Aldrich (Zwijndrecht, the Netherlands). The nuclease-free water was provided by Integrated DNA technologies (IDT, Leuven, Belgium). All of the enzymes unless stated otherwise, the protein stain reagents and the gels were acquired from Thermo (Fisher) Scientific (Landsmeer, the Netherlands). Primers for the PCR and for Sanger sequencing were provided by IDT (Leuven, Belgium). The single guide RNA (sgRNA) and the DNA template were acquired from Sigma-Aldrich (Haverhill, the United Kingdom). The lipids for the formation of the lipid nanoparticles, their manufacturers and the corresponding addresses of them can be found at "Impact of Formulation Conditions on Lipid Nanoparticle Characteristics and Functional Delivery of CRISPR RNP for Gene Knock-Out and Correction" (Walther et al., 2022). The 4-Azido-L-phenylalanine (pAzF), the DBCO-mPEG MW 30,000, the DBCO-PEG24-Maleimide and the DBCO-PEG24-NHS ester were ordered from BroadPharm (San Diego, USA).
In Figure 2.1 a representation of the general workflow of the project can be seen. The first step was to genetically incorporate the stop codon using PCR-mediated SDM (refer to section 2.1.), then introduce the mutated plasmid in *E. coli*, isolate the plasmid and verify its mutation (section 2.1.). Afterwards, the mutated plasmid of Cas9 was introduced into another strain of *E. coli*, which is suitable for protein production, and the variants of SpCas9 were produced and purified (section 2.2.). Then, the variants were characterized for their purity, activity and reactivity (section 2.3.), conjugated with various molecules, and each conjugate was also characterized (section 2.4. section 2.5.).

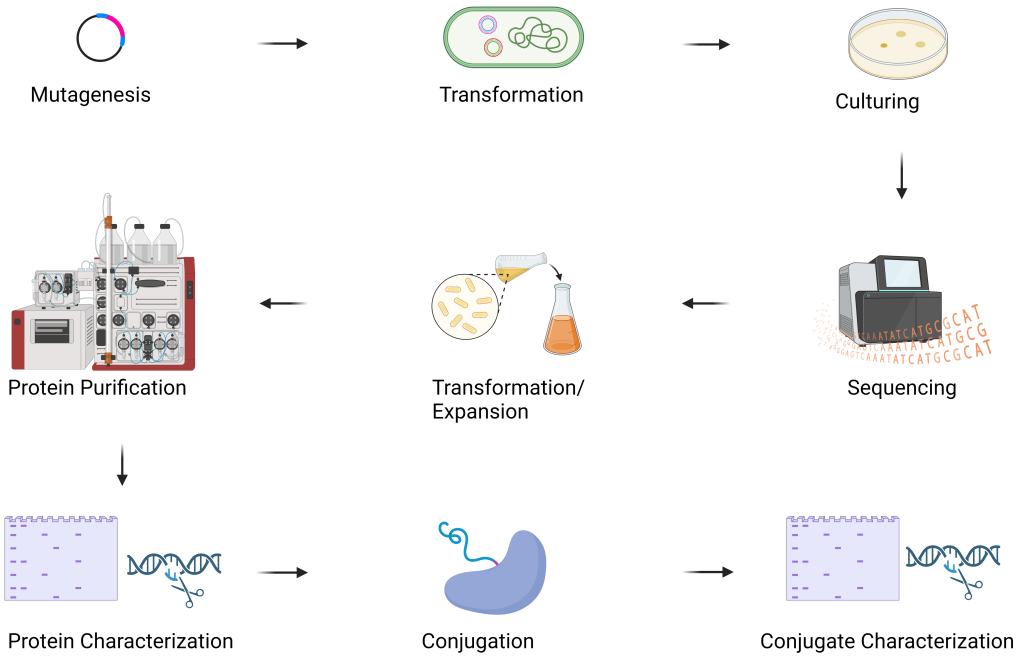

Figure 2.1: Schematic workflow of the project's experimental procedure

2.1. Production of the Mutated SP-Cas9 Sequences

The plasmid used for the production of SpCas9 (SP-Cas9) and also for downstream manipulation had an insert expressing the Streptococcus pyogenes serotype M1 Cas9 containing a nuclear localization signal and a His-tag at the C-terminal; SP-Cas9 was a gift from Niels Geijsen (Addgene

plasmid #62731)).

Site Directed Mutagenesis

In order to site-specifically insert the UAA sequence in the SP-Cas9, SDM was performed using the QuikChange XL Site-Directed Mutagenesis Kit (Catalog #200521) by Agilent Technologies (Amstelveen, the Netherlands), and the enzymes *PfuUltra* High-Fidelity DNA Polymerase (Catalog #600380) and *PfuTurbo* DNA Polymerase (Catalog #600250) by Agilent Technologies (Amstelveen, the Netherlands). The protocol followed was provided in the instructrion manual of the Agilent Kit, with the following adjustment: increasing the extension time to 2 minutes per kilobase of plasmid DNA. The primers used were designed using the QuikChange Primer Design Program, and had the characteristics shown in the Supplementary Materials, Table 4.1. The primers for mutation at position 196, 256, 446, 539, 688, 1010 and 1237 were designed in order to include a diagnostic restriction site that could verify the mutagenesis.

Bacterial Transformation, Plasmid Isolation and Verification of Sequence

After digestion of the parental DNA with DpnI, the PCR product was used to transform XL10-gold Ultracompetent cells ordered from Agilent (Catalog #200315, Amstelveen, the Netherlands), in order to ligase and amplify the mutated plasmid, using the Agilent transformation protocol. Some minor variations in the protocol were necessary: the recovery phase was done using SOC medium and shaking at 350-400 rpm, and after the recovery phase the cells were centrifuged, the supernatant was decanted and the cells were resuspended in the remaining medium before being spread in a Luria-Bertani (LB)-ampicillin plate.

After picking a number of colonies out of each plate and, following the addgene protocol, inoculating overnight the 5 ml culture containing ampicillin in final concentration 100 μg/mL and either chloramphenicol 25 μg/mL or tetracycline 10 μg/mL, a glycerol stock following addgene instructions was made out of each colony, while the rest of the culture was used to perform MiniPrep. The product used was GeneJET Plasmid Miniprep Kit (Catalog #K0502) by Thermo Scientific (Landsmeer, the Netherlands), and the manufacturer's protocol was followed with the exception of the elution step, which was done using nuclease-free water pre-warmed at 70°C. The concentration and purity of the extracted plasmid was determined using the NanoDrop One UV-Vis Spectrophotometer by Thermo Scientific (Landsmeer, the Netherlands).

Afterwards, a digestion of the plasmid DNA with, if applicable, a restriction enzyme that could specifically verify the mutation in order to confirm the successful mutagenesis of the plasmid, or else with a restriction enzyme that could linearize the DNA in order to confirm the length of the plasmid was performed, and the corresponding for each enzyme guidelines provived by the manufacturer were followed. The samples were then separated using electrophoresis (1% agarose gel with 0.005% final concentration of Midori Green by Nippon Genetics, Düren Germany, GeneRuler 1 kb DNA Ladder by ThermoFisher Scientific, Landsmeer, the Netherlands), while the running settings were 100 Volt and 50 minutes.

Following this preliminary screening method, the possibly correct samples were sent for Sanger sequencing, concerning the part where the mutation should be, at Macrogen Europe (Amsterdam, the Netherlands), with the final concentration of primers being 10 mM and the amount of DNA being 500 to 750 nanograms per sample. The results were analyzed using the Benchling platform (method: alignment, multisequence, with clustal omega platform (Sievers et al., 2011)). The sequence of the primers used for this experiment can be found in the Supplementary Material, Table 4.2.

Following the verification of the sequence, the correct plasmids were used to transform BL21 competent *E. coli* by New England Biolabs (Ipswich, MA, USA) that had been transformed previously with the pEVOL-pAzF, aliquoted and stored in -80°C. The pEVOL-pAzF plasmid was a gift from Peter Schultz (Addgene plasmid #31186). The bacterial transformation protocol used was provided by addgene, and the antibiotics used were ampicillin in final concentration 100 μg/mL and chloramphenicol in 25 μg/mL. Three colonies were picked out of each plate, to be expanded implementing the same procedure as for the XL10-gold bacteria, glycerol stocks were kept while the rest of the culture was used to extract the plasmids with the MiniPrep kit, and the presence of both plasmids (#62731, #31186) was verified with a double digestion, and using the same agarose electrophoresis method as mentioned before.

2.2. SpCas9 Variants Protein Production and Purification

In order to produce an adequate amount of the SpCas9 variants, the BL21 strain containing the pEVOL-pAzF and the mutated SP-Cas9 was expanded to a 500 ml culture, using LB-Broth, ampicillin and chloramphenicol. When the OD600 was between 0.6 and 0.8, Isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration 0.5 mM), and 4-Azido-L-phenyl-alanine (final concentration 2 mM, dissolved in 1M sodium hydroxide) were supplemented to the culture, while the fermentation was continued overnight at 18°C.

The cells were subsequently harvested by centrifugation and resuspended in 25 mL phosphate buffered saline (PBS) containing 25 mM imidazole and protease inhibitors (1 tablet of cOmplete™, EDTA-free Protease Inhibitor Cocktail, by Merck, Darmstadt, Germany) kept on ice. Lysis was performed on ice using a tip sonicator equipped with a 3 mm tip (Bandelin electronic GmbH & Co. KG, Berlin, Germany), and the lysate was then centrifuged, while the supernatant was kept on ice and filtered through a 0.45 μM MiniSart filter (Sartorius, Amersfoort, the Netherlands).

For the purification of the protein, immobilized metal affinity chromatography was implemented. The filtered lysate was applied to an 1 ml HisTrap HP column containing nickel by Cytiva (Catalog #17524701, Medemblik, the Netherlands), using the Äkta PURE chromatography system by Cytiva (Medemblik, the Netherlands), while the elution was performed using a stepwise gradient of imidazole dissolved in PBS, starting from 50 mM, escalating to 75 mM and finishing at 500 mM. Based on the UV (280nm) chromatogram, the fractions containing the eluted SpCas9 were collected, and were put on dialysis devices, namely Float-A-Lyzer G2, 100 kD by Repligen (Breda, the Netherlands), against storage buffer (300 nM NaCl, 0.1 mM EDTA, 10 mM Tris, pH 7.4) of a volume of at least 1000 times the sample volume, twice: once in room temperature for 2 hours and once at 4°C overnight. 10% glycerol was added to the proteins, which were subsequently snap-frozen in liquid nitrogen and then stored at -80°C.

2.3. SpCas9 Variants Protein Characterization

SpCas9 Variants Protein Concentration

To determine the concentration of the produced SpCas9 variants, the Bradford colorimetric assay was mainly implemented. The Pierce Coomassie Plus (Bradford) Assay Kit (Catalog #23236) by Thermo Scientific (Landsmeer, the Netherlands) was used, and the corresponding protocol was followed, noteworthy being that 20 μL of each standard or unknown sample were used, and the final albumin concentration was in the range of 5 to 500 μg/ml.

The plates were analyzed using the BMG SPECTROstar Nano wellplate reader (BMG Labtech, de Meern, the Netherlands), and the readings were analyzed with the MARS Data Analysis Software by the same company. The results were plotted using LibreOffice Calc, and the curve produced out of the data was based either on linear regression or on a point-to-point fit.

SpCas9 Variants Protein Size and Purity

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was the preliminary method used for the determination of the size and the purity of the isolated protein. The samples were treated with 1x Laemmli buffer by Bio-Rad Laboratories (Catalog #1610747, Veenendaal, the Netherlands), containing 55.0875 mM final concentration dithiothreitol (DTT), and denatured at 70°C for 10 minutes. Bolt 4 to 12%, Bis-Tris, 1.0 mm gels by Thermo Scientific (Landsmeer, the Netherlands) were used to separate the proteins, under the conditions of 150 Volt and 55 minutes. PageRuler Plus Prestained Protein Ladder, 10 to 250 kDa by Thermo Scientific (Catalog #26620, Landsmeer, the Netherlands) was the preferred ladder, while a previously produced SpCas9 was used as a control. The gels were then stained using the PageBlue Protein Staining Solution by Thermo Scientific (Catalog #24620, Landsmeer, the Netherlands) and following the instructions included in the kit, and subsequently imaged using the ChemiDoc Imaging System by Bio-Rad Laboratories (Veenendaal, the Netherlands). Whenever deemed necessary, western blot to detect the His-tag was performed: after the electrophoresis, the proteins were transferred to a PVDF membrane using the Trans-Blot Turbo Mini 0.2 µm PVDF Transfer Packs (Catalog #1704156) and the Trans-Blot Turbo Transfer System by Bio-Rad Laboratories (Veenendaal, the Netherlands, settings: Bio-Rad protocols, High Molecular Weight). 3% skim milk powder (Serva, Heidelberg, Germany) dissolved in 1x

Tris-Buffered Saline containing 0.1% Tween 20 Detergent (TBST) was used as blocking solution, and the membrane was incubated with agitation for 1 hour at room temperature. 1x TBST was used for the washing steps, while the HRP Anti-6X His tag antibody (ab1187) by Abcam (Amsterdam, the Netherlands), diluted 1:1000 in the blocking solution, was left to detect the His-tag under agitation at room temperature for 4 hours. The SuperSignal West Dura Extended Duration Substrate kit by Thermo Scientific (Catalog #34075, Landsmeer, the Netherlands), and then imaging using the ChemiDoc Imaging System by Bio-Rad Laboratories (Veenendaal, the Netherlands) were used to detect the chemiluminescence.

Incorporation of Azide Phenylalanine in SpCas9 Variants

To assess if the 4-Azido-L-phenylalanine was indeed incorporated in the structure of the SpCas9 variants, the protein variants were left to react for 1 hour at room temperature with the dissolved in PBS dye DBCO-AF647 by Jena Biosciene (Jena, Germany), with a ratio of 2:1 dye to protein. Then SDS-PAGE electrophoresis was performed as described in section 2.3., and afterwards fluorescence was detected using the ChemiDoc Imaging System by Bio-Rad Laboratories (Veenendaal, the Netherlands) and the Alexa Fluor 647 preset.

In Vitro Activity Assay of SpCas9 Variants

To assess the ability of the SpCas9 variants to cleave DNA *in vitro*, an activity assay was performed as described previously (Walther et al., 2022). The sequence of the sgRNA can be found in the Supplementary Materials Table 4.3, while buffer 3.1 was acquired by New England Biolabs (Ipswich, MA, USA), and the pMJ922 plasmid was a gift from Martin Jinek (Addgene plasmid #78312) (Burger et al., 2016). Gel densitometry was used to measure the activity of the variants (Villela et al., 2020): the area under the curve was specified using ImageJ (1.49p), and it was compared with the one of the control, that being in-house produced native SpCas9, in order to examine the relative activity.

Cell Line Experiments

Human embryonic kidney 293 cells of type T (HEK 293T) characterized by stable EGFP expression (HEK 293T-EGFP) were cultured in low-glucose Dulbecco's Modified Eagle Medium provided by ThermoFisher Scientific (Landsmeer, the Netherlands) containing 10% fetal bovine serum, at 37°C and 5% CO2, while the cell culture plastics were attained from Greiner Bio-One (Alphen aan de Rijn, the Netherlands). The cell line was a gift from Dr. Olivier de Jong, and further information about its creation, contents and selection procedure is described previously (Jong et al., 2016; Walther et al., 2022). Periodically, 1 mg/mL Gibco Geneticin Selective Antibiotic (G418 sulfate, by Fischer Scientific, Landsmeer, the Netherlands) was used when subculturing, in order to ensure the maintenance of the cell line.

To test the ability of the SpCas9 variants to induce gene editing in HEK 293T-EGFP, an assay detecting the mutation of EGFP towards BFP was used, which was based on literature (Glaser et al., 2016), and the cells were assayed using fluorescence-activated cell sorting (FACS). To introduce the SpCas9 variants inside the cells, lipid nanoparticles (LNPs) were used, made using nuclease-free water as a formulation buffer, and containing a final molar ratio of RNP/HDR template of 1:2, and DOTAP 0.25 mole%. The protocol for the production of the LNPs, as well as the protocol for the experimental procedure, the flow cytometry (also gating method and method validation) are described as previously (Walther et al., 2022). As a positive control, SpCas9 was used, and as a negative control PBS instead of protein was added to the mixture of LNPs. The BD Falcon U-bottom 96-well plate, as well as the BD FACSAria III cell sorter used were acquired from Becton Dickinson (Franklin Lakes, NJ, USA), while the generated data were analyzed using the Flowlogic software by Inivai Technologies (Mentone, Australia, version 7.3). The plasmid for BFP expression was ordered from Twist Bioscience (San Francisco, CA, USA).

2.4. SpCas9 Variants PEGylation

To determine the optimal reaction conditions between the SpCas9 variants and the DBCO-functionalized PEG, different sets of conditions regarding temperature (4°C, room temperature and 37°C), reaction time (ranging from 30 minutes to 7 days) and ratio of DBCO-mPEG (30 kDa) to F/Yx* SpCas9 (from 2.5:1 to 60:1) were studied, drawing inspiration from literature (Cook et al., 2016; Lühmann

et al., 2019; Ptacin et al., 2021; Simon et al., 2014; Zambrano et al., 2020; B. Zhang et al., 2021). Another aim of this experiment was to discover, beside the reaction yield, the possible factors that could render the SpCas9 variants inactive during the conjugation reaction. The conditions tested in this particular subexperiment were the 50:1 ratio and: 37°C for various durations of time, 1h at 37°C and then 7 days at 4°C, and 4 days at room temperature, while sgRNA against EFGP was provided since the start of the reaction in a 1 mM:1 mM ratio of RNA/SpCas9 variant to one sample out of each set of conditions, as it may change the reaction dynamics. The reaction for all samples was performed under gentle agitation.
The yield of the reaction was calculated by measuring the percentage of the SpCas9 variant that was converted into the conjugate. More precisely, the samples were electrophorized in SDS-PAGE gels as described previously with the exception of not using DTT, then the gels were stained with Coomasie, and the data were analyzed using ImageJ (1.49p) and densitometry (Villela et al., 2020). The yield was deducted by comparing the band of the PEGylated product with the band of the non-PEGylated SpCas9 in each sample (as more than one products were detected, the conversion rate could not be estimated only by the reduction of the SpCas9 band).
The SpCas9 variant mainly used in the PEGylation experiments was the F688AzF one. The conditions, under which all SpCas9 variants were subjected, were the ratio of 50:1 DBCO-mPEG/SpCas9-xAzF for 24 hours at 37°C, and the 4 days at room temperature 50:1 DBCO-mPEG/SpCas9-xAzF (providing sgRNA in a 1 mM:1 mM ratio of RNA/SpCas9-xAzF to one sample out of each set of the latest condition). A part of the samples of the subset "50:1 DBCO-mPEG/SpCas9-xAzF for 24 hours at 37°C" was used to perform the *in vitro* activity assay that was described previously. As the conjugate was not purified out of the sample, the activity of each sample could not be contributed only to the conjugate or the SpCas9 variant, but could be compared to the positive control while taking into account the percentage of the conjugate in the final sample.

2.5. Protein Domains Expression, Purification and Conjugation with SpCas9 Variants, Purification of Conjugates

The plasmids for the expression of the three protein domains were ordered from Twist Bioscience (San Francisco, CA, USA). Their general structure can be found in Supplementary Materials, Figure 4.3, Figure 4.4, and Figure 4.5. The sequences are encoded in the pET-21(+) vector, while the expressed proteins contain a Strep-tag II fused at the C-terminal.
The plasmids were used to transform Stbl3 Chemically Competent *E. coli* by Invitrogen (Waltham, MA, USA), with the purpose of safe-keeping the plasmids by making a glycerol stock of the cells, and competent BL21 *E. coli* by New England Biolabs (Ipswich, MA, USA) (made chemically competent in-house by Ator Ashoti), in order to express the protein domains. The transformation protocol was the one previously mentioned for the BL21 strain in section 2.2., and the same applies for the production method and downstream handling of the lysate, also referred in section 2.2., with the exceptions of using only ampicillin in the growth medium, and not adding 4-Azido-L-phenylalanine to the cells.
For the purification of the protein domains, affinity chromatography was implemented. The column used was the 1 ml Strep-Tactin XT 4Flow cartridge by IBA Lifesciences GmbH (Catalog #2-5023-001, Göttingen, Germany) containing a engineered form of streptavidin, while the purification was done using the Äkta PURE chromatography system by Cytiva (Medemblik, the Netherlands). The elution was performed using 1x Buffer BXT ordered from IBA Lifesciences GmbH (Göttingen, Germany), which contains biotin, and by pausing the elution for 30 minutes after 1 ml of elution buffer had saturated the column. Based on the UV (280nm) chromatogram, the fractions containing the eluted protein domains were collected, were left to react for 1 hour at room temperature with the heterobifunctional PEG linkers, namely DBCO-PEG24-Maleimide MW 1555.8 and DBCO-PEG24-NHS ester MW 1530.8, using a ratio of 2:1 regarding DBCO-PEG24-Maleimide/DN1S, 2:1 regarding DBCO-PEG24-Maleimide/IPP2 and 1:1 regarding DBCO-PEG24-NHS/Gem110, and then put on dialysis devices, the SLIDE-A-LYZER™ G3 DIALYSIS CASSETTES, 10K MWCO by Thermo Scientific (Landsmeer, the Netherlands), and following the same dialysis procedure as mentioned in section 2.2..
Using electrophoresis and Coomassie staining (refer to section 2.3., the proteins were characterized for their size and purity. The average size of the (PEGylated) DN1S domain was also characterized

using Dynamic Light Scattering (DLS): the Zetasizer Nano-S by Malvern Panalytical (Malvern ALV CGS-3, Malvern, UK) and a microcuvette were used, and each sample was subjected to 3 measurements.
The protein domains were then conjugated with one of the SpCas9 variants in a molar ratio of 4.5:1 linker-domain/SpCas9 variant, under gentle agitation for 4 days at room temperature.
A combination of the two chromatography methods mentioned previously in section 2.2. and section 2.5. was implemented for the purification of the final conjugates. The sample was first applied to either the HisTrap or the Strep-Tactin XT column, then the fractions corresponding to the elution peak were gathered and applied during a second run to the other column. As this combination proved inefficient, the samples were gathered by combining the eluted fractions, the proteins were denatured by urea in 3M final concentration, and then applied on the Strep-Tactin XT column in order to test if the folding of the conjugates hinders the binding to the streptavidin analogue. Each subsequent fraction of the whole purification was studied for its protein content, independent of the chromatogram, by firstly performing electrophoresis as described previously without the use of DTT, and then staining the gels using the Pierce Silver Stain Kit by Thermo Scientific (Landsmeer, the Netherlands) and the manufacturer's instructions.

Chapter 3

Results

3.1. Production of the SpCas9 Variants

To utilize click chemistry reaction tools on SpCas9, the pAzF amino acid should be incorporated into the protein sequence. After using SDM to introduce the TAG codon in the SpCas9 sequence, the produced DNA sequence was characterized, starting by an enzymatic digestion if a diagnostic restriction site was introduced during SDM, and then by partial sanger sequencing at the site of interest, as shown in Figure 3.1 for the representative F196 mutation. We managed to produce four out of the ten desired mutations on the plasmid level, namely the F196, F539, F688 and Y1036. The results of their Sanger sequencing can be found in Figure 3.1 for the F196 mutation and in the Supplementary materials Figure 4.1 for the rest of the mutations. The correctly mutated plasmids were then used to produce the four corresponding SpCas9 variants, which in turn were characterized regarding the incorporation of the pAzF in their sequence as illustrated in part C Figure 3.1, by conjugation with the DBCO-AF647 dye. For all four proteins, the pAzF was incorporated successfully in the amino acid sequence. Slight differences between the fluorescence emitted by the variants (under same conditions, regarding amount of protein, amount of dye and reaction conditions) can be attributed to different amount of dye binding to the variants. In particular, the F196AzF exhibits less fluorescence when having reacted with the DBCO-AF647. A weaker band can be seen at lower weight bands for some of the variants.

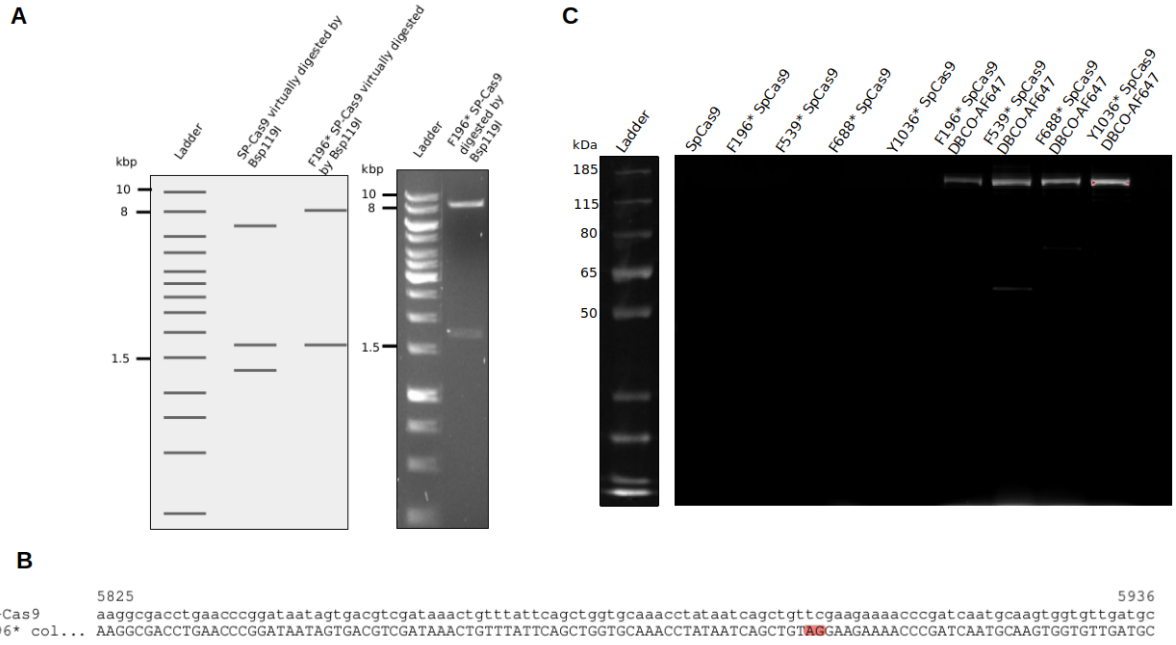

Figure 3.1: (A) Virtual digestion by Bsp119I of the SpCas9 plasmid and the SpCas9 plasmid bearing the correct mutation for F196* protein, and the digestion by Bsp119I of the in-house mutagenized SpCas9 plasmid towards the TAG codon at the 196 position; (B) Sequencing results analyzed in Benchling comparing the SpCas9 plasmid and the in-house mutagenized SpCas9 plasmid towards the TAG codon at the 196 position, showing the relevant for the mutation part of the DNA sequence. The differences between the two sequences are highlighted; (C) SDS-PAGE gel of native SpCas9 and F196AzF, F539AzF, F688AzF and Y1036AzF SpCas9 variants, plain and conjugated with the DBCO-AF647 dye with a molar ratio of 2:1 dye to protein

3.2. Activity of the SpCas9 Variants

In order to further characterize the produced SpCas9 variants, their ability to cleave DNA *in vitro* and in cells was tested. As can be seen in part A of Figure 3.2, the mutated SpCas9 variants were at

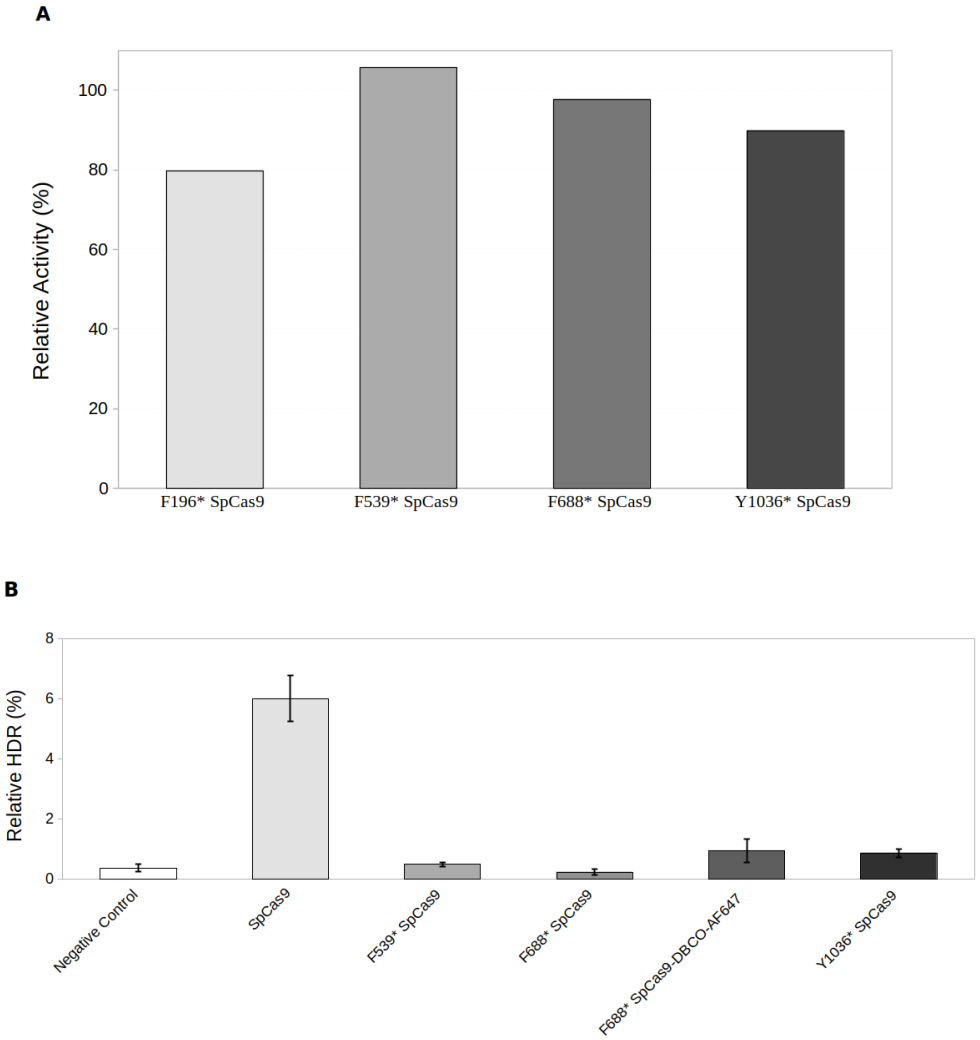

Figure 3.2: (A) *In vitro* relative activity of the F196AzF, F539AzF, F688AzF and Y1036AzF SpCas9 variants; (B) Relative HDR observed in HEK 293T-EGFP cells after addition of LNPs encapsulating the F539AzF, F688AzF, Y1036AzF SpCas9 variants and the conjugated with DBCO-AF647 F688AzF SpCas9 variant

least 80% active when compared to an internal SpCas9 control (thus the term relative activity) regarding their ability to cleave the pMJ922 plasmid *in vitro*, while the F539AzF SpCas9 variant performed even better than the control SpCas9, and the F688AzF SpCas9 was comparable to the control. The least performing variant would be the F196AzF one. Furthermore, the variants were formulated into LNPs, which were added to HEK 293T-EGFP cells, and the HDR efficiency was measured using the EFGP to BFP conversion assay and flow cytometry, and was compared to a negative (empty formulation) and positive (in-house produced SpCas9) control. The starting concentrations of the F539AzF SpCas9, F688AzF SpCas9 and Y1036AzF SpCas9 variants used to formulate the LNPs were correspondingly 0.06, 0.09, and 0.03 mg/ml. Part B of Figure 3.2 illustrates the HDR accomplished by the F196AzF, the F539AzF, the F688AzF plain and conjugated with DBCO-AF647 and the Y1036AzF SpCas9 variants compared to the total gene editing achieved in the cells (relative HDR). The F539AzF exhibited 12.36-fold decrease, the F688AzF 25.65-fold decrease, the F688AzF conju-

gated with DBCO-AF647 6.35-fold decrease, and the Y1036AzF SpCas9 6.93-fold decrease in relative HDR compared to the positive control. Of note is the fact that the conjugated with DBCO-AF647 F688AzF SpCas9 showed higher yield of HDR compared to the unconjugated same variant.

3.3. PEGylation of the F688AzF SpCas9 Variant

For the purpose of determining the ability of the variants to undergo click chemistry reactions with bigger molecules, and foreseeing the optimal reaction conditions for the bioconjugation with the protein domains, the reaction yield between the F688AzF SpCas9 variant and DBCO-mPEG (30 kDa) was tested. As evidenced in Figure 3.3, the higher the molar ratio of PEG to variant, the higher also the reaction yield, while a saturation can be observed at 50:1 molar ratio of PEG to variant. Furthermore, the reaction is more efficient under 37°C reaching a level of approximately 60% conversion in 24 hours, while under room temperature the yield is at most around 10% in 48 hours, and under 4°C is less than 5% even after 48 hours of reaction. For all the conditions, increasing the reaction time leads also to a higher reaction yield, but under 37°C, increasing the reaction time over 24 hours does not seem to affect the yield significantly. The data shown in Figure 3.3 were generated using the SDS-PAGE gels that can be found in Figure 4.8 of Supplementary Materials, the ImageJ (1.49p) program and the densitometry method.

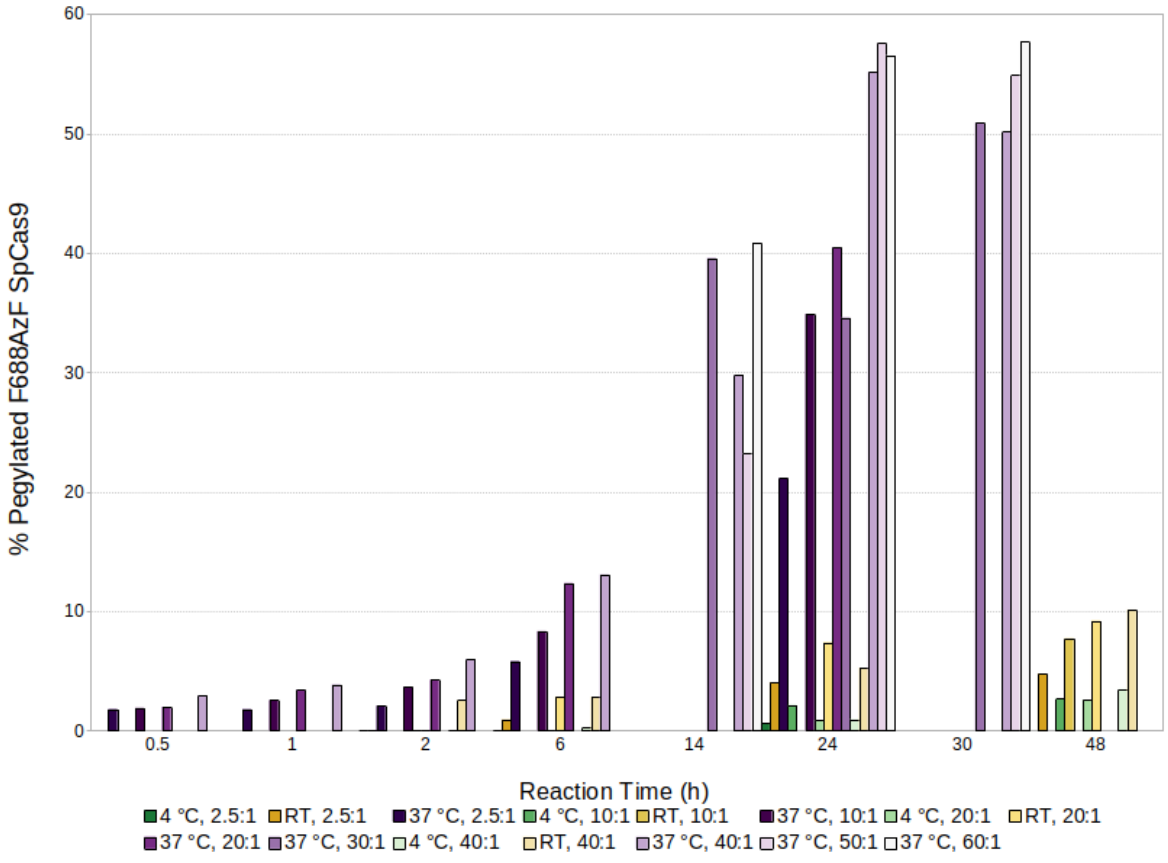

Figure 3.3: Reaction yield between DBCO-mPEG (30 kDa) and F688AzF SpCas9 variant under different conditions

3.4. PEGylation and Activity of the PEGylated and Conjugated with DBCO-AF6467 SpCas9 Variants

After observing that incubation at 37°C inactivates SpCas9 regardless of duration (data not shown), milder conditions for the conjugation of the variants to the DBCO-mPEG (30 kDa) molecule

and the protein domains had to be found. Having this in mind, the conditions of 4 days under room temperature, and 1 hour under 37°C followed by 7 days under 4°C, with the possible addition of sgRNA in both cases, were tested regarding the reaction yield and the activity of the products. The F688AzF SpCas9 was left to conjugate with the DBCO-mPEG (30 kDa) to test both sets of conditions: 1 hour under 37°C followed by 7 days under 4°C in the presence of 0.033 μM sgRNA led to a yield of 27% PEGylated SpCas9 and 45% relative activity *in vitro*, while 4 days under room temperature and in the presence of 0.033 μM sgRNA led to 31% PEGylated SpCas9 and 92% activity *in vitro*. Thus, 4 days under room temperature were chosen. In subsequent experiments, for the F539AzF SpCas9 variant the addition of sgRNA did not alter the reaction yield significantly (yield of 29% without sgRNA, and 33% with sgRNA), while for the Y1036AzF variant the reaction yield in the presence of sgRNA was approximately double. Regarding activity, it should be noted that the products were not purified out of the reaction tubes, and therefore only part of the relative activity can be attributed to the conjugated variant, while the conjugation with the DBCO-AF647 dye should be highly efficient and thus most of the SpCas9 in solution would be bound with dye, as the yield of SPAAC between protein and click chemistry functionalized dye is usually high (Jang et al., 2012). The F539AzF SpCas9 variant is 75% active when conjugated with the dye, while the F688AzF one is approximately 20% active under the same conditions. Incubation for 4 days under room temperature with DBCO-mPEG (30 kDa) without sgRNA led to approximately 100% relative activity of products concerning the F196AzF and the F539AzF variants, 85% for the Y1036AzF and 12% for the F688AzF, while the addition of sgRNA resulted in activity of 92 to 98% for F196AzF, F539AzF and F688AzF.

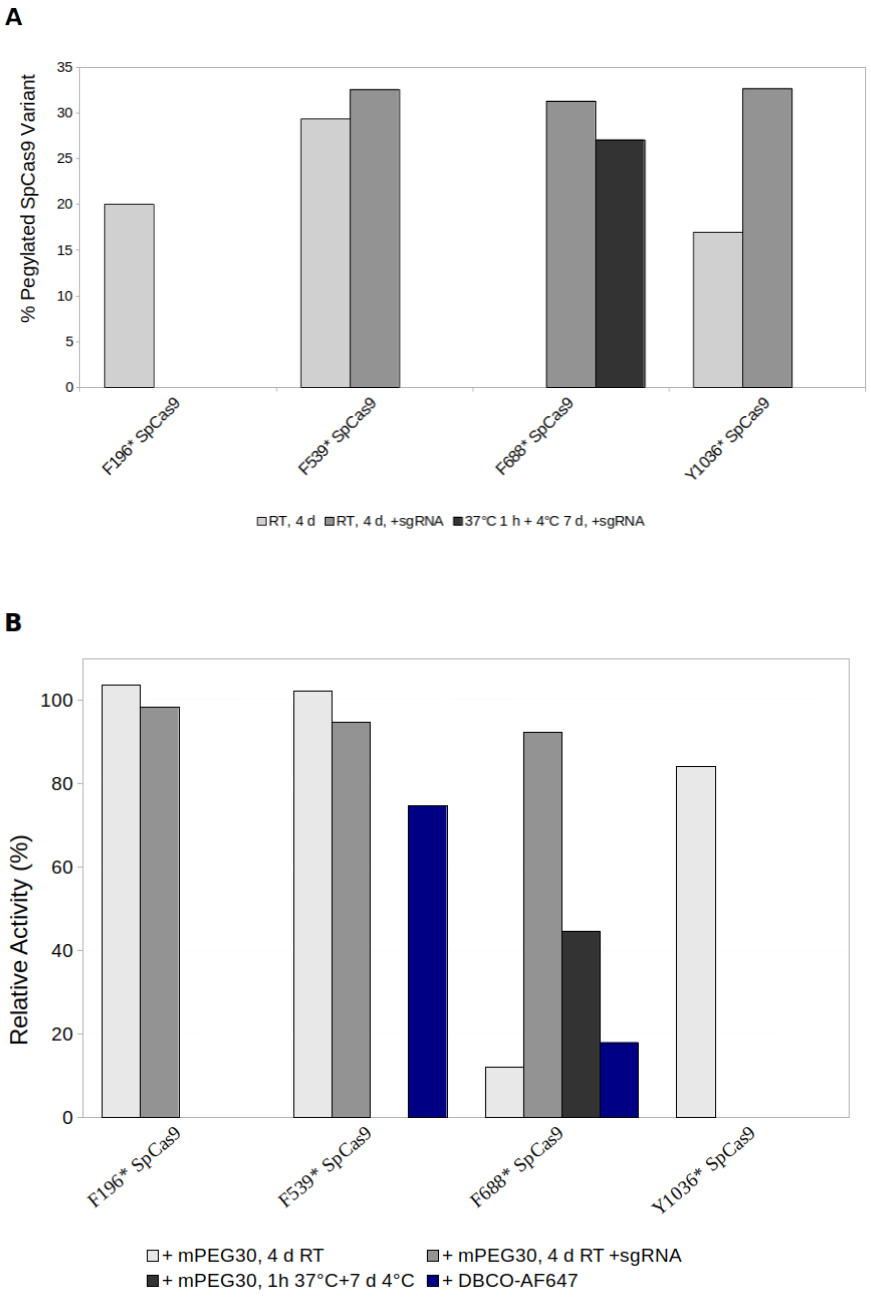

Figure 3.4: (A) Reaction yield between DBCO-mPEG (30 kDa) and F196AzF SpCas9, F539AzF SpCas9, F688AzF SpCas9, Y1036AzF SpCas9 under different conditions ; (B) *In vitro* activity of the conjugated under different conditions with DBCO-mPEG (30 kDa) F196AzF SpCas9, F539AzF SpCas9, F688AzF SpCas9, and Y1036AzF SpCas9, and *in vitro* activity of the conjugated with DBCO-AF647 F539AzF SpCas9 and F688AzF SpCas9

3.5. Characterization of the Protein Domains and Kinetics of the Variant:Domain Conjugation

To verify the presence of the protein domains after their production, the purified and dialysed samples were visualized on SDS-PAGE gels. In all three cases, a band at the correct height was present, while some impurities remained even after dialysis.

For the purpose of determining the stability of the protein domains, the size of the domain with the highest molecular weight was examined using DLS. The non-PEGylated DN1S seemed to aggregate over a period of 3 days, starting from an average size of 76.69 (d.nm) and ending at 128.8 (d.nm)

and also exhibiting a significant standard deviation betweeen the different measures, while the PEGylated DN1S was stable with minimal standard deviation over a period of 9 days, starting from 15.26 (d.nm) and ending at 13.35 (d.nm) as average size.
Afterwards, the F688AzF SpCas9 variant was left to conjugate with the earlier produced linker-domain complexes, under gentle agitation for 3 or 4 days in room temperature, and the kinetics of the reaction were examined by taking a sample every 24 hours and running it in a SDS-PAGE gel. The yield was determined by measuring the intensity of the band at the height of the conjugate and of the SpCas9 band and deriving the percentage of conversion, utilizing ImageJ (1.49p) and densitometry. For the conjugation of the variant with IPP2 and Gem110 the yield results can be seen at Figure 3.5 and the relevant gels can be found in Figure 4.12, while for the conjugation with DN1S the kinetics gel is depicted in Figure 3.5. As for the yield after 4 days of reaction, for the F688AzF-IPP2 it would be a conversion of 51% and for the F688AzF-Gem110 6%. Regarding the conjugation with the DN1S, the yield cannot be easily extracted from the gel. The highlighted in the figure band, that may be in the correct height, does not increase in intensity over time and conveys approximately a 10% yield from day 0, while the other band at a weight of 185 kDa that appears to increase over time reaches approximately 9% conversion over a period of 3 days.

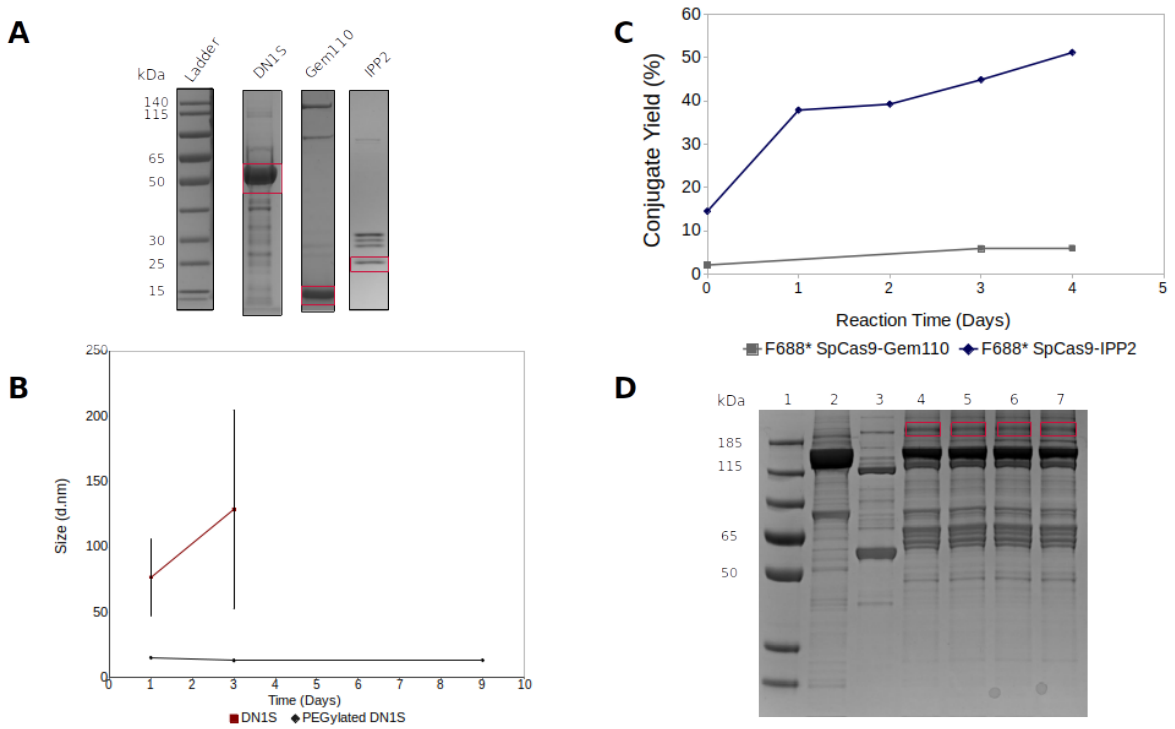

Figure 3.5: (A) SDS-PAGE gels of the produced domains, after dialysis, where DN1S, Gem110, and IPP2 are highlighted by the boxes; (B) The size stability of DN1S and DN1S conjugated with the DBC0-PEG24-maleimide linker over time, measured using DLS; (C) Reaction yield between F688AzF SpCas9 and IPP2, and between F688AzF SpCas9 and Gem110 over 4 days and under room temperature; (D) SDS-PAGE gel showing the kinetics and the products of the reaction between F688AzF SpCas9 and the conjugated with DBC0-PEG24-maleimide DN1S from day 0 to day 3 under room temperature: (1) Ladder; (2) 11.4 μg F688AzF SpCas9; (3) 4.5 μg freezed-thawed once DN1S; (4) Products of F668AzF SpCas9 and DBC0-PEG24-maleimide DN1S at day 0; (5) Products of F668AzF SpCas9 and DBC0-PEG24-maleimide DN1S at day 1; (6) Products of F668AzF SpCas9 and DBC0-PEG24-maleimide DN1S at day 2; (7) Products of F668AzF SpCas9 and DBC0-PEG24-maleimide DN1S at day 3

3.6. Purification of the Conjugates

To isolate the conjugation product, two-step affinity chromatography for the double-tagged conjugate was used. Representative chromatograms from the purifications are depicted in Figure 3.6, in which no peaks were observed when the elution for either purification started. Silver stained

SDS-PAGE gels of the purified samples and also from denatured purified samples are included in part C of Figure 3.6. Regarding the conjugate between F688AzF and IPP2, the product can possibly be found in the sample application fraction of the purification of the denatured product using the StrepTactin column, and not in the elution of either the normally folded or the denatured conjugate using the StrepTactin column. As for the F688AzF-DN1s conjugate, the product of correct height appears at the sample application fraction of the elution and at the elution fraction of the denatured product using the StrepTactin column in both cases, while the band at 185 kDa mentioned in section 3.5. can be found in the elution fraction of the normally folded product being purified with the HisTrap column. Finally, the F688AzF-Gem110 conjugate could not be detected in SDS-PAGE gels.

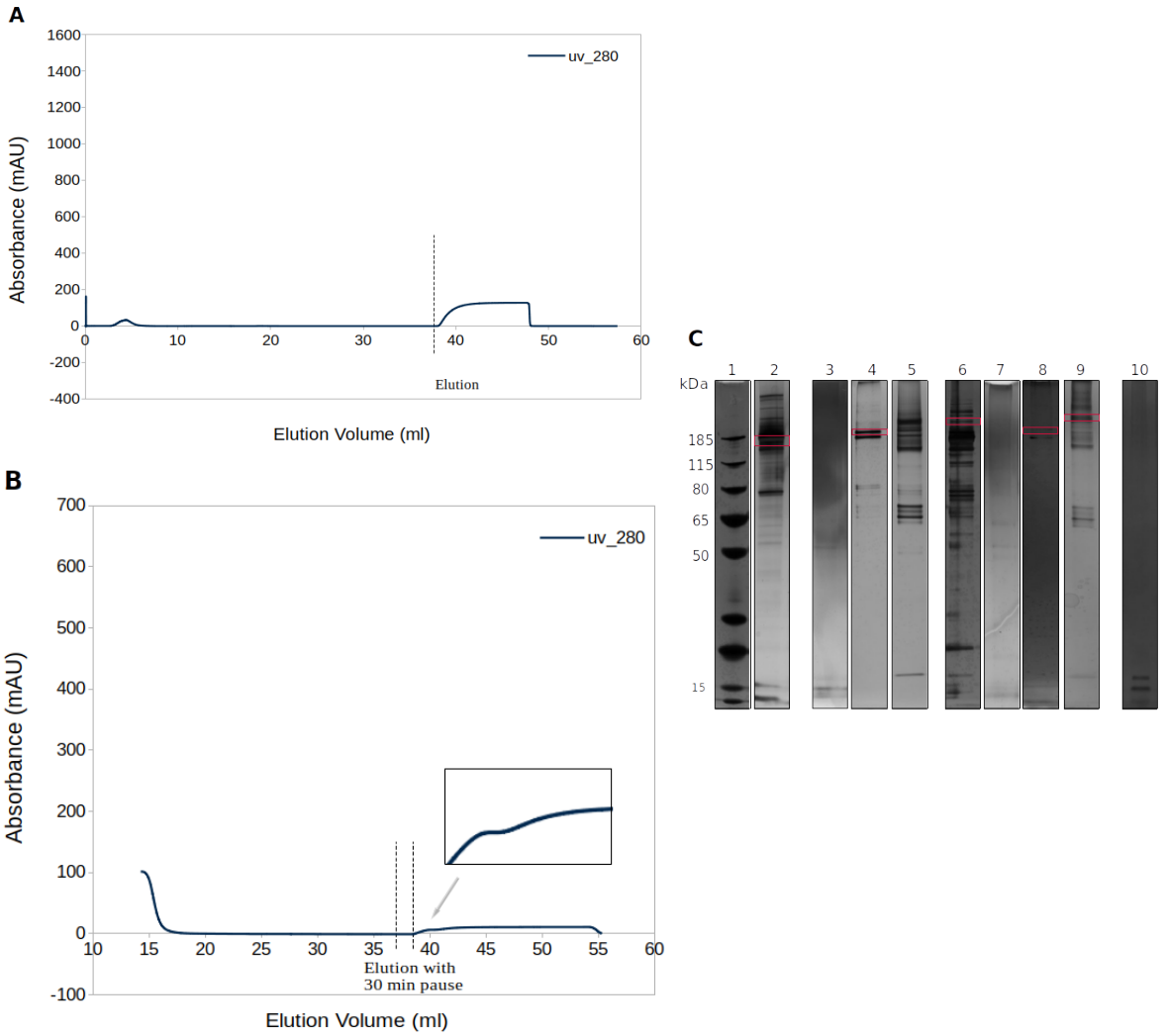

Figure 3.6: Purification of the conjugates (A) Elution Chromatogram of the F688AzF SpCas9-IPP2 conjugate durig His-Tag purification; (B) Elution Chromatogram of the F688AzF SpCas9-IPP2 conjugate durig Strep-Tag purification; (C) Silver stained SDS-PAGE gels of some fractions collected during His- or Strep-Tag purification of the conjugates: (1) Ladder; (2) F688AzF SpCas9 1 μg; (3) Elution Fraction during F688AzF SpCas9-IPP2 conjugate's Strep-Tag purification; (4) Sample Application Fraction during denatured F688AzF SpCas9-IPP2 conjugate's Strep-Tag purification; (5) Elution Fraction during denatured F688AzF SpCas9-IPP2 conjugate's Strep-Tag purification; (6) Sample Application Fraction during F688AzF SpCas9-DN1S conjugate's Strep-Tag purification; (7) Elution Fraction during F688AzF SpCas9-DN1S conjugate's Strep-Tag purification; (8) Elution Fraction during F688AzF SpCas9-DN1S conjugate's His-Tag purification; (9) Elution Fraction during denatured F688AzF SpCas9-DN1S conjugate's Strep-Tag purification; (10) Elution Fraction during denatured F688AzF SpCas9-Gem110 conjugate's Strep-Tag purification

Chapter 4

Discussion and Conclusion

As the CRISPR/Cas9 genome editing tool is being studied for its therapeutic use and potential in curing human genetic diseases (Sharma et al., 2021), the need to precisely manipulate the outcome of gene editing utilizing the HDR repair mechanism is evident. Various methods have been implemented in order to increase the efficiency of HDR after the DSB induced by Cas9, succeeding in varying degrees (Sun et al., 2022; Yang et al., 2020). This report aims to introduce another method of enhancing the HDR efficiency and frequency, by conjugating the Cas9 molecule with protein domains that intrinsically shift the balance of HDR to NHEJ towards the desired HDR. The protein domains have all been reported in the literature to act favorably regarding the goal of this project, and specifically IPP2 accumulates inside the nucleus during the S-phase of the cell cycle, when HDR is most prevalent (Kakinoki et al., 1997), DN1S increases locally HDR and inhibits NHEJ by interacting with other, related with DNA repair, proteins (Jayavaradhan et al., 2019), and Gem110 acts as a substrate for degradation during the late M and G1 phase of the cell cycle, when NHEJ is preferred over HDR (Gutschner et al., 2016). While DN1S and Gem110 have already been fused with SpCas9 to enhance HDR with positive results (Gutschner et al., 2016; Jayavaradhan et al., 2019), to the best of our knowledge, fusion or conjugation of IPP2 with SpCas9 has not been examined yet and thus the conjugation with this protein domain would be promising in its novelty.

Although fusion proteins have been traditionally used to engineer a heterobifunctional protein (Riggs, 2013), some N- or C-terminal Cas9 fusion proteins have been reported to be problematic due to steric hindrance in the final protein conformation, as the two terminals of Cas9 orient inward (Nishimasu et al., 2014). Having this in mind, intra-protein conjugation instead of terminal conjugation would seem like an exploitable way to make a multifunctional protein (S. Kim et al., 2016), in a site-specific manner that would not affect the functionality of Cas9 (Oakes et al., 2016). Based on the work of Oakes et al., several sites in the Cas9 sequence that would not alter its ability to bind to DNA were chosen, and in those sites an UAA, useful for the protein conjugation, was incorporated using genetic manipulation (C. H. Kim et al., 2013). 4-Azido-L-phenylalanine is the UAA of choice in this project based on its favorable characeristics (Debets et al., 2010), and in order to conjugate Cas9 with the previously mentioned protein domains, small heterobifunctional linkers are utilized, which are distinguished by having a DBCO functional group to react with the pAzF, and PEG of 24 repeats in length. According to literature, including a PEG of small length in the linker is beneficial for the solubility and spacial flexibility of the final construct (Pickens et al., 2017).

To begin with, taking into consideration the low efficiency of SDM (B. Zhang et al., 2021), we managed to generate four out of the ten variants of Cas9, with the correctly placed mutation of the stop codon being verified in the DNA level and the pAzF being incorporated in the protein sequence. To validate the incorporation of the pAzF, the DBCO-AF647 fluorescent dye was left to react with the Cas9 variants, as the DBCO functional group "clicks" specifically with the pAzF. For all four variants, namely the F196AzF, F539AzF, F688AzF and Y1036AzF Cas9, a fluorescent band at the molecular weight of SpCas9 was present, indicating the integration of the UAA in the protein. The difference in the fluorescent intensity of the bands between the SpCas9 variants conjugated with the dye could be attributed to the site of the mutation and consequently the position of the pAzF. The location could affect the accessibility towards the dye due to conformational reasons, as is the case for F196. Moreover, in the SDS-PAGE gel of the conjugated with dye variants, in the F539AzF SpCas9 sample a band of low intensity of fluorescence between 65 and 50 kDa is visible, and in the F688AzF SpCas9 sample a band of even lower intensity can be found between 80 and 65 kDa. In both cases these bands could be explained as being the truncated products of SpCas9, a reported incidence when the TAG codon is used for the incorporation of UAAs (Batjargal et al., 2015). A stop codon at position F539 would result in a product of 62 kDa and at position F688 in a product of 80 kDa.

Following the production of the variants, their ability to function as endonucleases had to be examined. The *in vitro* activity of the SpCas9 variants was firstly assessed, based on their ability to cleave the pMJ922 plasmid guided by an RNA complementary to the EGFP gene. The F539AzF Cas9 variant was superior to the native SpCas9 regarding activity. Alterations in this amino acid position are generally well-tolerated, and even increasing the specificity of SpCas9, as evidenced in Sniper

Cas9 (Lee et al., 2018). Following in order of performance is the F688AzF SpCas9 variant, while Y1036AzF SpCas9 and F196AzF SpCas9 exhibited lower percentages of activity, although not falling below 80% relative activity. With these results in mind, F539AzF SpCas9 and F688AzF SpCas9 conjugated with DBCO-AF647 were also examined regarding their relative activity *in vitro*. Both of the variants were active when conjugated with the fluorescent dye, with F539AzF SpCas9 performing approximately 4 times better than the F688AzF one. The activity of the variants in cells, and more precisely, the HDR efficiency in HEK 293T-EGFP after administration of the SpCas9 variants formulated in lipid nanoparticles, was the next measurement done. The variants performed poorly in this experiment, with the relative HDR percentage being less than 1% for all of them and the higher efficiency was attributed to the F688AzF SpCas9 variant conjugated with the DBCO-AF647, while the control SpCas9 exhibits a value of 6%. This observation may be attributed to the formulation of the LNPs: as the used in this experiment batches of SpCas9 variants were of low concentration (starting concentrations of variants ranged from 0.03 to 0.09 mg/ml, while the one of the positive control was 2.1 mg/ml), the formulation of the LNPs was not optimal regarding the amount of added dilution buffer, and this has been proved to affect encapsulation efficiency and delivery of the components to the cells (Walther et al., 2022). Based on these findings, in combination with the lower conjugation efficiency of F196AzF SpCas9 with the DBCO-AF647 and the fact that the F688 position has the best accessibility towards water, the project mainly focused on the F688AzF SpCas9 variant.
The second component of the project pursued the conjugation of the AzF containing SpCas9 with various molecules of higher molecular weight, particularly a DBCO-mPEG of 30 kDA and the three protein domains with an in-between heterobifunctional linker. Regarding the conjugation with the PEG, it was observed that the reaction between F688AzF SpCas9 and DBCO-PEG occured even at 4°C, although not readily and in low yields. Increasing the temperature had a positive impact in the yield, with the reaction exhibiting higher yield at room temperature and the highest at 37°C. Increased incubation time also led to higher yields, but a saturation can be observed after 24 hours only for the reaction at 37°C. In addition, another parameter checked was the molar ratio: increasing the ratio of PEG to protein gave rise to more PEGylated product, but again a saturation was obtained at 50:1 molar ratio. The final variable examined, which would possibly affect the yield, was the addition of sgRNA in the reaction. This addition can be justified as it forces the protein to take a conformation helpful for DNA-binding (Palermo et al., 2017) and possibly changes the accessibility of the AzF position, as conjugating the proteins while SpCas9 is in its almost active conformation may help avoid later steric hindrances. In our case, adding sgRNA did not alter the yield significantly for most variants, beside the Y1036AzF. These findings generally agree with other reports using different proteins and DBCO-PEG, but they do come in contrast in some aspects, such as the readily-occurring reaction at 4°C which was not observed in our case, or the declining conjugation after the optimal molar ratio, which was also significantly lower, i.e. 3:1 (van Moorsel et al., 2019) or 5:1, or the lower reaction time needed for high yields (Debets et al., 2010). The difference may be explained by the position of the azide in the protein structure: incorporating the azide at the terminal site of the protein may grant different accessibility to the other click chemistry reagent, as the C-terminal is generally more exposed and disordered (Lobanov et al., 2010), but even less accessible azides have proven to be able to click react with aza-dibenzocyclooctyne linked PEG (Debets et al., 2010). Also, it should be noted that staining using barium iodide is a better method to detect the PEG in an SDS-PAGE gel (Kurfürst, 1992), a method not used in this project, and that PEGylated products may migrate unexpectedly and thus their apparent molecular weight could be incorrect (Odom et al., 1997; Zheng et al., 2007). In regards to these remarks, future experiments to determine more precisely the yield of the PEGylation of the variants could use either native PAGE electrophoresis to avoid the SDS and PEG interference (Zheng et al., 2007), or use barium iodide staining and western blot in conjunction with coomassie blue staining of the SDS-PAGE gels.
As for the relative activity *in vitro*, incubation with DBCO-mPEG (30 kDa) at 37°C deactivated the protein regardless of incubation time, at 4°C resulted in low activity of the variants, while at room temperature yielded the most active protein. The addition of sgRNA since the beginning of the conjugation had a huge impact on the activity of most of the variants, and especially for the F688AzF one led to a more than 8-fold increase in activity.
The final experiment using the F688AzF SpCas9 would be its conjugation with IPP2, Gem110 and DN1S. Regarding the first two protein domains, visualization of the bioconjugation products in an

SDS-PAGE gel indicated the possible formation of the conjugate, with the yield of the reaction gradually increasing over days. About the DN1S conjugation, at the SDS-PAGE gel representing the kinetics of the reaction, a unaltered during the days band at the estimated correct molecular weight can be seen from point 0, while the same band is also found in the unconjugated DN1S when its freeze-thawed once. As the latter is the DN1S sample used for the bioconjugation, and as in the previous experiments using PEG and the other protein domains the reaction never occured readily, it is estimated that this band does not represent the conjugate. A band appearing at 185 kDa over the course of the reaction could be a conjugation product. One guess would be that the maleimide in the PEG-linker reacts with a cysteine of an impurity in the DN1S sample, such as the one appearing between 25 and 30 kDa, although unlikely to produce later obvious conjugation results with the SpCas9 due to its low amount. Another guess would be that the mobility of the DN1S-linker-F688AzF conjugate does not correlate perfectly with the molecular weight, as evidenced by reports using different protein-protein conjugates (Tacal and Ozer, 2002): SDS causes the protein to unfold and become linear (Turro et al., 1995; Winogradoff et al., 2020), while the conjugate cannot become linear as it is by definition branched even when the two distinct proteins, namely SpCas9 and the DN1S, are unfolded. Thus, the apparent molecular weight of the DN1S-linker-F688AzF conjugate could be lower. However, the logical continuation of this train of thought would be whether in the conjugation with the other two protein domains the band at the seemingly correct weight is indeed the conjugate. Our only proof regarding that is the gradual increase in the amount of product over the duration of the reaction. Moreover, the fact that DN1S is the bigger in size protein domain makes the branched aspect of the linearized with SDS conjugate more prominent, thus possibly affecting in higher degree the way the product runs in the SDS-PAGE gel.
Continuing with the purification of the conjugates, the two-step chromatography method proved inefficient. Regarding the His-Tag purification, which was usually done first, for all three conjugates the chromatogram did not reveal any peaks of either SpCas9 or the conjugate. It is probable that the protein peak is hidden, as imidazole also absorbs at 280 nm (https://pubchem.ncbi.nlm.nih.gov/compound/imidazole) and the amount of His-tagged proteins was low (approximately 350 μg of SpCas9 were added on Day 0). The following Strep-Tag purification also produced inconclusive results. Regarding the F688AzF-Gem110 conjugate, its formation was almost negligible, but as Strep-Tactin XT has affinity to Strep-Tag II even in the nM range according to IBA Lifesciences, it is expected that the conjugate should bind to the column. Yet, no visible peak in the chromatogram and no band in the silver-stained SDS-PAGE gel is present. As that is also the case for the two other conjugates, it is highly probable that the two-step purification method is faulty. The generated data is not convincing regarding the fate of the conjugates, as even the denatured with urea products (Strep-Tactin XT allows for the purification of proteins under denaturing conditions according to IBA Lifesciences) do not seem to end up in the elution fraction. A product with the correct apparent molecular weight (185 kDa) is eluted when the denatured F688AzF-IPP2 sample is applied to the Strep-Tactin column. Regarding the F688AzF-DN1S sample, the unknown conjugate appearing at 185 kDa can be hardly detected in the elution step using the HisTrap HP column, while the product of higher molecular weight that appears since day 0 can be found in the sample application in the purification using the Strep-Tactin and at the elution of the denatured product using again the Strep-Tactin column. These findings suggest that the Strep-Tag II is possibly inaccessible, but it cannot be degraded as the protein domains were purified using the same tag. Moreover, the Strep-Tactin column is compatible with an imidazole buffer of maximum 250 mM concentration, while the elution buffer used in this project during the first step of His-Tag purification is of 500 mM, and the eluate was then applied without a buffer exchange to the Strep-Tactin column, another factor possibly affecting the affinity of the Strep-Tag to the column. Again, reversing the order of the chromatography methods did not yield different results, as it was tried for the F688AzF-DN1S.
Future steps to resolve the purification problems would include to study the conjugation of the protein domain with the linker, and determine the stability of the maleimide-thiol bond, as it can degrade during electrophoresis (C. Zhang et al., 2014), and include also a cysteine in the Gem110 domain, so the same linker can be used for reasons of comparison. Performing the conjugation with bigger batches of variant and protein domains, and examining thoroughly if the conjugate is in the elution of the His-Tag purification are also rational steps. Another idea would be to use the Twin Strep-Tag in the protein domains, as it has higher affinity for the StrepTactin XT column according

to IBA Lifesciences. Lastly, it is deemed wise to examine in-depth other cyclooctyne-functionalised linkers (Chio and Bane, 2019), such as BCN, which is more hydrophilic and exhibits better reactivity towards azides compared to DBCO (van Geel et al., 2015), especially towards aromatic azides (Dommerholt et al., 2014).

In conclusion, the data we generated show that an unnatural amino acid can be introduced into the SpCas9 sequence, and the enzyme is functional after the incorporation or more active than native SpCas9, dependent on the incorporation site. Out the produced variants, the F688AzF and the F539AzF ones appear to be better-performing. The best one would be the F539AzF variant based on its higher than native SpCas9 *in vitro* activity, and 75% activity when conjugated with DBCO-AF647. The unnatural amino acid can then be utilized as a reactive handle, in order to conjugate SpCas9 with other molecules using strain promoted alkyne-azide cycloaddition. Further research could consider optimizing the copper-free click chemistry reaction, by examining the reactivity of SpCas9 containing 4-Azido-L-phenylalanine with different cyclooctyne reagents, and finding the ideal molecule to conjugate SpCas9 with, that being possibly a dye for labelling, or for example a small interfering RNA against a protein that promotes NHEJ.

The project succeed in producing four SpCas9 variants with pAzF provenly incorporated (F196pAzF, F539pAzF, F688pAzF, Y1036pAzF), all of them being active and some of them having comparable, if not higher, activity with native SpCas9. These variants can be conjugated with molecules such as a fluorescent dye or a PEG molecule that contain the necessary chemical groups for the click chemistry reaction to occur, but their ability to be bioconjugated with the protein domains remains elusive, as the conjugated could not be purified. Still, the variants can be proven useful in conjugating SpCas9 with other molecules.

Supplementary Materials

Information Relevant to SDM
Primers used for SDM: there was one primer pair for each mutation, and the melting temperature was calculated based on the equation in the Agilent's kit instruction manual. Their attributes are depicted in Table 4.1.
The results of the partial Sanger sequencing for the mutations F539, F688 and Y1036 are depicted in Figure 4.1.

Figure 4.1: Sequencing results analyzed in Benchling, comparing the SpCas9 plasmid and the in-house mutagenized SpCas9 plasmids towards the TAG codon at different positions, showing the relevant for the mutation part of the DNA sequence. The differences between the two sequences are highlighted. (A) The 539 position; (B) The 688 position; (C) The 1036 position

As only 4 out of the 10 mutated SP-Cas9 were produced, different PCR enhancers such as dimethylsulfoxide with 3% final concentration, formamide with 3% final concentration (Karunanathie et al., 2022) and tetramethylammonium chloride with 20 mM final concentration (Kovárová and Dráber, 2000) were also tested, to no avail.
The sequences of the primers used to verify the mutation after SDM can be found in Table 4.2.

Sequences of the sgRNA and HDR template, and Representative Plasmid Maps
Maps of some the plasmids used in this project are included in the following figures Figure 4.2, Figure 4.3, Figure 4.5, and Figure 4.4.

Nonsense Codon Position	Forward Primer Sequence	Reverse Primer Sequence	Melting Temperature	Length
196	GCATTGATCGGGTTTTCTTCCTACAGCTGATTATAGGTTTGC	GCAAACCTATAATCAGCTGTAGGAAGAAAACCCGATCAATGC	78.2°C	42
256	TCTTCTGCCAGGTCCTAATTCGATTTAAAGTTCGGCGTCAGG	CCTGACGCCGAACTTTAAATCGAATTAGGACCTGGCAGAAGA	80.2°C	42
446	GGACCAACGTAATACGGGATCCGCTACGTCAGGATTTTTTCGATT	AATCGAAAAAATCCTGACGTAGCGGATCCCGTATTACGTTGGTCC	79°C	45
539	TTTCTGTTCGCCACTCAGCTAAGCCGGTTTGCGCATACCC	GGGTATGCGCAAACCGGCTTAGCTGAGTGGCGAACAGAAA	79.7°C	40
688	GCAAAGCCGTCCGATTTCAGCTAATCTAGAATCGTTTTGCCTGATTG	CAATCAGGCAAAACGATTCTAGATTAGCTGAAATCGGACGGCTTTGC	79.1°C	47
943	TATCGTTTTCGTCCTATTTCGTGTTCATGCGTGAATCC	GGATTCACGCATGAACACGAAATAGGACGAAAACGATA	78.4°C	38
1010	GACTTTGTAATCGCCCTACACGAATTCGCTTTCCAGTTTCGGATA	TATCCGAAACTGGAAAGCGAATTCGTGTAGGGCGATTACAAAGTC	80.3°C	45
1036	CATGATGTTTGAATAGAAAAACTATTTGGCGGTCGCTTTGCC	GGCAAAGCGACCGCCAAATAGTTTTTCTATTCAAACATCATG	79.6°C	42
1237	AGTTTTTCATAATGGCTTGCCAGCTACAGAAAGTTAACATATTTACTCGG	CCGAGTAAATATGTTAACTTTCTGTAGCTGGCAAGCCATTATGAAAAACT	78.8°C	50

Table 4.1: Characteristics of Primers Used in Site-Directed Mutagenesis

Nonsense Codon Position	Sequencing Primer Sequence
196	GCCCAGTAGTAGGTTGAGGC
256	GGTTGGGCAGTGATTACCGA
446	TGACGATCTGGACAACCTGC
539	TGACGATCTGGACAACCTGC
688	AGTCAGACAGGCGGTTGATG
943	CAGACCGTGAAAGTCGTGGA
1010	CGAAAACACCCAGCTGCAAA
1036	CGAAAACACCCAGCTGCAAA
1237	CCGTTTAGAGGCCCCAAGG
1258	ATTTAAAAGCTGCCGGTGCG

Table 4.2: Sequence of the Primers Used for Partial Sanger Sequencing of the Mutated Plasmid of SpCas9

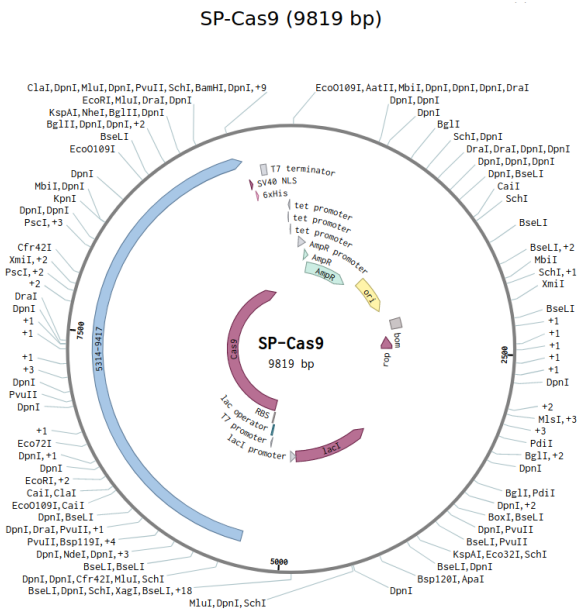

Figure 4.2: Plasmid map of the SP-Cas9 plasmid, which encodes the SpCas9

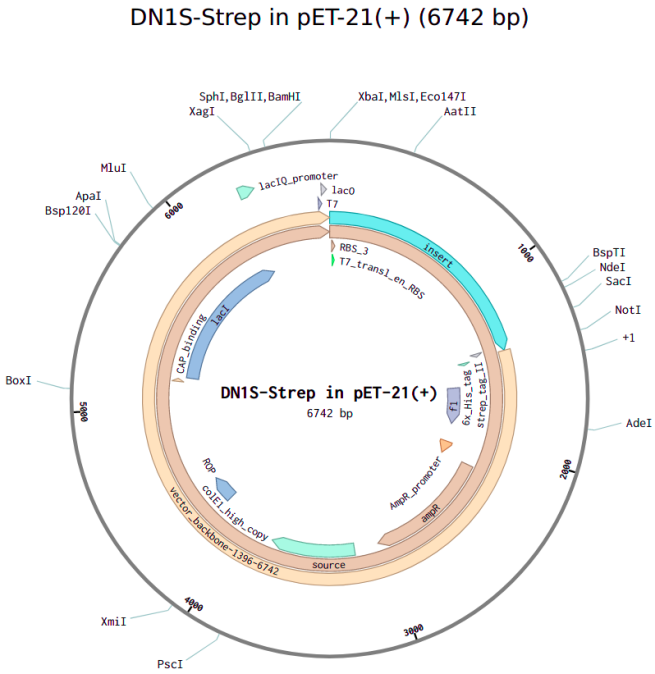

Figure 4.3: Plasmid map containing the DN1S, which encodes a dominant-negative mutant of human 53BP1

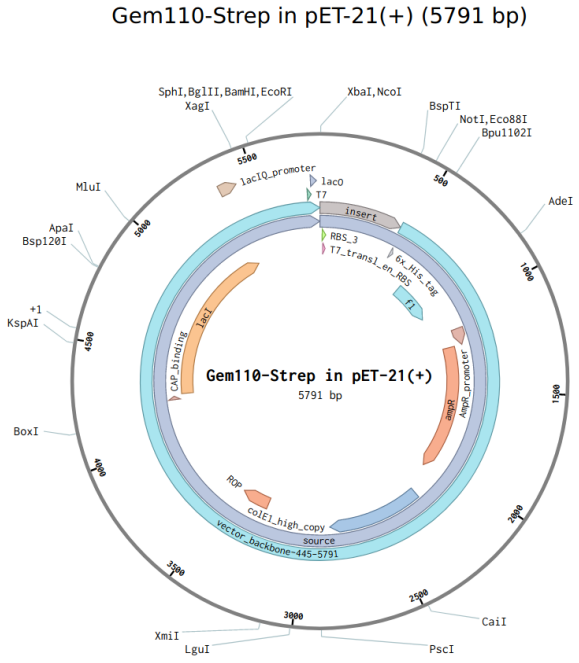

Figure 4.4: Plasmid map containing the Gem110, which encodes the first 110 amino acids from the N-terminal region of human Geminin

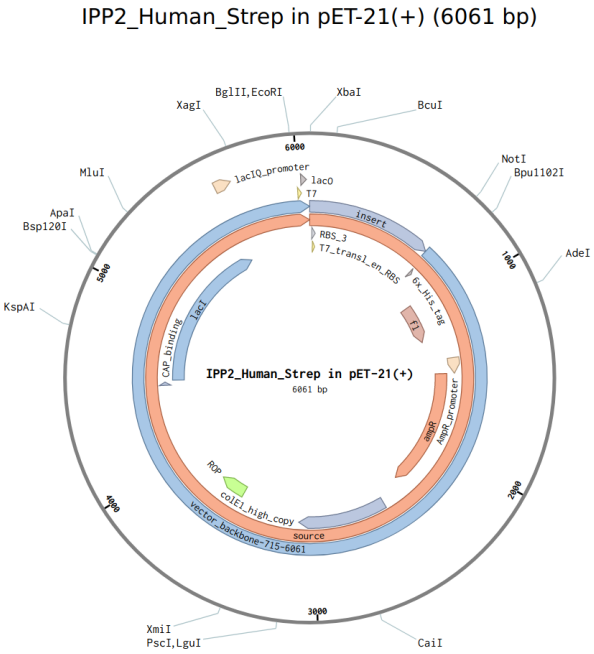

Figure 4.5: Plasmid map containing the IPP2, which encodes the full length of the human protein phosphatase inhibitor-2

The sequence of the sgRNA and the HDR template used during this project can be found in Table 4.3 and Table 4.4 respectively.

Target	20 nt Spacer Sequence
EGFP	GCUGAAGCACUGCACGCCGU

Table 4.3: Sequence of the guide RNA used in the experimental procedures

HDR Template	
86 bp (40bp homology arms)	CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGAGC CACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC

Table 4.4: Template DNA, referred to as HDR template, used in the EGFP to BFP mutation assay. The PAM sequence, needed for SpCas9 activity, is additionally mutated in this sequence

SpCas9 Variants Production

As mentioned in section 2.3., the Bradford assay was used to determine the concentration of the SpCas9 (variants). The concentration of the SpCas9 variants used in Figure 3.1, Figure 3.4, Figure 3.5 and Figure 3.6 was calculated based on the following curve depicted in Figure 4.6, and can be seen in Table 4.5, while the concentration of the SpCas9 variants used in Figure 3.2 and Figure 3.3 can be found at the section 3.2..

SpCas9 Variant	Concentration (mg/mL)
F196AzF	0.7172613
F539AzF	0.9138021
F688AzF	2.2814323
Y1036AzF	1.5196904

Table 4.5: Concentration of the SpCas9 Variants used to produce the results depicted in this report, beside the LNP formulation

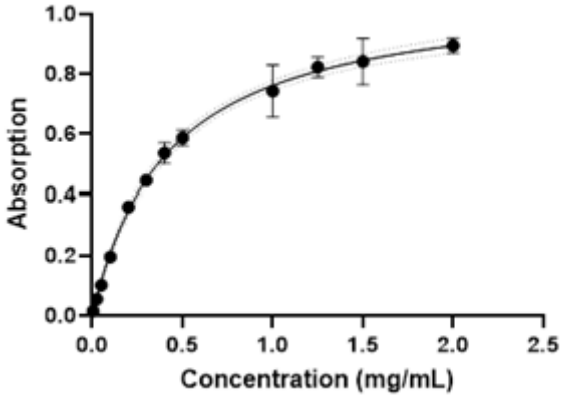

Figure 4.6: Bradford calibration curve. The error bars represent the standard deviation between the different measurements of the same sample.

To verify that the native SpCas9 was not produced by the mutated strains of BL21 *E. coli*, those strains were cultured as described in section 2.2. and normally induced with IPTG but the addition of 4-Azido-L-phenylalanine was skipped. Western blot was performed to detect the SpCas9, as described in section 2.3. with the exception of using a primary antibody (CRISPR/Cas9 Monoclonal Antibody [7A9] (Catalog #A-9000) from EpiGentek, Uden, the Netherlands, 1:300 dilution in 3% milk) left to incubate with agitation overnight at 4°C and a secondary antibody (Goat Anti-Mouse IgG H&L (HRP) (ab97023) from Abcam (Amsterdam, the Netherlands), 1:10.000 dilution in 3% milk) left to incubate with agitation for 1 hour at room temperature. The results can be seen in Figure 4.7, in which no band at the height of the SpCas9 is present in the lysates, indicating that the *E. coli* transformed with the mutated SP-Cas9 plasmids do not produce the native SpCas9.

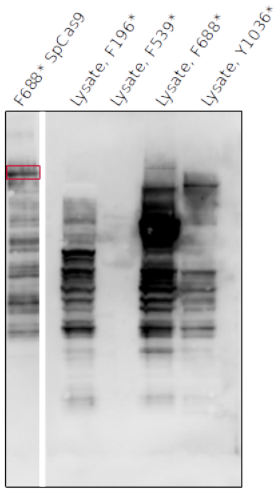

Figure 4.7: Western Blot of the lysates. From left to right, the samples are: 10 μg control SpCas9, 5 μL lysate of the BL21 *E. coli* containing the F196AzF SP-Cas9, 5 μL lysate of the BL21 *E. coli* containing the F539AzF SP-Cas9, 5 μL lysate of the BL21 *E. coli* containing the F688AzF SP-Cas9, 5 μL lysate of the BL21 *E. coli* containing the F1036AzF SP-Cas9. SpCas9 is marked with the red box in the control sample.

SpCas9 Variants PEGylation

The SDS-PAGE gels of all the PEGylation conditions mentioned in section 2.4. are included in Figure 4.8. The bands appearing only at the samples where sgRNA is added are due to bovine serum

albumin (66.4 kDa) and RiboLock RNase inhibitor (49.6 kDa), which are included in the buffers used concomitantly with sgRNA.
Another experiment of smaller scale was done using the bicyclononyne (BCN) as the click chemistry reagent (Canalle et al., 2011). More specifically, the endo-BCN-PEG (5 kDa) was produced in-house and gifted to us by Stefania Douka. This experiment was mainly done to compare the reactions kinetics and reaction yield between the two PEGx-molecules, and the conditions tested were 37°C, 6 to 30 hours, and 10:1 to 60:1 PEG:Variant ratio. The downstream analysis of the samples was the same (SDS-PAGE, etc.), and the only difference would be the polyacrylamide gel: the electrophoresis was done using 6% polyacrylamide 1.0 mm handcast gels and tris-glycine as running buffer, for 2 hours under 100 volt. The lowest conjugation efficiency was approximately 30% at 6 hours of reaction and 10:1 molar ratio of PEG to protein, while increasing the time and the molar ration led to a possible conversion of 80%. The bands in the SDS-PAGE gels are not easily interpreted and quantifiable due to the products producing smeared bands, as observed when PEG is run in an SDS-PAGE gel (Zheng et al., 2007).

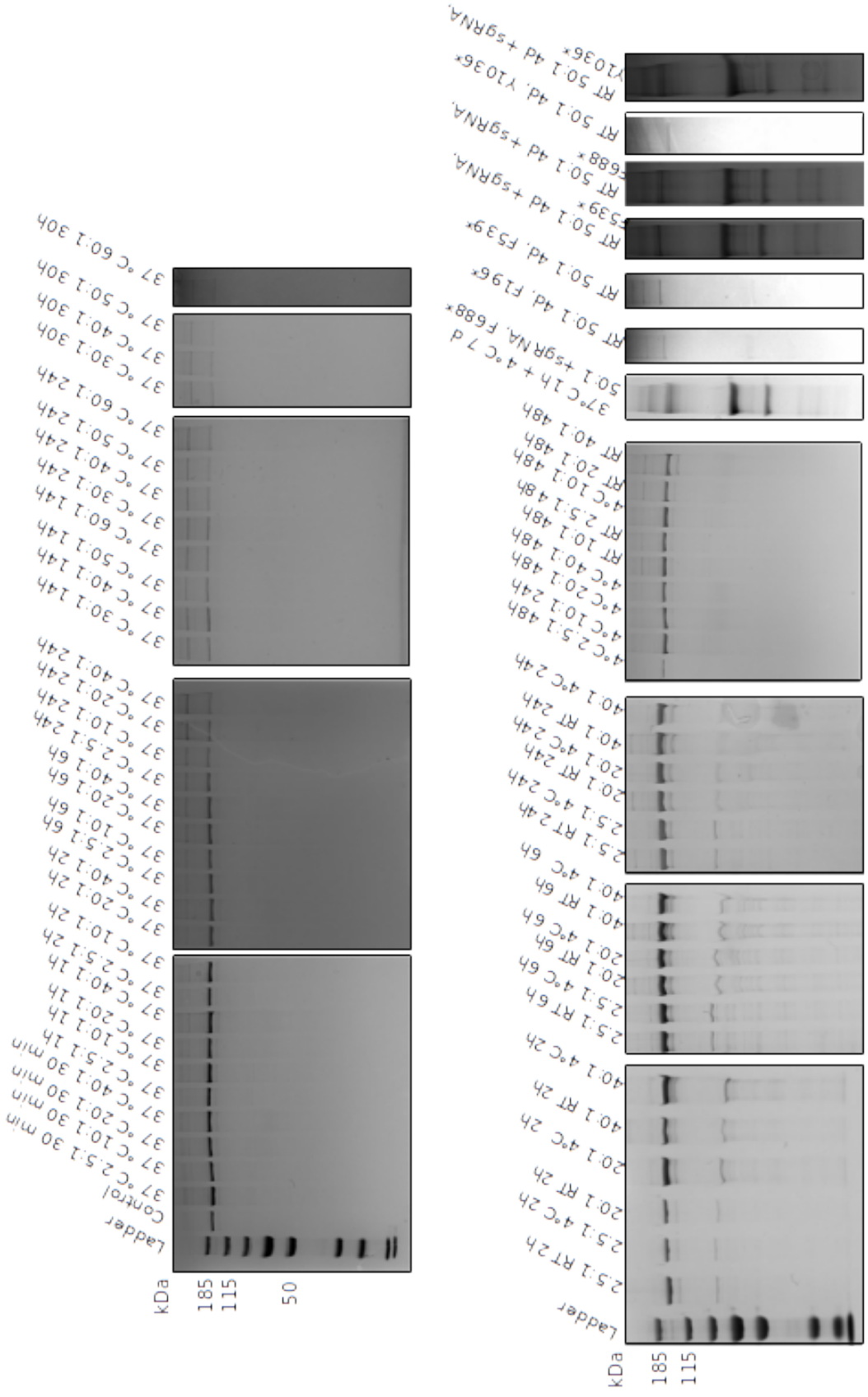

Figure 4.8: SDS-PAGE gels of the PEGylation conditions depicted in Figure 3.3 and Figure 3.4. Unless otherwise labelled, all the samples regard the F688AzF SpCas9 variant and DBCO-mPEG (30 kDa).

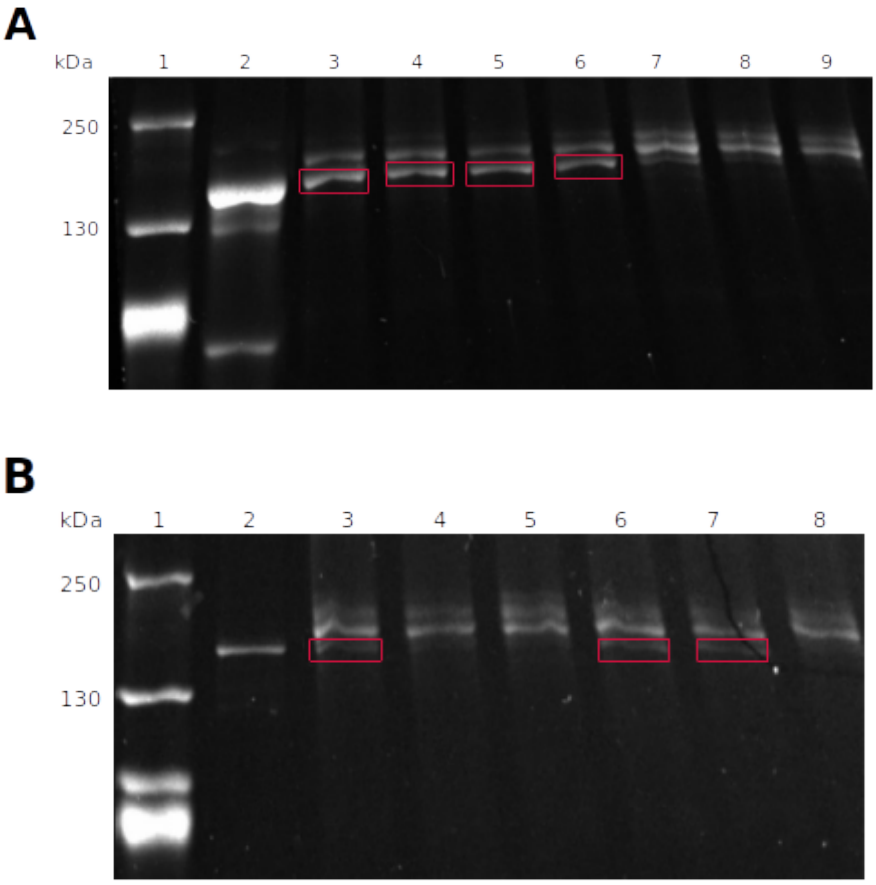

Figure 4.9: SDS-PAGE gels of the F688*AzF SpCas9 conjugated with endo-BCN-PEG of 5 kDa. Highlighted are the bands of unconjugated F688AzF SpCas9 (A) Gel with the products of 6 and 14 hours reaction. During electrophoresis, the loading buffer appeared to run in an inclined line instead of horizontically (1) Ladder; (2) 0.9 μg F688AzF SpCas9; (3) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 10:1 PEG to protein ratio, 6h; (4) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 20:1, 6h; (5) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 40:1, 6h; (6) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 60:1, 6h; (7) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 10:1, 14h; (8) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 20:1, 14h; (9) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 40:1, 14h; (B) Gel with the products of 24 and 30 hours reaction (1) Ladder; (2) 0.3 μg Y1036AzF SpCas9; (3) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 20:1 PEG to protein ratio, 24h; (4) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 40:1, 24h; (5) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 60:1, 24h; (6) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 10:1, 30h; (7) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 20:1, 30h; (8) Products of F688AzF SpCas9 and endo-BCN-PEG, reacted under 37°C, 40:1, 30h

SpCas9 Variants Activity

All the agarose gels analyzed to produce the *in vitro* relative activity results included in Figure 3.2 and Figure 3.4 can be found in Figure 4.10.

Representative Chromatogram for Strep Tag Purification

As can be seen in Figure 4.11, the purification method using the Strep-Tactin XT 4Flow column to isolate the IPP2 domain produced a clear peak, indicating that the Strep Tag is functional and accessible in the unconjugated domains.

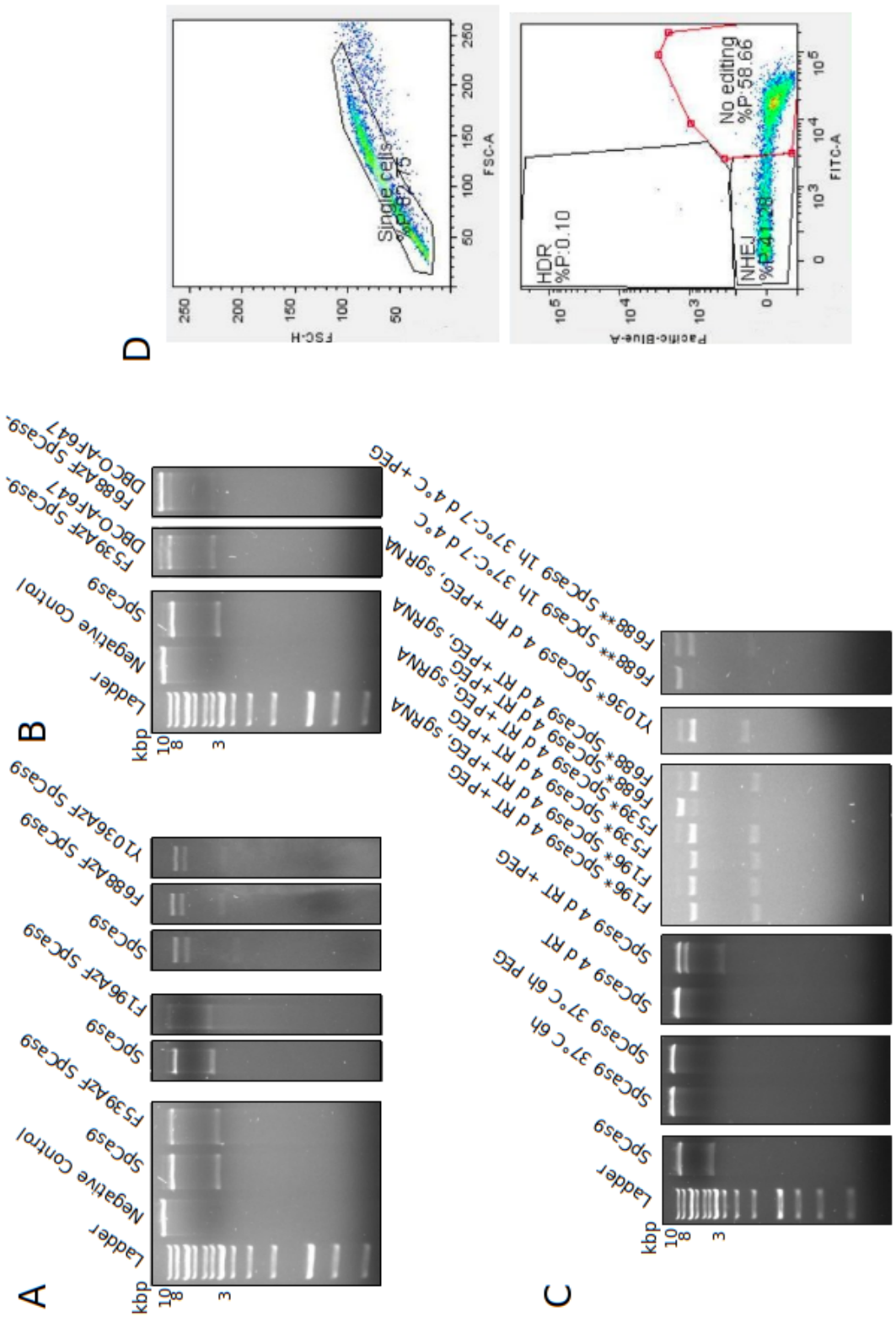

Figure 4.10: (A) Agarose gels of the *in vitro* activity assay done using the SpCas9 variants. The control used for each separate assay is grouped together with the corresponding tested variant; (B) Agarose gels of the *in vitro* activity assay done using the SpCas9 variants conjugated with DBCO-AF647; (C) Agarose gels of the *in vitro* activity assay done using the SpCas9 variants conjugated with PEG, and the plain SpCas9 under the conjugation conditions. The control used for each separate assay is grouped together with the corresponding tested variant; (D) The gating strategy used in flow cytometry, where the upper figure represents the single cells and the lower one the gating method.

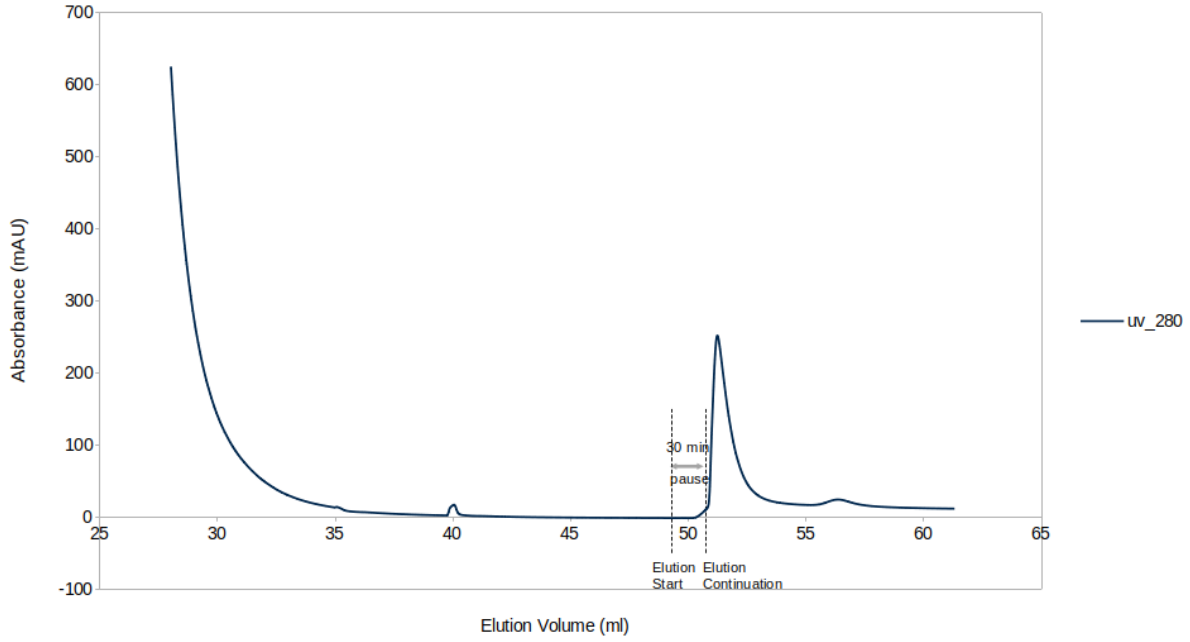

Figure 4.11: Chromatogram of the IPP2 purification, using the Strep-Tactin XT 4Flow column and the Äkta PURE chromatography

Conjugation of the F688AzF SpCas9 Variant with the IPP2 and the Gem110
The SDS-PAGE gels of the reaction kinetics between F688AzF SpCas9 and conjugated with the PEG linker IPP2 and Gem110, analyzed to produce the results in Figure 3.5, are included in Figure 4.12.

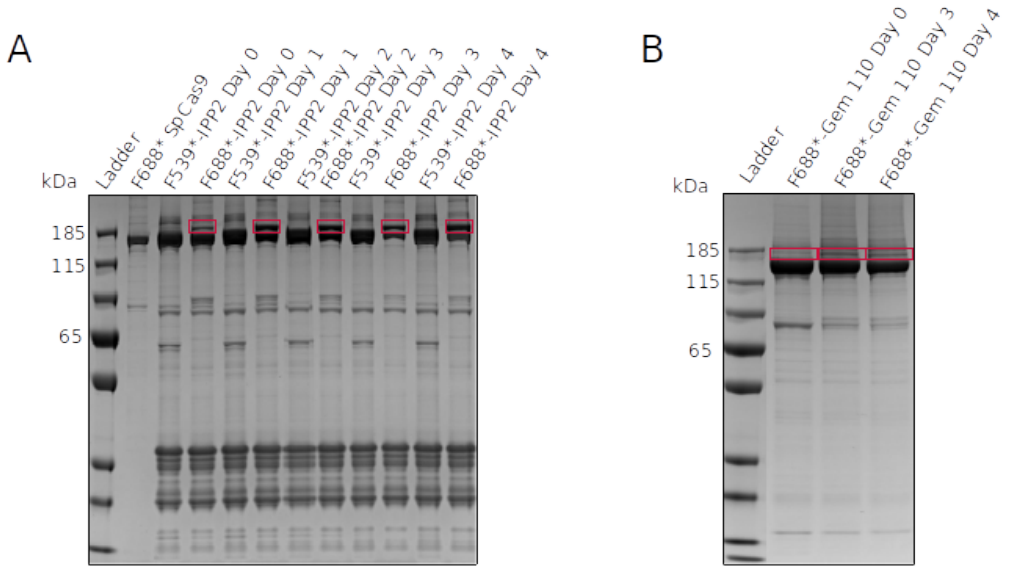

Figure 4.12: (A) SDS-PAGE gel of the products of the reaction between F539AzF or F688AzF SpCas9 with IPP2, from day 0 until dat 4. Highlighted is the conjugate between F688AzF SpCas9 and IPP2, with a DBCO-PEG24-maleimide linker; (B) SDS-PAGE gel of the products of the reaction between F688AzF SpCas9 and Gem110, on day 0, 3 and 4. Highlighted is the conjugate between F688AzF SpCas9 and Gem110, with a DBCO-PEG24-NHS linker

Bibliography

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04252-2

Agard, N. J., Prescher, J. A., & Bertozzi, C. R. (2004). A strain-promoted [3 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society*, *126*(46), 15046–15047. https://doi.org/10.1021/ja044996f

Anderson, G. W., Zimmerman, J. E., & Callahan, F. M. (1963). Bn-hydroxysuccinimide esters in peptide synthesis/b. *Journal of the American Chemical Society*, *85*(19), 3039–3039. https://doi.org/10.1021/ja00902a047

Bak, R. O., Gomez-Ospina, N., & Porteus, M. H. (2018). Gene editing on center stage. *Trends in Genetics*, *34*(8), 600–611. https://doi.org/10.1016/j.tig.2018.05.004

Barrangou, R. (2014). Cas9 targeting and the CRISPR revolution. *Science*, *344*(6185), 707–708. https://doi.org/10.1126/science.1252964

Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A., & Bertozzi, C. R. (2007). Copper-free click chemistry for dynamic iin vivo/i imaging. *Proceedings of the National Academy of Sciences*, *104*(43), 16793–16797. https://doi.org/10.1073/pnas.0707090104

Batjargal, S., Walters, C. R., & Petersson, E. J. (2015). Inteins as traceless purification tags for unnatural amino acid proteins. *Journal of the American Chemical Society*, *137*(5), 1734–1737. https://doi.org/10.1021/ja5103019

Bibikova, M., Beumer, K., Trautman, J. K., & Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science*, *300*(5620), 764–764. https://doi.org/10.1126/science.1079512

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J., Weber, L. M., Catena, R., Jinek, M., Robinson, M. D., & Mosimann, C. (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development*, *143*(11), 2025–2037.

Canalle, L. A., Vong, T., Adams, P. H. H. M., van Delft, F. L., Raats, J. M. H., Chirivi, R. G. S., & van Hest, J. C. M. (2011). Clickable enzyme-linked immunosorbent assay. *Biomacromolecules*, *12*(10), 3692–3697. https://doi.org/10.1021/bm2009137

Carey, M. F., Peterson, C. L., & Smale, S. T. (2013). PCR-mediated site-directed mutagenesis. *Cold Spring Harbor Protocols*, *2013*(8), pdb.prot076505. https://doi.org/10.1101/pdb.prot076505

Chen, H., Choi, J., & Bailey, S. (2014). Cut site selection by the two nuclease domains of the cas9 RNA-guided endonuclease. *Journal of Biological Chemistry*, *289*(19), 13284–13294. https://doi.org/10.1074/jbc.m113.539726

Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of p-azido-l-phenylalanine to the genetic code of escherichia coli. *J. Am. Chem. Soc.*, *124*(31), 9026–9027.

Chio, T. I., & Bane, S. L. (2019). Click chemistry conjugations. *Methods in molecular biology* (pp. 83–97). Springer US. https://doi.org/10.1007/978-1-4939-9929-3_6

Cook, B. E., Adumeau, P., Membreno, R., Carnazza, K. E., Brand, C., Reiner, T., Agnew, B. J., Lewis, J. S., & Zeglis, B. M. (2016). Pretargeted PET imaging using a site-specifically labeled immunoconjugate. *Bioconjugate Chemistry*, *27*(8), 1789–1795. https://doi.org/10.1021/acs.bioconjchem.6b00235

Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P. J. T., van Hest, J. C. M., & van Delft, F. L. (2010). Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (32) cycloaddition. *Chem. Commun.*, *46*(1), 97–99. https://doi.org/10.1039/b917797c

Dommerholt, J., van Rooijen, O., Borrmann, A., Guerra, C. F., Bickelhaupt, F. M., & van Delft, F. L. (2014). Highly accelerated inverse electron-demand cycloaddition of electron-deficient azides with aliphatic cyclooctynes. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms6378

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-cas9. *Science*, *346*(6213). https://doi.org/10.1126/science.1258096

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39). https://doi.org/10.1073/pnas.1208507109

Glaser, A., McColl, B., & Vadolas, J. (2016). GFP to BFP conversion: A versatile assay for the quantification of CRISPR/Cas9-mediated genome editing. *Mol. Ther. Nucleic Acids*, *5*(7), e334.

Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F., & Chin, L. (2016). Post-translational regulation of cas9 during G1 enhances homology-directed repair. *Cell Rep.*, *14*(6), 1555–1566.

Jang, S., Sachin, K., Lee, H.-j., Kim, D. W., & Lee, H. S. (2012). Development of a simple method for protein conjugation by copper-free click reaction and its application to antibody-free western blot analysis. *Bioconjugate Chemistry*, *23*(11), 2256–2261. https://doi.org/10.1021/bc300364z

Jayavaradhan, R., Pillis, D. M., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P. R., & Malik, P. (2019). CRISPR-cas9 fusion to dominant-negative 53bp1 enhances HDR and inhibits NHEJ specifically at cas9 target sites. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10735-7

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., & Doudna, J. A. (2014). Structures of cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, *343*(6176). https://doi.org/10.1126/science.1247997

Jong, O. G., Balkom, B. W. M., Gremmels, H., & Verhaar, M. C. (2016). Exosomes from hypoxic endothelial cells have increased collagen crosslinking activity through up-regulation of lysyl oxidase-like 2. *Journal of Cellular and Molecular Medicine*, *20*(2), 342–350. https://doi.org/10.1111/jcmm.12730

Kakinoki, Y., Somers, J., & Brautigan, D. L. (1997). Multisite phosphorylation and the nuclear localization of phosphatase inhibitor 2-green fluorescent protein fusion protein during s phase of the cell growth cycle. *Journal of Biological Chemistry*, *272*(51), 32308–32314. https://doi.org/10.1074/jbc.272.51.32308

Karimian, A., Azizian, K., Parsian, H., Rafieian, S., Shafiei-Irannejad, V., Kheyrollah, M., Yousefi, M., Majidinia, M., & Yousefi, B. (2019). CRISPR/cas9 technology as a potent molecular tool for gene therapy. *Journal of Cellular Physiology*, *234*(8), 12267–12277. https://doi.org/10.1002/jcp.27972

Karunanathie, H., Kee, P. S., Ng, S. F., Kennedy, M. A., & Chua, E. W. (2022). PCR enhancers: Types, mechanisms, and applications in long-range PCR. *Biochimie*, *197*, 130–143.

Kim, C. H., Axup, J. Y., & Schultz, P. G. (2013). Protein conjugation with genetically encoded unnatural amino acids. *Current Opinion in Chemical Biology*, *17*(3), 412–419. https://doi.org/10.1016/j.cbpa.2013.04.017

Kim, S., Ko, W., Sung, B. H., Kim, S. C., & Lee, H. S. (2016). Direct protein–protein conjugation by genetically introducing bioorthogonal functional groups into proteins. *Bioorganic &amp Medicinal Chemistry*, *24*(22), 5816–5822. https://doi.org/10.1016/j.bmc.2016.09.035

Kolb, H. C., Finn, M. G., & Sharpless, K. B. (2001). Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie International Edition*, *40*(11), 2004–2021. https://doi.org/10.1002/1521-3773(20010601)40:11<2004::aid-anie2004>3.0.co;2-5

Kovárová, M., & Dráber, P. (2000). New specificity and yield enhancer of polymerase chain reactions. *Nucleic Acids Res.*, *28*(13), E70.

Kurfürst, M. M. (1992). Detection and molecular weight determination of polyethylene glycol-modified hirudin by staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Analytical Biochemistry*, *200*(2), 244–248. https://doi.org/10.1016/0003-2697(92)90460-o

Lee, J. K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.-h., Lee, K., Jung, I., Kim, D., Kim, S., & Kim, J.-S. (2018). Directed evolution of CRISPR-cas9 to increase its specificity. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05477-x

Lobanov, M. Y., Furletova, E. I., Bogatyreva, N. S., Roytberg, M. A., & Galzitskaya, O. V. (2010). Library of disordered patterns in 3d protein structures (I. Simon, Ed.). *PLoS Computational Biology*, *6*(10), e1000958. https://doi.org/10.1371/journal.pcbi.1000958

Lühmann, T., Gutmann, M., Moscaroli, A., Raschig, M., Béhé, M., & Meinel, L. (2019). Biodistribution of site-specific PEGylated fibroblast growth factor-2. *ACS Biomaterials Science &amp Engineering*, *6*(1), 425–432. https://doi.org/10.1021/acsbiomaterials.9b01248

Memi, F., Ntokou, A., & Papangeli, I. (2018). CRISPR/cas9 gene-editing: Research technologies, clinical applications and ethical considerations. *Seminars in Perinatology*, *42*(8), 487–500. https://doi.org/10.1053/j.semperi.2018.09.003

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D., & Rebar, E. J. (2010). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, *29*(2), 143–148. https://doi.org/10.1038/nbt.1755

Ning, X., Guo, J., Wolfert, M., & Boons, G.-J. (2008). Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angewandte Chemie International Edition*, *47*(12), 2253–2255. https://doi.org/10.1002/anie.200705456

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., & Nureki, O. (2014). Crystal structure of cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001

Nwe, K., & Brechbiel, M. W. (2009). Growing applications of "click chemistry" for bioconjugation in contemporary biomedical research. *Cancer Biotherapy and Radiopharmaceuticals*, *24*(3), 289–302. https://doi.org/10.1089/cbr.2008.0626

Oakes, B. L., Nadler, D. C., Flamholz, A., Fellmann, C., Staahl, B. T., Doudna, J. A., & Savage, D. F. (2016). Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.*, *34*(6), 646–651.

Oakes, B. L., Nadler, D. C., & Savage, D. F. (2014). Protein engineering of cas9 for enhanced function. *Methods in enzymology* (pp. 491–511). Elsevier. https://doi.org/10.1016/b978-0-12-801185-0.00024-6

Odom, O. W., Kudlicki, W., Kramer, G., & Hardesty, B. (1997). An effect of polyethylene glycol 8000 on protein mobility in sodium dodecyl sulfate–polyacrylamide gel electrophoresis and a method for eliminating this effect. *Analytical Biochemistry*, *245*(2), 249–252. https://doi.org/10.1006/abio.1996.9993

Palermo, G., Miao, Y., Walker, R. C., Jinek, M., & McCammon, J. A. (2017). CRISPR-cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences*, *114*(28), 7260–7265. https://doi.org/10.1073/pnas.1707645114

Pickens, C. J., Johnson, S. N., Pressnall, M. M., Leon, M. A., & Berkland, C. J. (2017). Practical considerations, challenges, and limitations of bioconjugation via azide–alkyne cycloaddition. *Bioconjugate Chemistry*, *29*(3), 686–701. https://doi.org/10.1021/acs.bioconjchem.7b00633

Ptacin, J. L., Caffaro, C. E., Ma, L., Gall, K. M. S. J., Aerni, H. R., Acuff, N. V., Herman, R. W., Pavlova, Y., Pena, M. J., Chen, D. B., Koriazova, L. K., Shawver, L. K., Joseph, I. B., & Milla, M. E. (2021). An engineered IL-2 reprogrammed for anti-tumor therapy using a semi-synthetic organism. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24987-9

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-cas9 system. *Nature Protocols*, *8*(11), 2281–2308. https://doi.org/10.1038/nprot.2013.143

Ravasco, J. M. J. M., Faustino, H., Trindade, A., & Gois, P. M. P. (2018). Bioconjugation with maleimides: A useful tool for chemical biology. *Chemistry – A European Journal*, *25*(1), 43–59. https://doi.org/10.1002/chem.201803174

Riggs, P. (2013). Fusion protein. *Brenner's encyclopedia of genetics* (pp. 134–135). Elsevier. https://doi.org/10.1016/b978-0-12-374984-0.00565-9

Rouet, P., Smih, F., & Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology*, *14*(12), 8096–8106. https://doi.org/10.1128/mcb.14.12.8096-8106.1994

Rudin, N., Sugarman, E., & Haber, J. E. (1989). Genetic and physical analysis of double-strand break repair and recombination in saccharomyces cerevisiae. *Genetics*, *122*(3), 519–534. https://doi.org/10.1093/genetics/122.3.519

Salsman, J., Masson, J.-Y., Orthwein, A., & Dellaire, G. (2018). CRISPR/cas9 gene editing: From basic mechanisms to improved strategies for enhanced genome engineering in vivo. *Current Gene Therapy*, *17*(4). https://doi.org/10.2174/1566523217666171122094629

Sharma, G., Sharma, A. R., Bhattacharya, M., Lee, S.-S., & Chakraborty, C. (2021). CRISPR-cas9: A preclinical and clinical perspective for the treatment of human diseases. *Molecular Therapy*, *29*(2), 571–586. https://doi.org/10.1016/j.ymthe.2020.09.028

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, *7*(1), 539. https://doi.org/10.1038/msb.2011.75

Simon, M., Stefan, N., Borsig, L., Plückthun, A., & Zangemeister-Wittke, U. (2014). Increasing the antitumor effect of an EpCAM-targeting fusion toxin by facile click PEGylation. *Molecular Cancer Therapeutics*, *13*(2), 375–385. https://doi.org/10.1158/1535-7163.mct-13-0523

Sun, W., Liu, H., Yin, W., Qiao, J., Zhao, X., & Liu, Y. (2022). Strategies for enhancing the homology-directed repair efficiency of CRISPR-cas systems. *The CRISPR Journal*, *5*(1), 7–18. https://doi.org/10.1089/crispr.2021.0039

Tacal, O., & Ozer, I. (2002). A comparison between SDS-PAGE and size exclusion chromatography as analytical methods for determining product composition in protein conjugation reactions. *Journal of Biochemical and Biophysical Methods*, *52*(3), 161–168. https://doi.org/10.1016/s0165-022x(02)00070-2

Turro, N. J., Lei, X.-G., Ananthapadmanabhan, K. P., & Aronson, M. (1995). Spectroscopic probe analysis of protein-surfactant interactions: The BSA/SDS system. *Langmuir*, *11*(7), 2525–2533. https://doi.org/10.1021/la00007a035

van Geel, R., Wijdeven, M. A., Heesbeen, R., Verkade, J. M. M., Wasiel, A. A., van Berkel, S. S., & van Delft, F. L. (2015). Chemoenzymatic conjugation of toxic payloads to the globally conserved n-glycan of native mAbs provides homogeneous and highly efficacious antibody–drug conjugates. *Bioconjugate Chemistry*, *26*(11), 2233–2242. https://doi.org/10.1021/acs.bioconjchem.5b00224

van Moorsel, M. V., Urbanus, R. T., Verhoef, S., Koekman, C., Vink, M., Vermonden, T., Maas, C., Pasterkamp, G., & Schiffelers, R. M. (2019). A head-to-head comparison of conjugation methods for VHHs: Random maleimide-thiol coupling versus controlled click chemistry. *International Journal of Pharmaceutics: X*, *1*, 100020. https://doi.org/10.1016/j.ijpx.2019.100020

Villela, S. M. A., Kraïem, H., Bouhaouala-Zahar, B., Bideaux, C., Lara, C. A. A., & Fillaudeau, L. (2020). A protocol for recombinant protein quantification by densitometry. *MicrobiologyOpen*, *9*(6), 1175–1182. https://doi.org/10.1002/mbo3.1027

Walther, J., Wilbie, D., Tissingh, V. S. J., Öktem, M., van der Veen, H., Lou, B., & Mastrobattista, E. (2022). Impact of formulation conditions on lipid nanoparticle characteristics and functional delivery of CRISPR RNP for gene knock-out and correction. *Pharmaceutics*, *14*(1), 213.

Winogradoff, D., John, S., & Aksimentiev, A. (2020). Protein unfolding by SDS: The microscopic mechanisms and the properties of the SDS-protein assembly. *Nanoscale*, *12*(9), 5422–5434. https://doi.org/10.1039/c9nr09135a

Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J., & Xia, N. (2020). Methods favoring homology-directed repair choice in response to CRISPR/cas9 induced-double strand breaks. *International Journal of Molecular Sciences*, *21*(18), 6461. https://doi.org/10.3390/ijms21186461

Zambrano, G., Chino, M., Renzi, E., Girolamo, R. D., Maglio, O., Pavone, V., Lombardi, A., & Nastri, F. (2020). Clickable artificial heme-peroxidases for the development of functional nanomaterials. *Biotechnology and Applied Biochemistry*, *67*(4), 549–562. https://doi.org/10.1002/bab.1969

Zhang, B., Sun, J., Wang, Y., Ji, D., Yuan, Y., Li, S., Sun, Y., Hou, Y., Li, P., Zhao, L., Yu, F., Ma, W., Cheng, B., Wu, L., Hu, J., Wang, M., Song, W., Li, X., Li, H., … Zhang, X. (2021). Site-specific PEGylation of interleukin-2 enhances immunosuppression via the sustained activation of regulatory t

cells. *Nature Biomedical Engineering*, *5*(11), 1288–1305. https://doi.org/10.1038/s41551-021-00797-8

Zhang, C., Liu, Y., Feng, C., Wang, Q., Shi, H., Zhao, D., Yu, R., & Su, Z. (2014). Loss of PEG chain in routine SDS-PAGE analysis of PEG-maleimide modified protein. *ELECTROPHORESIS*, *36*(2), 371–374. https://doi.org/10.1002/elps.201400373

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., & Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnology*, *29*(2), 149–153. https://doi.org/10.1038/nbt.1775

Zheng, C. Y., Ma, G., & Su, Z. (2007). Native PAGE eliminates the problem of PEG–SDS interaction in SDS-PAGE and provides an alternative to HPLC in characterization of protein PEGylation. *ELECTROPHORESIS*, *28*(16), 2801–2807. https://doi.org/10.1002/elps.200600807