UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Medicine and Life Sciences
School for Life Sciences
Master of Biomedical Sciences

Master's thesis

Site-specific fluorescent labeling of human DNA ligases for single-molecule investigation of protein dynamics and substrate binding modes

Lambert-Paul Jorissen

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Bioelectronics and Nanotechnology

SUPERVISOR :

Prof. dr. Jelle HENDRIX

MENTOR :

De heer Tom KACHE

2022
2023

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Faculty of Medicine and Life Sciences
School for Life Sciences
Master of Biomedical Sciences

Master's thesis

Site-specific fluorescent labeling of human DNA ligases for single-molecule investigation of protein dynamics and substrate binding modes

Lambert-Paul Jorissen

Thesis presented in fulfillment of the requirements for the degree of Master of Biomedical Sciences, specialization Bioelectronics and Nanotechnology

SUPERVISOR :

Prof. dr. Jelle HENDRIX

MENTOR :

De heer Tom KACHE

Site-specific fluorescent labeling of human DNA ligases for single-molecule investigation of protein dynamics and substrate binding modes

Lambert-Paul Jorissen[1], Tom Kache[1] and Prof. dr. Jelle Hendrix[1]

[1]Dynamic Bioimaging research group, Biomedical Research Institute, Universiteit Hasselt, Campus Diepenbeek, Agoralaan Gebouw C - B-3590 Diepenbeek

*Running title: ***Site-specific fluorescent labeling of DNA ligases***

To whom correspondence should be addressed: prof. dr. Jelle Hendrix, Tel: +32 (11) 26 92 13; Email: jelle.hendrix@uhasselt.be

Keywords: DNA ligase, single-molecule, FRET, fluorescent labeling

ABSTRACT

DNA is vulnerable to strand breaks when exposed to DNA-damaging agents. Among the groups of enzymes involved in DNA replication and repair, DNA ligases bear the crucial role of (re)connecting the broken strands. The abnormal activity of these enzymes is associated with various cancers and immunodeficiency syndromes, making them potential therapeutic targets for small-molecule drugs. Therefore, understanding the structure-function relationship of DNA ligases is crucial for developing effective treatments. Although previous studies have unveiled the complex structural plasticity of the enzyme, the lack of real-time protein dynamics information has compromised the establishment of a direct link to enzymatic activity. Single-molecule Förster resonance energy transfer (smFRET) measurements offer a high-resolution approach to studying intra-molecular protein dynamics by measuring the distance between fluorescent molecules on a protein surface. To enable smFRET studies of DNA ligase activity, site-specific fluorescent labeling of the protein while preserving functional integrity is essential. In this research, purified DNA ligase was fluorescently labeled and characterized with ligation assays and fluorescence correlation spectroscopy. The purified enzyme demonstrated functional integrity through successful ligation of nicked DNA and reduced DNA substrate diffusion rates. Initial labeling reactions with a fluorescent dye indicated attachment to the protein surface. However, further optimization of the purification and labeling process is necessary to minimize potential non-site-specific binding of the dye and aggregate formation. This research presents preliminary evidence for purifying and labeling human DNA ligases for future smFRET studies and will help advance our understanding of their structure, function, and therapeutic potential.

INTRODUCTION

The continuation of life depends on maintaining the genetic sequence information of living organisms. While being the fundamental template encoding all genetic information, the genome is remarkably unstable [1, 2]. Due to the prevalence of DNA-damaging agents such as ultraviolet (UV) light from the sun and reactive oxygen species (ROS) from cellular metabolism, up to 10^5 DNA lesions form in an active mammalian cell every day [3]. DNA double-strand breaks (DSBs), arguably the most dangerous lesions, can cause chromosomal aberrations or even cell death due to the loss of DNA duplex integrity [2, 4, 5]. Hence, sophisticated **DNA repair mechanisms** have evolved to collectively function and minimize the negative effects of DNA damage. However, abnormalities in these maintenance mechanisms not only predispose cells to cancer but also support multiple forms of premature aging in humans [2]. Targeting enzymes involved in DNA repair defects is a promising therapy for various diseases and is still in development [6-12]. One of the most notable groups of enzymes in the DNA repair repertoire is the **DNA ligase** family. These essential proteins are found in all life forms and bear the crucial task of

connecting Okazaki fragments during DNA replication, as well as sealing strand breaks at the end of most DNA repair pathways [13-16]. Besides their importance in modern *in vitro* biochemical applications, DNA ligases are critical for maintaining genomic integrity and therefore cell viability. Genetic inactivation of the genes coding for DNA ligases results in pleiotropic cell phenotypes such as lethality and hypersensitivity to DNA-damaging chemicals [15, 17, 18]. Multiple human immunodeficiency disorders have been ascribed to the genetic inheritance of certain DNA ligase gene mutations. [13, 15, 19]. Moreover, the proliferation of cancer cells has been associated with abnormal regulation of DNA repair such as DNA ligase overexpression [12]. Therefore, to achieve the therapeutic goal of **targeting abnormal DNA ligase activity with small molecule inhibitors**, an extensive and detailed study of the enzyme's structure, kinetics and chemical function is essential.

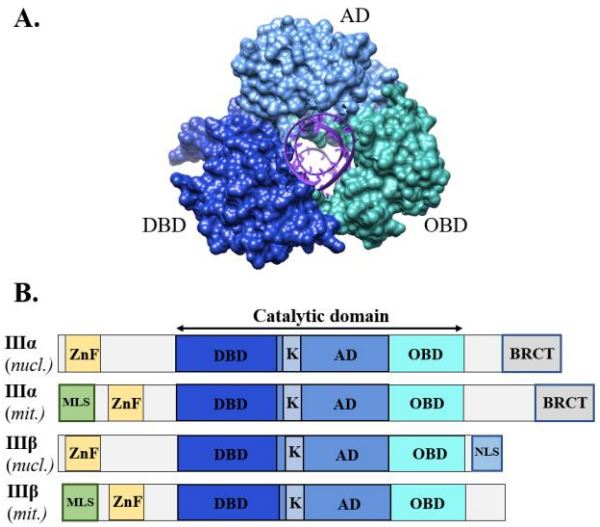


Figure 1. **A.** Space-filling representation of the conserved catalytic regions of the DNA ligase III species bound to nicked DNA. The DNA binding domain (DBD) is indicated in dark blue, the adenylation domain (AD) in light blue, the oligonucleotide/oligosaccharide binding-fold domain (OBD) in turquoise and the nicked DNA substrate in purple [20] **B.** Corresponding sequences of the DNA ligase III species isoforms. K: catalytically active lysine residue, ZnF: zinc-finger domain, MLS: mitochondrial localization signal, NLS: nuclear localization signal, BRCT: BRCA1 C-terminal repeat (BRCT) domain.

In humans, three DNA ligase genes have been found (*LIGI, LIGIII and LIGIV*) with common structural features (**fig. 1A**) [5, 13-16, 20]. A common catalytic core is found, consisting of an oligonucleotide/ oligosaccharide binding-fold domain (OBD) and an adenylation domain (AD) with a catalytically active lysine residue. While these two domains comprise the minimal unit for DNA ligase activity, the ligation reaction is greatly enhanced by an additional N-terminal DNA-binding domain (DBD) [21]. Besides the similarity in structure of the ligase family, unique differences between DNA LIG species have been found. For example, the conserved three-domain core of the DNA LIG III species is linked with the unique N-terminal zinc-finger domain that functions as a DNA strand break sensor, improving substrate recognition and binding (**fig. 1B**) [5, 22]. The germ-cell-specific DNA ligase IIIβ isoform is expressed in both the nucleus and mitochondria. This isoform lacks the BRCA1 C-terminal repeat (BRCT) domain that specifically interacts with the X-ray repair cross-complementing protein (XRCC1), which is otherwise crucial for the activity and stability of the DNA ligase IIIα isoform [13-16, 21].

The joining of breaks in the phosphodiester backbone of the duplex DNA ("nick") substrate also follows a conserved three-step mechanism (**fig. 2**) [5, 13-16, 21]. First, the catalytically active lysine residue within the catalytic domain is self-adenylated with adenosine monophosphate (AMP) to form a phosphamide bond using ATP. Then, the 5′ phosphate group of AMP is transferred from the active site lysine to a phosphorylated DNA 5′ end, forming a pyrophosphate linkage (5′P-5′P). Lastly, the nick is sealed by a nucleophilic attack of the juxtaposed 3'OH group and the release of AMP [5, 13-16, 21]. The understanding of DNA ligases' catalytic process in forming phosphodiester bonds has been documented for over five decades [16]. However, recent insights reveal that the nick-sensing, binding, and sealing mechanism of the enzyme is more complex than initially believed. Kinetic studies show complex behaviour concerning DNA ligase substrate specificity, binding, and the requirement for conformational transitions during the catalytic process [23, 24]. This is supported by structural studies, where the enzyme is captured in varying conformational states and thus indicates large structural flexibility during binding with DNA substrates [20, 25, 26]. The catalytic region can exist in an extended, open conformation during self-adenylyation, while the

interaction with nicked DNA is accompanied by a major change in protein conformation, turning the enzyme into a C-shaped clamp (**fig. 2**) [21]. Whilst crucial for its biological function, DNA ligase's complex structural plasticity poses obstacles to our understanding of its activity. The relationship between the observed conformational states, structure, function, and kinetics of human DNA ligases remains concealed by the ensemble average information collected so far.

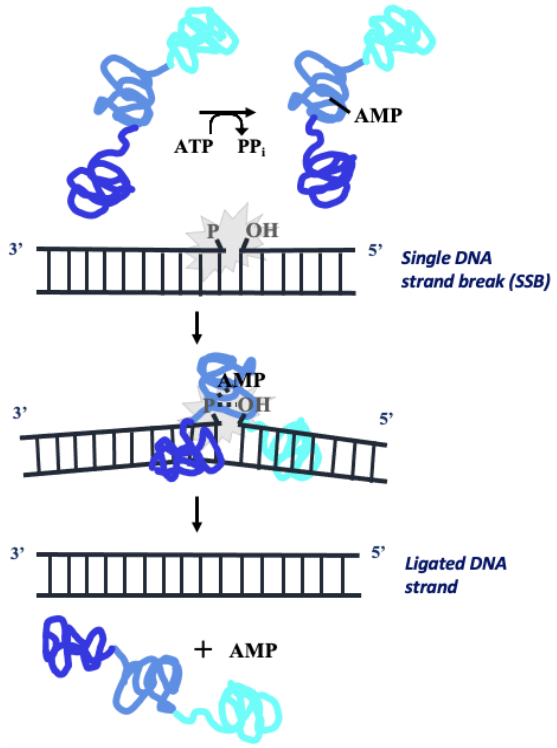


Figure 2: **Scheme of DNA ligase activity**. The catalytically active lysine residue within the catalytic domain is self-adenylyated with adenosine monophosphate (AMP) to form a phosphamide bond using ATP. The DNA single-strand break (SSB) is sealed by the formation of a phosphodiester link between the 5′ phosphate and 3' hydroxyl group at the DNA nick site.

Conventional kinetic and structural-based studies cannot define the necessary time constants needed to characterize the complex conformational heterogeneity of the protein [27]. Therefore, subtle protein domain movements and DNA substrate interactions exposing **hidden binding pockets** that are key for targeting with small-molecule drugs are possibly concealed by the ensemble [28]. Hence, the real-time investigation of DNA ligase activity entails the need for other, highly sensitive techniques. Rather than measuring averages of the vast number of molecules in a sample, information can be obtained from the molecules one by one. The **detection of single molecules** is achieved by limiting the number of molecules in the volume of observation, by separation and immobilization on a surface, or by studying freely diffusing molecules [29-33]. These single-molecule approaches, paired with **fluorescent energy transfer (FRET)** techniques, have been successfully employed for the study of protein-nucleic acid interactions, and various other structure-function relationships of biomolecules [31, 34-36]. Concisely, FRET is a process of energy transfer between two fluorescent molecules (fluorophores) based on nonradiative dipole–dipole interactions (**fig. 3**). For FRET to happen between two fluorophores, the emission spectrum of one fluorophore (donor) should overlap with the absorption spectrum of the other (acceptor) [27, 37]. Importantly, the efficiency of energy transfer between the donor and acceptor probes is highly sensitive to their proximity, whereby sub-nanometer accuracy (1-10 nm) can be achieved when measuring inter- and intramolecular distances [29, 35, 38]. For example, when there is a change in the distance between two fluorophores attached to a macromolecule(s), due to intrinsic dynamics or conformational changes, it is reflected in the change in energy transfer. The fluorescently labeled molecule undergoing conformational changes diffuses in the confocal excitation volume and emits bursts of photons that can be detected (**fig. 3**). Lowering the sample concentration to the pM regime (10^{-12} M) allows the burst-wise analysis of individual molecules [39]. Therefore, utilizing FRET grants the opportunity to detect single molecules and investigate their spatiotemporal dynamics that are otherwise lost by conventional tools. This has been previously demonstrated by Kilic and colleagues [35], where single-molecule FRET analysis was employed to detect flexible chromatin conformations that were not resolved before, although they are essential for their function. Furthermore, multiple studies employed FRET techniques to investigate the activity of DNA repair proteins, revealing complex domain movements and intermediate conformations during DNA substrate interactions [38, 40, 41]. Such knowledge is critical for resolving the molecular mechanisms of

underlying diseases related to abnormal DNA repair activity and the design of therapeutic drugs. To study conformational dynamics and individual DNA ligase-nucleic acid binding and dissociation events at the single-molecule level, fluorescent labeling of the protein will be necessary [24]. To realize this approach, a rational selection of the fluorophores and labeling methods are based on available structural data and computer simulations of DNA ligase and its DNA substrate. Typically, fluorophores are made up of two parts: the fluorescent moiety, which produces fluorescence, and the functional component, which reacts with amino acid residues [30]. Out of the 20 naturally occurring amino acids, cysteines (Cys) are the most popular for labeling purposes as they contain chemical handles (thiol group) for conjugation with maleimide dyes (**fig. 4A**) [30]. Since proteins generally contain fewer cysteines than other amino acids, donor-/acceptor (D/A) labeling can be accomplished by sequentially altering native amino acids to Cys residues in the polypeptide chain. The residues are selected so that the inter-dye distance lies within the FRET-sensitive range to detect conformational changes within the protein. However, placing dyes in catalytically active or structurally crucial parts of the protein should be avoided as it can interfere with enzymatic activity [42]. Furthermore, if the dye accessibility of the thiol side chains does not vary substantially, the introduced fluorophores can react with either of the two residues, resulting in a stochastic mixture of D/A-labeled molecules [43]. Since the conjugated fluorophore's photophysical characteristics may be affected by the local environment (charge, pH, or hydrophobicity), the use of cysteines may result in **sample heterogeneity** if product mixtures cannot be chromatographically separated. Consequently, the unique capability of single-molecule studies to interpret biologically relevant subpopulations and transient intermediate conformations is compromised [43]. However, the problem of stochastic labeling can be overcome by utilizing a bioorthogonal, **site-specific labeling** approach.

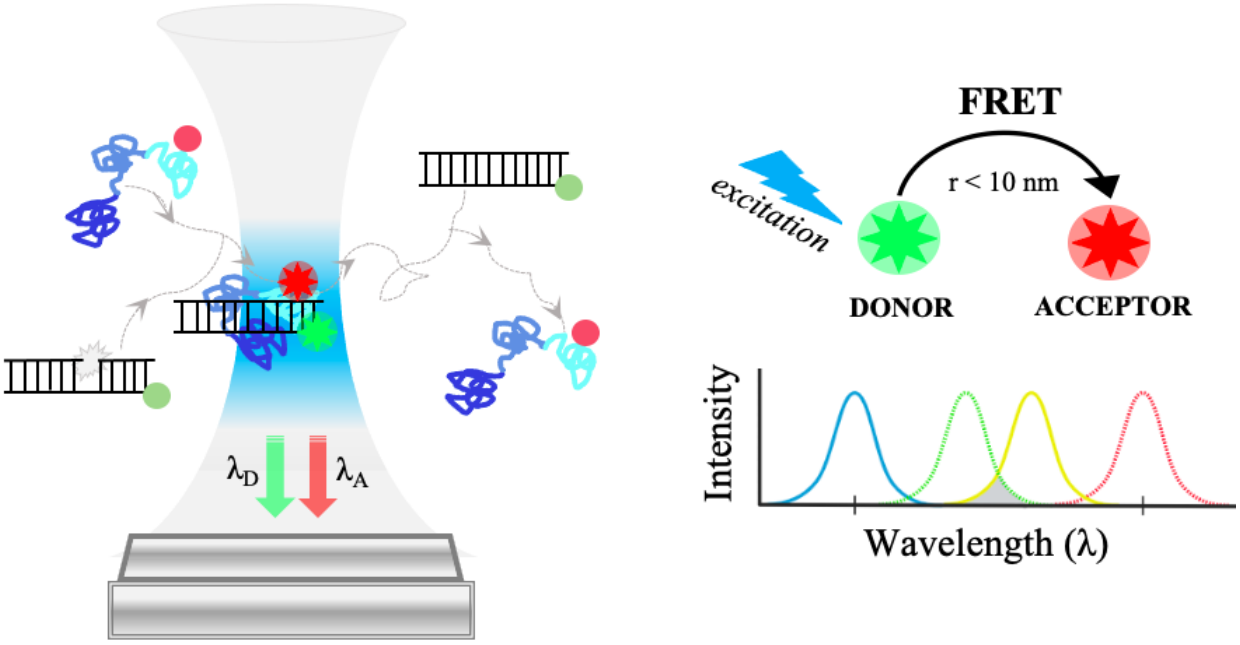


Figure 3: Single-molecule Förster resonance energy transfer spectroscopy. Single fluorescently labeled macro-molecules (DNA ligase and nucleic acids) randomly diffuse in and out of the confocal excitation volume. When the DNA ligase recognizes the nicked substrate, it will bind and perform its ligation activity. The enzyme-substrate pair diffuses into the excitation laser in the confocal volume, where the donor fluorophore will be excited and fluorescence resonance energy transfer (FRET) will occur when the donor/acceptor pair is in proximity (1-10 nm) and their emission/absorption spectra overlap. Every time a fluorescently labeled molecule diffuses into the confocal volume, a burst of photons is emitted that can be measured (λ_D: donor wavelength, λ_A: acceptor wavelength).

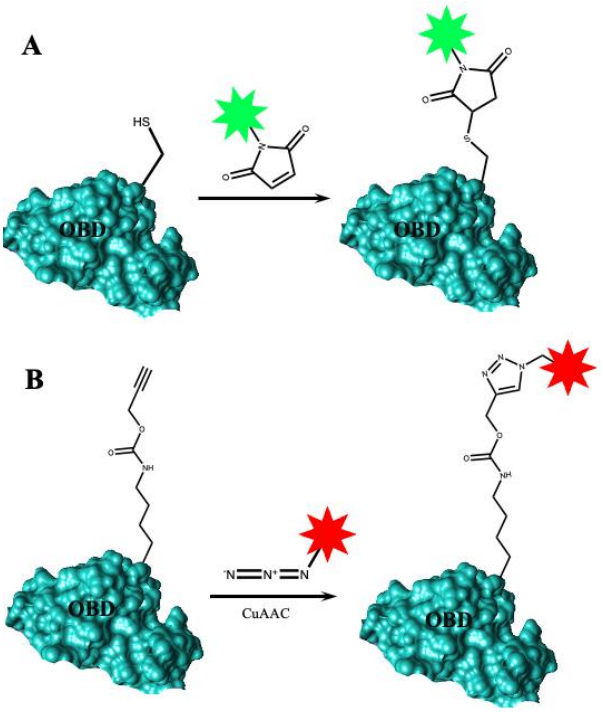


Figure 4: Fluorescent labeling of DNA ligase OBD domain. **A.** The protein is labeled with a maleimide dye (green) through the thiol of cysteine or **B.** The protein is labeled with an azide dye (red) through the azide functionality of genetically encoded Propargyl-L-Lysine (PrK), catalyzed by Copper (Cu). CuAAC: Copper-catalyzed alkyne-azide cycloaddition.

One of the most widely used bioorthogonal reactions to date falls under the section of "click chemistry" and was the topic of last year's Nobel Prize [44]. This reaction is called the Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC) and is biocompatible, highly selective, and can be performed in aqueous solutions at physiological pH [45]. Therefore, the reaction is considered an ideal candidate for protein engineering applications. **Artificial amino acids** (UAAs) with alkyne functionalities such as propargyl-L-lysine (PrK) can be conjugated with commercially available fluorescent azide dyes through the CuAAC reaction (**fig. 4B**). This site-specific labeling method with a unique reactive group can reduce sample heterogeneity during smFRET measurements and has already been successfully employed for various mammalian proteins [45, 46].

In this project, mutated human DNA ligase IIIβ was recombinantly expressed in *E. coli* host cells and purified using chromatographic techniques. The IIIβ isoform is stable in the absence of XRCC scaffold proteins and thus facilitates the purification process [47]. Furthermore, the recent association of the DNA ligase species with immunodeficiency syndromes makes it a valuable candidate for single-molecule investigations [19]. Genetically encoded PrK is introduced in the DNA ligase sequence through an engineered peptide translation machinery system [48]. Here, PrK is encoded by an amber stop codon (TAG) within the DNA sequence and is recognized by its orthogonal suppressor tRNA. The transfer of PrK to its specific tRNA is catalyzed by a unique pyrrolysyl-tRNA synthetase (PylRS) [48]. Successfully purified DNA ligase mutants were employed in fluorescent labeling reactions and the efficiency of the reaction was evaluated with fluorescence correlation spectroscopy (FCS). With FCS, the rate of fluorescence intensity fluctuations can be analyzed by transforming the fluorescence intensity signal into an auto-correlation function (ACF). The characteristic decay time of the ACF, known as the diffusion time, is inversely proportional to the diffusion coefficient of the fluorescently labeled molecule [49]. Not only does this allow us to investigate the binding of the fluorescent dye to the protein, but also to characterize interactions between the protein and fluorescently labeled DNA substrate. Moreover, the purified enzyme was subjected to DNA ligation assays to ensure its functional integrity. We hypothesized that DNA ligase mutants can be successfully purified and fluorescently labeled while maintaining their DNA-binding and ligation properties. Results obtained in this project will be useful for future studies focusing on the dual labeling of human DNA ligases to minimize smFRET sample heterogeneity. Ultimately, conformational studies will expand our fundamental understanding of DNA ligase activity and improve the design of specific inhibitors to target cancer cells based on their pathway defects.

EXPERIMENTAL PROCEDURES

Materials - All used chemicals and solvents were obtained from various commercial sources and used without additional purification. L-(+)- arabinose, Luria-Bertani (LB) - Broth and agar (Lennox), Sodium dodecyl sulfate (SDS), Phosphate buffered saline (PBS) (pH 7.4), and Sodium chloride (NaCl) were purchased from Sigma Aldrich (Diegem, Belgium). Terrific broth (TB), isopropyl-thio-galactoside (IPTG) (≥ 99%), and deionized formamide (≥ 99,5%) were purchased from Carl

Roth. Dithiothreitol (DTT) (≥ 99%) and Chloramphenicol (98%) were purchased from Fischer Scientific (Merelbeke, Belgium). ATTO 643 maleimide dye was purchased from ATTO-TECH. Propargyl-L-Lysine (PrK) was obtained from SiChem. Milli-Q water from a Sartorius Stedim Biotech machine (Goettingen, Germany) was used for all experiments.

Bacterial strains and plasmids - *E. coli* DH5α strain was used for cloning, and protein expression was carried out in *E. coli* BL21(DE3) or TOP 10 (BCCM/LMG). The pET-16c plasmid carrying the full-length histidine-tagged LIG3 β (906 amino acid residues) was a kind gift from Alan E Tomkinson (University of New Mexico). To incorporate PrK into the LIG3 β sequence, an orthogonal amber suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pair was used, encoded in the pEVOL PylRS WT and Y306A/Y384F plasmids. These plasmids, together with pBAD GFP(Y39TAG), encoding histidine-tagged green fluorescent protein (GFP) with a nonsense codon (TAG) at position 39, were a kind gift from Edward A Lemke (EMBL, Heidelberg)). pET-16c plasmids with a TAG or cysteine mutation at positions 410 and 705 respectively, were prepared by site-directed mutagenesis (Q5® Site-Directed Mutagenesis Kit (NEB)), yielding plasmids pET-16c LIG 3β (K410TAG) and pET-16c LIG 3β (S705C). Similarly, double-mutated pET-16c LIG 3β (S705C-S842C) and removal of the N-terminal zinc-finger domain in pET-16c LIG 3β ΔZnF were generated. Sequencing was done by LGC Genomics using appropriate primers (Sanger sequencing).

Expression of LIG 3β mutants - Calcium-competent *E. coli* strains BL21 (DE3) were typically transformed by the heat-shock method with 1-100 ng/μL of the expression plasmid pET-16c encoding ampicillin resistant, his-tagged LIG3 β S705C, LIG3 β S705C-C842S and LIG 3β ΔZnF and plated out on LB-agar plates with 100 μg/ml ampicillin. A single colony was picked for inoculation in 5-10 mL LB medium containing 100 μg/mL ampicillin and grown O/N at 37°C and 200 rpm. The O/N culture was diluted 1:250 in 1 L of TB medium containing 100 μg/mL ampicillin and incubated at 37°C and 200 rpm. At an OD_{600} of 0.6, IPTG was added to a final concentration of 1 mM, and incubation was continued for 4-6 h at 37°C or O/N at 25°C and 200 rpm. The bacteria were collected by centrifugation and resuspended in purification- or 1X PBS buffer. After recollection by centrifugation, the pellet was stored at -20°C for subsequent purification.

Expression of PrK mutants - Calcium-competent *E. coli* TOP10 or BL21 (DE3) was typically co-transformed by the heat-shock method with 1-100 ng/μL of the pEVOL PylRS Y306A/Y384F plasmid (chloramphenicol resistance) and the expression plasmid pET-16c encoding his-tagged LIG 3β K410TAG or pBAD GFP Y39TAG (ampicillin resistance) and plated out on LB-agar plates with 100 μg/ml ampicillin and 34 μg/mL chloramphenicol. A single colony was picked for inoculation in a 5 mL LB medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol and grown O/N at 37°C and 200 rpm. The O/N culture was diluted 1:250 in 50 mL of TB medium containing 100 μg/ml ampicillin and 34 μg/mL chloramphenicol and incubated at 37°C and 200 rpm. At an OD_{600} of 0.4-0.6, L-arabinose and PrK were added to final concentrations of 0.02% and 1 mM respectively for the expression of GFP. For LIG 3β, L-arabinose, PrK, and IPTG were added to final concentrations of 0.02%, 1 mM, and 1 mM respectively. Incubation was continued for 4-6 h at 37°C for GFP and O/N at 25°C and 200 rpm for LIG 3β K410. The bacteria were collected by centrifugation and resuspended in 1X PBS buffer or purification buffer and were stored at -20°C after recollection by centrifugation for subsequent purification.

Purification of LIG 3β variants - Cells were resuspended in 50 mL of ice-cold lysis buffer/ buffer A (50 mM Tris-HCl, 10% glycerol, 1 mM DTT, 50 mM NaCl [pH 7.4]) and sonicated (Branson Sonifier 150, Danbury, CT) for 10 min (pulse intervals of 0.5 seconds at an amplitude of 60% on an ice-water bath). The lysate was clarified by centrifugation at 8000 x g for 45 min at 4°C and filtered (0.2 μm) before loading onto a 5 mL Capto SP ImpRes column (Cytiva), pre-equilibrated with buffer A. After loading of the sample on the column, it was washed with buffer A, and bound proteins were eluted with a linear gradient of up to 1 M NaCl in buffer B (50 mM Tris-HCl, 10% glycerol, 1 mM DTT, 1000 mM NaCl [pH 7.4]).

The eluted fractions were assayed on SDS-PAGE and fractions containing LIG 3β were pooled and incubated by constant rotation at 4°C for 1 h with 0.5 mL of nickel Ni-NTA beads (Jena Bioscience) and 10 mM imidazole. After washing with buffer C (50 mM Tris-HCl, 10% glycerol, 1 mM DTT, 500 mM NaCl [pH 7.4]) containing 20 mM imidazole, bound proteins were eluted from the beads with elution buffer C containing 250 mM imidazole. Residual imidazole and DTT were removed by a HiTrap desalting column (5 mL, Amersham Biosciences) in degassed buffer C. The order of chromatographic purification techniques was changed throughout the work whenever mentioned. All purification steps were performed at 4°C. Approximately 1 to 10 mg of DNA ligase IIIβ was usually obtained from a 1 L culture. Purified DNA ligase IIIβ variants were aliquoted and stored at -80°C.

Cysteine labeling - Purified LIG3 β variants were mixed with a 5-fold molar excess (S705C) or 1.2x excess (S705C-C842S & ΔZnF) of maleimide ATTO 643 dye in degassed buffer C. The protein-dye mixture was incubated for 1 hour at 25°C (S705C) or O/N at 4°C (S705C-C842S & ΔZnF) in the dark. The fluorescently labeled protein was separated from excess dye using a PD-10 column (and Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences). Determination of protein concentration and degree of labeling (DOL) was performed by measuring the absorbances at 280 and 650 nm. Concentrations were calculated with measured absorbance at 1 mm path length and extinction coefficients (ε_{S705C}: 66800 $M^{-1}cm^{-1}$ and $\varepsilon_{\Delta ZnF}$: 52830 $M^{-1}cm^{-1}$). Pooled fractions were further concentrated using an Amicon 50 kDa spin filter.

Gel-based ligation assay - Purified DNA ligase 3β variants were mixed with labeled DNA substrate (Eurofins Genomics) at 37°C and/or 4°C in standard reaction buffer (50 mM Tris-HCl, pH 7.5, 10% glycerol, 20 mM $MgCl_2$, 40 mM NaCl, 1 mM DTT, 100 μg/mL BSA (MP Biomedicals) and 1 mM ATP) unless otherwise stated. The nicked DNA substrate was generated by annealing the following oligonucleotides: 5'-TCA GCC TAT CGG CTA TCG ATC CGA TCG-3', 5'-CGA TCG GAT CGA TAG C-3', and 5'P-CGA TAG GCT GA-3FAM' (P indicates the presence of a 5'-phosphate and FAM indicates the presence of 3'-fluorescein). The oligonucleotides were annealed at equimolar equivalents (20 μM) in annealing buffer (1 X TE buffer, pH 6.5, 150 mM NaCl) by heating the solution to 95 °C for 2 minutes and cooling to 25 °C at a rate of -4°C/min. To ensure 100% adenylyation, $MgCl_2$, and ATP were added to the concentrated DNA ligase fractions with final concentrations of 20 mM and 1 mM respectively, and incubated for 60 min on ice before the ligation assay unless otherwise stated. The ligation reactions were quenched with a 2-fold dilution at designated times (20-600 seconds, unless otherwise stated) into quench solution (90% formamide and (50 mM EDTA), before loading onto aan18% (w/v) polyacrylamide, 8 M urea, 1X TBE gel. The gels were scanned using an Amersham Imager 680. DNA substrate and product bands were quantified using ImageJ software and graph(s) were normalized.

Fluorescence correlation spectroscopy (FCS) - FCS measurements were performed on a homebuilt 483/635-nm alternating excitation (LDH-P-C-470, Picoquant)/LDH-P-C-635B, dual-polarization detection confocal microscope. Sample emission was transmitted through a pinhole and spectrally split. Photons were detected by four avalanche photodiodes (PerkinElmer or EG&G SPCM-AQR12/14, or Laser Components COUNT BLUE): B|| (blue-parallel), B⊥ (blue-perpendicular), R|| (red-parallel) and R⊥ (red-perpendicular), which were connected to a TCSPC device (SPC-630, Becker & Hickl GmbH). Microscope alignment was done by measuring freely diffusing ATTO488 and ATTO655 in water. The known diffusion coefficients of these fluorescent dyes were used to determine the confocal parameters ω_x and ω_z and were fixed throughout the fitting of the auto-correlation curve (ACF). DNA binding assays were performed with FAM-labeled DNA substrate concentrations of ~ 1 nM mixed with 0.25 to 5 μM purified DNA ligase. ATTO643 labeled DNA ligase mutants were diluted to the nM regime (0.5-2 nM) for measurements and compared to ATTO 655 (1nM). Typically, measurements of 60 s duration were performed in labeling buffer.

RESULTS & DISCUSSION

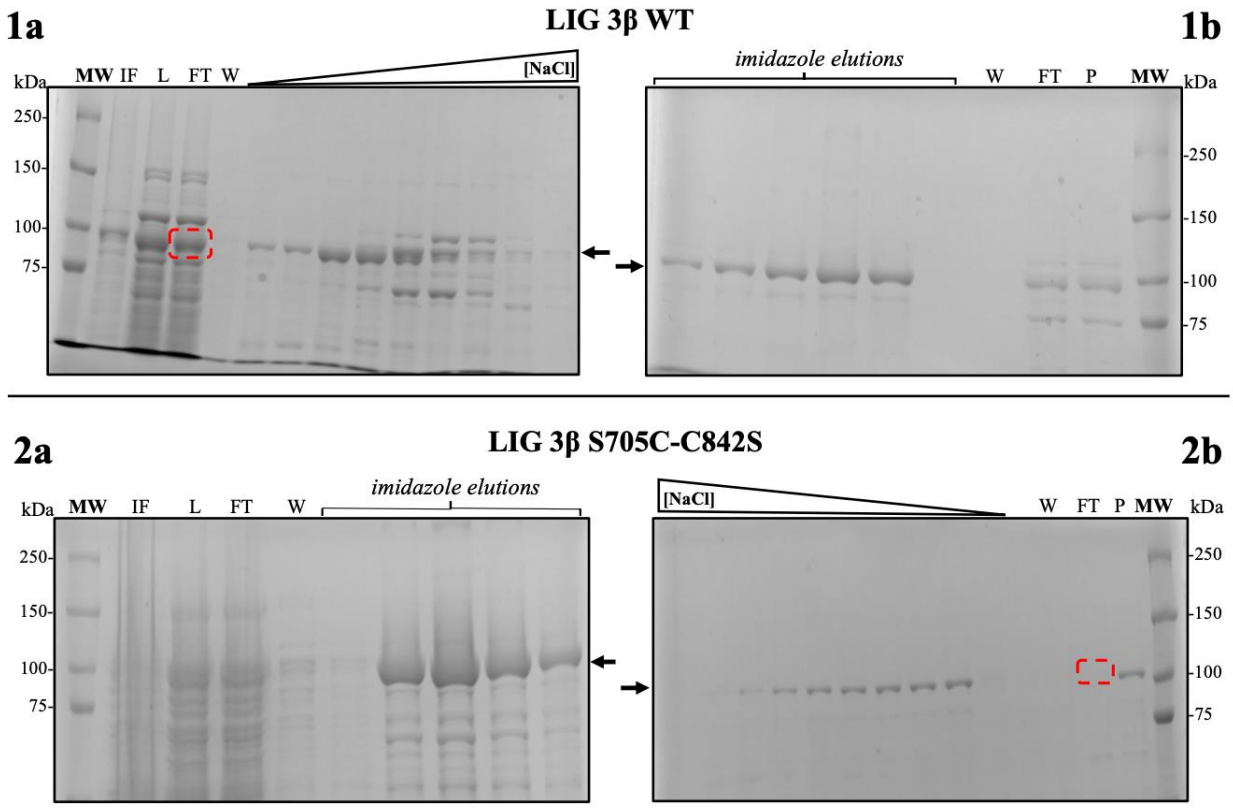


Figure 5: DNA ligase purification. 8% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) profiles of purified recombinant DNA ligase 3β wild-type (WT) and mutant S705C-C842S expressed in *E. coli* BL21(DE3) cells. NaCl gradient (1M) and imidazole elutions (250 mM) represent the fractions containing DNA ligase (indicated by arrow) collected from ion-exchange chromatography (IEXC) and immobilized metal affinity chromatography (IMAC) with Ni-NTA agarose beads (resin), respectively. **1**. LIG 3β WT purification profile where the enzyme was eluted with a NaCl gradient in IEXC column (**1a**) before IMAC His-tag purification (**1b**) **2.** Mutant LIG 3β S705C-C842S was purified in the reversed order (**2a-b**). Lane MW: molecular weight marker proteins (50-250 kDa), lane IF: insoluble fraction after sonication of *E.coli* cells, lane L: cleared lysate after sonication, lane FT: flow-through of column (non-binding), lane W: washing of column, lane P: pooled fractions containing DNA ligase. Unbound DNA ligase in the FT after IEXC is indicated by the red rectangle.

DNA ligase purification- To achieve site-specific fluorescent labeling of DNA ligases, it is crucial to minimize impurities through chromatographic purification. Additionally, obtaining sufficient yields of homogeneous protein is essential for the characterization and fluorescent labeling process. Published protocols describing the purification of DNA ligases typically include protease inhibitors and detergents to minimize proteolytic degradation and/or stabilize the proteins in solution throughout the purification process [20, 47, 50, 51]. In this work, purification reagents are minimized by eliminating both protease inhibitors and detergents to facilitate the purification process and investigate their necessity for the purification of functional DNA ligases. Here, the purification of DNA ligase 3β wild-type (WT) and mutants are discussed concerning yields, purity, and the order of purification.

Figure 5 shows the SDS-PAGE profiles after purification of both WT- and cysteine mutated (S705C-C842S) DNA ligase after expression in *E.coli* cells. Apparent bands at the 100 kDa molecular weight marker were visible that represent fractions containing the protein of interest (estimated molecular weight of human DNA LIG 3β ~ 101 kDa). Purifying the wild-type protein from the cleared bacterial lysate by ionic exchange

chromatography (IEXC) resulted in the elution of DNA ligase WT during the linear [NaCl] gradient (**Fig. 5.1a**). Typical impurities with molecular weights flanking the 100 kDa marker were observed. The question arose whether further removal of contaminants was feasible after additional chromatographic purification. As the elution fractions of IEXC contain contaminants close in size to the protein of interest, purification based on size exclusion was not recommended [52]. Therefore, a second round of purification was conducted using immobilized metal affinity chromatography (IMAC). This method of purification, based on the selection of the polyhistidine-tag in the DNA ligase protein sequence, resulted in further reduction of contaminants (**Fig. 5.1a-b**). During the initial IEXC purification, it appears that a notable amount of protein was lost (indicated by the red rectangle in **Fig. 5.1a**) in the flow-through (FT). This observation suggests unsuccessful binding of a significant portion of the protein of interest to the IEXC column. The order of purification was reversed for the mutant S705C-C842S (**Fig. 5.2a-b**) to investigate the effect of the chromatographic purification order on protein yield. The SDS-PAGE profile after an initial IMAC purification of the mutant (**Fig. 5.2a**) shows higher levels of contamination compared to the IMAC profile after IEXC of the WT DNA ligase (**Fig. 5.1b**). As the bacterial cells were lysed in a buffer with high ionic strength (500 mM NaCl), a 30-fold dilution was necessary to lower the ionic strength to ~50 mM for application on the IEXC column for the second round of purification. This is important to take into consideration, as the IEXC SDS-PAGE profile of S705C-C842S (**Fig. 5.2b**) seems to contain almost no contaminants. Although the dilution factor may explain the visible absence of other polypeptides, the 100 kDa band representing the highly concentrated protein of interest is visible in the diluted IMAC pooled fractions (P). The application of the IMAC fractions containing the DNA ligase mutant to the IEX column resulted in no visible protein loss in the FT (indicated by the red rectangle in **Fig. 5.2b**).

The impurities observed after chromatographic purification are possibly polypeptides of the *E. coli* host with a similar surface charge at pH 7.4 (IEXC), histidine-rich polypeptides (IMAC), or degradation products of the DNA ligase protein (< 101 kDa) [53]. The notable loss of protein observed during an initial IEXC purification could pose a challenge in achieving sufficient protein yields for the downstream characterization and fluorescent labeling process. This loss could be the result of non-optimized conditions of the purification process (e.g., column flow rate, loading capacity) or an unfavourable interaction between the enzyme and the negatively charged IEXC resin. As it is well established that DNA ligase molecules can exist in a DNA-bound state by both specific and/or non-specific interactions [20, 54], it is important to consider that the recombinant DNA-binding proteins might interact with intact bacterial DNA present after cell lysis [52, 55, 56]. This non-specific binding is primarily caused by electrostatic interactions between the negatively charged DNA backbone and the positively charged protein side chains [57]. As DNA ligase *in-vitro* studies show that unspecific binding to various non-nicked DNA substrates can occur at low ionic strength (< 80 mM NaCl) [51], it may be problematic in terms of purification if the bacterial cells are lysed in a low salt containing buffer (50 mM NaCl). This means that the protein-DNA complex would bind less efficiently with the negatively charged resin of the IEX column at pH 7.4 due to the change in net surface charge. This results in a loss of protein in the flow-through. Thus, lysing the cells in a buffer with higher ionic strength may remove most of the contaminant DNA and minimize protein-DNA interactions during IMAC purification. Consequently, a decreased amount of protein would be observed in the flow-through of IEXC. This was observed for the S705C-C842S DNA ligase mutant. Other researchers have previously followed a similar purification workflow with success [20, 47, 50, 51], although the order and selection of chromatographic techniques are not discussed in detail. The observations in our study support the idea of recombinant DNA ligase and bacterial lysate DNA interactions affecting the IEXC column binding. However, the argument is only based on observation of the SDS PAGE profiles and is therefore not conclusive. Further analysis of flow-through contents and optimization of both the IEXC and IMAC purification procedures is necessary to elucidate the role of ionic strength on the state of the protein during purification. Ideally, the IMAC resin protein-binding capacity should match the estimated

concentrations of DNA ligase protein after expression in *E.coli*. Higher levels of impurity during an initial IMAC purification are expected, as host proteins that are histidine-rich were previously eliminated by IEXC. When not removed, these proteins can compete for binding with the nickel ions of the column [58-60]. With a minimum amount of IMAC resin, the polyhistidine-tagged DNA ligase molecules will fill most of the available binding sites, reducing the number of contaminant proteins [60]. Thus, an initial, optimized IMAC protein purification may be sufficient in terms of yield and purity for the downstream characterization and fluorescent labeling of DNA ligase mutants. This method of purification was employed for the DNA ligase mutant ΔZnF (removal of the N-terminal Zinc-finger domain) and resulted in adequate purity and high yields based on a single IMAC SDS-PAGE profile (**Supplementary Fig. 1**). The purification procedure explored here shows promise for the fast and efficient purification of other DNA ligase mutants for future downstream applications.

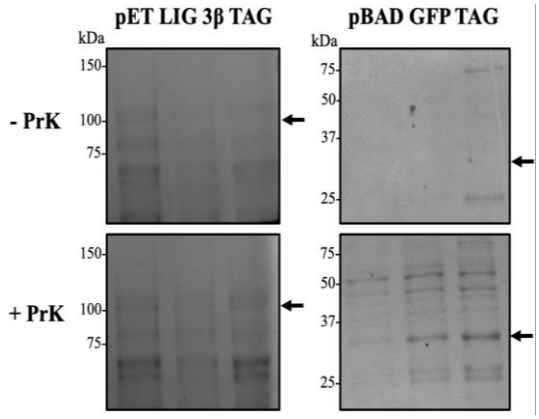


Figure 6: IMAC purification of PrK mutants. The expression of LIG 3β TAG in *E.coli* BL21(DE3) cells (pET expression vector) was induced after the addition of the artificial amino acid Propargyl-L-Lysine (PrK) (+ PrK) to the expression culture and resulted in no visible protein expression. GFP TAG was expressed in *E. coli* TOP10 cells (pBAD expression vector) after induction and the addition of PrK to the expression culture. Expression cultures without the addition of PrK (- PrK) served as a negative control. Proteins were purified with a small-batch IMAC purification protocol as described in the experimental procedure. Arrows indicate ~ 101 kDa and ~ 27 kDa for LIG 3β TAG and GFP TAG respectively. For the complete SDS-PAGE profiles with molecular weight markers, see **Supplementary Figure 2**.

The incorporation of unnatural amino acids (UAAs) into the protein sequence offers a promising alternative to cysteine labeling for future smFRET experiments. Introducing UAAs expands the fluorescent labeling repertoire and can minimize sample heterogeneity. In this study, both polyhistidine-tagged DNA ligase 3β and green-fluorescent protein (GFP) containing an Amber stop codon (TAG) were subjected to single-immobilized metal affinity chromatography (IMAC) small-batch purifications following expression in *E. coli*. The bacterial cultures were supplemented with the UAA Propargyl-L-Lysine (PrK) to express the mutant protein. No visible expression of DNA ligase 3β (K410TAG) was observed in the IMAC SDS-PAGE profiles (**Fig. 6**). Conversely, the expression of the GFP (Y39TAG) mutant, was shown by the presence of visible ~27 kDa bands on the SDS-PAGE profile (**Fig.6**) and through fluorescence imaging (**Supplementary Fig. 2**). The unsuccessful expression and/or purification of the DNA ligase mutant could be attributed to several factors. Firstly, the incorporation of UAAs in the protein sequence generally leads to reduced expression levels [61]. The binding of contaminants during the IMAC purification process can thus be more pronounced, as was discussed previously. Furthermore, the small-batch IMAC purifications performed in this study were conducted after protein expression in 50 mL cultures This is in contrast to the larger 0.5-1 L cultures used for the expression of the WT DNA ligase and other mutants. Consequently, the smaller culture volumes yield a lower amount of total protein, which may contribute to this observation. Again, this stresses the need for IMAC optimization and matching the resin bed volume with the expected concentration of the protein of interest. It is worth noting that the GFP mutant discussed in this work has been successfully expressed and purified in previous studies [62-64]. However, for the DNA ligase mutant, it remains unclear whether the efficiency of PrK incorporation is affected by the specific residue position (410) within the protein sequence. To address this, future attempts will involve generating additional DNA ligase mutants with an Amber stop codon positioned at different positions in the protein sequence. This will enable the exploration of the effect of the selected residue position and PrK incorporation efficiency on the expression and purification of the mutated DNA

ligase. Interestingly, another study focussing on the expression of a large human PrK-mutant protein in *E. coli* suggests the use of the pBAD expression system over pET [65]. The pBAD system allows for a more tightly controlled and titratable expression of the protein of interest, which is advantageous when dealing with potentially toxic recombinant DNA ligase expression in *E. coli*. As the expression and purification of the PrK-mutant were not successful, the following sections will focus only on the WT DNA ligase, ΔZnF, and other mutants relevant to (cysteine) fluorescent labeling. WT DNA ligase and mutants were subjected to purification which resulted in sufficient yields and purity for subsequent characterization and fluorescent labeling.

Characterization of purified DNA ligase activity – Following the successful purification of WT DNA ligase and mutants, it is important to assess their functional integrity. This critical step ensures the reliability of data obtained in future smFRET studies. By verifying their functional integrity after purification, we can preserve their complex dynamic behaviour and enzymatic activity. In this study, fluorescence correlation spectroscopy (FCS) was employed to characterize the DNA-binding ability of the purified enzyme, while DNA ligation assays (**Fig.7A**) were conducted to evaluate the ligation activity.

The purified WT DNA ligase was able to ligate the fluorescein (FAM)-labeled DNA substrate that contains a nick with a 3'-hydroxyl and a 5'-phosphate, resulting in a 27-nucleotide long ligation product (**Fig. 7B**). Notably, the product formation occurred even without the addition of ATP (**Fig. 7C**). Mixing the enzyme with the FAM-DNA mixture resulted in a rapid ligated product formation with the subsequent depletion of detectable nicked DNA substrate in the first 80 seconds. After this burst of product formation, it stabilizes to a maximum (**Fig. 7B**). It is worth noting that this maximum likely represents the complete depletion of the nicked substrate and, thus, the maximum amount of ligation product achieved in this experiment (190 nM).

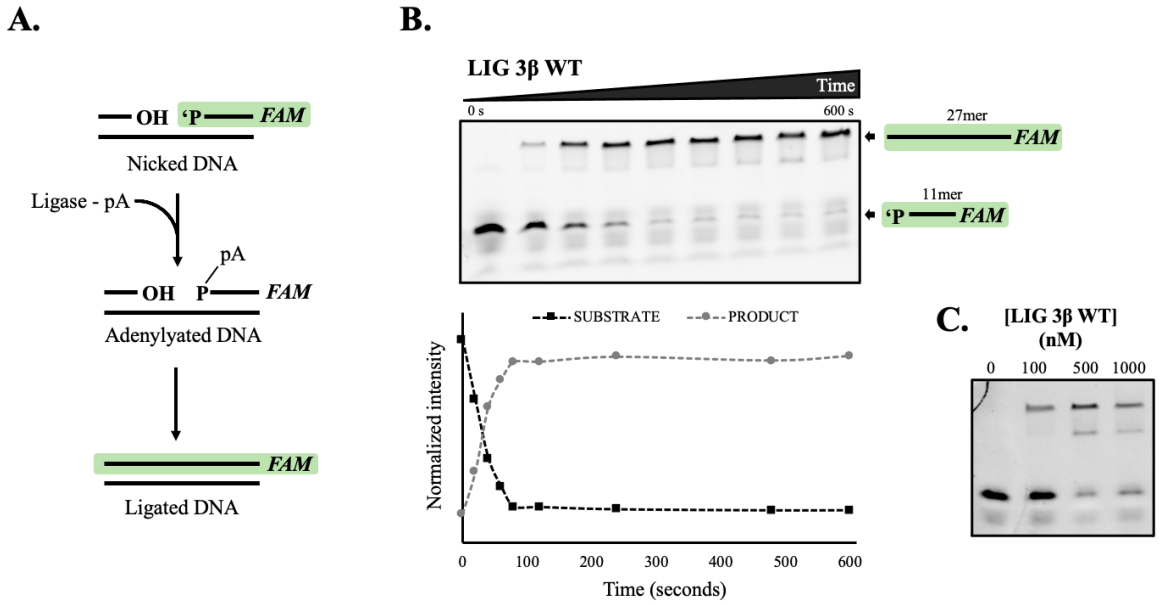


Figure 7: DNA ligase 3β ligation assay. **A**. Reaction schematic illustrating the ligation of 3'-fluorescein labeled (FAM) nicked DNA substrate by wild-type (WT) LIG 3β. The DNA species detected in the ligation assay are indicated in green. ***B.*** Representative denaturing 18 % polyacrylamide, 8 M urea gel with the denatured product (27-mer) and substrate (11-mer) indicated by arrows. The normalized graph below shows the product formation and substrate depletion intensities measured by Image J software. Reactions contained 190 nM FAM-DNA substrate, and 50 nM WT LIG 3β in standard reaction buffer at 150 mM ionic strength and were quenched at designated times (20, 40, 60, 80, 120, 240, 480, and 600 seconds) at 37°C. The purified ligase was pre-incubated with 20 mM $MgCl_2$ and 1 mM ATP to ensure 100% adenylyation. **C.** Denaturing 18 % polyacrylamide gel after mixing different LIG 3β WT concentrations (0-1000 nM) with 45-50 nM of FAM labeled nicked DNA and quenching after 10 minutes incubation at 23°C. Reactions were performed in reaction buffer containing 20 mM $MgCl_2$ and a final ionic strength of 150 mM.

Similar results were observed for the DNA ligase ΔZnF mutant (**Supplementary Fig. 3A**). Although these individual experiments are not sufficient to extract kinetic parameters regarding enzymatic activity, it demonstrates that the purified protein is functional. Thus, the results indicate the successful purification of functional human DNA ligase through a minimalistic purification protocol. Additionally, the enzymatic behaviour of both WT DNA ligase and the ΔZnF mutant aligns with findings reported in previous literature [20, 22, 51]. Even with the removal of the N-terminal Zinc-Finger domain, the enzyme can efficiently ligate nicked DNA substrate. It is worth noting that other studies have utilized ligation assays to identify an adenylyated intermediate product on the gel, which represents the formation of an abortive ligation state (**Fig. 7A**) [5, 50, 51, 66]. Although this intermediate product is not observed in the data presented here, it could be valuable for evaluating ligation activity after fluorescent labeling. One explanation for the visible absence of the intermediate product in this work is the employment of a low ligase/substrate ratio (50 nM/190 nM respectively), which does not accumulate visible intermediate product formation [50]. Interestingly, previous studies demonstrated that approx. 80% of recombinant human DNA LIG 3β purified from *E.coli* is adenylyated (**Fig. 2**) [50, 66]. Even in the absence of ATP in the ligation assays, nick ligation can still occur to some extent. This was demonstrated in **Figure 7C**, where different amounts of DNA ligase were mixed with nicked DNA substrate without the addition of ATP, resulting in the formation of ligated product. This observation suggests that a portion of the purified DNA ligase is in the adenylyated state, which is structurally distinct from its non-adenylyated state [67, 68]. Although the conformational changes induced by self-adenylyation are minimal, they may affect the accessibility of binding sites for fluorescent labeling. Note that if the site-specific fluorescent labeling of the protein is successful, it is presumably labeled in the adenylyated state. This information can be considered a priori knowledge for downstream data processing in future smFRET experiments. It might be of interest to consider deadenylyating the purified enzyme by incubation with sodium pyrophosphate followed by dialysis, as was previously described [54]. This would help homogenize the purified protein population and ensure its equilibrium for both the fluorescent labeling reaction and subsequent smFRET experiments.

In addition to screening the ligation activity of the purified enzyme, evaluating its DNA-binding ability provides an additional means to assess protein integrity after purification. For this reason, purified DNA ligase was incubated with the nicked FAM-DNA in FCS measurements. The measured rate of diffusion depends on both the molecular weight of the labeled molecule and the viscosity of the medium [49]. The binding of macromolecules to the FAM-DNA substrate is expected to result in a lower diffusion coefficient compared to free unbound FAM-DNA. Titrating purified WT DNA ligase to the FAM-DNA substrate mixture led to a concentration-dependent decrease in the diffusion rate of the DNA substrate (**Fig. 8A**). The change in diffusion coefficients was most pronounced at low WT DNA ligase concentrations (0.25-1 μM), whereafter the effect reaches a minimum at higher concentrations. A similar effect was observed for different NaCl concentrations (150 and 190 mM). The auto-correlation functions (ACFs) exhibited longer diffusion times compared to freely diffusing DNA substrate (right-shift) (**Fig. 8B**). Interestingly, a similar titration experiment (150 mM NaCl) for the purified ΔZnF mutant showed a notable deviation from WT DNA ligase behaviour. Here, titration of increasing protein concentrations led to an increase in measured diffusion coefficients of the FAM-DNA substrate (**Supplementary Fig.4**). The FCS data implies a stable interaction between the purified WT DNA ligase and DNA substrate. When the enzyme interacts with the DNA substrate, it forms a DNA-protein complex with a lower diffusion coefficient compared to the free DNA substrate (~17 kDa) due to the larger molecular weight of the enzyme (~101 kDa). The observed minimum in DNA diffusion rate for higher WT DNA ligase concentrations could be the result of a possible saturation effect, where all the DNA molecules are in continuous bound-unbound states and no additional binding of DNA ligase to free DNA substrate takes place. Moreover, the saturation effect suggests that the decrease in DNA substrate diffusion rate is not purely caused by other means than protein-substrate binding, such as a systematic increase in viscosity. However, the strange behaviour observed for the ΔZnF mutant cannot be explained by the same rationale.

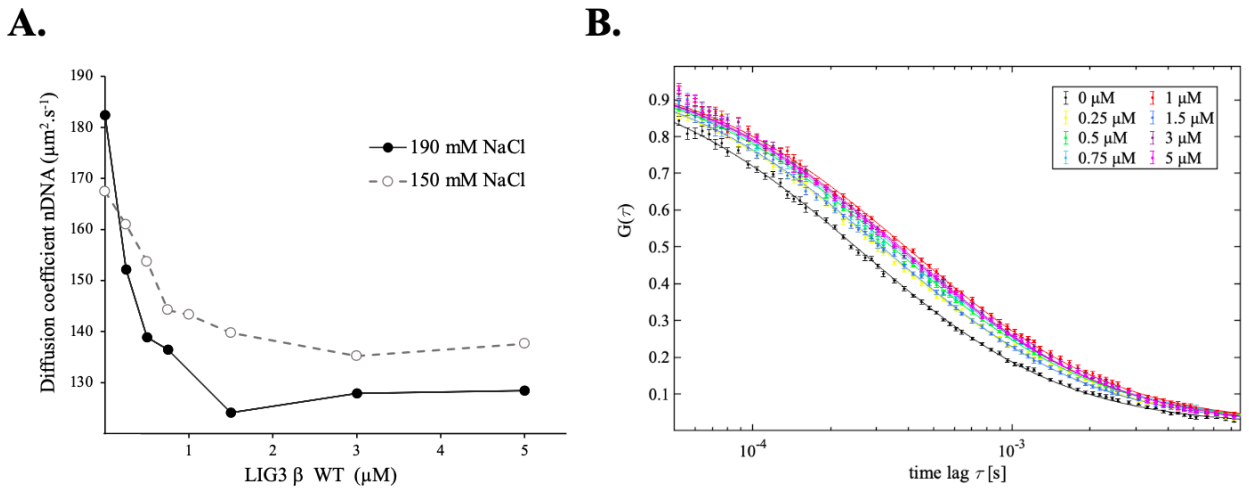


Figure 8: FCS DNA ligase 3β binding assay. Purified LIG 3β WT (0-5 µM) is titrated to 3' fluorescein-labeled (FAM) nicked DNA substrate and the rate of diffusion is measured with fluorescence correlation spectroscopy (FCS) in reaction buffer (150/190 mM NaCl and 20 mM $MgCl_2$). **A**. Diffusion coefficients of FAM-DNA after mixing with purified WT DNA ligase. **B.** Autocorrelation functions (ACFs) from the FCS experiment (0-5 µM DNA ligase: see legend). Note that the FCS titration experiments are performed in a way that for each increase of DNA ligase concentration, the FAM-DNA concentration decreases and results in different initial amplitudes between each titration. Therefore, the auto-correlation functions (ACFs) of different DNA ligase concentrations are normalized such that the initial amplitudes are equal to each other (=1) and the decay rates of the ACFs can be compared. Measurements were 60 seconds long and fluorescein was excited with 100 µW laser power.

The increasing diffusion rate of FAM-DNA of the ΔZnF titration experiment may be the result of inconsistent experimental conditions or the presence of unknown factors affecting the DNA-protein interactions. Although differences in ligation efficiency and DNA-binding behaviour between the WT and ΔZnF mutant DNA ligase are possible [51], a substantial increase in the diffusion rate of the DNA substrate was not expected. This highlights the need for further investigation to understand the underlying cause of this deviation compared to WT DNA ligase. Future work could consider using an electrophoretic mobility shift assay (EMSA) to validate the DNA-protein binding events observed in this study and gain additional insights into the unexpected behaviour of the purified ΔZnF mutant. Nonetheless, the assays demonstrated in this work offer a valuable method for evaluating the functional integrity of purified DNA ligase. The DNA ligation assay serves as a rapid and effective screening tool to assess the impact of the purification procedure on enzymatic activity. Additionally, the ligation assay can be used to evaluate the activity of DNA ligase after fluorescent labeling, ensuring the preservation of functional integrity.

***Fluorescent labeling of purified DNA* ligase** - To enable future single-molecule investigations of DNA ligase dynamics and substrate binding modes, a mutant library is designed for site-specific fluorescent labeling of the protein. The native amino acid residue 705 was selected for substitution with a cysteine (S705C) as it lies at the surface of the highly flexible OBD domain [26]. This positioning makes it an ideal site for conjugation with a fluorescent dye (ATTO 643) as it may allow us to capture the complex structural plasticity of the enzyme. To investigate undesired reactions between solvent-accessible native cysteines and the maleimide dye, two DNA ligase mutants, S705C-C842S and ΔZnF, were generated. The S705C-C842S mutant is designed to remove the cysteine residue at position 842, while the ΔZnF mutant lacks the Zinc-Finger domain, which also contains native cysteine residues. Ideally, only the S705C mutant will exhibit site-specific labeling with the fluorescent dye, while the S705C-C842S and ΔZnF mutants should show no evidence of fluorescent labeling due to the absence of cysteine residues accessible to the dye. This approach provides a means to indicate site-specific labeling and/or non-specific binding of the fluorescent dye. The mutants were subjected to a labeling reaction using the

fluorescent ATTO 643 dye, followed by characterization through FCS measurements. The attachment of the dye to the protein surface is expected to result in a lower diffusion rate compared to free dye in solution due to the difference in mass upon binding. Furthermore, the amount of fluorescent dye bound to the protein can be photospectromically estimated by measuring UV-Vis absorbances at 280 and 650 nm to determine the degree of labeling (DOL). This ratio represents the average number of dye molecules bound to each DNA ligase molecule. The FCS measurements in this study demonstrate the ACFs of DNA ligase mutants following a fluorescent labeling reaction (indicated by *) (**Fig. 9A**). All three mutants exhibited longer diffusion times (right-shift of ACF) compared to free dye ATTO 655 in solution. The diffusion coefficients of the labeled mutants (S705C*, S705C-C842S*, and ΔZnF*) were measured to be 55 $\mu m^2 s^{-1}$, 36 $\mu m^2 s^{-1}$ and 62 $\mu m^2 s^{-1}$, respectively. It is important to note that the diffusion coefficients are only reported to highlight the large difference in detected diffusion compared to free dye. However, no direct comparison between different mutants was made due to the reasons discussed below. Additionally, S705C* displayed a DOL of ~ 6% whereas ΔZnF* resulted in a remarkable DOL of ~ 89% (**Supplementary Fig. 5**).

The notable increase in diffusion time compared to free dye for all three mutants indicates the attachment of the fluorescent dye to the protein surface. Moreover, the increased diffusion times compared to free dye of both S705C-C842S* and ΔZnF* mutants (which do not contain the site-specific cysteine at residue 705) suggests that the fluorescent dye is attached to these mutants by means of non-site-specific binding. The calculated DOL values support the idea of fluorescent dye attached to the DNA ligase mutants S705C* and ΔZnF*. For the S705C* mutant, no conclusion can be made about the site-specific binding of the fluorescent dye or the accessibility of the labeling site. Moreover, it is not clear whether the DNA ligase mutant S705C-C842S* shows fluorescent labeling due to the reaction with native cysteines in the Zinc-Finger domain. Similarly, for the DNA ligase mutant ΔZnF*, it is unclear whether the dye has reacted to the protein via the cysteine at residue 842. Future studies should consider generating the respective DNA ligase mutant C842S-ΔZnF to determine if the removal of cysteine residues at position 842 and the zinc-finger domain eliminates unwanted dye reactions. It is important to note that non-site-specific binding of the maleimide dye could not only involve the conjugation to native cysteine residues but also reactivity towards amines and sticking to the protein by non-covalent interactions. Changes in the pH of the labeling buffer during the labeling reaction, especially at higher pH values (> pH 8), may favour the reaction between amines and maleimide dyes [69, 70]. The fluorescent dye employed in this work (ATTO 643) is a hydrophilic variant of the ATTO 647N dye and displays a decreased propensity for unspecific sticking [71]. However, the negative charge (-1) of the dye may interact electrostatically with the DNA ligase and cause non-site-specific binding (sticking). Interestingly, using high dye: thiol ratios during the fluorescent labeling procedure has been reported to result in non-site-specific binding of fluorescent dyes to proteins [69]. This should be taken into consideration for the 5-fold molar excess of dye that was initially used for the S705C* mutant compared to a 1.2-fold molar excess for the S705C-C842S* and ΔZnF* mutants during the labeling reaction. Nonetheless, the ACFs and measured diffusion coefficients are possibly affected by artifacts due to aggregate formation. When a single bright aggregate diffuses through the detection volume it can distort the autocorrelation function [49]. Although notable high photon counts from aggregates diffusing through the confocal volume are removed from the measurements [72], the possible formation of multiple, small protein oligomers cannot be excluded (**Supplementary Fig. 6**). Because of this reason, no quantitative information was extracted from the FCS measurements. Minimizing the formation of protein aggregates will thus be crucial to elucidate the state of the protein after fluorescent labeling and the degree of site- and/or non-site-specific binding. Modifying the ionic strength has been shown to stabilize proteins during the fluorescent labeling procedure [70]. Additionally, the use of detergent molecules that stabilize the proteins in solution can help overcome the formation of protein aggregates.

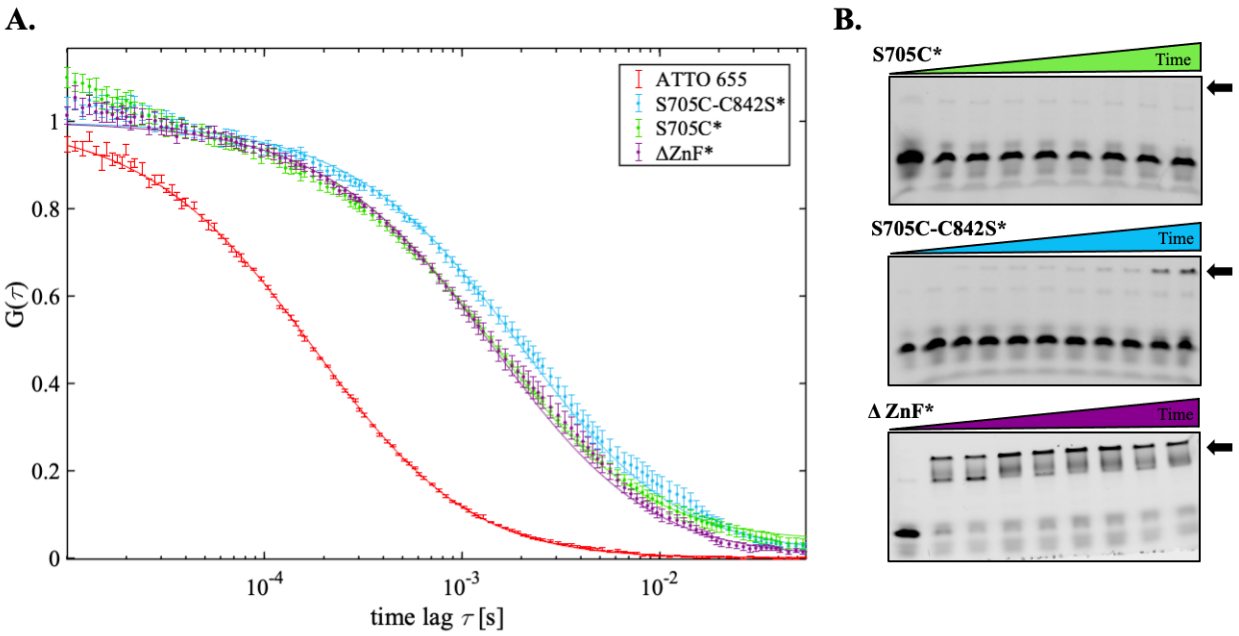


Figure 9: DNA ligase fluorescent labeling. A. Fluorescence correlation spectroscopy (FCS) measurements with auto-correlation functions (ACFs) of free dye (ATTO 655) compared to purified DNA ligase mutants following a fluorescent labeling reaction with ATTO 643 (S705C*, S705C-C842S* and ΔZnF*) in labeling buffer. All three DNA ligase mutants demonstrate an increased diffusion time compared to free dye. The ACF amplitudes are normalized (G (0)). Measurements were 60 seconds long and the dyes were excited with 50 µW laser power. **B.** Representative denaturing 18 % polyacrylamide, 8 M urea gels of all three mutants after the fluorescent labeling reaction, with the product formation indicated by an arrow. ATTO 655: red, S705C*: green, S705C-C842S*: blue and ΔZnF*: purple (see legend).

The binding of the fluorescent dye and the labeling procedure should not affect the DNA ligation activity of the enzyme. To investigate this, the DNA ligase mutants were screened for their DNA ligation activity following the fluorescent labeling reaction. The S705C* mutant showed no ligation activity after the fluorescent labeling procedure. Moreover, S705C-C842S* showed minimal activity at longer incubation times (8-10 min) whereas ΔZnF* showed high activity at all incubation times compared to WT DNA ligase (**Fig. 9B**). It should be mentioned that the mutant S705C-C842S showed a decreased ligation activity before the fluorescent labeling procedure compared to WT DNA ligase activity. The ligation activity of ΔZnF before the labeling reaction was described in the previous section.

Whether the disrupted ligation activity of both S705C* and S705C-C842S* was caused by non-site-specific dye binding, aggregation, and/or other factors is not conclusive. As the S705C mutant was not employed in the ligation assay before the fluorescent labeling procedure, no assumption can be made whether the ligation activity was disrupted before or/and after the labeling procedure. For further investigation, the DNA ligase mutants should be employed in the ligation assays, prior to the fluorescent labeling reaction for a direct comparison with WT DNA ligase activity. If the decreased activity is only observed after the fluorescent labeling procedure, it would suggest that the dye or the procedure is interfering with the protein's activity. In this work, only the S705C-C842S and ΔZnF mutants were employed in the DNA ligation assay prior to labeling. Comparing both the ligation activity of these mutants before (S705C-C842S, ΔZnF) (**Supplementary Fig. 3**) and after the fluorescent procedure (ΔZnF*, S705C-C842S*) (**Fig. 9B**), indicates that the enzymatic activity of these mutants was maintained throughout the labeling procedure. However, it is important to note that it is only assumed that the labeling procedure did not notably affect the ligation activity of these mutants. No conclusions can be made regarding the hindrance of the fluorescent dye with protein

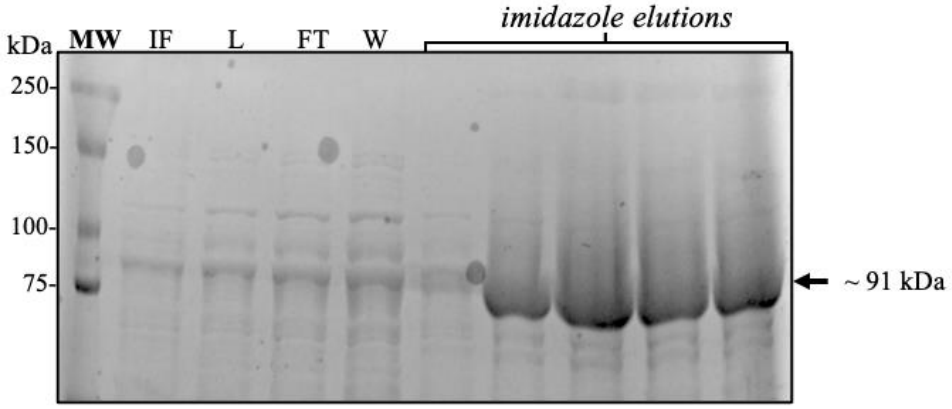
function based on the data presented here. It is important to ensure that the fraction of unsuccessfully labeled DNA ligase molecules is minimal and does not affect the evaluation of the ligation assay. If the fraction of labeled molecules is relatively low compared to the non-labeled molecules, the ligation assay could be dominated by the active non-labeled molecules and a change in ligation activity may be difficult to detect. In future work, it might be interesting to optimize the ligation assays and investigate the minimal concentrations of DNA ligase required for the visible detection of ligation activity. This stresses the importance of quantifying the fluorescently labeled DNA ligase molecules, which was not possible with the FCS measurements due to aggregate formation. If the estimated DOL of ~ 89% for ΔZnF* is accurate, it would imply that the attachment of the dye did not have a notable impact on its ligation function.

The separation of fluorescently labeled DNA ligase molecules from non-labeled molecules utilizing IEXC would facilitate the evaluation of ligation activity. As the ATTO 643 dye is negatively charged (-1), it may alter the net surface charge of the DNA ligase molecule upon binding. Consequently, the binding affinity of the DNA ligase molecule to the IEXC column may be affected. Other studies have isolated fluorescently labeled proteins from heterogeneous mixtures based on the labeling position and the number of fluorescent dyes [73, 74]. Employing this method to isolate fluorescently labeled DNA ligases will not only facilitate the evaluation of DNA ligase activity but also provide a homogeneous sample for future smFRET experiments. With additional optimization of the fluorescent labeling procedure and functionality assays, the enzyme can undergo a dual-labeling reaction allowing smFRET analysis. This will not only confirm the site-specific binding of the fluorescent dyes but also provide a first look into the structural plasticity of DNA ligases.

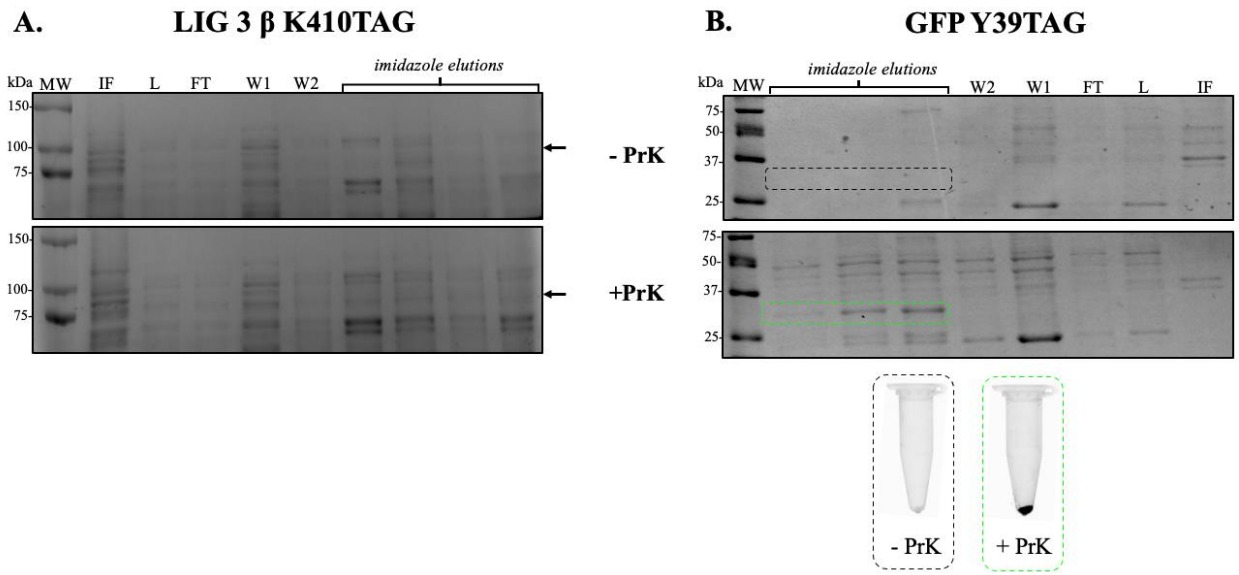
CONCLUSION

In this study, our primary objective was to purify, characterize, and site-specifically fluorescently label DNA ligases for future smFRET research. We successfully purified both wild-type DNA ligase and mutants suitable for fluorescent labeling. However, the incorporation of an unnatural amino acid into the protein sequence was unsuccessful due to non-optimized expression and purification conditions. Nonetheless, a minimalistic purification protocol without protease inhibitors maintained the functional integrity of the wild-type and mutant DNA ligase. This was demonstrated by the rapid ligation product formation in the DNA ligation assays and decreased DNA substrate diffusion rate in FCS measurements. Initial labeling reactions with a fluorescent dye showed potential dye attachment to the protein surface, aligning with our main research goal. However, non-site-specific binding and aggregate formation highlight the need for optimized purification and labeling procedures, such as using detergent molecules and adjusting the dye-to-protein ratio. Overall, this study provides evidence for the purification and labeling of human DNA ligases for smFRET research, advancing our understanding of their structure, function, and therapeutic potential.

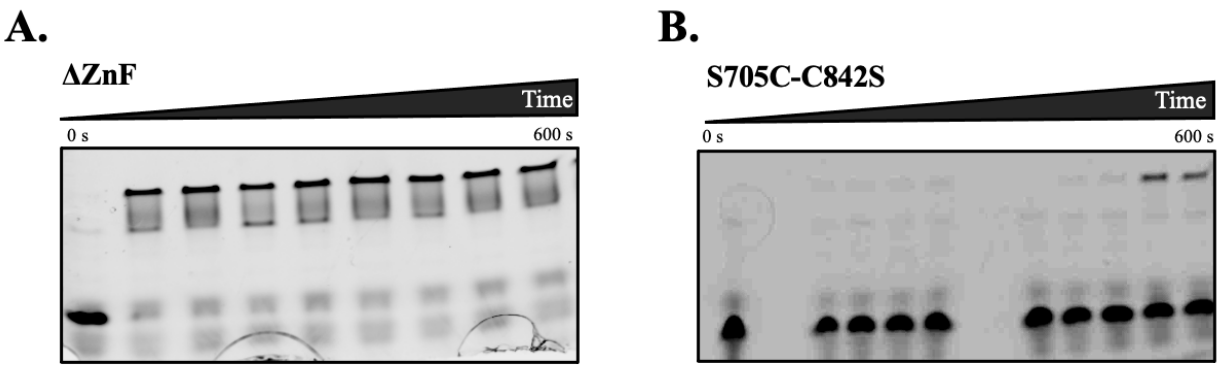
SUPPLEMENTARY INFORMATION



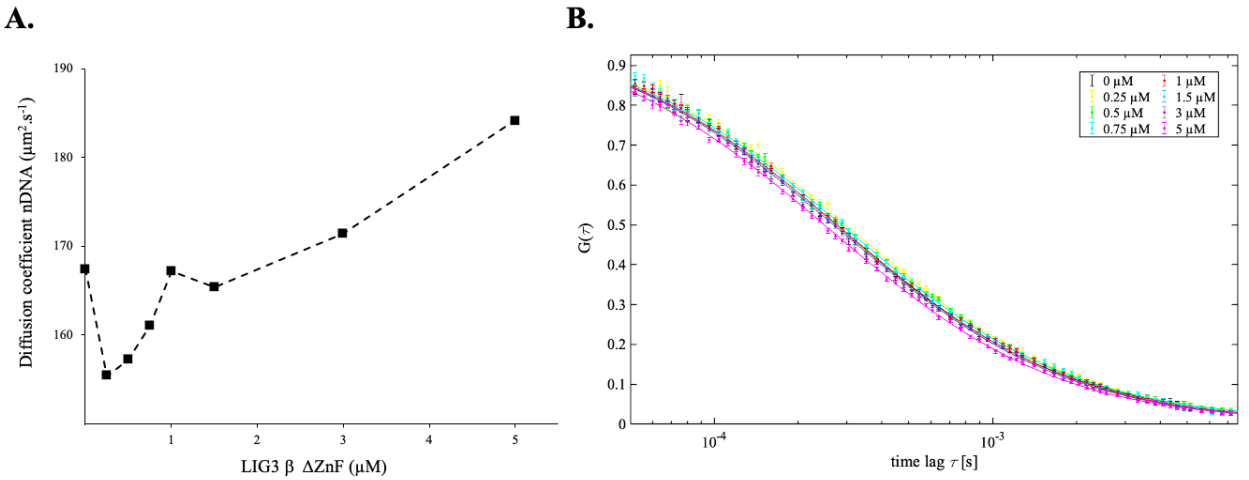
Supplementary Figure 1: **IMAC purification of ΔZnF.** 8% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) profile of purified recombinant DNA ligase 3β ΔZnF expressed in *E. coli* BL21(DE3). The 250 mM imidazole elutions represent the fractions containing the mutant (indicated by arrow: ~ 91 kDa) collected from immobilized metal affinity chromatography (IMAC) with Ni-NTA agarose resin. Lane MW: molecular weight marker proteins (75 – 250 kDa), lane IF: insoluble fraction after sonication of *E.coli* cells, lane L: cleared lysate after sonication, lane FT: flow-through of the column, lane W: washing of column.



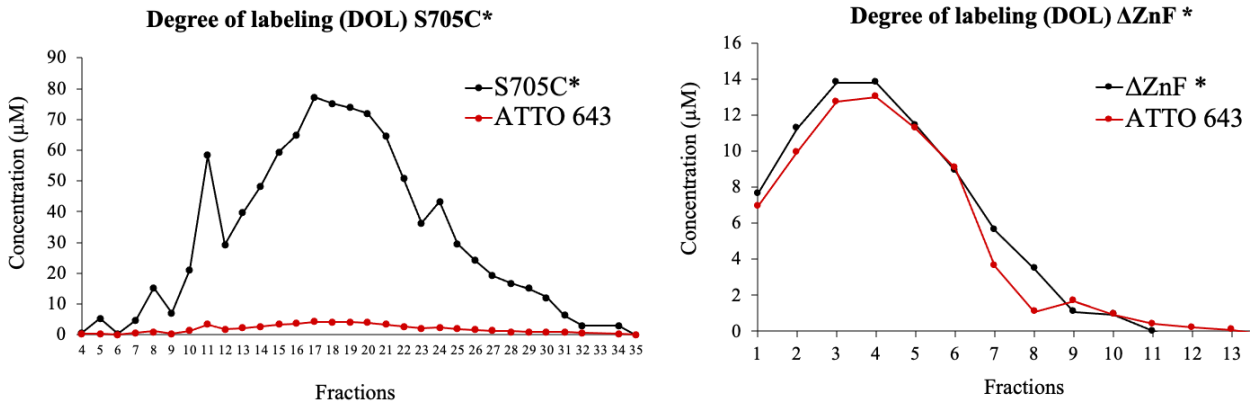
Supplementary Figure 2: IMAC purification of PrK mutants LIG 3β TAG and GFP TAG. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) profiles of purified recombinant DNA ligase 3β and mutant GFP with an Amber stop codon (TAG) expressed in *E. coli* BL21(DE3) and TOP10 cells. The proteins were expressed following the Propargyl-L-Lysine mutant expression protocol (see experimental procedure) and purified by a small batch immobilized metal affinity chromatography (IMAC) procedure with Ni-NTA agarose beads (resin). The protein was eluted with 250 mM imidazole. **A**. LIG 3β TAG SDS-PAGE profile (8%). **B.** GFP TAG SDS-PAGE profile (12%). Cells after expression were collected and imaged for fluorescence at 460/520 nm wavelengths. Lane MW: molecular weight marker proteins (50 – 250 kDa), lane IF: insoluble fraction after sonication of *E.coli* cells, lane L: cleared lysate after sonication, lane FT: flow-through of column (non-binding), lanes W1-W2: washing of column.



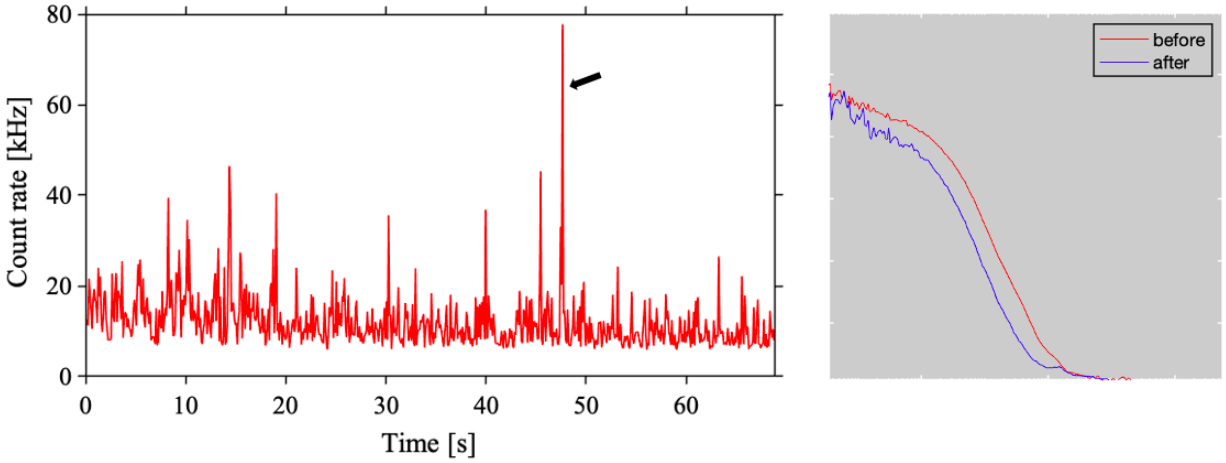
Supplementary Figure 3: DNA Ligase 3β ΔZnF and S705C-C842S mutant ligation assays. Representative denaturing 18 % polyacrylamide, 8 M urea gel with the denatured product (27-mer), and substrate (11-mer). Reactions contained 190 nM fluorescently labeled DNA substrate, and 50 nM LIG 3β in standard reaction buffer at 150 mM ionic strength and were quenched at designated times (20, 40, 60, 80, 120, 240, 480, and 600 seconds) at 37°C. The purified DNA ligase was pre-incubated with 20 mM $MgCl_2$ and 1 mM ATP to ensure 100% adenylyation. **A.** ΔZnF ligation assay. **B.** S705C-C842S ligation assay (DNA-containing wells are separated by two empty wells in 15-well gel).



Supplementary Figure 4: FCS DNA ligase 3β ΔZnF binding assay. Purified LIG 3β ΔZnF mutant (0-5 μM) was titrated to 3' fluorescein-labeled (FAM) nicked DNA substrate and the rate of diffusion was measured with fluorescence correlation spectroscopy (FCS) in reaction buffer (150 mM NaCl and 20 mM $MgCl_2$). **A**. diffusion coefficients of the FAM-DNA after titration of DNA ligase ΔZnF. **B.** autocorrelation functions (ACFs) from the FCS experiment (0-5 μM DNA ligase: see legend). Note that the FCS titration experiments are performed in a way that for each increase of DNA ligase concentration, the fluorescein-labeled DNA concentration decreases and results in different initial amplitudes between each titration. Therefore, the auto-correlation functions (ACFs) of different DNA ligase concentrations are normalized such that the initial amplitudes are equal to each other (=1) and the decay rates of the ACFs can be compared. Measurements were 60 seconds long and fluorescein was excited with 100 μW laser power.



Supplementary Figure 5: Degree of labeling (DOL). DNA ligase 3β S705C and ΔZnF mutants were employed in a labeling reaction (*) with ATTO 643 and collected in fractions after separation from free unbound dye utilizing gel filtration (PD-10). The DOL is calculated by measuring the absorbances at 280 nm (protein) and 650 nm (dye) and determination of the dye: protein ratio. Concentrations are calculated with measured absorbance at 1 mm path length and extinction coefficients (ε_{S705C}: 66800 $M^{-1}cm^{-1}$ and $\varepsilon_{\Delta ZnF}$: 52830 $M^{-1}cm^{-1}$).



Supplementary Figure 6: Aggregates affect FCS measurements. Aggregates diffusing through the confocal volume emit high bursts of photons and distort the autocorrelation function (ACF). The intensity trace shows large peaks (arrow) indicating the detection of an aggregate(s). Removal of the aggregate(s) results in a notable shift of the ACF curve. See PAM software [72].

REFERENCES

1. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. Nature Reviews Cancer. 2016;16(1):35-42.
2. Schumacher B, Pothof J, Vijg J, Hoeijmakers JHJ. The central role of DNA damage in the ageing process. Nature. 2021;592(7856):695-703.
3. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362(6422):709-15.
4. Schipler A, Iliakis G. DNA double-strand–break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Research. 2013;41(16):7589-605.
5. McNally JR, Ames AM, Admiraal SJ, O'Brien PJ. Human DNA ligases I and III have stand-alone end-joining capability, but differ in ligation efficiency and specificity. Nucleic Acids Research. 2023;51(2):796-805.
6. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361(2):123-34.
7. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434(7035):917-21.
8. Rosenthal AS, Dexheimer TS, Gileadi O, Nguyen GH, Chu WK, Hickson ID, et al. Synthesis and SAR studies of 5-(pyridin-4-yl)-1,3,4-thiadiazol-2-amine derivatives as potent inhibitors of Bloom helicase. Bioorg Med Chem Lett. 2013;23(20):5660-6.
9. Hopkins JL, Lan L, Zou L. DNA repair defects in cancer and therapeutic opportunities. Genes Dev. 2022;36(5-6):278-93.
10. Hewitt G, Borel V, Segura-Bayona S, Takaki T, Ruis P, Bellelli R, et al. Defective ALC1 nucleosome remodeling confers PARPi sensitization and synthetic lethality with HRD. Molecular Cell. 2021;81(4):767-83.e11.
11. Tomkinson AE, Howes TR, Wiest NE. DNA ligases as therapeutic targets. Transl Cancer Res. 2013;2(3).
12. Caracciolo D, Juli G, Riillo C, Coricello A, Vasile F, Pollastri S, et al. Exploiting DNA Ligase III addiction of multiple myeloma by flavonoid Rhamnetin. Journal of Translational Medicine. 2022;20(1):482.
13. Tomkinson AE, Naila T, Khattri Bhandari S. Altered DNA ligase activity in human disease. Mutagenesis. 2019;35(1):51-60.
14. Sallmyr A, Rashid I, Bhandari SK, Naila T, Tomkinson AE. Human DNA ligases in replication and repair. DNA Repair (Amst). 2020;93:102908.
15. Altmann T, Gennery AR. DNA ligase IV syndrome; a review. Orphanet Journal of Rare Diseases. 2016;11(1):137.
16. Tomkinson AE, Vijayakumar S, Pascal JM, Ellenberger T. DNA Ligases: Structure, Reaction Mechanism, and Function. Chemical Reviews. 2006;106(2):687-99.
17. Riballo E, Critchlow SE, Teo SH, Doherty AJ, Priestley A, Broughton B, et al. Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. Curr Biol. 1999;9(13):699-702.
18. Barnes DE, Tomkinson AE, Lehmann AR, Webster AD, Lindahl T. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. Cell. 1992;69(3):495-503.
19. Bonora E, Chakrabarty S, Kellaris G, Tsutsumi M, Bianco F, Bergamini C, et al. Biallelic variants in LIG3 cause a novel mitochondrial neurogastrointestinal encephalomyopathy. Brain. 2021;144(5):1451-66.
20. Cotner-Gohara E, Kim I-K, Hammel M, Tainer JA, Tomkinson AE, Ellenberger T. Human DNA Ligase III Recognizes DNA Ends by Dynamic Switching between Two DNA-Bound States. Biochemistry. 2010;49(29):6165-76.
21. Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T. Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature. 2004;432(7016):473-8.

22. Taylor RM, Whitehouse CJ, Caldecott KW. The DNA ligase III zinc finger stimulates binding to DNA secondary structure and promotes end joining. Nucleic Acids Res. 2000;28(18):3558-63.
23. Bauer RJ, Evans TC, Jr., Lohman GJS. The Inhibitory Effect of Non-Substrate and Substrate DNA on the Ligation and Self-Adenylylation Reactions Catalyzed by T4 DNA Ligase. PLOS ONE. 2016;11(3):e0150802.
24. Bauer RJ, Jurkiw TJ, Evans TC, Jr., Lohman GJS. Rapid Time Scale Analysis of T4 DNA Ligase–DNA Binding. Biochemistry. 2017;56(8):1117-29.
25. Kaminski AM, Tumbale PP, Schellenberg MJ, Williams RS, Williams JG, Kunkel TA, et al. Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. Nature Communications. 2018;9(1):2642.
26. Sverzhinsky A, Tomkinson AE, Pascal JM. Cryo-EM structures and biochemical insights into heterotrimeric PCNA regulation of DNA ligase. Structure. 2022;30(3):371-85.e5.
27. Mazal H, Haran G. Single-molecule FRET methods to study the dynamics of proteins at work. Curr Opin Biomed Eng. 2019;12:8-17.
28. Wu X, Xu G, Li X, Xu W, Li Q, Liu W, et al. Small Molecule Inhibitor that Stabilizes the Autoinhibited Conformation of the Oncogenic Tyrosine Phosphatase SHP2. Journal of Medicinal Chemistry. 2019;62(3):1125-37.
29. Bartnik K, Barth A, Pilo-Pais M, Crevenna AH, Liedl T, Lamb DC. A DNA Origami Platform for Single-Pair Förster Resonance Energy Transfer Investigation of DNA–DNA Interactions and Ligation. Journal of the American Chemical Society. 2020;142(2):815-25.
30. Yang Z, Xu H, Wang J, Chen W, Zhao M. Single-Molecule Fluorescence Techniques for Membrane Protein Dynamics Analysis. Applied Spectroscopy. 2021;75(5):491-505.
31. Feng XA, Poyton MF, Ha T. Multicolor single-molecule FRET for DNA and RNA processes. Current Opinion in Structural Biology. 2021;70:26-33.
32. Lerner E, Barth A, Hendrix J, Ambrose B, Birkedal V, Blanchard SC, et al. FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. Elife. 2021;10.
33. Gryczynski ZK. FCS imaging--a way to look at cellular processes. Biophys J. 2008;94(6):1943-4.
34. Lee G, Yoo J, Leslie BJ, Ha T. Single-molecule analysis reveals three phases of DNA degradation by an exonuclease. Nat Chem Biol. 2011;7(6):367-74.
35. Kilic S, Felekyan S, Doroshenko O, Boichenko I, Dimura M, Vardanyan H, et al. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. Nat Commun. 2018;9(1):235.
36. Maslov I, Volkov O, Khorn P, Orekhov P, Gusach A, Kuzmichev P, et al. Sub-millisecond conformational dynamics of the A2A adenosine receptor revealed by single-molecule FRET. Communications Biology. 2023;6(1):362.
37. Bartels K, Lasitza-Male T, Hofmann H, Löw C. Single-Molecule FRET of Membrane Transport Proteins. ChemBioChem. 2021;22(17):2657-71.
38. Sefer A, Kallis E, Eilert T, Röcker C, Kolesnikova O, Neuhaus D, et al. Structural dynamics of DNA strand break sensing by PARP-1 at a single-molecule level. Nat Commun. 2022;13(1):6569.
39. Yukhnovets O, Höfig H, Bustorff N, Katranidis A, Fitter J. Impact of Molecule Concentration, Diffusion Rates and Surface Passivation on Single-Molecule Fluorescence Studies in Solution. Biomolecules. 2022;12(3).
40. Spinks RR, Spenkelink LM, Stratmann SA, Xu Z-Q, Stamford N Patrick J, Brown SE, et al. DnaB helicase dynamics in bacterial DNA replication resolved by single-molecule studies. Nucleic Acids Research. 2021;49(12):6804-16.
41. Fijen C, Mahmoud MM, Kronenberg M, Kaup R, Fontana M, Towle-Weicksel JB, et al. Using single-molecule FRET to probe the nucleotide-dependent conformational landscape of polymerase β-DNA complexes. J Biol Chem. 2020;295(27):9012-20.
42. Sánchez-Rico C, Voith von Voithenberg L, Warner L, Lamb DC, Sattler M. Effects of Fluorophore Attachment on Protein Conformation and Dynamics Studied by spFRET and NMR Spectroscopy. Chemistry. 2017;23(57):14267-77.

43. Lemke E. Site-Specific Labeling of Proteins for Single-Molecule FRET Measurements Using Genetically Encoded Ketone Functionalities. Methods in molecular biology (Clifton, NJ). 2011;751:3-15.
44. Wu P. The Nobel Prize in Chemistry 2022: Fulfilling Demanding Applications with Simple Reactions. ACS Chemical Biology. 2022;17.
45. Swiderska KW, Szlachcic A, Czyrek A, Zakrzewska M, Otlewski J. Site-specific conjugation of fibroblast growth factor 2 (FGF2) based on incorporation of alkyne-reactive unnatural amino acid. Bioorg Med Chem. 2017;25(14):3685-93.
46. Zhou H, Cheung JW, Carpenter T, Jones SK, Luong NH, Tran NC, et al. Enhancing the incorporation of lysine derivatives into proteins with methylester forms of unnatural amino acids. Bioorganic & Medicinal Chemistry Letters. 2020;30(2):126876.
47. Chen X, Pascal J, Vijayakumar S, Wilson GM, Ellenberger T, Tomkinson AE. Human DNA ligases I, III, and IV-purification and new specific assays for these enzymes. Methods Enzymol. 2006;409:39-52.
48. Wandrey G, Wurzel J, Hoffmann K, Ladner T, Büchs J, Meinel L, et al. Probing unnatural amino acid integration into enhanced green fluorescent protein by genetic code expansion with a high-throughput screening platform. Journal of Biological Engineering. 2016;10(1):11.
49. Yu L, Lei Y, Ma Y, Liu M, Zheng J, Dan D, et al. A Comprehensive Review of Fluorescence Correlation Spectroscopy. Frontiers in Physics. 2021;9.
50. McNally JR, O'Brien PJ. Kinetic analyses of single-stranded break repair by human DNA ligase III isoforms reveal biochemical differences from DNA ligase I. J Biol Chem. 2017;292(38):15870-9.
51. Cotner-Gohara E, Kim I-K, Tomkinson AE, Ellenberger T. Two DNA-binding and Nick Recognition Modules in Human DNA Ligase III*. Journal of Biological Chemistry. 2008;283(16):10764-72.
52. Wingfield PT. Overview of the purification of recombinant proteins. Curr Protoc Protein Sci. 2015;80:6.1.-6.1.35.
53. Doherty AJ, Ashford SR, Wigley DB. Characterization of Proteolytic Fragments of Bacteriophage T7 DNA Ligase. Nucleic Acids Research. 1996;24(12):2281-7.
54. Bauer RJ, Evans TC, Jr., Lohman GJ. The Inhibitory Effect of Non-Substrate and Substrate DNA on the Ligation and Self-Adenylylation Reactions Catalyzed by T4 DNA Ligase. PLoS One. 2016;11(3):e0150802.
55. Fykse EM, Olsen JS, Skogan G. Application of sonication to release DNA from Bacillus cereus for quantitative detection by real-time PCR. J Microbiol Methods. 2003;55(1):1-10.
56. Wingfield PT, Stahl SJ, Payton MA, Venkatesan S, Misra M, Steven AC. HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. Biochemistry. 1991;30(30):7527-34.
57. Marcovitz A, Levy Y. Frustration in protein-DNA binding influences conformational switching and target search kinetics. Proc Natl Acad Sci U S A. 2011;108(44):17957-62.
58. Andersen KR, Leksa NC, Schwartz TU. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins. 2013;81(11):1857-61.
59. Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. Biochimica et Biophysica Acta (BBA) - General Subjects. 2006;1760(9):1304-13.
60. Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 2000;326:245-54.
61. Young TS, Ahmad I, Yin JA, Schultz PG. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. Journal of Molecular Biology. 2010;395(2):361-74.
62. Koehler C, Sauter PF, Wawryszyn M, Girona GE, Gupta K, Landry JJM, et al. Genetic code expansion for multiprotein complex engineering. Nature Methods. 2016;13(12):997-1000.
63. Milles S, Tyagi S, Banterle N, Koehler C, VanDelinder V, Plass T, et al. Click Strategies for Single-Molecule Protein Fluorescence. Journal of the American Chemical Society. 2012;134(11):5187-95.
64. Plass T, Milles S, Koehler C, Szymański J, Mueller R, Wießler M, et al. Amino Acids for Diels–Alder Reactions in Living Cells. Angewandte Chemie International Edition. 2012;51(17):4166-70.

65. Ponzar N, Pozzi N. Probing the conformational dynamics of thiol-isomerases using non-canonical amino acids and single-molecule FRET. Methods. 2023;214:8-17.
66. Taylor MR, Conrad JA, Wahl D, O'Brien PJ. Kinetic mechanism of human DNA ligase I reveals magnesium-dependent changes in the rate-limiting step that compromise ligation efficiency. The Journal of biological chemistry. 2011;286(26):23054-62.
67. Tomkinson AE, Della-Maria JA. DNA Ligases: Mechanism and Functions. In: Lennarz WJ, Lane MD, editors. Encyclopedia of Biological Chemistry (Second Edition). Waltham: Academic Press; 2013. p. 28-32.
68. Doherty AJ, Suh SW. Structural and mechanistic conservation in DNA ligases. Nucleic Acids Res. 2000;28(21):4051-8.
69. Tyagarajan K, Pretzer E, Wiktorowicz JE. Thiol-reactive dyes for fluorescence labeling of proteomic samples. ELECTROPHORESIS. 2003;24(14):2348-58.
70. Modesti M. Fluorescent Labeling of Proteins. Methods Mol Biol. 2018;1665:115-34.
71. Hübner K, Joshi H, Aksimentiev A, Stefani FD, Tinnefeld P, Acuna GP. Determining the In-Plane Orientation and Binding Mode of Single Fluorescent Dyes in DNA Origami Structures. ACS Nano. 2021;15(3):5109-17.
72. Schrimpf W, Barth A, Hendrix J, Lamb DC. PAM: A Framework for Integrated Analysis of Imaging, Single-Molecule, and Ensemble Fluorescence Data. Biophys J. 2018;114(7):1518-28.
73. Teske CA, Simon R, Niebisch A, Hubbuch J. Changes in retention behavior of fluorescently labeled proteins during ion-exchange chromatography caused by different protein surface labeling positions. Biotechnol Bioeng. 2007;98(1):193-200.
74. Zosel F, Holla A, Schuler B. Labeling of Proteins for Single-Molecule Fluorescence Spectroscopy. Methods Mol Biol. 2022;2376:207-33.

Acknowledgements – First and foremost, I would like to thank my daily supervisor and mentor, Tom Kache. Besides being a role model for every student who aspires to become a scientist, he is also one of the kindest and most generous people I have ever met. Truly a person that I look up to. Being his student was an extremely rewarding experience and an opportunity that I will remember and cherish. Also, I want to express my gratitude towards my promotor, Jelle Hendrix. He has given me the opportunity to join this project and broaden my knowledge and expertise. Although sometimes scary, his critical attitude and directness are traits that I value and respect. A special thanks to Nastya and Pedro for helping me out in the lab and providing me with useful knowledge when needed. Of course, the work that I have delivered would not be feasible without the strength of Igna, Petra, Regine, Ellen, and all the other people working in the lab, as they keep the institution from collapsing. Thank you to Kristina and Veronique for the lab tours provided at the start of the internship and for opening the door every time I forgot my badge. Also, thank you to Wouter Maes and Geert-Jan Graulus for taking the time to enlighten me with knowledge that I could integrate during my internship. All the love towards my friends, family, and beautiful girlfriend Lobke for always supporting me. Last but not least, I want to thank Sam, Stijn, Ivan, Laura, Reindert, Kinga, and Gabriele for giving me a good time and providing a helping hand when needed. Maybe the most important thing that I have learned during my internship is that, as long as you are with the right people, sitting in a basement for half a year is actually not that bad.

Author contributions – Tom Kache and Prof. Dr. Jelle Hendrix conceived and designed the research. Tom Kache and Lambert-Paul Jorissen performed experiments. Anastasiia Smolentseva helped with the DNA isolation and bacterial transformations. Lambert-Paul Jorissen performed the data analysis and wrote the paper.