Norwegian University
of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Biosciences

Characterization and Optimization of the ParB-INT Imaging Tool

Martine Mesel Isom
Biotechnology, Molecular Biology

Acknowledgements

The work presented in the following thesis was carried out in the laboratory of Ragnhild Eskeland at the Department of Molecular Medicine, University in Oslo, from August 2019 through May 2020. The project was funded by FRIPRO, Norwegian Research Council. The following experiments could not have taken place were it not for the generous gift of ParB constructs from Prof. Kerstin Bystricky, Paul Sabatier University Tolouse III, France.

My deepest gratitude goes to my main external supervisor at UiO, Ragnhild Eskeland, for her deep involvement in my project, for believing in me, and for constantly encouraging me. The passion you have for your work is inspiring, and I am thankful for the rewarding experience of working in your group.

I also want to thank my main internal supervisor at NMBU, Dag Inge Våge, for his support throughout this project, and for always being available to answer my many questions.

I am forever thankful to my co-supervisors at UiO, Naima Azouzi and Beata Nadratowska-Wesolowska. Naima, thank you for teaching me so much, for constantly encouraging me, and always sharing of your time. You always knew how to put a smile back on my face when I was discouraged. Beata, thank you for sharing of your knowledge and experience, and for working tirelessly to guide and support me through the final part of my project.

I want to thank Marit Ledsaak, with whom I have been fortunate to spend many days in the lab, and who always so kindly shares of her knowledge and experience.

I also want to thank Ankush Sharma for his support and guidance in the bioinformatics part of my project.

I am grateful to the entire Chromatin Biology Group for interesting conversations, laughs, and for promoting a good social and scientific environment. I especially want to express my gratitude to Aruna and Tasmia, whose friendship has meant so much to me during this time.

I want to thank my amazing family for their unfailing support and love. Last but not least, I want to thank my ever-supporting husband Alex for his patience and love.

Martine Mesel Isom

Martine Mesel Isom

July 11th 2020

Oslo

Abstract

Partitioning B (ParB) proteins are key to the prokaryotic partitioning machinery which ensures proper segregation of replicated plasmid DNA strands. ParB proteins bind specifically to a centromere-like *parS* sequence, from which dimerized partition complexes distribute to flanking regions over several kbp in a mechanism termed spreading. These proteins and *parS* sequence have been repurposed for the ParB_INT imaging system to visualize single gene loci in living cells by Bystricky and colleagues. By integrating a relatively short *parS* sequence close to a sequence of interest and ectopically expressing a fluorescent ParB protein, the spreading of ParB upstream and downstream on the DNA creates a visible spot for the locus detectable by light microscopy.

ParB proteins may possess nuclease activity, which could have a negative impact on the genomic region when ParB spreads over *parS* and neighboring sequences. We have predicted the structure of ParB proteins and performed an evolutionary analysis of ParB proteins. We suggest that the nuclease activity is likely located to the two highly conserved regions, BOX I and BOX II.

I constructed 15 different ParB clones. Different ParB3 proteins were tested for nuclease activity: the original prokaryotic protein, a humanized version, and a truncated core domain. I also show that the original prokaryotic protein and the humanized version may possess nuclease activity, whereas the structural core domain of ParB3 does not exhibit the same activity in the presence of plasmid DNA. The results from the nuclease assay need further validation after improvement of the recombinant purification protocol. Moreover, we found that the humanized ParB protein did not improve the ParB-INT imaging system in OS25 cells.

Sammendrag

Partisjoneringsprotein B (ParB) er et nøkkelprotein for bakteriers evne til å segregere replikerte plasmider til cellepolene under celledeling. ParB-proteiner binder til en sentromer-liknende *parS*-sekvens. Fra denne sekvensen distribueres dimeriserte ParB-proteiner i begge retninger over flere kbp. Denne «spredningss»-mekanismen har blitt utnyttet som et verktøy til å visualisere enkelt genlokus in levende celler av Bystricky og kollegaer. Ved å integrere den kilobaser-lange *parS* -sekvensen i nærheten av en sekvens av interesse, og ektopisk uttrykke et fluoriserende ParB-protein, vil ParB koblet til et fluoriserende protein gjøre lokuset synlig gjennom lysmikroskopi.

Tidligere studier har beskrevet ParB som en nuklease, noe som kan ha negative innvirkninger på genomisk DNA i cellen når ParB binder seg til de integrerte *parS*-sekvensene. Vi har i denne studien predikert strukturen til ParB-proteiner og utført en evolusjonær analyse av ParB-homologer, hvor vi observerte at nuklease-aktiviteten til proteinet kan tilegnes aminosyrer som befinner seg i to svært konserverte områder, BOX I og BOX II. Disse områdene kan benyttes til videre mutasjonsstudier.

Jeg konstruerte 15 forskjellige ParB-kloner. Ulike ParB3-proteiner ble undersøkt for nuklease-aktivitet: det originale prokaryote proteinet, en humanisert versjon og et protein bestående av kjernedomenet.

Jeg detekterte nuklease-aktivitet hos både det originale prokaryote proteinet og den humaniserte versjonen. Derimot virker ikke kjernedomenet til å uttrykke liknende aktivitet med plasmid-DNA tilstede. Nuklease-analysen krever flere underbyggende analyser for å bekrefte resultatene, etter en forbedret protokoll for opprensing er utarbeidet. Videre så vi ingen forskjell på humanisert og prokaryot versjon av ParB3 ved visualisering i stamceller fra muse-embryo med en integrert *parS* sekvens.

Table of Contents

1 Introduction

1.1 The Eukaryotic Genome

Cells in a multicellular eukaryotic organism contain the same genetic code. The genome stores all information that is needed for development, sustainability, and reproduction. The relatively large size of the eukaryotic genome poses a great challenge to the cells and requires a highly organized and sophisticated packaging mechanism that still allows for access to coding and non-coding parts of the genome. The eukaryotic genome is well structured, highly variable between species, and unique for each individual. Each cell contains a complete set of chromosomes encoded with this unique information necessary for the survival of the organism. Tissue formation relies on well-regulated expression of the tissue-specific genes, and the complexity of transcriptional regulation is continuously becoming unveiled.

1.2 Organization of DNA in the Nucleus

Unlike the prokaryotic cells, eukaryotic cells protect their genomes inside a cellular compartment enclosed by a porous nuclear envelope. The nuclear envelope is a lipid bilayer membrane interspaced with large pores for nuclear transport. DNA is generally visualized as densely packaged mitotic chromosomes **(Figure 1.1)**. To achieve this density, the linear DNA molecule is tightly wound around a histone core, forming a nucleoprotein complex referred to as the nucleosome. The histone core is an octamer consisting of four positively charged proteins: H2A, H2B, H3, and H4. The histone proteins are bound together as two copies of a H2A-H2B and H3-H4 dimer, assembled into the compacted octamer which DNA is wrapped around (McGinty and Tan, 2015; Onufriev and Schiessel, 2019). Another histone, the linker histone H1, is not a part of the core of the nucleosome but will associate with the linker DNA (DNA between two nucleosome cores) and wrap the DNA molecule more tightly around the histone core, in an effort to further compact the DNA (McGinty and Tan, 2015). Histones, in particular the tails that protrude out of the nucleosome, are subject to post-translational modifications (PTM's), including methylation, acetylation, phosphorylation, and ubiquitinylation (Bannister and Kouzarides, 2011; Tessarz and Kouzarides, 2014). The purpose these modifications is to regulate histone-histone and histone-DNA interactions, which further act as binding platforms to regulate the accessibility of DNA sequences for other proteins like the transcription machinery components (Fraser et al., 2015). The nucleosome is the lowest

level of chromosome packaging. These are stacked on top of each other to further condense the DNA strands and form a packed 30 nm in diameter chromatin fiber as shown in **Figure 1.1**, although this strict assembly has been a topic of debate and the packing density is likely more variable (McGinty and Tan, 2015; Wako et al., 2020). These fibers will again loop and condense into the chromatin fiber we know as chromosomes (Alberts et al., 2015, pp. 193–209). However, this compacted structure does not form until the metaphase of mitosis. Interphase is the cell cycle phase where cell growth and DNA synthesis occur, when chromatin is in the least condensed state, uncoiled into loose strands that are distributed in the nuclear space. During interphase, the genomic DNA localize into two distinct functional compartments reflecting the level of activity in the respective area of the molecules. Heterochromatin are the predominantly transcriptionally inactive chromatin molecules. It is densely packed, gene poor, and reside largely in the periphery of the nuclear envelope and around the nucleolus (Bannister and Kouzarides, 2011; Bickmore, 2013). Euchromatin, the active chromatin, is loosely packed to give access to DNA binding proteins, gene rich, and localized more to the center of the nucleus although accessible to the nuclear pore complexes (Fraser et al., 2015; Hoffbrand et al., 2019). The nucleolus also has an important organizational function in the nucleus. This structural region contains the ribosomal genes, and is the site of ribosomal RNA (rRNA) synthesis and assembly of the ribosome from its subunits (Hoffbrand et al., 2019). Heterochromatin and euchromatin are the most easily distinguished ways that chromatin is organized in the nucleus. Nevertheless, a strict organization of individual chromosomes have been observed where chromosomes reside in territories and have specific radial positions within the nucleus (Bickmore, 2013; Cremer and Cremer, 2001; Kempfer and Pombo, 2020).

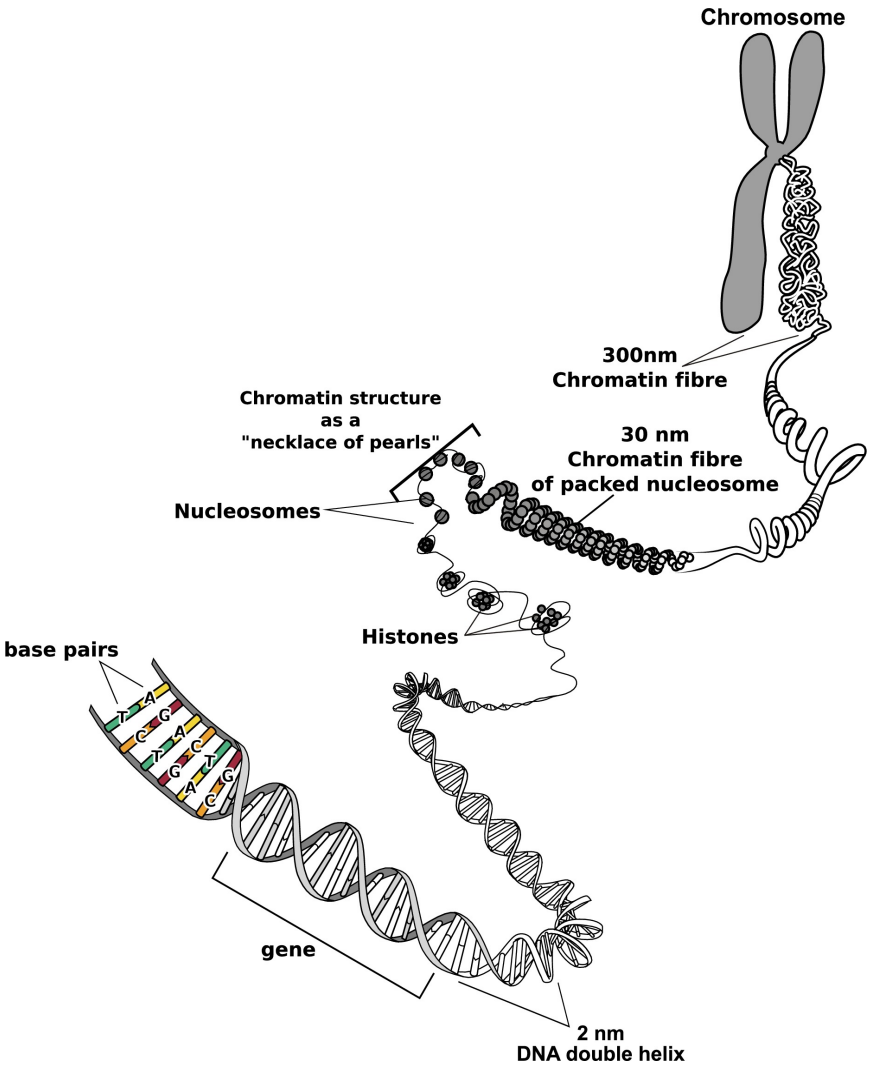


Figure 1.1. Schematic overview of the packaging of DNA into chromosomes. DNA is wound around histones to form the nucleosomes. The nucleosomes are further assembled into higher order structures called chromatin fibers that are densely packaged into mitotic chromosomes. Image is adapted from Giancarlo, Rombo & Utro (2019).

1.3 Chromatin-Chromatin Interactions

Individual chromosomes locate to discreate areas of the nucleus and form chromosomal territories, where the location of diploid chromosomes can be far from each other in this rearrangement. Further knowledge about the organization of chromatin launched extensive research on how interactions between and within chromosomes occur. Distribution of chromosomes into distinct regions makes interchromosomal interactions less available, although specific sequences have been observed to be looping out of their territory into another. However, intrachromosomal activity has been an area of growing interest. The spatial organization of chromatin has been demonstrated to be highly relevant for transcriptional

regulation (Fraser et al., 2015). Enhancers are regulatory elements that act in *cis* on promoters, and their spatial organization appears to be instrumental to transcriptional regulation by forming loops engaging enhancer and promoter interactions (Bickmore, 2013; Grosveld et al., 1998). Moreover, the physical movement of enhancer regions in relation to promoter regions appear to play a significant role in tissue formation through regulation of transcription (Souaid et al., 2018). The quantification of interactions between regulatory enhancers and promoters can largely be attributed to the emergence of the Chromosome Conformation Capture (3C) technologies (Dekker et al., 2002) in combination with fluorescence *in situ* hybridization (FISH) techniques (Langer-Safer et al., 1982). These techniques have enabled identification and analysis of interacting sequences in the genome. This has revealed that promoters are usually regulated by more than one enhancer and one enhancer can regulate multiple genes (Fraser et al., 2015). Moreover, enhancers rarely act on their most proximal genes, but rather on the more distal genes (Souaid et al., 2018). By gaining more insight into how promotors are activated and the general dynamic changes in the chromatin structures, deeper understanding can be gained about, for example, disease-causing genes and treatment.

1.4 Visualization of Genomic Sites in Eukaryotic Cells

The ability to visualize DNA in eukaryotic cells allows for studies of the dynamic movement of chromatin and spatial organization in the nucleus, which is vital to our understanding of how transcription is regulated and how dynamic changes are involved metabolic processes (Bystricky, 2015; Germier et al., 2017). Many methods have been developed for this purpose, and can be divided into two main categories based on the condition of the cells used: fixed cells or live cells.

1.4.1 Visualizing DNA in Fixed Eukaryotic Cells

Fluorescent *in-situ* hybridization (FISH) is a technique that emerged in the early 1980's and is still considered an essential tool in cytogenetics today (Langer-Safer et al., 1982). A common application for FISH is to study the spatial organization of chromatin in the nucleus. The method is based on the inherent property of DNA to denature and then renature (Pernthaler et al., 2001). A fluorescently labeled DNA probe hybridizes to a target DNA sequence, which creates a fluorescent focus detectable by microscopy **(Figure 1.2)**. The DNA probe can be labeled either directly, where the probe itself is built with fluorescently tagged nucleotides, or indirectly, where a non-fluorescent hapten is incorporated on the probe and fluorescently

labeled antibodies are bound to the hapten at a later stage (Fraser et al., 2015). Drawbacks to this technique is that fixed cells provide limited temporal dynamic information, it may cause tissues to exhibit autofluorescence that interferes with detection of the probe fluorescence, and the architectural integrity of DNA following denaturation may be compromised (Pernthaler et al., 2001; Wu et al., 2019). Several new technologies have emerged from the traditional FISH for a variety of applications, including 2D FISH, 3D FISH, and cryoFISH (Fraser et al., 2015).

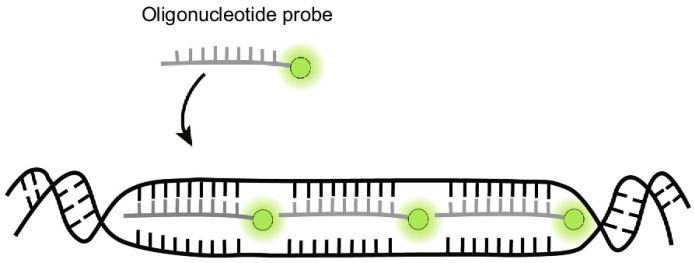


Figure 1.2. Schematic overview of hybridization of fluorescently tagged probe to a target DNA sequence. Image adapted from X. Wu et al. (2019).

The chromosome conformation capture (3C) technology enabled a new way to study chromatin contacts in the nucleus (Dekker et al., 2002). The 3C technology identifies and isolates interacting sequences by crosslinking proximal chromatin regions with formaldehyde. The cross-linked sequences can be determined by PCR and subsequently gel electrophoresis (Han et al., 2018). A range of methods based on the 3C technology has emerged, greatly improved by the application of next generation sequencing. For example, 4C, 5C, and Hi-C have greatly expanded our knowledge of chromatin architecture, organization, and interaction. Particularly important was the emergence of Hi-C (Lieberman-Aiden et al., 2009), which allows for high-throughput data and analysis of a whole range of interactions at the same time (Han et al., 2018; Sparks et al., 2020). However, 3C and derived techniques are limited to visualization of fixed cells, although new 3C-derived technologies are continuously developed and improved (Sparks et al., 2020).

Recently, new methods for studying chromatin interactions have emerged (Kempfer and Pombo, 2020; Sparks et al., 2020), including Genome Architecture Mapping (GAM) (Beagrie et al., 2017) and Split-Pool Recognition of Interactions by Tag Extension (SPRITE) (Quinodoz et al., 2018).

1.4.2 Visualizing DNA in Live Eukaryotic Cells

Many different approaches to visualize chromatin in living cells are available. Living cells give a unique insight into the dynamic and temporal organization of chromatin, which is not available after fixation. Fixation may cause formation of structural artefacts that interfere with interpretation of results, an issue that is resolved by working with live cells (Zink et al., 2003). In this section the most widely used techniques for live cell imaging using fluorescent proteins (FPs) will be presented together with some of the emerging cutting-edge techniques. These techniques are divided into two subcategories: bulk and specific labeling.

1.4.2.1 Bulk Labeling

Bulk labeling refers to the unspecific labeling of chromatin. Two such methods are nucleotide labeling and histone labeling. Labeling nucleotides with fluorophores and injecting them into cell nuclei is a way to visualize newly synthetized DNA. During mitosis a new sister chromatid will be synthesized, and if provided with fluorophore-tagged dNTPs the replication machinery has been shown to incorporate these into the new strand. One of the sister chromatids will therefore be fluorescently labeled and can be visualized by microscopy (Zink et al., 2003). Histone labeling is achieved by fusing an FP to one of the histone proteins, for which the core histone H2B has been frequently applied (Récamier et al., 2014; Zink et al., 2003). When looking at the distribution of H2B proteins in the nucleus, one simultaneously looks at the spatial distribution of DNA because of the high affinity of histone proteins to DNA.

These two bulk labeling techniques are useful when looking at the large-scale genome organization in the nucleus and the compaction of chromatin. For specific observation of individual DNA, other methods need to be considered.

1.4.2.2 Specific Labeling of genomic loci

1.4.2.2.1 Fluorescent repressor operator system

The fluorescent repressor operator system (FROS) was developed for chromatin studies in yeast. A 256 tandem repeat of a prokaryotic *lac* operator were integrated close to the centromere in a yeast chromosome by restriction enzyme digestion and homologous recombination (Straight et al., 1996). Because the Lac repressor protein has specific binding to the *lac* operon, Lac repressor proteins fused to a FP-tag and an NLS sequence were transformed

into the yeast cells where they localized to the repeats in the nucleus and generated visible fluorescence signals (Straight et al., 1996). The disadvantage to this method is that it depends on integration of long repetitive sequences for foci formation, which due to their bulky intrusion in the DNA sequence could interfere in the transcription and replication processes, and potentially disrupt local chromatin organization (Saad et al., 2014).

1.4.2.2.2 Imaging by zinc finger proteins

Zinc fingers (ZFs) are a diverse group of transcription factors that are highly abundant in the cell, where they perform a range of tasks as transcriptional regulators. ZFs comprise a DNA-binding helix folded over a zinc ion. They are generally characterized based on the amino acids (aa) interacting with the zinc ion, and the most studied and best characterized ZF is the double cysteine and double histidine (C_2H_2) type (Isalan, 2013). Specific aa residues in the helix recognize a specific trinucleotide on the DNA strand. By fusing the ZF to a fluorescent protein and engineering a sequence of ZF motifs with specific aa to recognize a specific sequence in the DNA. In order to visualize the loci by microscopy, multiple FP-tagged ZF arrays need to bind to repetitive sequences in the genome **(Figure 1.3)** (Wu et al., 2019). This tool needs careful evaluation of the ZF array for each application due to evidence that neighboring ZFs could affect each other's target specificity (Anton et al., 2014). Because the tool only can visualize repetitive sequences, its application is limited.

1.4.2.2.3 Imaging using TALEs

Transcription activator-like effectors (TALEs) are proteins secreted by the bacterial genus *Xanthomonas* to infect plants (Miyanari et al., 2013). The TALEs attach to DNA with a DNA binding domain consisting of 33-35 aa repeats. Two of the central aa residues in each repeat, the repeat variable diresidues (RVDs), recognize a specific single base pair in the DNA sequence **(Figure 1.3)** (Ma et al., 2013; Miyanari et al., 2013). The TALEs can be tailored in each individual repeat to the recognition and binding of a specific DNA sequence. By fusing the TALEs to FPs and a nuclear localization signal, the protein array will localize to the nucleus and bind to the specifically tailored sequence, which forms a fluorescent focus by attachment of a series of TALEs to a repetitive sequence (Ma et al., 2013; Miyanari et al., 2013; Wu et al., 2019). A disadvantage to the technique is the elaborate effort involved in generating this array (Anton et al., 2014). It is also still limited to imaging of repetitive sequences. Methods to

circumvent insertion of repeats is in high demand and new techniques are continuously developed.

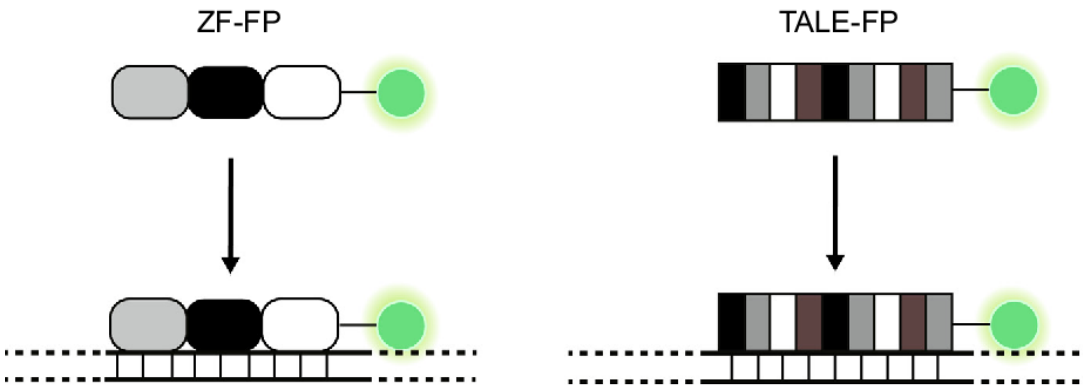


Figure 1.3. Schematic overview of fluorescently labeling chromatin with the zinc finger (ZF) and transcription activation-like effector (TALE) approaches. Image adapted from Wu (2019).

1.4.2.2.4 Imaging using CRISPR-Cas9 system

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein 9 (Cas9) technology was derived from the adaptive immune system of prokaryotes. The Cas9 nuclease works in coordination with two small RNA components: CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), which form the single guide RNA (sgRNA) (Faure et al., 2019). crRNA is specifically designed for the target sequence, to which Cas9 is directed and performes a double stranded cut (Jinek et al., 2012; Mali et al., 2013). The generation of a nuclease deficient Cas9 (dCas9), enabled a range of possible new applications, including in imaging (Qi et al., 2013). The mutant was directly fused to an FP for imaging of not only repetitive regions, but also non-repetitive regions **(Figure 1.3 A)**. Advantages to this system includes easier design than ZFs and TALEs, and a target flexibility that makes this tool hihgly versatile (Chen et al., 2013). The drawback is that many different sgRNAs (at least 26) need to be generated for a sufficiently high signal emission, which would make a large intrusion by the Cas9 complexes in the target region. High background signals have also been observed in the nucleolus, where Cas9 tends to be enriched (Chen et al., 2013).

The supernova tagging system (SunTag) was made as an attempt to improve the Cas9-FP by using less sgRNAs and improve signal strength (Tanenbaum et al., 2014). The SunTag system consists of a scaffolding peptide fused to a dCas9 that is directed to a specific locus in

the genome, and a cognate FP-fused antibody which is recruited to the site and form a high-signal fluorescent focus **(Figure 1.4 B)** (Tanenbaum et al., 2014; Wu et al., 2019). The advantage to using antibodies is their very high affinity and specificity towards their epitope, although they are large and bulky structures, and usually express poorly in the cytoplasm (Tanenbaum et al., 2014). A study has also shown that unbound dCas9-SunTag gives high background signal (Hong et al., 2018).

Another use of the CRISPR/dCas9 mediated fluorescence is the MS2 system. MS2 is an RNA aptamer derived from bacteriophage MS2, which forms a stem loop structure that specifically and with high affinity binds to the MS2 coat protein (MCP) (Wu et al., 2012, 2019). This system is utilized by attaching the MS2 to the sgRNA of the Cas9-complex, which attracts multiple MCP-FP fusion proteins to the site where Cas9 is bound **(Figure 1.4 C)**. However, this method also experiences high background due to unbound MCP-FP (Wu et al., 2012).

Other similar approaches which employ CRISPR/dCas9 have been developed for live cell imaging. However, a recently emerged tool adopts an entirely different approach: the ParB-INT single gene loci imaging system. This system will be discussed in detail in the following chapter.

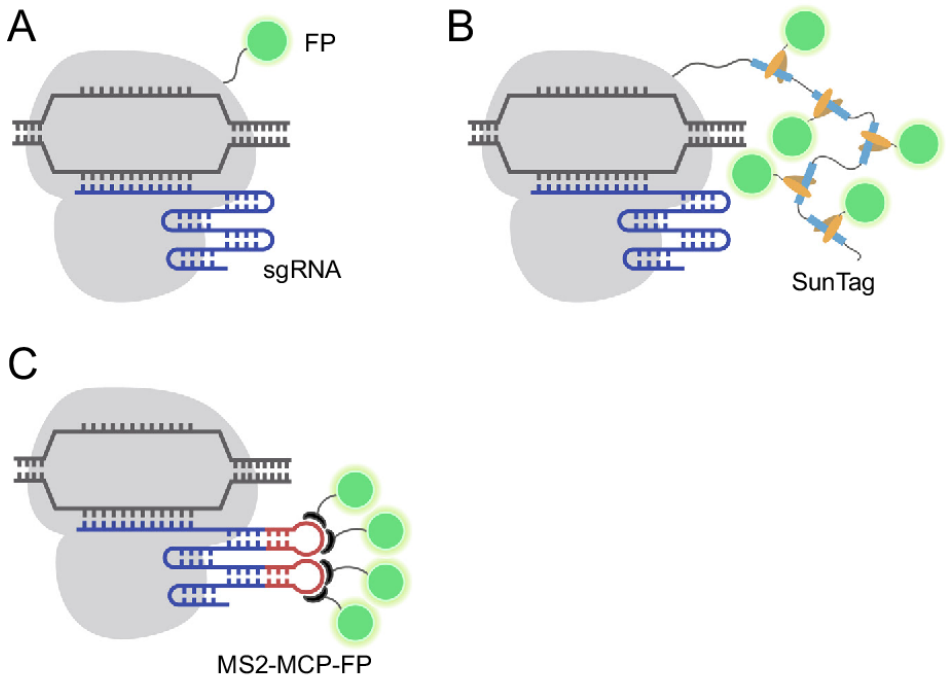


Figure 1.4. Schematic illustration of novel live cell imaging technologies using the CRISPR/dCas9 system. **(A)** dCas9 directly fused to a FP. **(B)** dCas9 fused to the SunTag system GCN4 proteins with epitopes to which the scFv-FP antibodies bind. **(C)** dCas9 with an RNA amptemer stem loop structure integrated in the sgRNA with binding site for the fusion coat protein MCP-FP. Image adapted from X. Wu et al. (2019).

1.5 The ParB-INT Single Gene Loci Imaging System

The ParB-INT imaging system was modified from the prokaryotic partitioning system for visualization of single gene loci (Bystricky, 2015).

1.5.1 The Bacterial Partitioning Machinery

A process of vital importance to all life, is the reliable transmission of hereditary genetic information from a parent cell to a progenitor cell. Central to this process is proper segregation of chromosomes to each new cell. In eukaryotes, this process is complex and chromosomes are faithfully replicated, divided, and transported to each cell pole by the mitotic spindles (Ovejero et al., 2020). In bacteria, this process is less understood and variable between species. Three main categories of chromosome and plasmid segregation in bacteria have been described: Type I, II, and III NTPase systems (Baxter and Funnell, 2014; Funnell, 2016). The type I ATPase system comprises the most well-known partitioning systems and the only ones known to operate on bacterial chromosomes. The partitioning machinery typically consists of three components: an ATPase (ParA), a DNA-binding partitioning protein (ParB), and a number of *cis*-acting centromere-like partition sites (*parS*) (Baxter and Funnell, 2014). Together they form the ParABS partitioning system. The *parAB* genes are arranged in an operon under the control of a single promoter, and the *parS* partition site is a 16 bp site of inverted repeats located in the centromere-like region close to the origin of replication in the chromosomes. When the ParB protein binds to *parS*, a nucleoprotein-complex referred to as the segrosome is formed (Kawalek et al., 2020). Through a phenomenon referred to as “spreading”, the *parS* bound ParB proteins recruit hundreds of other ParB proteins to form DNA-interactions in the vicinity of the *parS*. They then become distributed both upstream and downstream of the *parS* sequence (Breier and Grossman, 2007; Funnell, 2016; Murray et al., 2006). The spreading mechanism recruits the structural maintenance of chromosomes (SMC) proteins which unwind and organize the DNA replicates (Jalal et al., 2020; Sullivan et al., 2009). ParA is also recruited to the segrosome, where it provides energy obtained by ATP hydrolysis to drive the active segregation of the chromosomes. The segrosome pairs attached to each allele are separated and moved to each cell pole, ensuring that both progenitor cells inherit a copy of each chromosome and at least one copy of each plasmid (Baxter and Funnell, 2014; Kawalek et al., 2020; Passot et al., 2012).

The mechanism by which spreading occurs has not been fully understood, although recently three studies uncovered an additional ParB function which takes us closer to understanding the binding and spreading capacity of ParB (Jalal et al., 2020; Osorio-Valeriano et al., 2019; Soh et al., 2019). ParB proteins were shown to bind and hydrolyze cytidine triphosphate (CTP) **(Figure 1.5)**. In the presence of CTP ParB dimers nucleate around the *parS* site, and will upon CTP binding form a clamp-like structure around the *parS* site and slide in a spreading manner to flanking DNA (Audibert et al., 2020; Jalal et al., 2020; Soh et al., 2019). Phosphorylation of CTP causes the ParB proteins to release the DNA-clamp and dissociate from the DNA. It has also been demonstrated that the spreading mechanism can be perturbed by roadblocks such as transcription factors which causes ParB proteins to be released from DNA (Jalal et al., 2020; Soh et al., 2019).

ParB proteins are comprised of three domains: The N-terminal domain (NTD) which is involved in protein-protein interactions with ParA and other ParB proteins, the DNA-binding domain (DBD) which binds ParB to *parS,* and the C-terminal domain (CTD) responsible for the dimerization of ParBs. In the model in **Figure 1.5**, Jalal et al. (2020) shows how these domains likely act in relation to each other *in vivo*. Several articles have reported that ParB proteins exhibit nuclease activity (Grohmann et al., 1997; Johnson et al., 1999; Shaw et al., 2008). If and how this activity is involved in protein partitioning or there is any relation to the CTP activity, has not yet been uncovered.

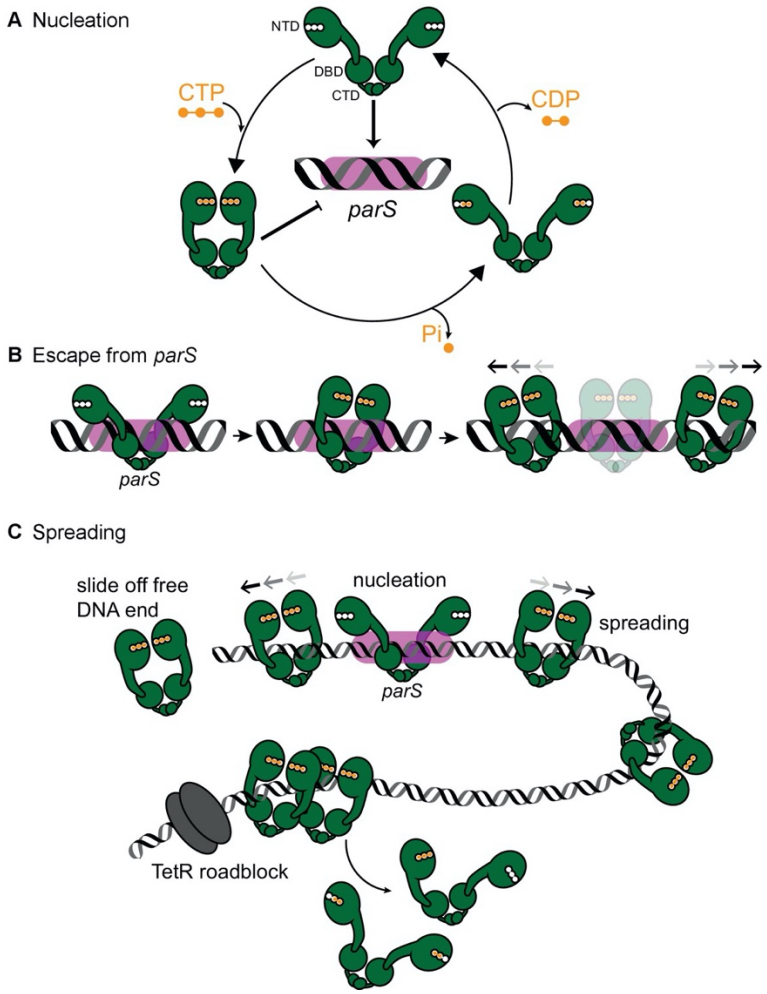


Figure 1.5. Spreading model of ParB proteins on DNA. **(A)** The nucleation and following binding of the ParB dimer is initiated by CTP binding, and ParB release from the DNA follows CTP hydrolysis to CDP. **(B)** The spreading mechanism happens when CTP binds and the ParB dimer forms a clamp-like structure around the DNA molecule. **(C)** Spreading occurs both upstream and downstream of the *parS* site. Roadblocks can be added for containment of the signal. Image adapted from Jalal (2020).

1.5.2 Repurposing of ParB-INT for Imaging in Eukaryotes

The ParB-INT imaging tool comprises a repeated number of integrated *parS* sequences (INT) and its cognate ParB protein, derived from several bacterial species, including pathogenic bacteria of the Burkholderiales order (Bystricky, 2015; Germier et al., 2018; Saad et al., 2014). The ParB protein NTD is fused to a FP, which makes it possible to visualize in the microscope. The INT sequence can be inserted by CRISPR-Cas9 into a specific gene locus, for instance an enhancer or promoter of interest. The imaging technique takes advantage of the inherent property of the bound ParB dimer to further recruit other ParB dimers to the INT site. Seeing that all ParB dimers are fluorescently tagged, the area around the INT site will be enriched in protein and create a fluorescent focus strong enough to detect by microscopy. Because each

ParB protein has a strict specificity to its cognate INT, even on different chromosomes within the same species, a number of ParB systems can be used simultaneously for different target regions with distinguishable fluorescent tags (Dubarry et al., 2006; Germier et al., 2018). It enables visualization of colocalized DNA regions and interactions between them. For example, promoter and enhancer interactions can be observed by inserting a INT1 segment next to the enhancer with a red fluorescent tag and an INT2 next to the promoter region with green fluorescent tag **(Figure 1.6)**.

By applying this system to live cells, the spatiotemporal localization of the foci can be analyzed to see if interactions form and in which cellular conditions they form. Because the ParB binding to DNA is loose and unspecific, the proteins assemble and disassemble on the DNA stand continuously, even when encountering roadblocks like other DNA binding proteins. The intrusion of the ParB-INT system on the INT-surrounding regions is therefore less than for other imaging methods previously mentioned (Saad et al., 2014). The ParB-INT system thus enables a feasible method for imaging of single gene loci in live eukaryotic cells. The technology has been applied to various different organisms, presented in **Table 1.1**, proving this is a versatile system. To further uncover how this system works in a cell, protein structure can provide valuable insight into how this protein interacts with genomic structures.

Table 1.1. Overview of studies that have applied the ParB-INT imaging system in different organisms.

Organism	Purpose	Reference
Eukaryotes		
Yeast, *Saccharomyces cerevisiae*	Study dynamic of double strand breaks	(Saad et al., 2014)
Fruit fly, *Drosophila melanogaster*	Visualize transcription and physical interactions of promoter-enhancer	(Chen et al., 2018)
Human breast cancer cells, MCF7	Study chromatin dynamics in transcription	(Germier et al., 2017)
Fruit fly, *Drosophila melanogaster*	Colocalization of loci in Notch signaling	(Gomez-Lamarca et al., 2018)
Yeast, *Saccharomyces cerevisiae*	Observation of ParB spreading mechanism	(Audibert et al., 2020)
Prokaryotes and viruses		
Bacteria, *Deinococcus radiodurans*	Study chromosome organization during double strand break repairs	(Passot et al., 2015)
Human cytomegalovirus	Observe infection and replication of virus in live human cells	(Mariamé et al., 2018)

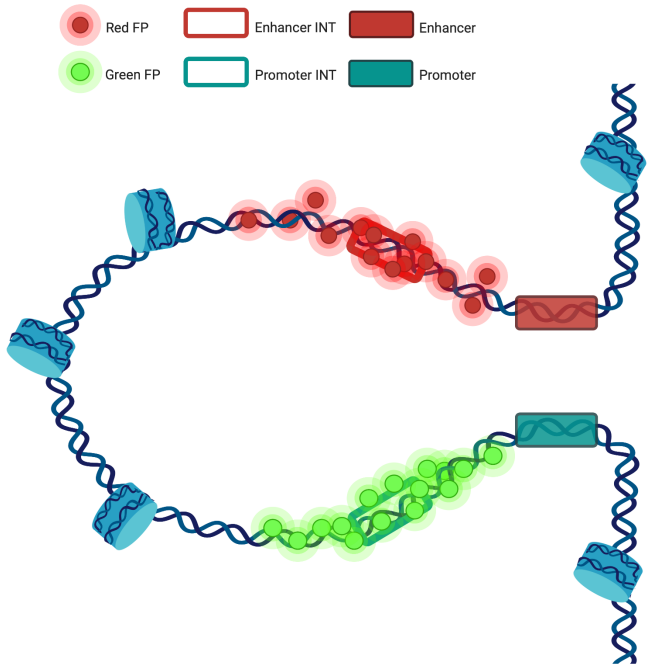


Figure 1.6. Schematic illustration of ParB-INT system applied to looping promoter and enhancer region. To study the interactions between two specific loci, the ParB-INT system enables visualization by integration of two different INT sequences next to the target sequences. ParB-FP fusion proteins will localize to the INT sequences and spread bidirectionally, forming a visible fluorescent focus. Image created with BioRender (https://biorender.com).

1.6 Protein Structure and Function

To better understand protein function and genomic interactions, like transcription factors involved in promoter-enhancer interactions, a key step is to resolve the protein structure. Because the structure of proteins is highly determinant to the protein function, solving the structure of proteins are crucial to unlock their functional properties. Due to a rapidly increasing number of discovered protein sequences, time consuming and costly methods for experimental protein structure determination by high-resolution methods such as x-ray crystallography (Kendrew et al., 1958) and nuclear magnetic resonance spectroscopy (NMR) (Dötsch and Wagner, 1998; Serber et al., 2001) can not keep up the pace (Braitbard et al., 2019). Fast and free *in silico* methods for protein structure prediction are becoming increasingly popular, even as a supplement to other experimental structure modeling techniques (Braitbard et al., 2019; Liu et al., 2019). A large collection of all experimentally determined structures is available in the Protein Data Bank (https://www.rcsb.org).

1.6.1 Molecular Structure Prediction

There is a vast range of tools to use for computationally determining the structure of a protein. They can be based on the observation that homologous proteins often adopt the same structural fold and therefore will search databases to find experimentally solved protein structures of homologous proteins (Khanna et al., 2019). Other tools randomly search databases for a fitting fold. When a structure has been determined for a protein, a lot of new possible information can be gained. Multiple tools for cavity searches, ligand binding site analysis, and active pocket recognition have been developed taking into account the relative orientation of individual atoms in the structure (Khanna et al., 2019). The secondary, tertiary, quaternary structures of a protein can also give more information about how the protein binds to for example a DNA molecule or interacts with other proteins like cofactors or allosteric regulators (Joyce et al., 2015). The protein structure is also highly relevant when designing mutational studies (Joyce et al., 2015).

1.6.2 Evolutionary History

The evolution of genomic sequences is tightly connected to their spatial organization and function. By also studying how evolutionary well-conserved certain regions of genes are, further insight can be gained into which regions are functionally crucial for the activity of a protein in a cell. Aa involved in the same enzymatic site or binding site, rely on coevolution to maintain proper function of the activity (Braitbard et al., 2019). Such sequences can often be uncovered by multiple sequence alignments of homologous sequences. Discovery of such evolutionary conserved aa residues is valuable to for instance prediction of protein-protein interactions or provide valuable information in mutational studies or protein function (Braitbard et al., 2019).

1.7 Aims of the Study

Well-working non-intrusive systems for live-cell imaging of single gene loci are vital to expand our current knowledge on nuclear organization, interactions between genomic regions, and transcriptional regulation. The recently emerged tool based on the ParB-INT system is one of the live cell imaging tools that have the potential to meet such demands (Bystricky, 2015; Saad et al., 2014). Unpublished data from our lab shows that transfection of mouse embryonic stem cells (mESCs) with prokaryotic fluorescent ParB proteins, which propagate fluorescent signals on genomic loci visible by microscopy, results in unhealthy cells and a high degree of apoptosis. Moreover, ParB proteins have been observed to act as nucleases on prokaryotic DNA (Grohmann et al., 1997; Johnson et al., 1999; Shaw et al., 2008). We wondered whether the ParB nuclease activity could be the cause of cell mortality observed in mESCs.

In this study the overall aim was to characterize two ParB proteins for improvement of the ParB-INT imaging system. The aim is further divided into three parts. (1) Gain structural and functional insights into the ParB proteins in order to identify potential location of amino acids crucial to nuclease activity. (2) Construct a selection of clones for the investigation of potential nuclease activity of the prokaryotic ParB proteins including humanized versions, and the core domain of ParB3. (3) Test the humanized ParB proteins in the ParB-INT imaging tool for comparison with the original prokaryotic protein in mESCs.

Contributing to new knowledge about the function of the ParB proteins will provide new insights into how the ParB-INT imaging system can be further optimized for imaging of single gene loci in live cells.

2 Methods

Commercial kits were used according to the protocol provided by the manufacturer. Recipes and a complete list of materials and reagents used in the experiments are provided in appendix 5 and 6.

2.1 Bacterial Work

In this project, two types of bacteria have been used depending on the aim of the experiment. DH5α cells are specifically designed for plasmid maintenance and amplification and has therefore been used in cloning experiments. DH5α is an *Escherichia coli* strain with a mutation in the *recA1* recombinase gene to improve plasmid stability and in the *endA* gene to prevent the EndA endonuclease from degrading a transformed plasmid. BL21(DE3)pLysS is another *E. coli* derived strain of bacteria, that has been used for protein expression experiments due to its special properties that allow for overexpression of genes connected to a T7 promotor. The cells are specifically designed for protein expression as they carry the chromosomally integrated DE3 prophage which contains the T7 RNA polymerase gene controlled by a *lac UV5* promoter. The polymerase is highly specific for its own promoters, which do not exist natively in *E. coli* cells and can therefore be used as a promoter to the gene of interest. The *lac* operon repressor inactivates transcription of the T7 polymerase gene in normal conditions, but the promoter is leaky which means that some transcription of genes downstream of the T7 promoter will occur at all times. If the gene product is toxic, cells might die from even small concentrations. Therefore BL21(DE) carry the pLysS plasmid, which encodes the T7 lysozyme. The lysozyme forms a complex with the T7 polymerase, inhibiting its binding to the T7 promoter, but does not interfere with induction of protein expression by IPTG (Moffatt and Studier, 1987).

2.1.1 Culturing Bacteria

Bacterial cultures have been used in this work to either amplify a plasmid or express a protein. LB medium was added to a sterile flask or tube in desired volumes with a proper antibiotic. Transformed bacteria was spread on LB agar plates with antibiotic for selective growth and single colony formation. The stock concentrations of antibiotic used in this study was 100 mg/ml ampicillin, 50 mg/ml kanamycin, and 25 mg/ml chloramphenicol. The working

concentrations were a 1/1000 dilution of the antibiotic stocks in medium. The culture was incubated at 37°C in a shaking incubator for 16-18 hours at 200 rpm. The optical density of the bacterial culture was measured with a spectrophotometer at 600 nm with fresh LB medium as blanking solution.

2.1.2 Cryopreservation and stock

Bacterial stocks were made for each clone construct made. The bacterial cells were stored at -80°C in a 15% glycerol suspension. To use the stock, a droplet size portion of bacteria was added to a flask with LB medium and proper antibiotics for culturing.

2.1.3 Preparation of BL21(DE3)pLysS Chemically Competent Cells

BL21(DE3)pLysS stock was streaked on an agar plate and incubated for 16-18 hours at 37°C. 5-10 colonies were added from the plate to a flask of 100 ml SOB medium, and incubated at 37°C for 3 hours shaking on 200 rpm. The optical density of the culture was measured with a spectrophotometer. The amount of culture needed in a total volume of 250 ml SOB solution to reach OD_{600} of 0.05 was added. The appropriate volume was added to a 2L flask and incubated at 18°C for 16-18 hours shaking on 200 rpm. The OD_{600} was measured and the culture removed from incubation when in the range of 0.3-0.6. The culture was cooled on ice for 10 minutes and transferred to five falcon tubes of 50 ml. The tubes were centrifuged at 2500 rpm for 10 min at 4°C to pellet the bacteria. The supernatant was carefully discarded, and the pellet resuspended in 16 ml TB buffer. The solutions of the five tubes were separated into two tubes and chilled on ice for 10 minutes. The centrifugation step was redone for the two tubes: 2500 rpm for 10 minutes at 4°C. The supernatant was carefully discarded, and each pellet resuspended in 10 ml TB buffer. DMSO with a final concentration of 3.5% were added to each tube. The solution was gently mixed by swirling the tubes and the left for 5 minutes on ice. DMSO was again added to reach a final concentration of 7%, then gently swirled and left on ice for 10 minutes. 200 µl portions of the solution were transferred to cold 0.5 ml tubes and transferred to an ice box with liquid nitrogen for shock freezing. The tubes were subsequently stored at -80°C.

The chemically competent cells were frozen down as 200 µl stocks in a state of high membrane permeability. This permeability was induced through specific handling of the cells prior to freezing, and also by adding dimethyl sulfoxide (DMSO) to the solution. DMSO is a molecule that works as a cryoprotectant in frozen cells and will also disrupt the integrity of the

cell membrane making it more permeable to hydrophilic molecules like DNA (Anchordoguy et al., 1992; Cheng et al., 2015; Gurtovenko and Anwar, 2007).

2.1.4 Transformation of Chemically Competent Cells

In a transformation experiment the aim is to introduce foreign DNA into a bacterial cell. In order for the bacteria to allow DNA to pass through the cell membrane, we use chemically competent cells. When a 42°C heat shock is applied to the suspension of plasmid and cells, the plasmid is able to cross the cell membrane into the cytoplasm.

1 pg – 100 ng plasmid was added to a tube together with 50 µl competent cells and kept on ice for 20 minutes. The suspension was heat shocked at 42°C for 90 seconds, then cooled on ice for 2 minutes. 1 ml of LB medium was added to the solution and incubated at 37°C for 1 hour at 200 rpm. The cells were pelleted by centrifugation at 4000 rpm for 4 minutes and the pellet resuspended in 50 µl of LB medium. The bacteria were spread out on agar plates made with a proper selective antibiotic. The plates were incubated at 37°C for 16-18 hours.

2.2 DNA Work

2.2.1 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) is a commonly applied tool for amplification of DNA sequences. In this project the tool was used for pre-cloning amplification of a sequence of interest and mutagenesis purposes. A pair of synthetically designed primers flanking the region of interest were added to a reaction mix of MQ H_2O, deoxyribonucleotide triphosphates (dNTPs), a polymerase, a reaction buffer, and a target DNA template **(Table 2.1)**. A negative control without DNA template was always included to detect contamination in the samples. A PCR reaction runs in multiple cycles, usually 25-35, where each cycle is divided into three important steps: denaturation, annealing, and elongation **(Table 2.2)**. During the denaturation step the temperature increases to >90°C in order to denature the double stranded helix into single stranded templates. In the annealing step, the temperature is lowered to the point where the primers will hybridize with the denatured DNA template. The temperature where primers will anneal is specific for each primer. In the elongation phase, the polymerase will synthesize a new complementary template strand starting at the primer ends.

Table 2.1. Reaction mix volumes and concentrations. A reaction mix with concentrations for a Q5 PCR reaction using a Q5 High-Fidelity DNA Polymerase.

Reagents	Final Concentration
5X Q5 reaction buffer	1X
5 mM dNTP	200 µM
10 µM Primer F	0.5 µM
10 µM Primer R	0.5 µM
DNAtemplate	1 pg-10 ng
Q5 High-Fidelity DNA Polymerase	0.02 U/µl
5X Q5 High GC Enhancer	1X
Sterile MQ	x
Final volume	25-50 µl

Table 2.2. PCR thermocycler conditions. Conditions for the PCR reaction with primer specific melting temperatures (T_m) and number of cycle repeats for the individual experiments, as recommended by the manufacturer.

Temperature (°C)	Time	Cycles	Step	Description
98	30 sec	1	1	Initial denaturation
98	10 sec	28-35	1	Denaturation of template
T_m	30 sec	28-35	2	Annealing of primers
72	40 sec	28-35	3	Elongation by polymerase
72	2 min	1	1	Final elongation
4-12	∞	1	1	Hold on low temperatures for preservation

2.2.2 Agarose Gel Electrophoresis

Agarose gel electrophoresis is a molecular technique used to separate DNA fragments according to size in a gel. The gel is made from the polysaccharide agarose dissolved in 1X Tris-acetate-EDTA (TAE) buffer to form a polysaccharide network with pores of varying size depending on the desired concentration of the gel. DNA was loaded into wells created in the gel, placed in an electrophoresis chamber filled with 1X TAE buffer, and an electric field was applied. Because DNA has a negative charge it will migrate from the negative cathode towards the positive anode. Depending on the size of the loaded DNA fragments, they will migrate to the anode in varying speed. The longer DNA fragments will be more deterred by the density of the polysaccharide network and migrate more slowly than shorter DNA fragments. Prior to

casting, an intercalating fluorescent nucleic acid gel stain GelRed (Biotium) was added to visualize the DNA bands under UV light. Agarose gel electrophoresis was performed after a PCR run or an enzymatic digestion to validate the results, or to isolate DNA fragments for downstream cloning purposes.

A 1% agarose gel was made by mixing 1g agarose with 100mL 1X TAE in an Erlenmeyer flask of 2-4 times the volume of the solution and heated until the agarose was dissolved. The solution was cooled to ~50°C and GelRed was added in a 10,000X dilution. The solution was poured into a casting tray and a gel comb inserted for well formation. After 20-30 minutes the gel solidified and was placed in the electrophoresis chamber. The DNA samples were mixed with 6X Gel Loading Dye Purple to a final concentration of 1X. The gel comb was removed, and the DNA samples were loaded to the wells. 1.5-2 µl (75-100 ng/µl) of 1 kb Plus DNA ladder was loaded to one or more wells as a reference guide for size determination. The gel ran at 100V for 45-60 minutes.

2.2.3 PCR Clean-up and Gel Extraction

PCR samples and DNA fragments in agarose gel were cleaned with the PCR Clean-up Kit and NucleoSpin® Gel from Macherey-Nagel. Instructions provided by the manufacturer was followed. The sample was mixed with NTI buffer, which contains a chaotropic salt. The chaotropic salt (guanidinium thiocyanate) interferes with the water molecules surrounding the DNA, destroying the water shell that makes DNA soluble in aqueous solutions. A hydrophobic environment is created instead, which makes DNA bind to the silica membrane, while all other contaminants are washed out with the ethanol containing buffer NT3. Examples of contaminants that are removed are: nucleotides, primers, DMSO, dyes, detergents, etc. The DNA is released from the silica membrane by the elution buffer, which provides low salt and alkaline conditions.

2.2.4 Plasmid DNA Purification

The Mini prep (NucleoSpin® Plasmid) and Maxi prep (NucleoBond® Xtra Maxi) kits by Macherey-Nagel were used for plasmid purification from DH5α bacterial cultures. The protocol was followed according to the instructions given by the manufacturer. The Mini prep was used for small scale DNA purifications, which is useful for subcloning experiments and verification by sequencing. The Maxi prep was used mostly for the end products of a cloning

reaction or mutagenesis that had been confirmed by sequencing for larger yields of DNA product.

The Mini and Maxi prep kits are based on the same concepts. A transformed bacterial culture grown with antibiotic selection is pelleted to remove the supernatant and resuspended in a resuspension buffer containing RNase to remove RNA contamination. The bacterial cells are lysed by an alkaline NaOH/SDS buffer, releasing the contents of the cell. A neutralization buffer is added to bind the proteins, genomic DNA, and other cellular components in precipitating SDS, leaving the plasmid DNA in the supernatant. The supernatant is loaded onto the column and the precipitate is removed either as a pellet through centrifugation (Mini prep) or in a filter (Maxi prep). The silica column is hydrophilic and positively charged, and thus binds the negatively charged DNA, retaining it in the column. Contaminants left on the column is removed by washing with an ethanol containing buffer. The plasmid DNA is released from the column by the addition of an elution buffer that neutralizes the positive charge of the column thus disrupting DNA binding conditions. The Mini prep will have elution as a last step, but the eluted Maxi prep sample will also be treated with isopropanol and ethanol, so that the DNA will precipitate from the solution and can be centrifuged into a pellet and salts be removed. The pellet is resuspended in 1X TE-buffer and stored at -20°C.

2.2.5 Cloning Techniques

Molecular cloning is a set of techniques which are used to create recombinant DNA molecules to study their function. There are several types of cloning that have been used in this project.

2.2.5.1 Traditional Cloning with Restriction Enzymes

Traditional cloning involves two main steps, (1) cutting out the desired sequence from a parent vector and (2) ligating it into a linearized destination vector. Restriction endonucleases, referred to as restriction enzymes, are used to cut the parent vector backbone at specific recognition sites on each side of the sequence of interest, so that it is isolated from the rest of the vector. The same enzymes are used to linearize the destination vector so that the ends of the sequence of interest match the vector. Vectors are often designed to contain sites for selected restriction enzymes to facilitate this type of cloning. In this project only double digest cloning with incompatible sticky ends were performed, which decreases the possibility that the vector will self-ligate, as opposed to cutting with one enzyme or using enzymes that create blunt ends.

The vectors were treated with calf intestinal alkaline phosphatase (CIP) which dephosphorylates the ends, further preventing self-ligation and making the insert and vector more compatible.

The digestions were carried out according to the protocol provided by the supplier of the enzymes, New England Biolabs. A 50 µl reaction mix was made with 1 µg template DNA, 1X reaction buffer specified for each restriction enzyme, 20 units restriction enzyme and H_2O. The reaction was incubated at 37°C for 10-15 minutes, or overnight for enzymes that are time-saver qualified. Some enzymes also required heat inactivation.

Prior to the ligation reaction, the digested DNA samples were run on a gel to separate the fragments of interest and linearized destination vector, then they were cut out and purified with the Gel Extraction Kit by Macherey-Nagel and quantified with a NanoDrop spectrophotometer. The ligation reaction was adapted from the NEB protocol. The amount of insert DNA added to the reaction was calculated according to the following formula: ng(insert) = [ng(vector) x bp(insert)]/bp(vector) x ratio. Thus, the amount of insert varies for each reaction according length and desired ratio. A sequence of interest to vector ratio of 1:3 is preferred in most cases. A 20 µl reaction was prepared with a calculated amount of insert, 50 ng vector, 1X DNA T4 ligase buffer, and 40 U T4 DNA ligase. The ligase reaction was incubated in room temperature for 1 hour, followed by heat inactivation at 65°C for 10 minutes. After the solution has chilled a few minutes on ice, 5 µl of the ligation product was transformed into DH5α and grown on agar plates with antibiotic selection. Bacteria transformed with a control reaction without insert was spread on a plate to account for the number of plasmids that self-ligate. Specific cloning details for individual reactions are given in **Table 2.3**.

Table 2.3. Overview of clones created by restriction cloning methods and individual reaction conditions.

Protein sequence in parent vector	ParB3Hu in pUC57		ParC in pUC57
Destination vector	pET15b	pDasher	pET28a
Restriction enzymes	KpnI NheI	KpnI NheI	NcoI XhoI
Buffer	1.1 (NEB)	1.1 (NEB)	CutSmart (NEB)
CIP	No	Yes	No

2.2.5.2 Gateway Cloning

The Gateway Cloning Kit (Invitrogen) facilitates cloning through a series of recombination reactions. The process involves two main steps referred to as the BP and LR reactions **(Figure 2.1)**. The sequence of interest is PCR amplified with special Gateway primers that attach attB adapters to the flanking regions of the sequence, forming a gene cassette. The BP reaction is made by adding the following to a tube of 1X TE buffer: the sequence of interest with attB adapters, a donor vector with corresponding attP adapters, and a special BP enzyme mix. The enzyme catalyzes the recombination of the adapter sites, adding the gene cassette into the donor vector and creating an entry clone. From the entry clone, the gene cassette can easily be flipped into any number of available destination vectors that is suitable to downstream experiments. This is done by the LR reaction, where an LR enzyme mix is mixed with the entry clone and a destination vector. The combinations of adapters in the entry clone is now labelled attL and the corresponding adapters in the destination vectors are attR, and the gene cassette from the entry clone is flipped into the destination vector. In both the BP and LR reaction, a negative selection of donor and destination vectors that have not taken up the gene of interest is ensured by the formation of a toxic by-product. The gene cassette which should be replaced by the gene of interest is the *ccd*B gene, which if transformed will produce the CcdB protein that interferes with the topoisomerase gyrase in *E. coli* and thereby blocks cell growth.

An initial Q5 PCR reaction was performed according to the manufacturer's protocol, to attach the *att* sites to each end of the gene of interest (see section 2.2.1).

The BP and LR reactions were made according to **Table 2.4**, gently mixed, spun down in a microcentrifuge, and left to react in room temperature overnight. 2 µg Proteinase K was added to the samples which were incubated at 37°C for 1 hour to stop the reaction and transformed into DH5α (see section 2.1.3).

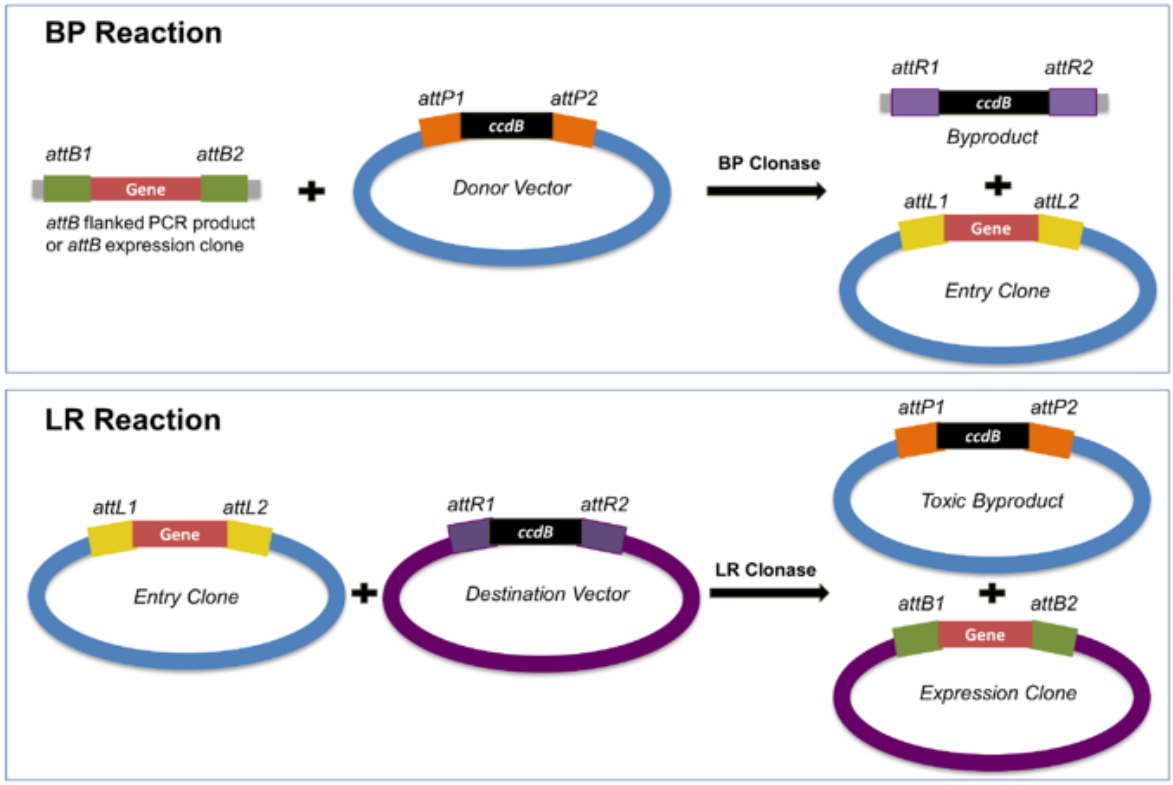


Figure 2.1. Schematic overview of the Gateway recombination system. The att-flanked entry clone created in the BP reaction was used to clone the gene of interest into a number of destination vectors through the LR reaction. The figure is adapted from blog.addgene.org.

Table 2.4. Gateway cloning reactions: BP and LR reaction components and concentrations. Table modified from the protocol provided by the manufacturer and adjusted for smaller volumes.

BP reaction	Total concentration	LR reaction	Total concentration
PCR product	75 ng	Entry clone	25 ng
Donor vector	75 ng	Destination vector	50 ng
1X TE buffer, pH 8	x	1X TE buffer, pH 8	x
BPClonase Enzyme Mix	1 µl	LRClonase Enzyme Mix	1 µl
Total volume	5 µl	Total volume	5 µl

2.2.6 Site Directed Mutagenesis by PCR

Making a point mutation in a plasmid DNA sequence can be done by a PCR reaction with primers that have been designed with the desired mutation. The PCR reaction will have to

amplify the entire plasmid, so it is vital to ensure that the correct conditions and reactants are used to avoid off-target single nucleotide polymorphisms (SNPs). The 5x HOT FIREPol Blend Master Mix (Solis BioDyne) was used for this reaction. This ready-to-use master mix contains two enzymes, a DNA polymerase for synthesising new DNA strands and a proofreading enzyme which will decrease the likelihood of an unintended SNP.

The reaction mix was made according to **Table 2.5**. A negative control without DNA template was added to each PCR run. The annealing temperature (T_m) was adjusted to fit the primer characteristics, and the cycling conditions are outlined in **Table 2.6**. Following the PCR run, the sample was run on a gel and the band corresponding to the size of the plasmid was cut out and purified with the PCR Clean-up Kit and NucleoSpin® Gel from Macherey-Nagel (see section 2.2.3).

Table 2.5. Overview of reaction mix components and corresponding concentrations. The reaction mix protocol is adapted from the manufacturer recommendations.

Components	Final concentration
5x HOT FIREPol® Blend Master Mix Ready to Load	1X
Forward primer (10 pmol/µl)	0.25 µM
Reverse primer (10 pmol/µl)	0.25 µM
DNA template	50 ng
MQ H_2O	To 10 µl

Table 2.6. PCR thermocycler conditions. Adapted from protocol given by the manufacturer.

Temperature (°C)	Time	Cycles	Step	Description
95	12-15 min	1	1	Initial polymerase activation
95	10-20 sec	25-30	1	Denaturation of template
54-66	30-60 sec	25-30	2	Annealing of primers
72	20 sec-4 min	25-30	3	Elongation by polymerase
72	5-10 min	1	1	Final elongation
4-12	∞	1	1	Hold on low temperatures for preservation

2.2.7 Sequencing

End products of the cloning experiments were sequenced by Eurofins Genomics to verify the proper insertion of the sequence of interest into the vector. 5 µl of 80-100 ng/µl vector

was added to an Eppendorf tube along with 5 µl of a 5 µM primer. The sequencing data was analyzed and aligned with the reference sequence in CLC Main Workbench and can be found in appendix 3**, Figures S5-S8**.

2.3 Protein Expression and Purification

2.3.1 Recombinant Expression of ParBs

ParB and ParC proteins have been cloned into pET plasmids under control of a T7 promoter. The T7 RNA polymerase gene controlled by a *lacUV5* promoter is repressed in standard growth conditions. By adding IPTG to the culture, a molecular mimic of allolactose, the repressor will detach from the promoter and the *E. coli* RNA polymerase will transcribe the gene. Since the T7 polymerase is highly specific for its own promotor, it will exclusively transcribe the gene of interest and at higher transcription rates than the native polymerase of *E. coli*, which leads to an expression of this specific gene product (Davanloo et al., 1984; Studier and Moffatt, 1986).

The pET15b_ParB and pET28a_ParC plasmids were transformed into BL21(DE3) pLysS and plated on LB agarose plates with selective antibiotics and chloramphenicol. A night culture was made by streaking an inoculation loop across the plate a few times so that many colonies are collected by the loop. A 1/100 dilution of the night culture was added to a 1000 ml Erlenmeyer flask with a total of 250 ml LB medium together with selective antibiotic. The culture was incubated at 37°C until OD_{600} was within the range of 0.5-0.6, and 1 ml of the culture was saved for gel separation (T_0). IPTG with a final concentration of 1 mM was added to the flasks for induction of protein expression, and incubated at 37°C for 3 hours, shaking at 200 rpm. The OD_{600} was measured and the same number of bacteria as the T_0 sample was collected for the much denser T_3 sample. After collecting the T_0 and T_3 samples, they were spun down at 11 0000 rpm for 2 minutes to pellet the bacteria. The supernatant was carefully discarded, and the pellet resuspended in 10 µl H_2O. 80 µl of SDS loading buffer, to which 10 % DTT was added freshly, was mixed with each sample. The samples were then boiled at 90°C for 5 minutes to denature the proteins, sonicated for 5 minutes to reduce viscosity, and spun down at 11 0000 rpm for 5 minutes. The excess induced culture was spun down at 4000 rpm for 10 minutes in 4°C, then the pellet was washed in 1X dPBS and spun down again in 50 ml falcon tubes. The supernatant was removed, and bacterial pellets stored in -80°C for further protein purification experiments.

2.3.2 SDS-PAGE and Staining

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis, usually referred to as SDS-PAGE, is a widely used molecular technique for separation of denatured proteins on a gel according to mass. An electrical field is applied to the gel, and the proteins move through the pores according to their molecular weight (kDa). Prior to loading, the sample is mixed with a SDS loading buffer. SDS will bind to the aa chain and impart a net negative charge on the proteins in addition to denaturing the secondary and most tertiary protein structures. Heating enables SDS to bind to more areas of the protein and keep them permanently denatured. The negative charge drives all proteins in the same direction in the gel, towards the anode. The dithioreitol (DTT) denatures the remaining tertiary structure that will not be affected by SDS, the disulfide bridges. Glycerol allows for the sample to drop to the bottom of the wells, and bromphenol blue dye colors the sample so that it can easily be visualized when moving through the gel.

Protein expression was checked by SDS-PAGE. 6 µl of the supernatant was loaded to a 12% polyacrylamide gel (Bio-Rad) and 5 µl of Precision Plus Protein Dual Color Standards (BioRad) was used as a reference. An electrical current was applied to the chamber at 160V and the gel ran for 60 minutes. The gel was cleared from the casting tray and transferred to a plastic box with 30 ml fixing solution. After agitation for 15 minutes the fixing solution was discarded, and the gel washed two times in MQ H_2O. The gel was stained with 30 ml PageBlue protein staining solution (Thermo Fisher Scientific) and agitated on the orbital shaker for at least 1 hour or overnight. The gel was rinsed two times with MQ H_2O and pictures were taken with the ChemiDoc™ Touch Imaging System (Bio-Rad). Molecular weights of proteins were determined with the ExPASy tool ProtParam (https://web.expasy.org/protparam/).

2.3.3 Protein Purification

2.3.3.1 Protein Purification by Precipitation

Protein purification is performed to purify a protein of interest from the total protein content of a cell. Protocol was modified from Johnson et al. (1999). The first step in a protein purification process is to isolate the total protein content in a cell from the dispensable cell debris. Because the BL21(DE3)pLysS cells express lysozyme, the bacterial cells will start to lyse when thawed. A few minutes on the bench will allow this process to start, although sonication is needed for full cell lysis. The sonicator will by means of high frequency sound

waves disrupt cell membranes and cause cells to lyse and DNA to fragment. Poly(ethyleneimine) (PEI) is added for the removal of DNA by precipitation. PEI is a cation and will bind DNA and other negatively charged molecules in the solution, forming a white precipitate that can be removed through centrifugation. The salt concentration needs to be adjusted according to the acidity of the proteins. If the proteins of interest are acidic, the salt concentration has to be increased in order to avoid precipitation (Burgess, 2009). Once the DNA is removed from solution, the proteins have to be separated from the PEI which can interfere with downstream applications. This is done by 80% saturated ammonium sulphate precipitation (Burgess, 2009). By adding a highly soluble compound like ammonium sulphate, soluble proteins will become less soluble because the salt ions will compete with the proteins in forming hydrogen bonds with water. Dehydrated proteins will aggregate and form a white precipitate that can be collected as a pellet and separated from the PEI solution. Throughout the entire purification process, it is vital to include protease inhibitors in all buffers to avoid protein degradation.

The pellet collected from the protein expression culture was thawed in room temperature for 5-10 minutes to induce the lysozyme, then transferred to an ice container. A 1g pellet was dissolved in 10 ml cold buffer A by vortexing. 20 µl of the dissolved pellet was saved in a 1.5 ml Eppendorf tube with 30 µl 2X SDS loading dye for further analysis. The rest of the resuspended pellet was added to a 15 ml falcon tube and sonicated in ice water with a LABSONIC M (B. Braun Biotech International) probe sonicator (50% amplitude, 30 seconds on/off, 6 repetitions). Cell debris was removed by centrifugation at 15 0000 g for 30 minutes at 4°C. 50 µl of the supernatant was saved in a 1.5 ml Eppendorf tube with 12.5 µl 4X SDS loading dye, and a small sample of the pellet collected with an inoculation loop was saved in 50 µl 2X SDS loading dye. The supernatant was transferred to a new centrifuge tube and 10% PEI was added to a final concentration of 0.75%. The tube was rotated for 5 minutes at 4°C for the formation of a white precipitate of PEI and DNA. The tube was centrifuged at 8000 x g for 10 minutes at 4°C. 50 µl of the supernatant was saved in a 1.5 ml Eppendorf tube with 12.5 µl 4X SDS loading dye, and a small sample of the pellet collected with an inoculation loop was saved in 50 µl 2X SDS loading dye. DNA has now been removed by PEI precipitation. To remove the rest of PEI in solution, two successive 80% saturation ammonium sulphate precipitation steps were performed. Four times the volume of the PEI supernatant was added of 100% saturated ammonium sulphate. The solution was left to chill on ice for 20 minutes, then spun down at 18000 x g for 15 minutes at 4°C. 50 µl of the supernatant was saved in a 1.5 ml Eppendorf tube with 12.5 µl 4X SDS loading dye, and a small sample of the pellet collected

with an inoculation loop was saved in 50 µl 2X SDS loading dye. The proteins in the solution should have precipitated together with the ammonium sulphate, and the supernatant was thus discarded. The pellet was resuspended in buffer B (volume equal to supernatant prior to addition of ammonium sulphate). All steps from addition of ammonium sulphate was repeated again. Pellet and supernatant were collected only from the first step. The resuspended pellet was added and sealed in a dialysis tube with an 8000 kDa mass cut-off. Dialysis tubes were prepared by soaking 10 minutes in MQ H_2O and then washing in buffer B. The tubes were added to a beaker with 500 ml buffer B on a magnet stirrer for circulation at 4°C. The buffer was changed every hour for 3 hours. The solution was then centrifuged for 10 minutes at 15000 rpm at 4°C. 50 µl of the supernatant was saved in a 1.5 ml Eppendorf tube with 12.5 µl 4X SDS loading dye, and a small sample of the pellet collected with an inoculation loop was saved in 50 µl 2X SDS loading dye. The supernatant was added to 15 ml falcon tubes, flash frozen in dry ice, and transferred to -80°C. 12 µl of the pellet and supernatant samples were loaded on a 12% polyacrylamide gel together with 5 µl Precision Plus Protein Dual Color Standards (BioRad), fixed, stained with Coomassie, and destained with MQ H_2O. Specific modifications to protocol for each purified protein sample is outlined in appendix 3, **Table S1**.

2.3.3.2 Protein Purification by Nickel Affinity Chromatography

Fused to the N-terminal domain of the ParB proteins is a polyhistidine-tag (his-tag) consisting of six consecutive histidine aa, which allows specific separation of these proteins from the rest of the protein content of a bacterial cell through nickel affinity chromatography. A HisPur™ Ni-NTA Purification Kit (Thermo Scientific) was used to purify the his-tagged ParB proteins. The column consists of agarose beads coupled to nickel (Ni^{2+}) ions by a chelating agent, nitrilotriacetic acid (NTA). The aromatic imidazole ring on the histidine side chain has high affinity for nickel and will bind strongly to the column resin. Other proteins without a histidine tag will be washed out through multiple washing steps with buffers containing low concentrations of imidazole (10-25mM). Low concentrations of imidazole will outcompete proteins that are weakly attached to the nickel but not the tightly bound ParB proteins. When other proteins have been washed out, the elution buffer, which contains high concentrations of imidazole (250mM) will outcompete tightly bound ParB proteins which will then be eluted from columns. In order to remove the imidazole, a dialysis against buffer C was performed.

The column was equilibrated to 4°C. The samples were thawed in ice water and mixed with equilibration buffer in a 1:1 ratio. The storage buffer was removed from the columns by

centrifugation at 7000 x g for 2 minutes at 4°C in a centrifuge tube. The column was equilibrated with two resin-bed volumes of equilibration buffer (400 µl) and centrifuged on the same settings. Two resin-bed volumes of sample mixed with equilibration buffer was added to the column, mixed by shaking and inverting the tube, and incubated for 5 minutes on ice. The columns were centrifuged, and the process was repeated until the columns were saturated. 10 resin bed volumes of each sample mixed with one resin bed volume of equilibration buffer (10 mM imidazole) was run through the column in this project. The saturation can be monitored by measuring the absorbance of the wash flow-through with a spectrophotometer at 280 nm, or by running the flow-through on an SDS polyacrylamide gel to check if the protein of interest is no longer retained in the column but is washed out with the flow-through. The resin was then washed with two resin-bed volumes wash buffer (25 mM imidazole) and centrifuged. The washing step was repeated two more times, and the flow through (fractions) were kept from each round to run SDS-PAGE. The his-tagged proteins now retained in the columns were released by the addition of one resin-bed volume of elution buffer (250 mM imidazole), incubated at 5 minutes, and centrifuged. This step was repeated two more times. Protein presence in eluted fractions was confirmed by a filter paper dye binding assay. A small piece of 3mm Whatman Grade 17 Chr Cellulose Chromatography paper was marked with pencil and a drop (2.5 µl) from each of the elution fractions were applied to the paper. Then Coomassie dye was added and the solution agitated for 5 minutes. The paper was destained in 10% acetic acid, and dried. If there is protein present in the sample, it will be dyed blue. The higher the concentration, the stronger the color. Selected samples were dialyzed against 500 ml buffer C for 3 hours at 4°C on a magnetic stirrer, with buffer refreshed every hour. The sample was centrifuged for 10 minutes at 15000 rpm and 4°C. The supernatant was transferred to tubes and flash frozen in dry ice and transferred to -80°C for storage.

2.3.3.3 Bradford Protein Assay

A Bradford assay is a way to determine the total protein concentration in a sample using a dye that reacts with protein in the sample and changes color according to the concentration. Subsequently, the absorbance at 595 nm by a spectrophotometer will vary depending on the color change. The acidic dye itself will be absorbed at 470 nm due to its protonated and unstable state, recognized by its red-brown color. When the dye binds to proteins by hydrophobic and ionic interactions with aa side chains, it will become unprotonated and stable and can be recognized by a shift to blue color which is absorbed at 595 nm. Depending on how much

protein has interacted with the dye, the color change gives an indication of how much protein is present in the sample. The absorbance levels of the unknown protein sample concentrations are compared to a standard curve made from a of dilution series of bovine serum albumin (BSA) proteins of a known concentration. It is important that the standards are made together with the unknown samples to ensure that all conditions are as similar as possible so that the only difference between the samples are in fact due to variable protein concentrations.

A stock solution of 10 mg/ml BSA diluted in MQ H_2O were used to make the standard dilution series. To five plastic cuvettes 1, 2.5, 5, 7.5, and 10 µg of BSA was added. Buffer C was added to a total volume of 10 µl. To a control sample 10 µl of buffer C was added. To all six cuvettes 790 µl MQ H_2O was added, and then 200 µl Quick Start™ Bradford 1X Dye Reagent (Bio-Rad). Unknown protein samples were added to cuvettes in volumes of 1, 5, and 10 µl and adjusted to 10 µl with buffer C. MQ H_2O and Bradford reagent was added as for the standards. The samples were incubated at room temperature for 5 minutes to let the dye form interactions with the proteins, and then the cuvettes were analyzed with a spectrophotometer at 595 nm. Two measurements were performed for each cuvette. A standard curve with a trend line is made from the standard serial dilution, and the concentrations of the unknown samples can be determined based on the trend line function y = ax + b or simply by extrapolating from the standard curve.

2.3.3.4 Nuclease Assay

400 ng his-purified protein samples were added to 200 ng pUC19 plasmid, together with 5 mM $CaCl_2$, 10% BSA, 0.2 µl EcoRI, 3 µl EcoRI Buffer (NEB), and MQ H_2O to 30 µl. An uncut pUC19 sample, a cut pUC19 with EcoRI in buffer C, and an EcoRI cut sample in MQ H_2O were included as controls. The samples were incubated at 37°C for 45 minutes, shaking at 300 rpm. The reactions were stopped with 3 µl EDTA, and the samples were run on a 0.8% agarose gel with GelRed dye, and a 1 kb plus ladder.

2.4 Stem Cell Work

The cells used in this project are transgene OS25_INT3 (INT = 6 x *parS*) mouse embryonic stem cells (mES cells). Beata Nadratowska-Wesolowska has inserted, mediated by CRISPR/Cas9 technology, an INT3 sequence close to the Nanog promoter and an INT1 sequence close to the Nanog enhancer for the purpose of using the ParB-INT system to study promoter and enhancer interactions for the Nanog gene.

2.4.1 Culturing of OS25 Mouse Embryonic Stem Cells

The cells were cultured in T25 and T75 cell culture flasks with stem cell medium and kept at 37°C in a sterile CO_2 incubator, to mimic native temperature and pH conditions. Medium used for culturing of OS25 cells was Glasgow Modified Essential Medium (GMEM) BHK-21 (Gibco) supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin (P/S), 1 mM sodium pyruvate, 1% non-essential amino acids, 1000 U leukemia inhibitory factor (LIF), and 0.1 mM β-mercaptoethanol. Medium was changed and cell culture was split into subcultures every second day.

Subculturing is necessary to avoid overgrowth of cells and consequently differentiation and cell death. A new flask was coated in 0.1% gelatine for a minimum of five minutes. Old medium was aspirated from the cell culture, and the cells were briefly washed with 1X dPBS to remove proteins and debris. dPBS was aspirated from the flask and 1X Typsin-EDTA solution was added to dissociate the cells from the surface of the flask. The flask was incubated with the solution for 2 minutes. Medium was added to dilute the trypsin 10-fold. The solution was resuspended a few times to split large colonies into single cells and then spun down at 1000 rpm for 4 minutes to pellet the cells. The supernatant was removed, and the cells were resuspended in 1-2 ml medium. The cells were counted with the Countess instrument (Invitrogen), and an appropriate number of cells were seeded with fresh medium **(Table 2.7)**.

Table 2.7. Overview of volumes and specificities for subculturing vessels. The volume of different solutions varies depending on the size of the culturing vessel, refer to this overview for reference.

Cell culture vessel	0.1% gelatin (ml)	Trypsin-EDTA (ml)	1X dPBS (ml)	Culturing medium (ml)	Cells to seed
T25 flask	1-2	1	5	5	5×10^5
T75 flask	2-3	2	15	15	1.5×10^6
Slides in quadriperm dish	2	N/A	N/A	5	2×10^6

2.4.2 Quantifying Cells with Countess

The Countess™ II Automated Cell Counter (Invitrogen) was used for counting the number of cells in the culture. To determine the viability of the cells in the solution, a 1:1 mix of cells and Trypan blue dye was added to a cell counting slide and inserted into the Countess.

The dye is impermeable to the cell membrane of live cells, but dead cells have ruptured cell membranes and therefore will retain some of the dye in the cytoplasm. Based on the sample, the Countess estimated the number of cells/ml in the cell suspension as well as the number of live and dead cells/ml. The concentration was used to calculate how many ml of cell suspension should be added to a new cell flask for subculturing.

2.4.3 Transfection

Transfection is the process of introducing foreign DNA into eukaryotic cells. To study the expression of protein constructs in cells, the plasmid that contains the gene of interest needs to be fed into the cell for transcription and translation. In this project the transfecting reagent DreamFect Stem was used, which is specifically designed for use in stem cells. The reagent contains cationic lipids that bind the negatively charged DNA and enclose it in a liposome structure and will release the enclosed DNA into the cytoplasm by endocytosis.

The day before transfection, cells were seeded on microscopy slides. The slides were first sterilized in 70% EtOH and dried. Then they were placed in quadriperm dishes and coated with 0.1 % gelatin. The cells were split as described in section 2.4.1 and ~ 2x 10^6 cells were seeded on each slide in 5 ml of fresh medium. Next day, cells were transfected. 3 µg plasmid DNA was diluted in 150 µl Opti-MEM reduced serum medium and mixed with 9 µl of DreamFect transfection reagent diluted in the same volume of Opti-MEM. The contents of the two tubes were mixed, gently vortexed, spun down, and incubated at room temperature for 20 minutes. The medium on the slides were aspirated, and 4 ml fresh medium was added. The 300 µl transfection mix was added to the slide dropwise, carefully covering the slide surface. The dish was gently agitated before returning it to the incubator for a 24-hour incubation step.

2.4.4 Cell Fixation on Slides

Fixation of cells is done to preserve the cells components for further analysis. It stops activity in the cell, which means that we will only get to analyze what is happening in the cell at the exact time of fixation. It allows us to look at a protein that is transiently expressed and visible for a short time. Formaldehyde has been used as a fixative as it covalently cross-links aa to nucleic acid bases and therefore importantly preserve the ParB-INT interactions (Hoffman et al., 2015). Cell fixation is performed for a number of applications, such as preparation for immunostaining, imaging, long-term preservation etc. In this study, the cells were fixed to

preserve the fluorescent signal for imaging. DAPI was used as a fluorescent stain to color DNA and nuclear content, to visualize the nucleus in a microscope.

Medium was aspirated from transfected cells, and the slides were gently washed three times with 5 ml 1X dPBS. 5 ml 4 % formaldehyde was added and incubated for 10 minutes in room temperature. The formaldehyde was removed, and slides were washed two times with 5 ml 1X PBS. 5 ml 50 mM NH_4Cl in 1X PBS was added to each slide and incubated for 10 minutes. NH_4Cl works as a quencher of background signals, because it binds to the autofluorescent free aldehyde groups. The slides were then washed three times with 5 ml 1X PBS and mounted for use in microscope. Cover slips were cleaned in 100 % EtOH and air dried. VectaShield was supplemented with DAPI for visualization of nuclei. 40 µl of the Vectashield/DAPI was applied to each cover slip. The cell side of the slides were placed down on the cover slip surface and attached. The slide was left to dry for a few minutes, then the edges of the cover slip were sealed with nail polish. The slides were dried in a cool and dark place and stored at 4°C.

2.4.5 Microscopy

2.4.5.1 Light Microscopy for Live Cell Visualization

To monitor the daily growth and condition of the OS25_INT3 cells, the cell cultures were visually inspected with the Olympus CK40 Phase Contrast Tissue Culture Microscope.

2.4.5.2 Fluorescence Imaging

Fluorescent imaging was done on the Zeiss 710 confocal microscope. Fast 3D Z—stack images were captured. Images were analyzed with the FIJI version of ImageJ (version 2.0.0-rc-69/1.52p). All visualization on fluorescent microscopy was performed by Beata Nadratowska-Wesolowska.

2.5 Protein Structure Prediction

2.5.1 Molecular Modeling

2.5.1.1 I-TASSER

I-TASSER (iterative threading assembly refinement) is a template based tool for *ab initio* prediction of protein structure and function (Roy et al., 2010; Zhang, 2008). I-TASSER is subjected with an aa sequence for which it identifies structural templates from the Protein Data Bank (PDB) library utilizing a method called Threading. Threading, also known as fold recognition, is a method of protein structure prediction which models the query sequence to known structural protein folds. The method is based on the assumption that only a small number of folds exist in nature, and that most of them have already been discovered and are registered in the PDB library. Most likely a fold will be suggested for the core domain and the flanks can be predicted based on that fold. The threading method aligns each aa to a target template and evaluates how well the sequence aligns to the structure. Based on the templates with the best fit, i.e. with the lowest energy conformation, the rest of the model that has not aligned will be built *ab initio* to fit the sequence. The resulting models will be clustered based on the hypothesis that the structures that are closest to the lowest energy state, the native state, will be the most populated states. From this, the structures are again re-assembled for refinement, and finally matched with the BioLip database for insights into the function of the modeled protein (Zhang, 2008). The top 10 models are the outputs from the I-TASSER pipeline, and the top five are automatically selected ranked by cluster size. Each model is given a confidence score (C-score). The C-score estimates the overall quality of each model. The higher the score, the better the quality of the model is. Typically, the C-score of predicted models are found in the range of -5 to 2. Each model I-TASSER was used in this thesis for an *ab initio* modeling approach where disordered regions are modeled and to compare it with the homology modelling templates from SWISS-MODEL and ModWeb to get a better guess of the native protein structure.

2.5.1.2 SWISS-MODEL

Homology modeling was performed with the SWISS-MODEL modeling pipeline as a complement to the I-TASSER fold recognition modeling approach (Bertoni et al., 2017; Bienert et al., 2017; Guex et al., 2009; Studer et al., 2020; Waterhouse et al., 2018). Homology

modeling is based on the presumption that a closely related protein, a homolog, has an experimentally solved structure that is available. The modeling starts with a search of the structural database for candidates that are comparably similar to the query sequence. Aa residues that are not conserved between the template and the candidates are modeled based on a rotamer library, and optimal configuration is determined by energy minimization algorithms. Energy minimizing algorithms are applied for the model in its entirety to resolve unfavorable interactions and distortions of aa residues. All models generated by the SWISS-MODEL pipeline are given a GMQE (Global Model Quality Estimate) score, which estimates the quality of the given model based on the expected model accuracy calculated from how well the template aligns with the query. The score is value between 0-1, the higher value the better the model quality. The QMEAN score, which gives both a local and global estimate of the expected accuracy of the model, can also be compared between the models to determine model quality. A score of ~0 indicates a comparatively good model to the experimental structures, and a QMEAN value of -4 or below indicates low agreement with experimental structures.

2.5.1.3 Modeller

The Modeller web server, ModWeb version r217, is an automated web server that performs protein structure prediction by comparative structure modeling (Eswar et al., 2003) with a homology modeling approach. The target sequence is run through PSI-BLAST (position specific iterative basic local alignment search tool), for homologous sequence alignments and comparison with a PSSM (protein specific scoring matrix) which will recognize a conservation pattern in even distantly related proteins. These matches are aligned with template structures in the PDB, creating sequence-structure matches. These matches are also scanned against a database of PSI-BLAST profiles (IMPALA). The output is three predicted protein structures based on the fed aa sequence, which are given a MPQS quality score comprising of sequence identity to the template, how much of the query sequence is covered by the template, the number of gaps in the alignment to the template, energy scores, and model compactness. A value greater than 1.1 is considered reliable. ModWeb was used in this thesis as an additional homology-based prediction server to compare with the results of SWISS-MODEL to further aid the search for the best model prediction.

2.5.2 ProSA

ProSA-web server is a protein structure analysis tool that is utilized for refinement and validation of structure prediction models. ProSA-web provides a 3-dimentional structure viewer, a quality score, and energy scores for each aa position. The Z-score is an overall model quality score.

It calculates the deviation between the total energy score of the model and various random conformations (Sippl, 1993; Wiederstein and Sippl, 2007). A model Z-score that is not within the range that is calculated for native proteins, suggest a misfolding of the protein. The model score is indicated in a plot of the Z-scores of other proteins of similar size that have been experimentally determined by NMR or X-ray crystallography, for comparison with native protein scores (Wiederstein and Sippl, 2007). ProSA-web scores and plots further highlight structure prediction quality of input models as well as potential problems in protein structures and is useful for comparing the quality of different models.

2.5.3 Rampage

Rampage is a tool that gives a stereochemical quality analysis of the protein structure (Lovell et al., 2003). The model of interest is fed into the tool and the output contains a range of information pertaining to the stereochemical quality of the structure, presented in a Ramachandran plot. The Ramachandran plot details the psi and phi torsion angles of the individual aa. The phi angles are plotted on the x-axis and the psi angles on the y-axis. There is only a certain number of angles that would actually be stereochemically possible in real life proteins. A Ramachandran plot shows the aa residue conformations that are disallowed in nature (outlier) because of a stereochemical hinderance between two aa. The plot has four different regions in regard to the torsion angles: favored, allowed, and outlier regions. Glycine is more commonly found in outlier and favored regions than other aa. Glycine only has a single hydrogen on its side chain, which enables it to resume conformations that would be disallowed for other aa with longer side chains. The glycine aa in outlier and favored regions that are in fact favored will be labelled accordingly.

2.5.4 DISOPRED3 Disorder Prediction Server

Intrinsically disordered regions are stretches of aa in a protein recognized by a lack of stable secondary and/or tertiary structure. These regions will exist in a continuous

conformationally interchanging condition, in which no single Ramachandran angle description will suffice (Li et al., 2015). The DISOPRED3 tool uses machine learning to predict disordered regions in proteins. It has been trained by a series of conserved features of intrinsically disordered regions to recognize these regions in a protein structure based on its aa sequence (Jones and Cozzetto, 2015; Li et al., 2015). The aa sequences of the ParB proteins were subjected to disorder prediction by the DISOPRED3 tool.

2.5.5 UCSF Chimera

Visualization of protein structures and analyses are performed with UCSF Chimera version 1.14 (Pettersen et al., 2004). This is a modeling tool with which you can visualize a protein and make changes to highlight selected regions and interactions. Chimera supports different input formats, including PDB. Superimposition of structures for comparison is possible with the MatchMaking command, which superimposes the structures based on a pairwise sequence alignment and then uses the structural information for superimposition. A root mean square deviation (RMSD) value can be calculated from the superimposition. The RMSD value describes the average distance between the atoms of the superimposed protein structures, thereby giving a measure of how much the protein structures differ. An RMSD value of 0 Å (angstrom) describes identical structures. The value increases with less similarity.

2.6 Homology Search

2.6.1 ConSurf

ConSurf is a server that calculates evolutionary conservation between homologous sequences based on their phylogenetic relationships (Ashkenazy et al., 2016, 2010; Celniker et al., 2013). The tool includes sequence information as well as structural information in the search for homologs, if a known structure exists. Sequences with structural and sequence similarity are compiled in a multiple sequence alignment created with MAFFT. In this project, select models predicted by SWISS-MODEL were subjected to the ConSurf server. Protein sequence can also be used as an input in Consurf to find structural and sequential homologs. The Bayesian method for estimation of conservation was used for all searches. The Clean Uniprot database was searched, which is a modified version of Uniprot with more reliable sequences, with a HMMER search algorithm. Based on a generated multiple sequence alignment (MAFFT), ConSurf provides homologs identified by the selected parameters.

2.6.2 MAFFT Alignment Tool

MAFFT (Multiple Alignment using Fast Fourier Transform) version 7 is a multiple sequence alignment tool which can align a number of homologous sequences (Katoh et al., 2019; Kuraku et al., 2013). It uses by default the BLOSUM62 scoring matrix for aa sequences. The BLOSUM62 is chosen in this project because sequences that are both closely and more distantly related protein sequences will be aligned in the same alignment. The alignment is run in the high-accuracy L-INS-I MAFFT program, which aligns the sequences around one well-aligned domain (Katoh and Standley, 2013). All alignments made in this project were visualized with CLC Main Workbench. MAFFT can also construct a phylogenetic tree based on the alignment. The tree is made with the Neighbor Joining method (NJ) and the output format is Newick. The phylogenetic tree in this project was visualized with ITOL (https://itol.embl.de).

Table 2.8. List of bioinformatics tools applied.

Tool	Application purpose	URL	References
I-TASSER	Modeling by threading	https://zhanglab.ccmb.med.umich.edu/I-TASSER/	(Roy et al., 2010; Yang and Zhang, 2015; Zhang, 2009)
SWISS-MODEL	Homology modeling	https://swissmodel.expasy.org/interactive	(Benkert et al., 2010; Bertoni et al., 2017; Bienert et al., 2017; Guex et al., 2009; Studer et al., 2020; Waterhouse et al., 2018)
ModWeb	Homology modeling	https://modbase.compbio.ucsf.edu/modweb/	(Eswar et al., 2003)
ProSA	Energy analysis of protein structure	https://prosa.services.came.sbg.ac.at/prosa.php	(Sippl, 1993; Wiederstein and Sippl, 2007)
Rampage	Stereochemical analysis of residues	http://mordred.bioc.cam.ac.uk/~rapper/rampage.php	(Lovell et al., 2003)
DISOPRED3	Prediction of disordered	http://bioinf.cs.ucl.ac.uk/psipred/	(Jones and Cozzetto, 2015)
UCSF Chimera	Molecular modeling software	http://www.cgl.ucsf.edu/chimera/ (Software download)	(Pettersen et al., 2004)
ConSurf	Homology search	https://consurf.tau.ac.il	(Ashkenazy et al., 2016, 2010; Celniker et al., 2013)
MAFFT	Multiple sequence alignment	https://mafft.cbrc.jp/alignment/server/	(Katoh et al., 2019; Kuraku et al., 2013)

3 Results

In this project a bioinformatics analysis of sequence-structure-function relationships of the ParB protein family is presented. The structural features, sequence specificities, and nuclease activity of two partitioning proteins from the Burkholderiaceae family were studied, species *Burkholderia cenocepacia* and an undisclosed exotic species (appendix 1), and will be referred to as ParB1 and ParB3, respectively.

3.1 *In Silico* Protein Structure Prediction: Homology modeling and Threading of ParB1

Protein function is largely dictated by its three-dimensional structure and conformational dynamics. Protein folding is an essential process thus making structure a very significant protein trait. The spatial folding of ParB proteins enables us to understand the biological function and location of the active site. The ParB proteins have large stretches of disordered regions, making it challenging to fully capture the structure of the whole protein by present technologies such as X-ray crystallography. However, many ParB homologs with resolved core structures are available. We performed structure prediction analysis for ParB proteins to get more insight into the protein function and features. In the following, a visualization of the full predicted protein structure with its disordered regions along with other sequence-based analysis is presented.

The two proteins ParB1 and ParB3 were modeled with the modeling tools SWISS-MODEL (SM), ModWeb (MOD), and I-TASSER (IT) **(Table 2.8)**. Homology modeling with MOD and SM was performed to identify well-fitted folds in homologous proteins. I-TASSER was used to model the intrinsically disordered N-terminal and C-terminal domains by *ab initio* modeling, and to model the folds of the proteins from a threading approach. All models were compared and analyzed to find the best possible model predicting the protein structure of these two prokaryotic proteins.

Protein structures modeled for ParB1 were evaluated and compared according to their quality scores **(Table 3.1)**. MOD generated three models. Model 1 was the only complete core model of amino acids (aa) 85-286 and had an MPQS model quality score of 0.78, while model 2 and 3 showed protein structure models only for aa numbers 7 to 83 with MPQS score of 0.41

and from aa 196 to 298 with MPQS score of 0.46, respectively (section 2.5.1.2). The low quality scores were also indicative of low model confidence.

Eleven models were calculated by the SM tool, ranked by their GMQE model quality score (section 2.5.1.1). The top five models were selected for further analysis, and the six other models were not further considered owing to their low scores. In addition to the GMQE score, a QMEAN quality score was also included for each model **(Table 3.1)**. The GMQE score ranks model 1 and 2 on top with values of 0.38 and 0.37, and with top QMEAN scores of -1.08 and -1.35, respectively. The model ranked third has a lower GMQE score of 0.33 and a much lower QMEAN score of -5.47. Model 1 and 2 ranked highest in overall quality, and therefore will be structurally compared to the selected MOD model. MOD model 1 was superimposed with SM model 1 and 2 using the MatchMaker function in UCSF Chimera for comparison and visualization of protein structures. The MOD model aligned very well with SM model 2 (RMSD of 0.790 for 142 pruned atom pairs), but not well with SM model 1 (RMSD of 1.235 for 73 pruned atom pairs). SM model 2 only predicts the highly stable core region from aa residues 84 to 288.

The core of the ParB homolog of *Bacillus subtilis* has been solved by x-ray crystallography (PDB: 6SDK), and was the homolog protein used for structure prediction of both MOD model 1 and SM model 2. Intrinsically disordered regions are not readily solved by x-ray crystallography as they do not have a single, stable conformation that can be plotted on an electron density map (Wójcik et al., 2018). The lack of structured N-terminal and C-terminal regions from the resolved *B. subtilis* protein could indicate that these regions are disordered in ParB proteins (Audibert et al., 2020; Shaw et al., 2008).

The ParB1 protein sequence was analyzed with the DISOPRED3 tool (part of the PSIPRED Protein Analysis Workbench) **(Table 2.8)**. The tool predicted which aa residues of ParB1 were part of intrinsically disordered regions **(Figure 3.1 A)**. The aa residues above the threshold value of 0.5, are above the false positive cutoff and predicted to be disordered. The N-terminal region was predicted to be intrinsically disordered from aa 1 to 73. There is no further predicted disorder in this protein, although two peaks of aa just below the threshold at ~290-300 and ~350 may indicate otherwise.

Ab initio modeling with IT was necessary to find a structure for the N- and C-terminal regions, which in sequence-based analysis with DISOPRED3 appears to be caused by intrinsic disorder. The IT tool computes 10 models and automatically selects the top five models ranked by their confidence score (C-score) **(Table 3.1)**. The IT C-scores for the models were either very similar or identical, leading to a necessity for further analysis of the models for selection

of the best representative models. All five models were analyzed with ProSA-webserver, which created an energy score (Z-score) for each model (section 2.5.2) (Graphs available in appendix 3, **Figure S1 A-E**). A low Z-score shows that the protein is more stable, and the models with lowest Z-score is more likely to represent the native structure. Model 5 had the lowest Z-score and model 1 had the highest. The models were further analyzed with Rampage, which generates Ramachandran Plots (Ramachandran et al., 1963) by calculating the stereochemical probability of each aa residue position according to the backbone positions (plots available in appendix 3, **Figure S2**). The output shows each residue classified as either existing in a favored, allowed, or outlier region. The less aa residues in the outlier region and the more in favored region, the more likely the structure is to exist in nature. The top model based on the Rampage analysis is model 4 with 69.2% of its aa residues in the favored region and 11.7% in the outlier region. The second best is model 5 with 67.2% favored and 12.0% outlier aa residues. In total IT model 4 appears to be of highest quality based on the analysis results. All five IT models were superimposed with the 11 SM models in UCSF Chimera. All IT models, including model 4 (RMSD of 0.734 Å for 158 pruned atom pairs), aligned very well with SM model 2, which also aligned well with MOD model 1, and poorly with all of the others.

All three modeling tools generated the highest quality models from the fold of the homolog ParB homo-dimer from *Bacillus subtilis*. I decided that IT model 4 was the best attempt at describing the structure of these intrinsically disordered regions.

Table 3.1. Analysis of predicted models from SWISS-MODEL (SM), ModWeb (MOD), and I-TASSER (IT) for ParB1.

Model	Mod Web[a]	SWISS-MODEL[b]		Model	I-TASSER[c]	ProSA[d]	Rampage[e]		
-	MPQS	GMQE	QMEAN	-	C-score	Z-score	Favored (%)	Allowed (%)	Outlier (%)
1	**0.78**	0.38	-1.08	1	-3.21	-3.64	67.0	19.1	14.0
2	0.41	**0.37**	**-1.35**	2	-3.21	-4.38	67.2	17.7	15.1
3	0.46	0.33	-5.47	3	-3.56	-4.78	64.4	21.1	14.5
4	-	0.31	-5.47	4	**-3.45**	**-4.45**	**69.2**	**19.1**	**11.7**
5	-	0.27	-4.48	5	-3.60	-5.08	67.2	20.8	12.0

[a] The ModWeb MPQS quality score is ~1 for high quality models.

[b] SM analysis by two quality scores, GMQE is the template to query score between 0 and 1, and QMEAN is the local and global quality score, which should be ~0 for high quality models. Models are ranked by the GMQE score.

[c] IT evaluates the models by a quality C-score that estimates the overall quality of the prediction, usually in the range -5 to 2.

[d]IT models were analyzed with ProSA which gave an energy Z-score. Graphs can be found appendix 3 Figure S1 (A-E).

[e]IT models are analyzed by Rampage, which gives stereochemical classification of aa residues. Ramachandran plots are found in appendix 3, Figure S2.

3.2 *In Silico* Protein Structure Prediction: Homology Modeling and Threading of ParB3

The similar procedure as explained in section 3.1 was applied to the ParB3 protein. MOD yielded one complete core model of aa 68 to 265 with a high-quality score of 0.90 and two incomplete and poor-quality models for aa 167 to 280 with an MPQS of 0.49 and aa 261 to 344 with a score of 0.47 **(Table 3.2)**. SM generated 12 models where the top five were selected and compared by their GMQE and QMEAN scores. The GMQE score ranked model 1 as the best quality model with a score of 0.39. The QMEAN score is also highest for this model with -1.57. Model 2 was ranked second best with a GMQE score of 0.36 and a QMEAN score of -1.74. These two models were therefore selected for further comparison. The MOD model 1 was superimposed with SM model 1 and 2. The MOD structure aligned very well with model 1 (RMSD of 0.774 for 144 pruned atom pairs) and poorly with model 2 (RMSD of 1.171 for 73 pruned atom pairs). Both MOD model 1 and SM model 1 was modeled from the fold of

B. subtilis, identical to the ParB1 models. SM model 1 covers aa residues 65 to 269 of the total sequence.

A sequence analysis was performed for ParB3 by DISOPRED3 **(Figure 3.1 B)**. For this protein, DISOPRED3 predicted three intrinsically disordered regions. In this case, disordered aa were predicted in both the N- and C-terminal regions, which further supports the presumption that SM generates core models because the terminal regions are disordered.

Ab initio modeling with IT is necessary for a complete model. The top five models generated by IT were more comparable for this protein with a higher discrepancy between the C-scores **(Table 3.2)**. A ProSA-webserver analysis gave a very low energy score of -6.0 for model 5 and -5.3 for model 2 (graphs in appendix 3, **Figure S1 F-J**).

The stereochemical analysis with Rampage classified 78.1% of aa residues in favored region and 7.6% in outlier region for model 5. Model 2 ranked second with 72.3% of aa residues in favored region and 9.6% in outlier region. These two models score much higher in the analysis than the remaining three models (**Table 3.2**) (plots not disclosed). Because the IT C-score is higher for model 2 and ProSA and Rampage scores model 5 higher, both models are kept as potentially good estimates of the native structure. The five IT models were superimposed with the 12 SM models. IT model 1 and 3 were well superimposed with SM model 7, while IT models 2, 4, and 5 aligned very well with SM model 1, the same model that aligned very well with MOD model 1. This superimposition supports the selection of IT models 2 (RMSD of 0.9 Å for 140 pruned atom pairs) and 5 (RMSD of 0.863 Å for 163 pruned atom pairs) as good candidates.

Table 3.2. Analysis of predicted models from SWISS-MODEL (SM), ModWeb (MOD), and I-TASSER (IT) for ParB3.

Model	ModWeb[a]	SWISS-MODEL[b]		Model	I-TASSER[c]	ProSA[d]	Rampage[e]		
-	MPQS	GMQE	QMEAN	-	C-score	Z-score	Favored (%)	Allowed (%)	Outlier (%)
1	**0.90**	**0.39**	**-1.57**	1	-3.28	-3.97	64.4	23.0	12.5
2	0.49	0.36	-1.74	2	**-3.82**	**-5.3**	**72.3**	**18.1**	**9.6**
3	0.47	0.34	-4.58	3	-3.92	-3.19	71.7	16.3	12.0
4	-	0.32	-4.58	4	-4.20	-2.17	69.4	16.0	14.6
7	-	0.26	-4.28	5	**-4.06**	**-6**	**78.1**	**14.3**	**7.6**

[a] The ModWeb MPQS quality score is ~1 for high quality models.

[b] SM analysis by two quality scores, GMQE is the template to query score between 0 and 1, and QMEAN is the local and global quality score, which should be ~0 for high quality models. Models are ranked by the GMQE score.

[c] IT evaluates the models by a quality C-score that estimates the overall quality of the prediction, usually in the range -5 to 2.

[d]IT models were analyzed with ProSA which gave an energy Z-score. Graphs can be found appendix 3 Figure S1 (F-J).

[e]IT models are analyzed by Rampage, which gives stereochemical classification of aa residues. Ramachandran plots remain undisclosed.

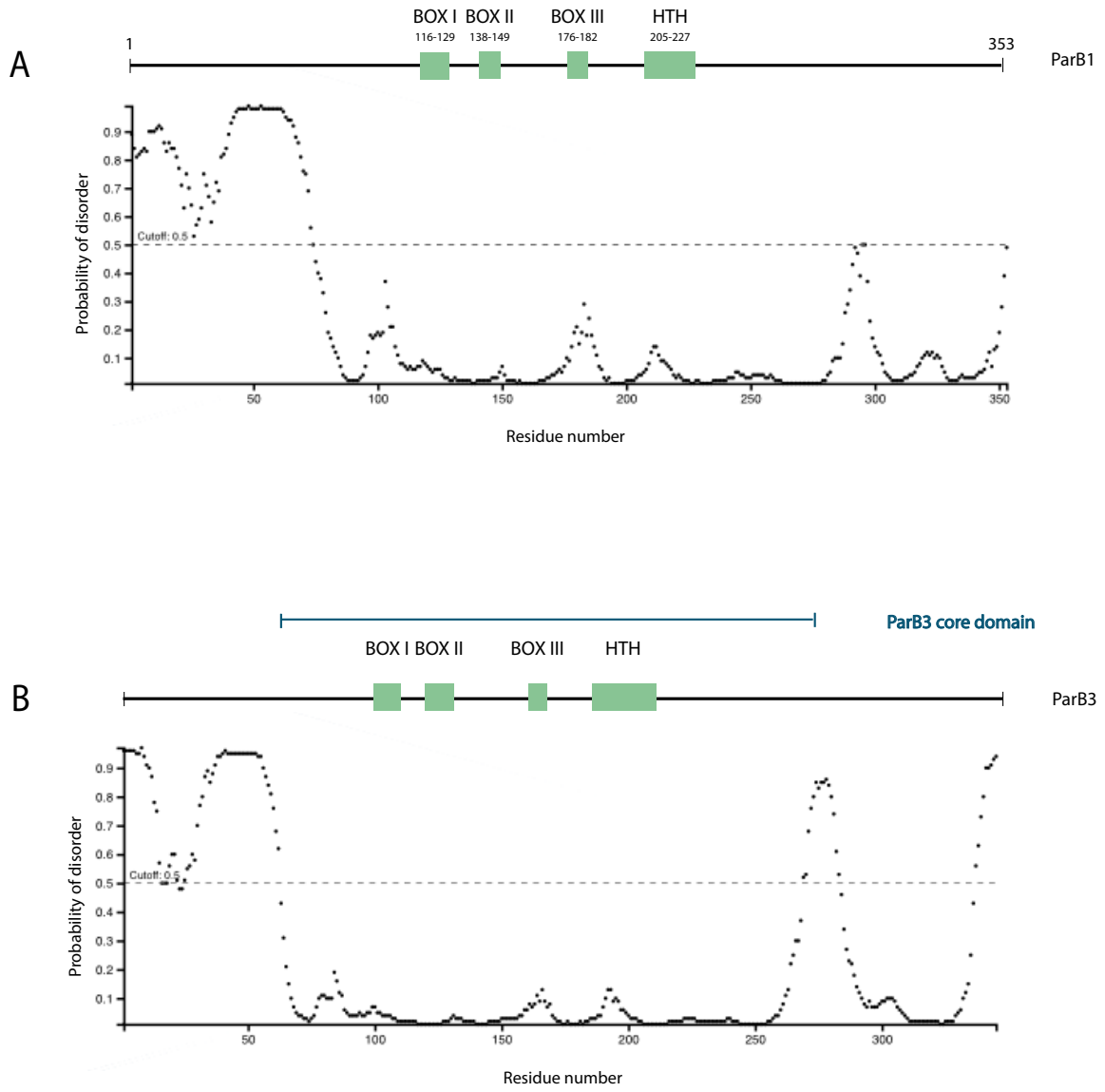


Figure 3.1. Prediction of intrinsically disordered regions. The protein sequences of ParB1 **(A)** and ParB3 **(B)** were analyzed for disorder with DISOPRED3 (PSIPRED). Aa residue number on the x axis is plotted against probability of disorder on the y axis. An automatically selected cutoff value of 0.5 separates ordered aa residues from disordered aa residues. The signature motifs of well-conserved regions are indicated by green boxes (corresponding aa are confidential for ParB3). A core sequence of the ParB3 protein is shown in blue.

The structures of the selected IT models, ParB1 model 4 and ParB3 model 2 and 5, are illustrated in **Figure. 3.2 A, C and D**. The alignment with the SM models is shown in the ribbon model where the SM models in red are well aligned with the selected IT models in blue, which again indicates that the protein structures are modeled from the same fold. The same selected fold for all models indicates that the structures have a stable core that is well modeled. The stability of the core is supported by the Rampage results for each of the three models. In **Figure 3.2 B** aa residues for ParB1 model 4 that are in outlier regions are colored red. These are located

primarily in the N- and C-terminal regions, which supports the findings from the DISOPRED3 analysis where residues in the N-terminal region are predicted to be intrinsically disordered. According to this figure, a larger number of red residues is found in the C-terminal region, indicating as previously suggested that there are disordered amino acids in this region as well. Aa residues in allowed regions marked in yellow follow the same trend and are mainly located in the terminal regions. The core has fewer colored residues, indicating higher stability.

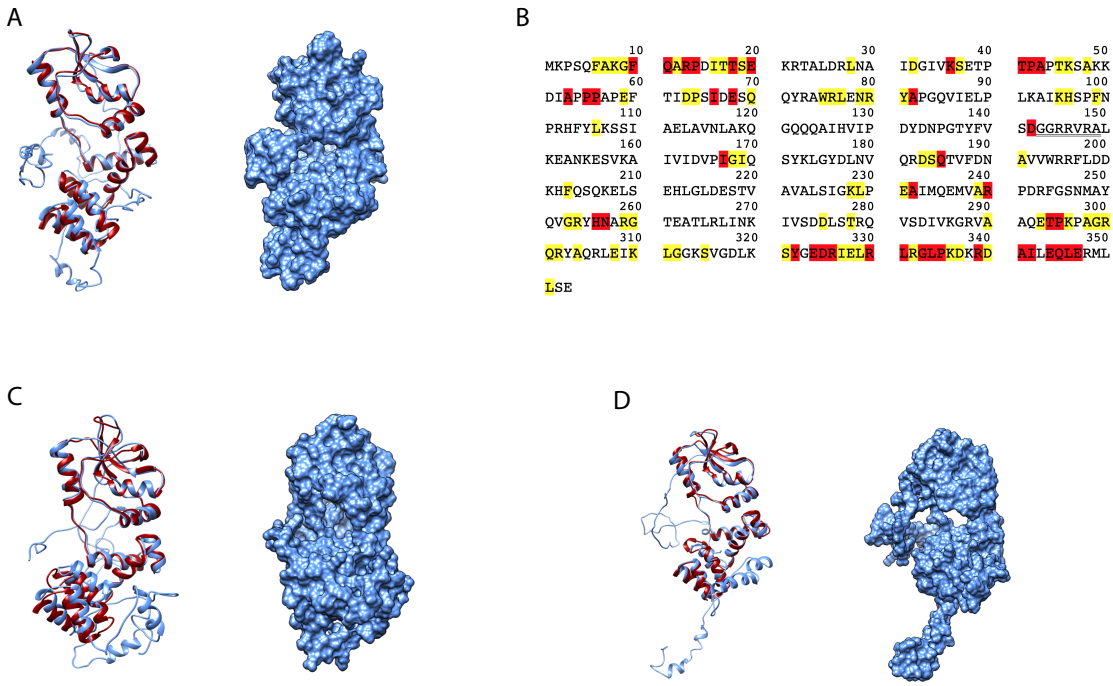


Figure 3.2. Predicted models of ParB1 and ParB3 proteins. Model predictions made with IT is shown in blue ribbon and surface structures, superimposed with SM models in red ribbon structures. **(A)** IT model 4 and SM model 2 predicted for ParB1. **(C)** IT model 2 and SM model 1, and **(D)** IT model 5 and SM model 1 predicted for ParB3. Structural visualization and superimposition was carried out using UCSF Chimera. **(B)** Ramachandran plot classification of aa residues for ParB1 IT model 4. Non-colored aa residues are in favored regions, aa residues in yellow are in allowed regions, red aa residues are in outlier regions.

3.3 Homology Search and Evolutionary Relationship

3.3.1 Evolutionary Relationship of ParB Family Proteins

Studying the evolutionary relationship between proteins within a family can expand our knowledge about the characteristics of a protein. SM ParB1 model 2 and ParB3 model 1 were used as input in the ConSurf server to search for homologs (section 2.6.1; **Table 2.8**). The server created a MAFFT alignment from a homology search, then removed redundant sequences. A HMMR search algorithm was selected and both one and two iterations were

performed to accumulate closely related homologs (one iteration) and more distantly related homologs (two iterations). The selected e-value cutoff was 0.001. Conservation scores were calculated with the default Bayesian method. The Clean Uniprot search database was selected, which is a collection of more reliable sequences from the Uniprot database. The default maximal and minimal sequence identity was 95% and 35%, respectively. A 95% maximal identity was set to remove identical sequences, and 35% minimal identity was selected to minimize the number of false positives in the alignment. From the search with two iterations, ten homologs for ParB1 and nine homologs for ParB3 were manually selected based on the criteria that no identical species should be included, and full species names had to be given. Five homologs for ParB1 and two for ParB3 was selected from the single iteration search. A literature search was also performed with the help of Ragnhild Eskeland where 15 sequences discussed in relevant articles were added to the ConSurf and BLAST compilation (Basu and Koonin, 2005; Dubarry et al., 2006; Passot et al., 2012; Shaw et al., 2008; Song et al., 2017). Three sulfiredoxin (Srx) proteins from species *Homo sapiens*, *Arabidopsis thaliana*, and *Saccharomyces cerevisiae* were included from this literature search. This was based on the articles from Basu et al. (2005), Chi et al. (2012), and Soh et al. (2019), which show the evolutionary relationship of the functionally unrelated eukaryotic Srx protein to the bacterial partitioning proteins. The Srx enzyme has gained sulfinic acid reductase activity, but has also been shown to possess nuclease activity (Chi et al., 2012).

Each protein in the final compilation was given a shortened name of the three first letters of the genus name and the first two letters of the species name. The full sequence names including Uniprot entry names of the species in the alignment can be found in appendix 9. ParB3 was omitted in the following analysis due to confidentiality.

A multiple sequence alignment was generated for the collected sequences with MAFFT as described in section 2.6.2, using default parameters with a BLOSUM62 scoring matrix (**Table 2.8**). The automatically selected alignment method was MAFFT-L-INS-i, which is an iterative method that creates local pairwise alignments for small data sets. The alignment presented in **Figure 3.3** shows highly conserved aa residues in dark blue. BOX I-III are regions of high conservation that have been recognized in earlier studies and are likely important to protein function (Bartosik et al., 2004; Osorio-Valeriano et al., 2019; Yamaichi and Niki, 2000). The framed red region in BOX II is referred to as the GxxRxxA site, named by its highly conserved residues. Recently, the GxxRxxA site has been recognized as the main site for CTP hydrolysis activity in Srx proteins (Jalal et al., 2020; Osorio-Valeriano et al., 2019; Soh et al., 2019). Conservation of this site in prokaryotes could suggest that the ParB proteins also catalyze

CTP to CDP with H2O and inorganic phosphate as by-products. The region has also been studied as a potential site for nuclease activity (Chi et al., 2012). The background color of the aa residues indicates high conservation in all homologs, including the Srx proteins. The nuclease activity site is likely well conserved in the Srx proteins because the activity has been conserved. Aa residues in BOX I and BOX II are well conserved in all proteins. BOX III residues are highly conserved in all ParB proteins, but this box is absent in the Srx protein sequences. BOX I-III has been suggested to play a crucial role in ParB spreading, complex formation, ParA positioning, segregation, and recently, CTPase activity (Bartosik et al., 2004; Graham et al., 2014; Osorio-Valeriano et al., 2019; Soh et al., 2019; Song et al., 2017). The HTH motif has been established as the DNA binding site and has high sequence conservation in the ParB proteins. The aa in this region are largely absent in the Srx proteins, although DNA-binding property has been observed (Chi et al., 2012).

Figure 3.3.

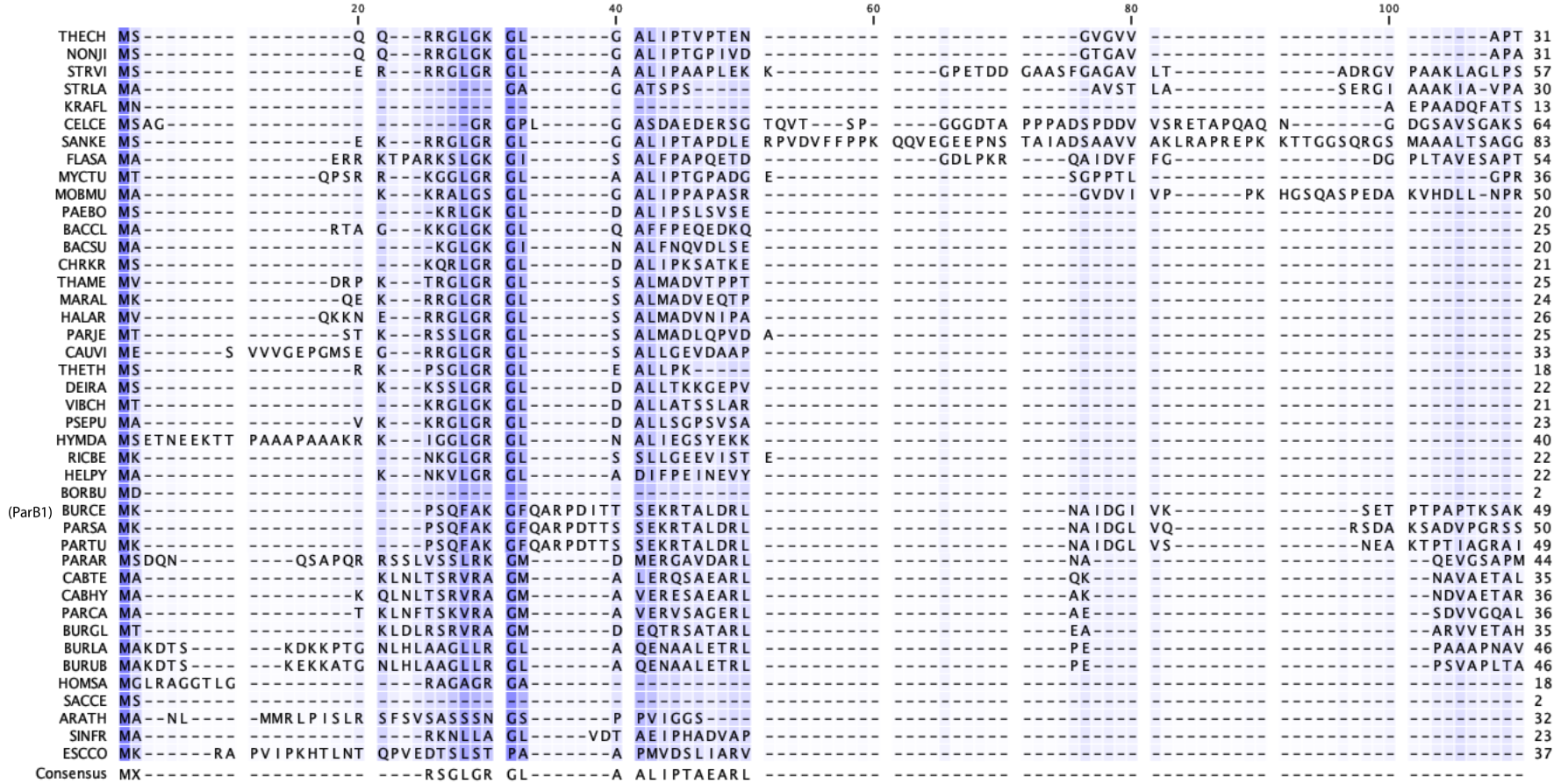

20
40
60
80
100
THECH MS-------- ---------Q Q---RRGLGK GL-------G ALIPTVPTEN ---------- ---------- -----GVGVV ---------- ---------- -------APT 31
NONJI MS-------- ---------Q Q---RRGLGK GL-------G ALIPTGPIVD ---------- ---------- -----GTGAV ---------- ---------- -------APA 31
STRVI MS-------- ---------E R---RRGLGR GL-------A ALIPAAPLEK K--------- ----GPETDD GAASFGAGAV LT-------- -----ADRGV PAAKLAGLPS 57
STRLA MA-------- ---------- ---------- GA-------G ATSPS----- ---------- ---------- ------AVST LA-------- -----SERGI AAAKIA-VPA 30
KRAFL MN-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------A EPAADQFATS 13
CELCE MSAG------ ---------- --------GR GPL------G ASDAEDERSG TQVT---SP- ----GGGDTA PPPADSPDDV VSRETAPQAQ N--------G DGSAVSGAKS 64
SANKE MS-------- ---------E K---RRGLGR GL-------G ALIPTAPDLE RPVDVFFPPK QQVEGEEPNS TAIADSAAVV AKLRAPREPK KTTGGSQRGS MAAALTSAGG 83
FLASA MA-------- -------ERR KTPARKSLGK GI-------S ALFPAPQETD ---------- ----GDLPKR ----QAIDVF FG-------- --------DG PLTAVESAPT 54
MYCTU MT-------- ------QPSR R---KGGLGR GL-------A ALIPTGPADG E--------- ---------- ----SGPPTL ---------- ---------- -------GPR 36
MOBMU MA-------- ---------- K---KRALGS GL-------G ALIPPAPASR ---------- ---------- -----GVDVI VP------PK HGSQASPEDA KVHDLL-NPR 50
PAEBO MS-------- ---------- -----KRLGK GL-------D ALIPSLSVSE ---------- ---------- ---------- ---------- ---------- ---------- 20
BACCL MA-------- -------RTA G---KKGLGK GL-------Q AFFPEQEDKQ ---------- ---------- ---------- ---------- ---------- ---------- 25
BACSU MA-------- ---------- -----KGLGK GI-------N ALFNQVDLSE ---------- ---------- ---------- ---------- ---------- ---------- 20
CHRKR MS-------- ---------- ----KQRLGR GL-------D ALIPKSATKE ---------- ---------- ---------- ---------- ---------- ---------- 21
THAME MV-------- -------DRP K---TRGLGR GL-------S ALMADVTPPT ---------- ---------- ---------- ---------- ---------- ---------- 25
MARAL MK-------- --------QE K---RRGLGR GL-------S ALMADVEQTP ---------- ---------- ---------- ---------- ---------- ---------- 24
HALAR MV-------- ------QKKN E---RRGLGR GL-------S ALMADVNIPA ---------- ---------- ---------- ---------- ---------- ---------- 26
PARJE MT-------- --------ST K---RSSLGR GL-------S ALMADLQPVD A--------- ---------- ---------- ---------- ---------- ---------- 25
CAUVI ME-------S VVVGEPGMSE G---RRGLGR GL-------S ALLGEVDAAP ---------- ---------- ---------- ---------- ---------- ---------- 33
THETH MS-------- ----------R K---PSGLGR GL-------E ALLPK----- ---------- ---------- ---------- ---------- ---------- ---------- 18
DEIRA MS-------- ---------- K---KSSLGR GL-------D ALLTKKGEPV ---------- ---------- ---------- ---------- ---------- ---------- 22
VIBCH MT-------- ---------- ----KRGLGK GL-------D ALLATSSLAR ---------- ---------- ---------- ---------- ---------- ---------- 21
PSEPU MA-------- ---------V K---KRGLGR GL-------D ALLSGPSVSA ---------- ---------- ---------- ---------- ---------- ---------- 23
HYMDA MSETNEEKTT PAAAPAAAKR K---IGGLGR GL-------N ALIEGSYEKK ---------- ---------- ---------- ---------- ---------- ---------- 40
RICBE MK-------- ---------- ----NKGLGR GL-------S SLLGEEVIST E--------- ---------- ---------- ---------- ---------- ---------- 22
HELPY MA-------- ---------- K---NKVLGR GL-------A DIFPEINEVY ---------- ---------- ---------- ---------- ---------- ---------- 22
BORBU MD-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 2
(ParB1) BURCE MK-------- ---------- ----PSQFAK GFQARPDITT SEKRTALDRL ---------- ---------- ----NAIDGI VK-------- -------SET PTPAPTKSAK 49
PARSA MK-------- ---------- ----PSQFAK GFQARPDTTS SEKRTALDRL ---------- ---------- ----NAIDGL VQ-------- -------RSDA KSADVPGRSS 50
PARTU MK-------- ---------- ----PSQFAK GFQARPDTTS SEKRTALDRL ---------- ---------- ----NAIDGL VS-------- -------NEA KTPTIAGRAI 49
PARAR MSDQN----- ----QSAPQR RSSLVSSLRK GM-------D MERGAVDARL ---------- ---------- ----NA---- ---------- ---------- --QEVGSAPM 44
CABTE MA-------- ---------- KLNLTSRVRA GM-------A LERQSAEARL ---------- ---------- ----QK---- ---------- ---------- --NAVAETAL 35
CABHY MA-------- ----------K QLNLTSRVRA GM-------A VERESAEARL ---------- ---------- ----AK---- ---------- ---------- --NDVAETAR 36
PARCA MA-------- ----------T KLNFTSKVRA GM-------A VERVSAGERL ---------- ---------- ----AE---- ---------- ---------- --SDVVGQAL 36
BURGL MT-------- ---------- KLDLRSRVRA GM-------D EQTRSATARL ---------- ---------- ----EA---- ---------- ---------- --ARVVETAH 35
BURLA MAKDTS---- ---KDKKPTG NLHLAAGLLR GL-------A QENAALETRL ---------- ---------- ----PE---- ---------- ---------- --PAAAPNAV 46
BURUB MAKDTS---- ---KEKKATG NLHLAAGLLR GL-------A QENAALETRL ---------- ---------- ----PE---- ---------- ---------- --PSVAPLTA 46
HOMSA MGLRAGGTLG ---------- ----RAGAGR GA-------- ---------- ---------- ---------- ---------- ---------- ---------- 18
SACCE MS-------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 2
ARATH MA--NL---- -MMRLPISLR SFSVSASSSN GS-------P PVIGGS---- ---------- ---------- ---------- ---------- ---------- 32
SINFR MA-------- ---------- ----RKNLLA GL-----VDT AEIPHADVAP ---------- ---------- ---------- ---------- ---------- ---------- 23
ESCCO MK------RA PVIPKHTLNT QPVEDTSLST PA-------A PMVDSLIARV ---------- ---------- ---------- ---------- ---------- ---------- 37
Consensus MX-------- ---------- ----RSGLGR GL-------A ALIPTAEARL ---------- ---------- ---------- ---------- ---------- ----------

Figure 3.3.

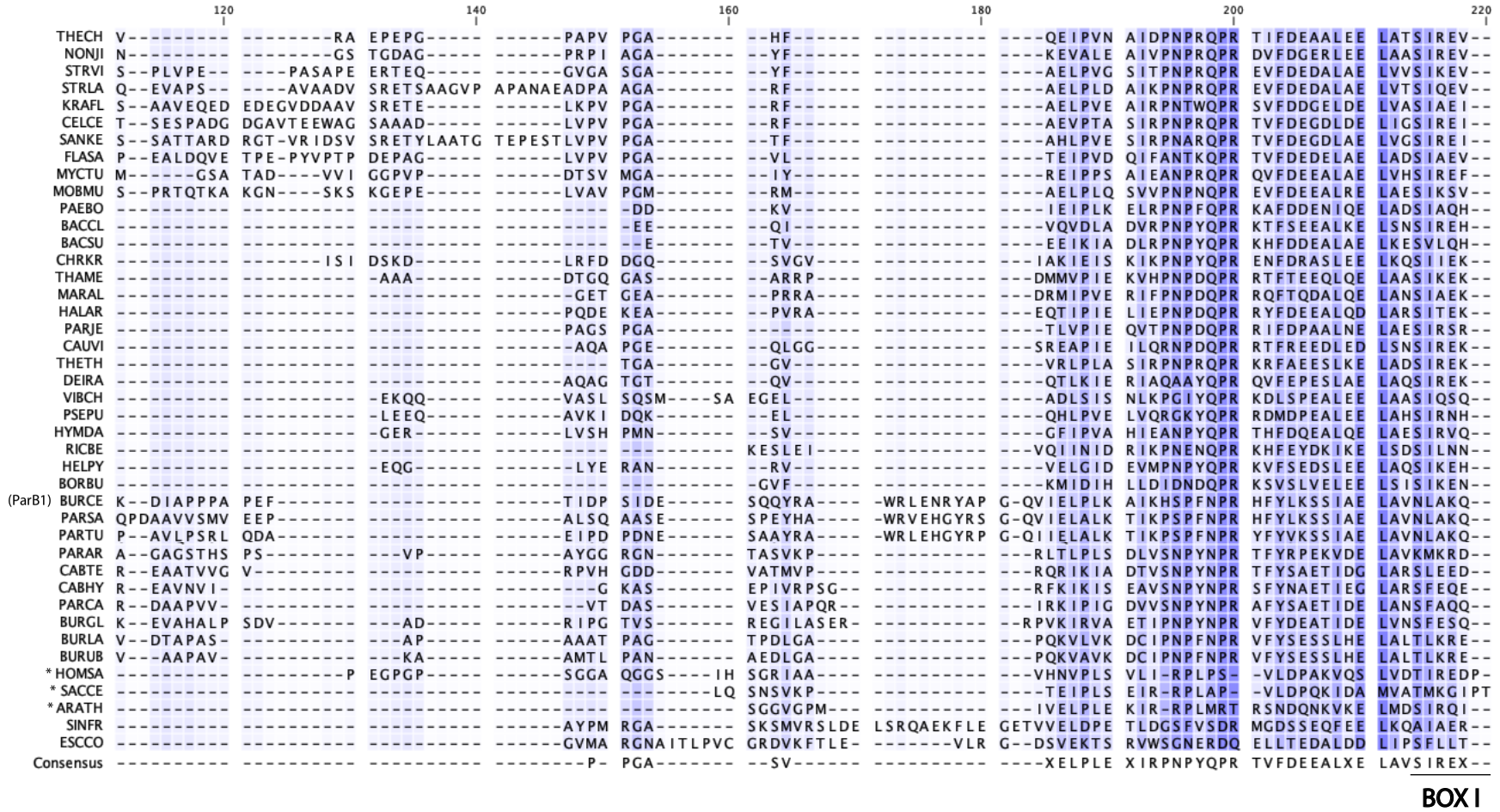

120
140
160
180
200
220
THECH V--------- --------RA EPEPG----- ------PAPV PGA------- --HF------ ---------- ----QEIPVN AIDPNPRQPR TIFDEAALEE LATSIREV-- 82
NONJI N--------- --------GS TGDAG----- ------PRPI AGA------- --YF------ ---------- ----KEVALE AIVPNPRQPR DVFDGERLEE LAASIREV-- 82
STRVI S--PLVPE-- ----PASAPE ERTEQ----- ------GVGA SGA------- --YF------ ---------- ----AELPVG SITPNPRQPR EVFDEDALAE LVVSIKEV-- 117
STRLA Q--EVAPS-- ----AVAADV SRETSAAGVP APANAEADPA AGA------- --RF------ ---------- ----AELPLD AIKPNPRQPR EVFDEDALAE LVTSIQEV-- 101
KRAFL S--AAVEQED EDEGVDDAAV SRETE----- ------LKPV PGA------- --RF------ ---------- ----AELPVE AIRPNTWQPR SVFDDGELDE LVASIAEI-- 79
CELCE T--SESPADG DGAVTEEWAG SAAAD----- ------LVPV PGA------- --RF------ ---------- ----AEVPTA SIRPNPRQPR TVFDEGDLDE LIGSIREI-- 130
SANKE S--SATTARD RGT-VRIDSV SRETYLAATG TEPESTLVPV PGA------- --TF------ ---------- ----AHLPVE SIRPNARQPR TVFDEGDLAE LVGSIREI-- 159
FLASA P--EALDQVE TPE-PYVPTP DEPAG----- ------LVPV PGA------- --VL------ ---------- ----TEIPVD QIFANTKQPR TVFDEDELAE LADSIAEV-- 119
MYCTU M------GSA TAD----VVI GGPVP----- ------DTSV MGA------- --IY------ ---------- ----REIPPS AIEANPRQPR QVFDEEALAE LVHSIREF-- 94
MOBMU S--PRTQTKA KGN----SKS KGEPE----- ------LVAV PGM------- --RM------ ---------- ----AELPLQ SVVPNPNQPR EVFDEEALRE LAESIKSV-- 112
PAEBO ---------- ---------- ---------- ---------- -DD------- --KV------ ---------- ----IEIPLK ELRPNPFQPR KAFDDENIQE LADSIAQH-- 58
BACCL ---------- ---------- ---------- ---------- -EE------- --QI------ ---------- ----VQVDLA DVRPNPYQPR KTFSEEALKE LSNSIREH-- 63
BACSU ---------- ---------- ---------- ---------- --E------- --TV------ ---------- ----EEIKIA DLRPNPYQPR KHFDDEALAE LKESVLQH-- 57
CHRKR ---------- --------ISI DSKD------ ------LRFD DGQ------- --SVGV---- ---------- ---IAKIEIS KIKPNPYQPR ENFDRASLEE LKQSIIEK-- 74
THAME ---------- ---------- -AAA------ ------DTGQ GAS------- --ARRP---- ---------- ---DMMVPIE KVHPNPDQPR RTFTEEQLQE LAASIKEK-- 74
MARAL ---------- ---------- ---------- -------GET GEA------- --PRRA---- ---------- ---DRMIPVE RIFPNPDQPR RQFTQDALQE LANSIAEK-- 69
HALAR ---------- ---------- ---------- ------PQDE KEA------- --PVRA---- ---------- ---EQTIPIE LIEPNPDQPR RYFDEEALQD LARSITEK-- 72
PARJE ---------- ---------- ---------- ------PAGS PGA------- ---------- ---------- ----TLVPIE QVTPNPDQPR RIFDPAALNE LAESIRSR-- 66
CAUVI ---------- ---------- ---------- -------AQA PGE------- --QLGG---- ---------- ---SREAPIE ILQRNPDQPR RTFREEDLED LSNSIREK-- 78
THETH ---------- ---------- ---------- ---------- TGA------- --GV------ ---------- ----VRLPLA SIRPNPRQPR KRFAEESLKE LADSIREK-- 57
DEIRA ---------- ---------- ---------- ------AQAG TGT------- --QV------ ---------- ----QTLKIE RIAQAAYQPR QVFEPESLAE LAQSIREK-- 65
VIBCH ---------- ---------- -EKQQ----- ------VASL SQSM----SA EGEL------ ---------- ----ADLSIS NLKPGIYQPR KDLSPEALEE LAASIQSQ-- 73
PSEPU ---------- ---------- -LEEQ----- ------AVKI DQK------- --EL------ ---------- ----QHLPVE LVQRGKYQPR RDMDPEALEE LAHSIRNH-- 70
HYMDA ---------- ---------- -GER------ ------LVSH PMN------- --SV------ ---------- ----GFIPVA HIEANPYQPR THFDQEALQE LAESIRVQ-- 86
RICBE ---------- ---------- ---------- ---------- ---------- KESLEI---- ---------- ---VQIINID RIKPNENQPR KHFEYDKIKE LSDSILNN-- 63
HELPY ---------- ---------- -EQG------ -------LYE RAN------- --RV------ ---------- ----VELGID EVMPNPYQPR KVFSEDSLEE LAQSIKEH-- 67
BORBU ---------- ---------- ---------- ---------- ---------- -GVF------ ---------- ----KMIDIH LLDIDNDQPR KSVSLVELEE LSISIKEN-- 39
(ParB1) BURCE K--DIAPPPA PEF------- ---------- ------TIDP SIDE------- SQQYRA---- -WRLENRYAP G-QVIELPLK AIKHSPFNPR HFYLKSSIAE LAVNLAKQ-- 120
PARSA QPDAAVVSMV EEP------- ---------- ------ALSQ AASE------ SPEYHA---- -WRVEHGYRS G-QVIELALK TIKPSPFNPR HFYLKSSIAE LAVNLAKQ-- 123
PARTU P--AVLPSRL QDA------- ---------- ------EIPD PDNE------ SAAYRA---- -WRLEHGYRP G-QIIELALK TIKPSPFNPR YFYVKSSIAE LAVNLAKQ-- 120
PARAR A--GAGSTHS PS--------- ---VP----- ------AYGG RGN------- TASVKP---- ---------- ---RLTLPLS DLVSNPYNPR TFYRPEKVDE LAVKMKRD-- 104
CABTE R--EAATVVG V--------- ---------- ------RPVH GDD------- VATMVP---- ---------- ---RQRIKIA DTVSNPYNPR TFYSAETIDG LARSLEED-- 92
CABHY R--EAVNVI- ---------- ---------- ----------G KAS------- EPIVRPSG-- ---------- ---RFKIKIS EAVSNPYNPR SFYNAETIEG LARSFEQE-- 90
PARCA R--DAAPVV- ---------- ---------- ---------VT DAS------- VESIAPQR-- ---------- ---IRKIPIG DVVSNPYNPR AFYSAETIDE LANSFAQQ-- 91
BURGL K--EVAHALP SDV------- ---AD----- ------RIPG TVS------- REGILASER- ---------- --RPVKIRVA ETIPNPYNPR VFYDEATIDE LVNSFESQ-- 100
BURLA V--DTAPAS- ---------- ---AP----- ------AAAT PAG------- TPDLGA---- ---------- ---PQKVLVK DCIPNPFNPR VFYSESSLHE LALTLKRE-- 103
BURUB V---AAPAV- ---------- ---KA----- ------AMTL PAN------- AEDLGA---- ---------- ---PQKVAVK DCIPNPFNPR VFYSESSLHE LALTLKRE-- 102
*HOMSA ---------- ---------P EGPGP----- ------SGGA QGGS----IH SGRIAA---- ---------- ---VHNVPLS VLI-RPLPS- -VLDPAKVQS LVDTIREDP- 73
*SACCE ---------- ---------- ---------- ---------- ---------- --------LQ SNSVKP---- ---------- ---TEIPLS EIR-RPLAP- -VLDPQKIDA MVATMKGIPT 43
*ARATH ---------- ---------- ---------- ---------- ---------- ---------- SGGVGPM--- ---------- ---IVELPLE KIR-RPLMRT RSNDQNKVKE LMDSIRQI-- 73
SINFR ---------- ---------- ---------- ------AYPM RGA------- SKSMVRSLDE LSRQAEKFLE GETVVELDPE TLDGSFVSDR MGDSSEQFEE LKQAIAER-- 88
ESCCO ---------- ---------- ---------- ------GVMA RGNAITLPVC GRDVKFTLE- -------VLR G--DSVEKTS RVWSGNERDQ ELLTEDALDD LIPSFLLT-- 99
Consensus ---------- ---------- ---------- --------P- PGA------- --SV------ ---------- ----XELPLE XIRPNPYQPR TVFDEEALXE LAVSIREX--
BOX I

Figure 3.3.

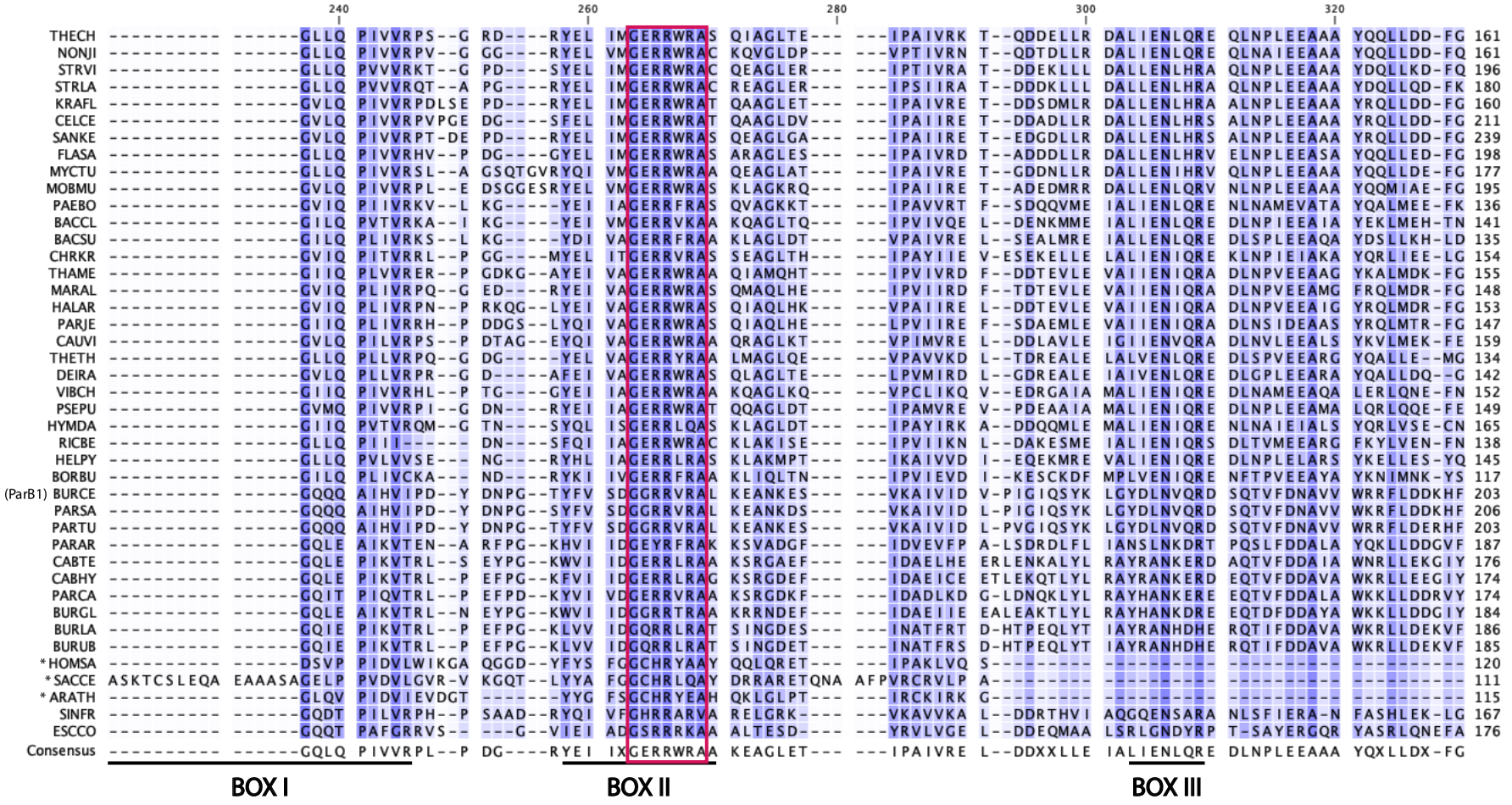

240
260
280
300
320
THECH
NONJI
STRVI
STRLA
KRAFL
CELCE
SANKE
FLASA
MYCTU
MOBMU
PAEBO
BACCL
BACSU
CHRKR
THAME
MARAL
HALAR
PARJE
CAUVI
THETH
DEIRA
VIBCH
PSEPU
HYMDA
RICBE
HELPY
BORBU
(ParB1) BURCE
PARSA
PARTU
PARAR
CABTE
CABHY
PARCA
BURGL
BURLA
BURUB
*HOMSA
*SACCE
*ARATH
SINFR
ESCCO
Consensus
BOX I
BOX II
BOX III

Figure 3.3.

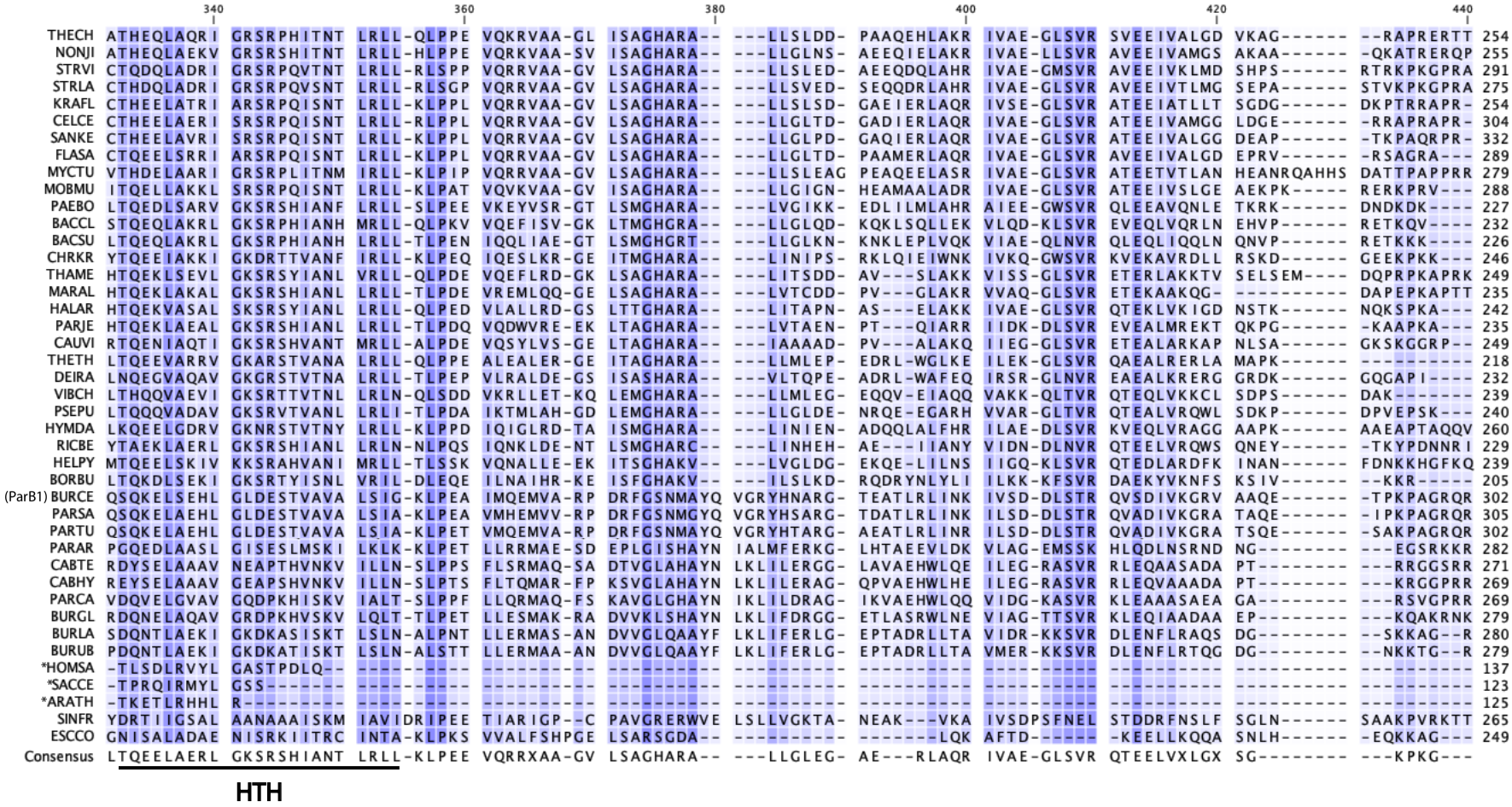

340
360
380
400
420
440
THECH ATHEQLAQRI GRSRPHITNT LRLL-QLPPE VQKRVAA-GL ISAGHARA-- ---LLSLDD- PAAQEHLAKR IVAE-GLSVR SVEEIVALGD VKAG------ --RAPRERTT 254
NONJI ATHEQLAEKV GRSRSHITNT LRLL-HLPPE VQRKVAA-SV ISAGHARA-- ---LLGLNS- AEEQIELAKR IVAE-LLSVR AVEEIVAMGS AKAA------ -QKATRERQP 255
STRVI CTQDQLADRI GRSRPQVTNT LRLL-RLSPP VQRRVAA-GV LSAGHARA-- ---LLSLED- AEEQDQLAHR IVAE-GMSVR AVEEIVKLMD SHPS------ RTRKPKGPRA 291
STRLA CTHDQLADRI GRSRPQVSNT LRLL-RLSGP VQRRVAA-GV LSAGHARA-- ---LLSVED- SEQQDRLAHR IVAE-GLSVR AVEEIVTLMG SEPA------ STVKPKGPRA 275
KRAFL CTHEELATRI ARSRPQISNT LRLL-KLPPL VQRRVAA-GV LSAGHARA-- ---LLSLSD- GAEIERLAQR IVSE-GLSVR ATEEIATLLT SGDG------ DKPTRRAPR- 254
CELCE CTHEELAERI SRSRPQISNT LRLL-RLPPL VQRRVAA-GV LSAGHARA-- ---LLGLTD- GADIERLAQR IVAE-GLSVR ATEEIVAMGG LDGE------ -RRAPRAPR- 304
SANKE CTHEELAVRI SRSRPQISNT LRLL-KLPPL VQRRVAA-GV LSAGHARA-- ---LLGLPD- GAQIERLAQR IVAE-GLSVR ATEEIVALGG DEAP------ -TKPAQRPR- 332
FLASA CTQEELSRRI ARSRPQISNT LRLL-KLPPL VQRRVAA-GV LSAGHARA-- ---LLGLTD- PAAMERLAQR IVAE-GLSVR AVEEIVALGD EPRV------ -RSAGRA--- 289
MYCTU VTHDELAARI GRSRPLITNM IRLL-KLPIP VQRRVAA-GV LSAGHARA-- ---LLSLEAG PEAQEELASR IVAE-GLSVR ATEETVTLAN HEANRQAHHS DATTPAPPRR 279
MOBMU ITQELLAKKL SRSRPQISNT LRLL-KLPAT VQVKVAA-GV ISAGHARA-- ---LLGIGN- HEAMAALADR IVAE-GLSVR ATEEIVSLGE AEKPK----- RERKPRV--- 288
PAEBO LTQEDLSARV GKSRSHIANF LRLL-SLPEE VKEYVSR-GT LSMGHARA-- ---LVGIKK- EDLILMLAHR AIEE-GWSVR QLEEAVQNLE TKRK------ DNDKDK---- 227
BACCL STQEQLAKRL GKSRPHIANH MRLL-QLPKV VQEFISV-GK LTMGHGRA-- ---LLGLQD- KQKLSQLLEK VLQD-KLSVR EVEQLVQRLN EHVP------ RETKQV---- 232
BACSU LTQEQLAKRL GKSRPHIANH LRLL-TLPEN IQQLIAE-GT LSMGHGRT-- ---LLGLKN- KNKLEPLVQK VIAE-QLNVR QLEQLIQQLN QNVP------ RETKKK---- 226
CHRKR YTQEEIAKKI GKDRTTVANF IRLL-KLPEQ IQESLKR-GE ITMGHARA-- ---LINIPS- RKLQIEIWNK IVKQ-GWSVR KVEKAVRDLL RSKD------ GEEKPKK--- 246
THAME HTQEKLSEVL GKSRSYIANL VRLL-QLPDE VQEFLRD-GK LSAGHARA-- ---LITSDD- AV---SLAKK VISS-GLSVR ETERLAKKTV SELSEM---- DQPRPKAPRK 249
MARAL HTQEKLAKAL GKSRSHIANL LRLL-TLPDE VREMLQQ-GE LSAGHARA-- ---LVTCDD- PV---GLAKR VVAQ-GLSVR ETEKAAKQG- ---------- DAPEPKAPTT 235
HALAR HTQEKVASAL SKSRSYIANL LRLL-QLPED VLALLRD-GS LTTGHARA-- ---LITAPN- AS---ELAKK IVAE-GLSVR QTEKLVKIGD NSTK------ NQKSPKA--- 242
PARJE HTQEKLAEAL GKSRSHIANL LRLL-TLPDQ VQDWVRE-EK LTAGHARA-- ---LVTAEN- PT---QIARR IIDK-DLSVR EVEALMREKT QKPG------ -KAAPKA--- 235
CAUVI RTQENIAQTI GKSRSHVANT MRLL-ALPDE VQSYLVS-GE LTAGHARA-- ---IAAAAD- PV---ALAKQ IIEG-GLSVR ETEALARKAP NLSA------ GKSKGGRP-- 249
THETH LTQEEVARRV GKARSTVANA LRLL-QLPPE ALEALER-GE ITAGHARA-- ---LLMLEP- EDRL-WGLKE ILEK-GLSVR QAEALRERLA MAPK------ ---------- 218
DEIRA LNQEGVAQAV GKGRSTVTNA LRLL-TLPEP VLRALDE-GS ISASHARA-- ---VLTQPE- ADRL-WAFEQ IRSR-GLNVR EAEALKRERG GRDK------ GQGAPI---- 232
VIBCH LTHQQVAEVI GKSRTTVTNL LRLN-QLSDD VKRLLET-KQ LEMGHARA-- ---LLMLEG- EQQV-EIAQQ VAKK-QLTVR QTEQLVKKCL SDPS------ DAK--------- 239
PSEPU LTQQQVADAV GKSRVTVANL LRLI-TLPDA IKTMLAH-GD LEMGHARA-- ---LLGLDE- NRQE-EGARH VVAR-GLTVR QTEALVRQWL SDKP------ DPVEPSK--- 240
HYMDA LKQEELGDRV GKNRSTVTNY LRLL-KLPPD IQIGLRD-TA ISMGHARA-- ---LINIEN- ADQQLALFHR ILAE-DLSVR KVEQLVRAGG AAPK------ AAEAPTAQQV 260
RICBE YTAEKLAERL GKSRSHIANL LRLN-NLPQS IQNKLDE-NT LSMGHARC-- ---LINHEH- AE---IIANY VIDN-DLNVR QTEELVRQWS QNEY------ -TKYPDNNRI 229
HELPY MTQEELSKIV KKSRAHVANI MRLL-TLSSK VQNALLE-EK ITSGHAKV-- ---LVGLDG- EKQE-LILNS IIGQ-KLSVR QTEDLARDFK INAN------ FDNKKHGFKQ 239
BORBU LTQKDLSEKI GKSRTYISNL VRIL-DLEQE ILNAIHR-KE ISFGHAKV-- ---ILSLKD- RQDRYNLYLI ILKK-KFSVR DAEKYVKNFS KSIV------ --KKR----- 205
(ParB1) BURCE QSQKELSEHL GLDESTVAVA LSIG-KLPEA IMQEMVA-RP DRFGSNMAYQ VGRYHNARG- TEATLRLINK IVSD-DLSTR QVSDIVKGRV AAQE------ -TPKPAGRQR 302
PARSA QSQKELAEHL GLDESTVAVA LSIA-KLPEA VMHEMVV-RP DRFGSNMGYQ VGRYHSARG- TDATLRLINK ILSD-DLSTR QVADIVKGRA TAQE------ -IPKPAGRQR 305
PARTU QSQKELAEHL GLDESTVAVA LSIA-KLPET VMQEMVA-RP DRFGSNMAYQ VGRYHTARG- AEATLRLINR ILSD-DLSTR QVADIVKGRA TSQE------ -SAKPAGRQR 302
PARAR PGQEDLAASL GISESLMSKI LKLK-KLPET LLRRMAE-SD EPLGISHAYN IALMFERKG- LHTAEEVLDK VLAG-EMSSK HLQDLNSRND NG-------- ---EGSRKKR 282
CABTE RDYSELAAAV NEAPTHVNKV ILLN-SLPPS FLSRMAQ-SA DTVGLAHAYN LKLILERGG- LAVAEHWLQE ILEG-RASVR RLEQAASADA PT-------- ---RRGGSRR 271
CABHY REYSELAAAV GEAPSHVNKV ILLN-SLPTS FLTQMAR-FP KSVGLAHAYN LKLILERAG- QPVAEHWLHE ILEG-RASVR RLEQVAAADA PT-------- ---KRGGPRR 269
PARCA VDQVELGVAV GQDPKHISKV IALT-SLPPF LLQRMAQ-FS KAVGLGHAYN IKLILDRAG- IKVAEHWLQQ VIDG-KASVR KLEAAASAEA GA-------- ---RSVGPRR 269
BURGL RDQNELAQAV GRDPKHVSKV LQLT-TLPET LLESMAK-RA DVVKLSHAYN LKLIFDRGG- ETLASRWLNE VIAG-TTSVR KLEQIAADAA EP-------- ---KQAKRNK 279
BURLA SDQNTLAEKI GKDKASISKT LSLN-ALPNT LLERMAS-AN DVVGLQAAYF LKLIFERLG- EPTADRLLTA VIDR-KKSVR DLENFLRAQS DG-------- --SKKAG--R 280
BURUB PDQNTLAEKI GKDKATISKT LSLN-ALSTT LLERMAA-AN DVVGLQAAYF LKLIFERLG- EPTADRLLTA VMER-KKSVR DLENFLRTQG DG-------- --NKKTG--R 279
*HOMSA -TLSDLRVYL GASTPDLQ-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 137
*SACCE -TPRQIRMYL GSS------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 123
*ARATH -TKETLRHHL R--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 125
SINFR YDRTIIGSAL AANAAAISKM IAVIDRIPEE TIARIGP--C PAVGRERWVE LSLLVGKTA- NEAK---VKA IVSDPSFNEL STDDRFNSLF SGLN------ SAAKPVRKTT 265
ESCCO GNISALADAE NISRKIITRC INTA-KLPKS VVALFSHPGE LSARSGDA-- ---------- -------LQK AFTD------ -KEELLKQQA SNLH------ -EQKKAG--- 249
Consensus LTQEELAERL GKSRSHIANT LRLL-KLPEE VQRRXAA-GV LSAGHARA-- ---LLGLEG- AE---RLAQR IVAE-GLSVR QTEELVXLGX SG-------- ---KPKG---
HTH

Figure 3.3.

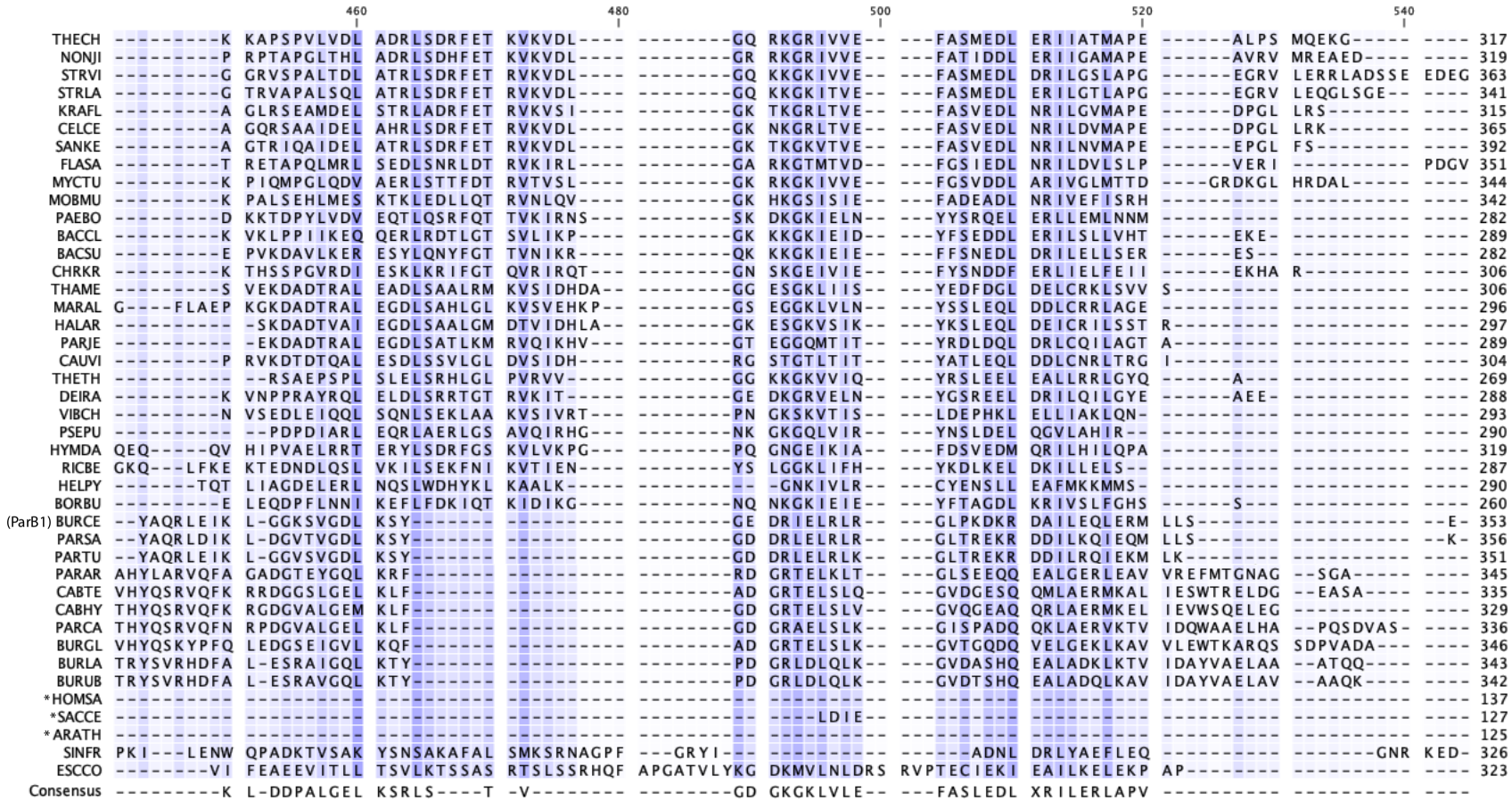
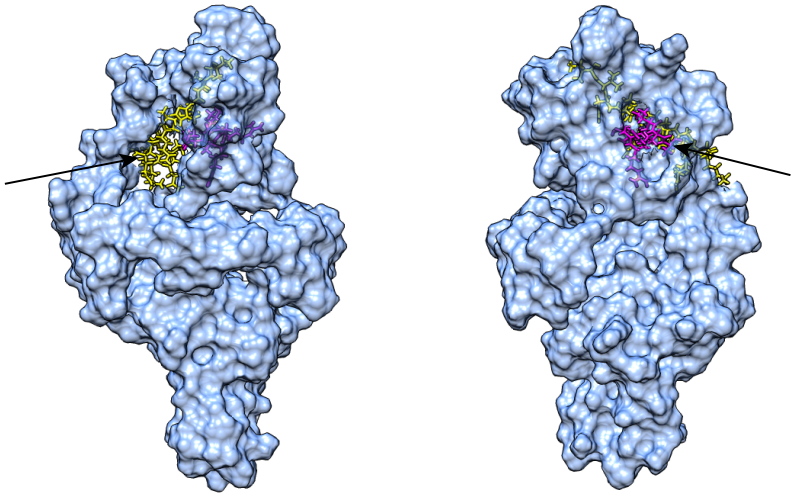

460
480
500
520
540
THECH
NONJI
STRVI
STRLA
KRAFL
CELCE
SANKE
FLASA
MYCTU
MOBMU
PAEBO
BACCL
BACSU
CHRKR
THAME
MARAL
HALAR
PARJE
CAUVI
THETH
DEIRA
VIBCH
PSEPU
HYMDA
RICBE
HELPY
BORBU
(ParB1) BURCE
PARSA
PARTU
PARAR
CABTE
CABHY
PARCA
BURGL
BURLA
BURUB
*HOMSA
*SACCE
*ARATH
SINFR
ESCCO
Consensus

Figure 3.3. Multiple sequence alignment of homolog proteins with conserved aa residues. ParB1 is labelled next to the sequence and Srx proteins are marked with an asterisk. Dark blue background indicates highly conserved aa residues. Lighter color indicates less conservation. Previously recognized regions of high conservation are underlined in black and labelled. The signature motif GxxRxxA is highlighted in a red frame. The alignment was created with MAFFT and modified in CLC Main Workbench. Abbreviations of species names are as follows: ARATH*, Arabidopsis thaliana*; BACCL, *Bacillus clausii*; BACSU, *Bacillus subtilis*; BORBU*, Borrelia burgdorferi*; BURCE, *Burkholderia cenocepacia*; BURGL, *Burkholderia glumae*; BURLA, *Burkholderia lata*; BURUB, *Burkholderia ubonensis*; CABHY, *Caballeronia hypogeia;* CABTE, *Caballeronia temeraria*; CAUVI, *Caulobacter vibrioides*; CELCE, *Cellulosimicrobium cellulans*; CHRKR, *Chrysopegis kryptomonas*; DEIRA, *Deinococcus radiodurans*; ESCCO, *Escherichia coli*; FLASA, *Flaviflexus salsibiostraticola*; HALAR, *Halocynthiibacter arcticus*; HELPY, *Helicobacter pylori*; HOMSA, *Homo sapiens*; HYMDA, *Hymenobacter daecheongensis*; KRAFL, *Krasilnikoviella flava*; MARAL, *Marinovum algicola*; MOBMU, *Mobiluncus mulieris*; MYCTU, *Mycobacterium tuberculosis*; NONJI, *Nonomuraea jiangxiensis*; PAEBO, *Paenibacillus bovis*; PARJE, *Paracoccus jeotgali*; PARAR, *Paraburkholderia aromaticivorans*; PARCA, *Paraburkholderia caballeronis*; PARSA, *Paraburkholderia sartisoli*; PARTU, *Paraburkholderia tuberum*; PSEPU, *Pseudomonas putida*; RICBE, *Rickettsia bellii*; SACCE, *Saccharomyces cerevisiae*; SANKE, *Sanguibacter keddieii*; SINFR, *Sinorhizobium fredii*; STRLA, *Streptomyces lavendulae*; STRVI, *Streptomyces violaceoruber*; THAME, *Thalassobius mediterraneus*; THECH,*Thermostaphylospora chromogena*; THETH, *Thermus thermophilus*; VIBCH, *Vibrio cholerae*.

The GxxRxxA site was inspected in the structure of IT model 4 made for ParB1 in **Figure 3.4 A**. The model shows that the site is located in a large cavity in the protein core, which may indicate an active site. **Figure 3.4 B** shows a pairwise alignment of the ParB1 and ParB protein for *B. subtilis* made with the pairwise local alignment tool EMBOSS Water (https://www.ebi.ac.uk/Tools/psa/emboss_water/) with default parameters (Smith-Waterman, BLOSUM62, gap open = 10, gap extend = 0.5) and modified in CLC Main Workbench. As the proteins belong to different prokaryotic species and bind different *parS* sequences, the sequences show little overall similarity seen by conserved aa shown in blue. The GxxRxxA site is highly conserved, but only a few aa in each of the BOX-regions are conserved. This could point to the aa that are crucial for function and would be relevant to study further for location of nuclease activity.

A



B

```
ParB1_B.cenocepacia VIELPLKAIK HSPFNPRHFY LKSSIAELAV NLAKQGQQQA IHV---IPDY DNPGTYFVSD GGRRVRALKE ANKES
ParB_B.subtilis     VEEIKIADLR PNPYQPRKHF DDEALAELKE SVLQHGILQP LIVRKSLKGY D------IVA GERRFRAAKL AGLDT
Consensus           VXEXXXXXXX XXPXXPRXXX XXXXXAELXX XXXXXGXXQX XXVRKSXXXY DNPGTYFXXX GXRRXRAXKX AXXXX
                                                      BOX I                    BOX II
ParB1_B.cenocepacia VKAIVIDVPI GIQSYKLGYD LNVQRDSQTV FDNAVVWRRF LDDKHFQ-SQ KELSEHLGLD ESTVAVALSI GKLPE
ParB_B.subtilis     VPAIVRELSE ALMR-EIALL ENLQREDLSP LEEAQAYDSL L--KHLDLTQ EQLAKRLGKS RPHIANHLRL LTLPE
Consensus           VXAIVXXXXX XXXXYXXXXX XNXQRXXXXX XXXAXXXXXX LDDKHXXLXQ XXLXXXLGXX XXXXAXXLXX XXLPE
                              BOX III                                      HTH
ParB1_B.cenocepacia AIMQEMVARP DRFGSNMAYQ VGR----YHN ARGTEATLRL INKIVSDDLS TRQVSDIVK- -GRVAAQETP KP---
ParB_B.subtilis     NI-QQLIAE- ----GTLSMG HGRTLLGLKN KNKLEP---L VQKVIAEQLN VRQLEQLIQQ LNQNVPRETK KKEPV
Consensus           XIMQXXXAXP DRFGXXXXXX XGRTLLGXXN XXXXEXTLRL XXKXXXXXLX XRQXXXXXXQ LXXXXXXETX KXEPV

ParB1_B.cenocepacia -----AGRQR YAQRLEIKLG GKSVGDLKSY GEDRIELRLR GLPKDKRDAI LEQLERMLLS E
ParB_B.subtilis     KDAVLKERES YLQ----NYF GTTVNIKRQK KKGKIEIEF- -FSNEDLDRI LE-----LLS E
Consensus           KDAVLXXRXX YXQRLEIXXX GXXVXXXXXX XXXXIEXXXR GXXXXXXDXI LEQLERMLLS E
```

Figure 3.4. Graphical representation of signature ParB GxxRxxA site. Predicted SM model 4 of ParB1 presented with a 50% transparency surface structure with the GxxRxxA site highlighted by magenta stick molecules and BOX I with yellow stick molecules indicated with arrows **(A)**. Models presentation created with UCSF Chimera. A pairwise alignment of ParB1 and ParB of *Bacillus subtilis* (PDB: 6SDK) displaying conserved aa in blue, the GxxRxxA site in a red frame, and conserved regions underlined in black **(B)**. A pruned local pairwise alignment was created with EMBOSS Water (https://www.ebi.ac.uk/Tools/psa/emboss_water/) and modified in CLC Main Workbench.

3.3.2 Evolutionary Relationships Within the ParB Family

Large regions of conserved residues between homologs indicate a close evolutionary relationship. As there exists about 60 000 ParB proteins, it is of interest to study the evolutionary relationships between the proteins selected for the multiple sequence alignment. To study their evolutionary relationships, a phylogenetic tree was generated by MAFFT based on the multiple sequence alignment (shown in **Figure 3.3**) with default parameters and a bootstrapping resampling number of 1000 **(Figure 3.5)**. The tree outlines the expected

evolutionary relationship between the proteins based on the amount of genetic change that has accumulated and caused divergence. Branch lengths indicate the number of substitutions that has occurred since divergence, which can be estimated with the scale bar of 0.1 (1 substitution per 10 aa). The longer the branches are the more genetic changes have accumulated throughout evolution. Many of the homologs have short branches, and have thus accumulated little genetic change, while two monophyletic groups are distinguished by longer branch lengths. One of these groups comprises the Srx homologs **(Figure 3.5, A)**, which would be expected to have long branches based on the new function the protein has gained since divergence (Basu and Koonin, 2005; Soh et al., 2019).

The second group contains the ParB1 protein sequence **(Figure 3.5, B)**. The monophyletic group indicated by a broken black frame was observed in this construction in 47-100% of the reconstructed trees. ParB1 is placed on a branch with the specific construction reoccurring in 99% of the resamplings. It is closely connected to the PARSA (*Paraburkholderia sartisoli*) and PARTU (*Paraburkholderia tuberum*) sequences. The BACSU (*Bacillus subtilis*) protein, from which ParB1 was modeled, branches off from ParB1 only a few nodes earlier. In this tree, they are relatively closely connected but from the branch lengths we observe that they are separated by significant evolutionary distance.

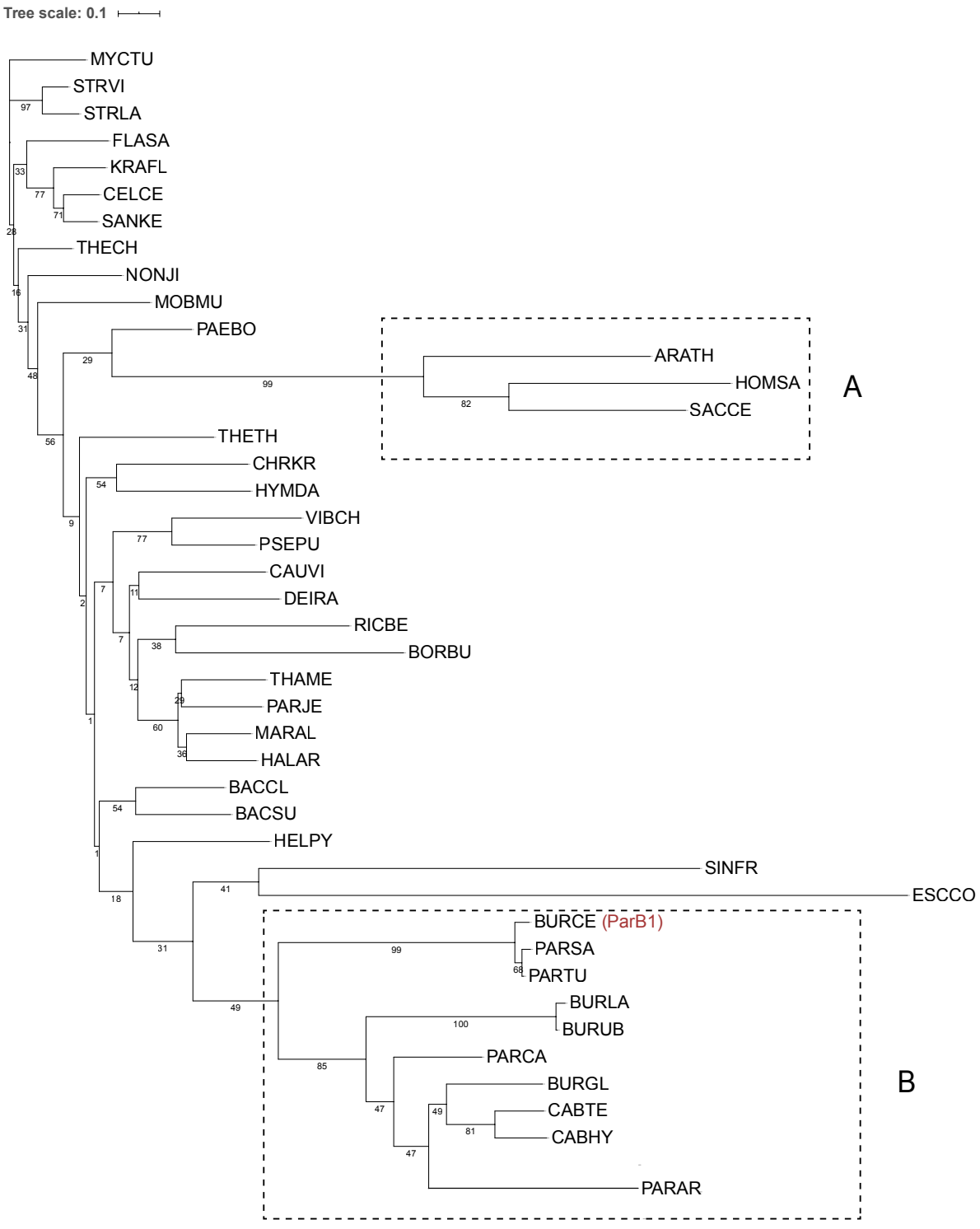

Tree scale: 0.1
MYCTU
STRVI
STRLA
FLASA
KRAFL
CELCE
SANKE
THECH
NONJI
MOBMU
PAEBO
ARATH
HOMSA
SACCE
A
THETH
CHRKR
HYMDA
VIBCH
PSEPU
CAUVI
DEIRA
RICBE
BORBU
THAME
PARJE
MARAL
HALAR
BACCL
BACSU
HELPY
SINFR
ESCCO
BURCE (ParB1)
PARSA
PARTU
BURLA
BURUB
PARCA
BURGL
CABTE
CABHY
PARAR
B

Figure 3.5. A phylogenetic tree of ParB homologs. Two monophyletic groups are highlighted by black frames: the clade containing the Srx homologs **(A)** and the clade containing ParB1, shown in red **(B)**. Bootstrapping values (1000 replicates) are displayed with branch color, referenced in the color scheme. The phylogenetic tree is made with MAFFT and visualized in ITOL. Abbreviations of species names are as follows: ARATH*, Arabidopsis thaliana*; BACCL, *Bacillus clausii*; BACSU, *Bacillus subtilis*; BORBU*, Borrelia burgdorferi*; BURCE, *Burkholderia cenocepacia*; BURGL, *Burkholderia glumae*; BURLA, *Burkholderia lata*; BURUB, *Burkholderia ubonensis*; CABHY, *Caballeronia hypogeia;* CABTE, *Caballeronia temeraria*; CAUVI, *Caulobacter vibrioides*; CELCE, *Cellulosimicrobium cellulans*; CHRKR, *Chrysopegis kryptomonas*; DEIRA, *Deinococcus radiodurans*; ESCCO, *Escherichia coli*; FLASA, *Flaviflexus salsibiostraticola*; HALAR, *Halocynthiibacter arcticus*; HELPY, *Helicobacter pylori*; HOMSA, *Homo sapiens*; HYMDA, *Hymenobacter daecheongensis*; KRAFL, *Krasilnikoviella flava*; MARAL, *Marinovum algicola*; MOBMU, *Mobiluncus mulieris*; MYCTU, *Mycobacterium tuberculosis*; NONJI, *Nonomuraea jiangxiensis*; PAEBO, *Paenibacillus bovis*; PARJE, *Paracoccus jeotgali*; PARAR, *Paraburkholderia aromaticivorans*; PARCA, *Paraburkholderia caballeronis*; PARSA, *Paraburkholderia sartisoli*; PARTU, *Paraburkholderia tuberum*; PSEPU, *Pseudomonas putida*; RICBE, *Rickettsia bellii*; SACCE, *Saccharomyces cerevisiae*; SANKE, *Sanguibacter keddieii*; SINFR, *Sinorhizobium fredii*; STRLA, *Streptomyces lavendulae*; STRVI, *Streptomyces violaceoruber*; THAME, *Thalassobius mediterraneus*; THECH,*Thermostaphylospora chromogena*; THETH, *Thermus thermophilus*; VIBCH, *Vibrio cholerae*. Full protein sequences can be found in appendix 9. Position of ParB3 protein in the phylogenetic tree is not revealed.

3.4 *In Vitro* Analysis of Nuclease Activity of ParB1 and ParB3

3.4.1 Generation of the plasmid toolbox for ParB1 and ParB3 expression

In order to express, purify, and analyze activity of the ParB proteins, they first have to be cloned into suitable expression vectors. ParB1 and ParB3 were cloned into four Gateway vectors: pDEST14, pDEST17, pDEST566, and pGEX2T1GW. A core domain of ParB3, ParB3Core, was included in this project to observe the functionality of the core domain alone. The core domain includes 207 aa of the original sequence (**Figure 3.1 B)**. ParB3Core was cloned into the same Gateway vectors.

The four Gateway expression vectors were chosen for the purpose of having alternative approaches for recombinant expression and purification of the inserted proteins. pDEST14, pDEST17, and pDEST566 each have an IPTG-inducible T7 promoter which controls transcription of the inserted gene. pDEST14 does not contain a tag for purification of the insert protein. pDEST17 contains a histidine 6 (His)-tag which will attach to the N-terminal end of the protein, generating a fusion protein. A His6-tag comprises six consecutive histidine aa with a high affinity for metal ions. This enables purification by metal affinity chromatography

methods. pDEST566 has a combined His and MBP-tag. The maltose binding protein (MBP)-tag is often included in protein purification experiments because it enhances the solubility of overexpressed proteins in the cytoplasm that tend to become insoluble and form aggregates. pGEX2T1GW has a different expression system with an IPTG-inducible tac-promoter controlling the transcription of the insert. This vector uses a glutathione S-transferase (GST)-tag, which encodes a 211 aa long protein which promotes protein solubility and allows for purification by a pulldown assay due to its high affinity for glutathione. A thrombin site is included between the GST-tag and the sequence of interest, so that the tag can be cleaved off by a thrombin protease. This toolbox allows for testing of different approaches for recombinant expression and purification of the ParB proteins.

pDONR221 was used as an intermediate entry clone in the Gateway cloning process. The reaction conditions and primers can be found in appendix 7 **Table S4** and appendix 8 **Table S6-S11**, and plasmid map for ParB1 in appendix 4, **Figure S11**. The constructed clones were verified with restriction enzyme digestion with one cutting site in both the vector and insert to determine the presence of each (restriction enzymes in appendix 3, **Table S2**).

Digested ParB1 Gateway clones are presented in **Figure 3.6 A.** pDEST14_ParB1 and pDEST17_ParB1 were cut with EcoRI and correct clones give bands of 1427 and 4263 bp, and 1427 and 4336 bp. pDEST566_ParB1 was cut with SalI and XbaI to give bands of 1864 and 5980 bp. pGEX2T1GW_ParB1 was cut with BtgI and correct clones give bands of 1387 and 4726 bp. The expected bands are indicated by black arrows on the gel. In lane 4 the extra band of ~6 kb is likely linearized plasmid. The expected bands seem to be present and the constructs were validated by sequencing (appendix 3, **Figure S6**). Plasmid maps are available in appendix 4, **Figure S12**. Digested ParB3 Gateway clones are presented in **Figure 3.6 B**. pDEST14_ParB3 and pDEST17_ParB3 were cut with PvuII and give correct bands of 2572 and 3094 bp, and 2645 and 3094 bp. pDEST566_ParB3 was digested with PstI and XbaI to give bands of 2002 and 5818 bp. pGEX2T1GW_ParB3 was cut with PstI to give bands of 1268 and 4821 bp. The clones were validated by sequencing (appendix 3, **Figure S7**). Digested ParB3Core Gateway clones are presented in **Figure 3.6 C**. pDEST14_ParB3Core and pDEST17_ParB3Core were cut with PvuII and give bands of 2380 and 2845 bp, and 2452 and 2876 bp. pDEST566_ParB3Core was digested with PstI and XbaI and give bands of 1810 and 5599 bp. pGEX2T1GW_ParB3Core was cut with PstI and give bands of 1049 and 4629 bp. The clones were validated by sequencing (appendix 3, **Figure S8**).

Further DNA constructs were made with traditional restriction enzyme cloning. A humanized version of ParB1 and ParB3, ParB1Hu and ParB3Hu, was designed by Ragnhild

Eskeland by interchanging common bacterial codons with common human codons to improve protein expression in mammalian cells. Synthesized sequences were purchased from GenScript. To create a ParB protein expression vector, ParB3Hu was moved into pET15b as referred to in methods section 2.2.5.1 with details given in **Table 2.3**.

To create the construct for ParB3 expression in mammalian cells, ParB3Hu was further cut from pET15b and ligated into the fluorescently tagged vector pDasher under control of a cytomegalovirus (CMV) expression promoter (section 2.2.5.1, **Table 2.3**). The positive clones were verified with a restriction enzyme digestion **(Figure 3.6 D**) (restriction enzymes in appendix 3, **Table S2**). pET15b_ParB3Hu was digested with PstI and to generate bands of 1343 and 5425 bp. pDasher_ParB3Hu was digested with BstXI which gives bands of 722 and 5005 bp. Expected fragments were obtained and are indicated by black arrows. Positive clones were sequenced (appendix 3, **Figure S5, B** and **C**).

Two more constructs were planned: ParB1Hu in pET15b for protein purification and ParB1Hu in pOriGFP to express ParB1Hu in mammalian cells. Attempts for cloning ParB1Hu from vector pUC57_ ParB1Hu into expression vector pET15b were carried out without success. The ligation into pET15b was unsuccessful likely due to a vector contamination with another plasmid DNA. ParB1Hu was therefore later cloned into pET16b by Naima Azouzi.

Another obstacle was discovered during cloning of ParB1Hu fragment into mammalian expression vector pOriGFP. A NdeI restriction site was present in the CMV promoter, which made the pOriGFP vector not suitable for cloning. Partial restriction enzyme cutting attempts were performed without success. Finally, the NdeI restriction site present in the pOriGFP_ParB1 CMV promoter was changed into MfeI restriction site by site directed mutagenesis - with the help of Naima Azouzi. Primers for the PCR reaction and cycling conditions can be found in appendix 7 in **Table S4** and appendix 8 in **Table S5**, and plasmid maps in appendix 4 **Figure S10**. To verify positive clones of the mutagenesis reaction, plasmids were purified and cut with MfeI enzyme and separated by agarose gel electrophoresis **(Figure 3.6 F)**. Successful clones are indicated by green arrows (lanes 17, 19, 21, and 22). The selected plasmids were checked with double digestion by EcoRI where bands of 1647 and 4294 bp were expected to form and are indicated by black arrows (**Figure 3.6 E**, lane 14). The mutation site was also validated by sequencing (appendix 3, **Figure S5 D)**.

Another construct was made to improve ParB protein purification. An *Escherichia coli* partitioning protein C (ParC) topoisomerase was found to increase the solubility of an *E. coli* ParB protein upon coexpression (Johnson et al., 1999). A truncated ParC sequence (aa 1-490) was synthesized by GenScript and cloned into pET28a with a C-terminal Hemagglutinin (HA)-

tag. For co-expression with ParBs, ParC was cloned into a vector with kanamycin resistance, whereas ParB constructs have ampicillin resistance. The construct was validated by digestion with EcoRV (**Figure 3.6 E**, lane 15). The correct fragments were obtained as indicated by black arrows with expected fragment sizes of 1898 and 4841 bp. The construct was further validated by sequencing (appendix 3, **Figure S5 A)** and plasmid map is presented in appendix 4, **Figure S9**.

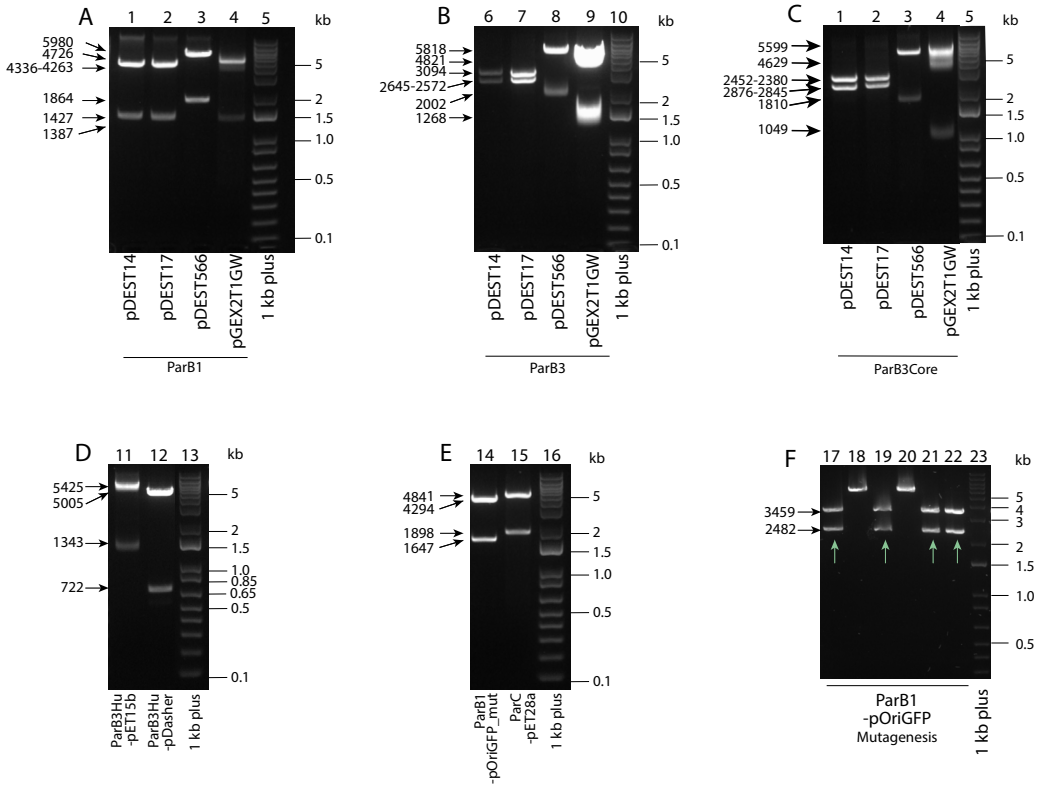


Figure 3.6. Restriction analysis of constructed clones. The ParB1 **(A)**, ParB3 **(B)**, and ParB3Core **(C)** sequences were cloned into Gateway vectors pDEST14, pDEST17, pDEST566, and pGEX2T1GW with entry vector pDONR221 as an intermediate step. **(D)** ParB3Hu was cloned into pET15b and pDasher using KpnI and NheI restriction enzymes. **(E)** ParC was cloned into vector pET28a using XhoI and NcoI restriction enzymes, and the vector backbone of pOriGFP_ParB1 was mutated for further cloning. **(F)** Restriction analysis of mutagenesis reaction of restriction site NdeI into MfeI in the pOriGFP vector. Positive clones were confirmed by digestion with MfeI and should give two bands, indicated by green arrows. 1% Agarose gel electrophoresis in 1X TAE. 1 kb plus DNA ladder (lanes 5, 10, 15, 18, and 21). Fragment sizes are indicated by black arrows.

3.4.2 Recombinant expression of ParB1, ParB3, and ParC proteins

A selection of the constructed clones was expressed in BL21 cells: pDEST17_ParB1, pDEST17_ParB3, pDEST17_ParB3Core, pET15b_ParB3Hu, and pET28a_ParC. An empty pET16b plasmid was also included as a control for expression of his-tag. The pDEST17

gateway vector was chosen for purification of Gateway constructs, because it contains a T7 promoter controlling the transcription of the integrated protein sequence and an N-terminal his-tag. The selected constructs were transformed into BL21(DE3)pLysS bacteria. Each construct was expressed alone or cotransformed with ParC to check the effect on solubility of ParB and give a higher protein yield as observed by Johnson et al. (1999).

The samples collected before and after induction were analyzed by SDS-PAGE. **Figure 3.7 A** shows samples from cultures coexpressing ParB's and ParC (+ ParC) and **Figure 3.7 B** shows ParB's alone (- ParC). Three proteins were highly expressed after IPTG induction and are indicated by red arrows: ParB1 at 42 kDa (lane 3), ParB3 at ~40 kDa (lane 5), and ParB3Core at ~25 kDa kDa (lane 7). Expression is higher in the - ParC samples than + ParC samples. ParB3Hu at ~40 kDa (lane 9) was not well expressed in either, although a faint band is visible. The pET16b should express his-tag alone, but this protein is too small to appear on the gel. The expression level of the ParC protein at 55 kDa (lane 13 **Figure 3.7 A**) is relatively low. It is also low in the coexpressed samples in lanes 3, 5, 7, and 9.

Due to low expression of ParB3Hu, an attempt was made to check single colonies for expression to select the one with the highest expression. A 50 ml expression culture was made for five selected colonies. Unfortunately, none of the selected colonies had markedly better expression levels than the others, so colony 1 from the – ParC samples and colony 3 from the + ParC samples were selected at random, although avoiding colonies that ran poorly on the 12% polyacrylamide gel (appendix 3, **Figure S3)**.

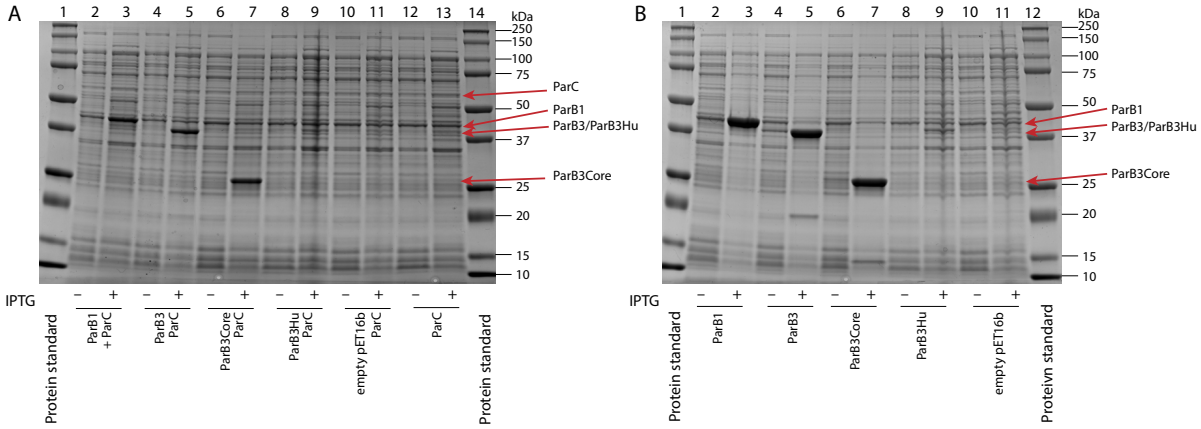


Figure 3.7. Expression of ParB1 and ParB3 +/- ParC in BL21 cells. Proteins were overexpressed from plasmids: pDEST17_ParB1 (42kDa), pDEST17_ParB3 (~40 kDa), pDEST17_ParB3Core (~25 kDa kDa), pET15b_ParB3Hu (~40 kDa), and pET28a_ParC (57 kDa) in BL21(DE) cells. Empty pET16 vector +/- ParC was included as a control. Samples collected from uninduced cultures (-) and induced cultures (+) analyzed by 12% SDS-PAGE. Bands of expected protein sizes are indicated by red arrows in samples – ParC **(A)** and samples + ParC **(B)**. Molecular weight proteins standards are in lanes 1 and 14 **(A)** and lanes 1 and 12 **(B)**.

3.4.3 Recombinant His-tag Purification of ParB3 proteins

3.4.3.1 Protein Purification by Precipitation

Successful induction was seen in the expression experiment resulting in prominent bands on the gel, which indicates sufficient amounts of protein present for purification. Outlined in **Figure 3.8** are the steps of the protein purification by precipitation process (section 2.3.3.1). Indicated in purple and red boxes are the collection points for samples analyzed by SDS-PAGE.

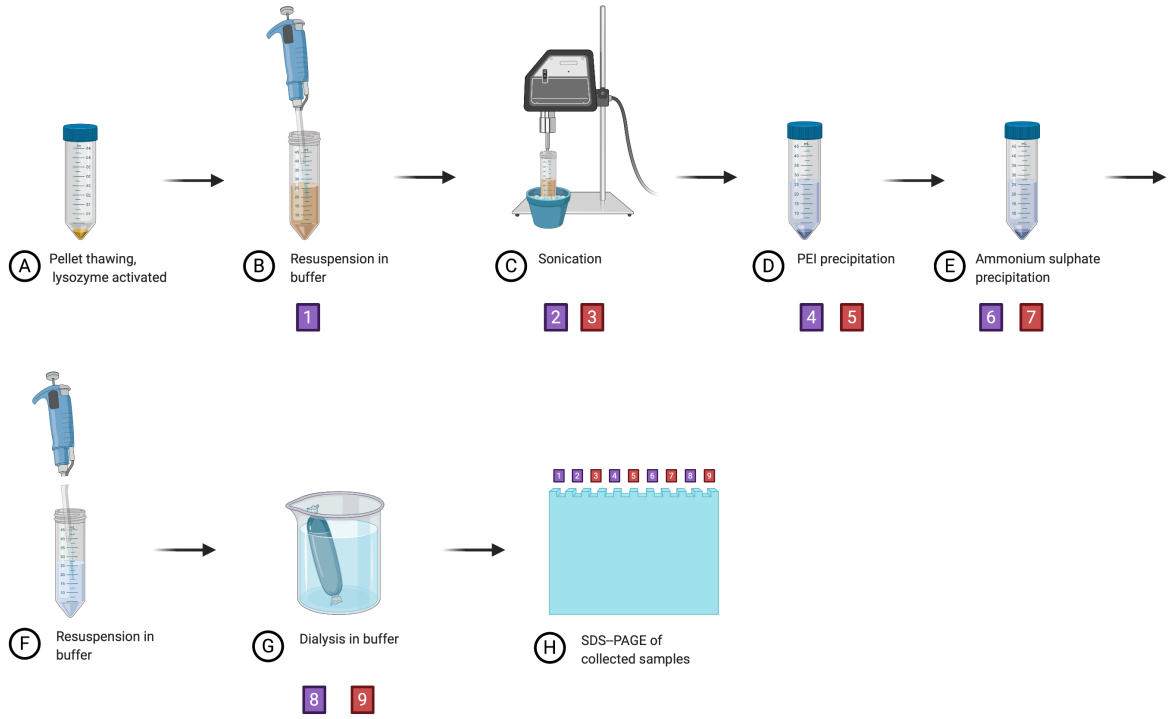


Figure 3.8. Illustration of purification by precipitation experimental process. The steps are marked A-H. Steps where samples were collected is indicated by purple (supernatant) and red (pellet) boxes. For protocol modifications for each experiment see appendix 3, **Table S1**. Image was created with BioRender (https://biorender.com).

Of the 12 plasmid samples recombinantly expressed in BL21 bacteria, eight were purified (pDEST17_ParB3, pDEST17_ParB3Core, pET15b_ ParB3Hu, and empty pET16b +/- ParC). For each step in the purification process samples were collected: the resuspended pellet from the expression experiment, pellet and supernatant after sonication, supernatant and pellet after PEI precipitation, supernatant and pellet after ammonium sulphate precipitation, supernatant and pellet after dialysis. The samples were analyzed by 12% SDS-PAGE and expected protein sizes are indicated by red arrows **(Figure 3.9)**, showing molecular weight of ~40 kDa for ParB3 and ParB3Hu, ~ 25 kDa for ParB3Core, and ~55 kDa for ParC.

The first purification was done for ParB3Hu +/- ParC. In **Figure 3.9 A**, it appears the sonication was successful with little protein lost in the pellet of ParB3Hu - ParC (lane 11). The pellet sample of + ParC was not included on the gel. The PEI pellet contained a large amount of precipitated protein (lanes 14 and 15). The protein content of PEI supernatant is difficult to discern as it ran poorly on the gel. The ammonium sulphate supernatant samples were separated poorly on the gel due to a large concentration of salt and the amount of protein lost to the supernatant cannot be discerned either. There is still a decent amount of protein left in the ammonium sulphate pellet (lanes 18 and 19). A large amount of protein precipitated during dialysis into buffer B, which can be seen in the dialysis pellet (lanes 22 and 23). Still, some protein is present in the supernatant (lanes 20 and 21). Overall, the sample containing coexpression with ParC has a lower yield of ParB3Hu than the sample with ParB3Hu expressed alone. There is also less ParB3Hu recovered in the coexpressed sample after precipitation and following the final dialysis step, indicating that ParC does not help solubilize ParB.

In the second purification experiment, ParB3Core +/- ParC samples were purified. Samples were collected from each step of the purification and analyzed by 12% SDS-PAGE **(Figure 3.9 B** and **C).** From the gels it appears that the sonication was not completely successful in either sample because a large portion of the ParB3Hu protein content was lost in the pellet together with the cellular debris (lanes 6). A similar observation as in the first purification, the PEI pellet contains a large amount of precipitated protein (lane 8), seemingly identical to the supernatant (lane 7). The ammonium sulphate pellet shows that a significant amount of protein is still left in the pellet. Much protein precipitation was detected after the dialysis step, although a band is visible in the supernatant sample (lane 11). The initial concentration of protein after expression was higher in the – ParC sample. The collected samples throughout show that – ParC samples have higher protein concentration than the + ParC samples, even in the final supernatant sample.

In the third purification experiment, ParB3 and pET16b +/- ParC samples were purified. Samples collected from each purification step was analyzed by 12% SDS-PAGE. The ParB3 samples are presented in **Figure 3.9 D** and **E**, while pET16b control samples are included in appendix 3, **Figure S4**. Only gels in **Figure 3.9 D** and **E** are addressed in the following analysis. Sonication in this experiment was not sufficient, leading to loss of protein in the sonication pellet (lane 6). Especially large amounts were lost from the – ParC sample. After PEI precipitation, some protein was lost to the pellet sample (lane 8). Following centrifugation after ammonium sulphate precipitation, the ammonium sulphate pellet was resuspended in buffer A instead of previously used buffer B, to reduce salt concentration prior to dialysis. The

ammonium sulphate supernatant was attempted improved by two-times centrifugation followed by separation of supernatant and addition of loading dye. This had no enhanced effect on the sample when run on the gel. The ammonium sulphate pellet shows a clear band for the ParB3 protein. The resuspended ammonium sulphate pellet was centrifuged for 10 min at 23 000 x g at 4°C, to remove precipitated proteins. The pellet that formed for the ParB3 – ParC samples was larger than the + ParC pellet. This might indicate that the ParC protein could enhance solubility of the ParB proteins as observed by Johnson et al. (1999). The dialysis step was performed with a modification in buffer B content, with 150 mM KCl. Only a small pellet of precipitate was collected after dialysis for the pET16b + ParC, while the other samples had no post dialysis pellet. Overall, more protein is present in the ParB3 – ParC sample although that could be attributed to higher yield in the expression experiment.

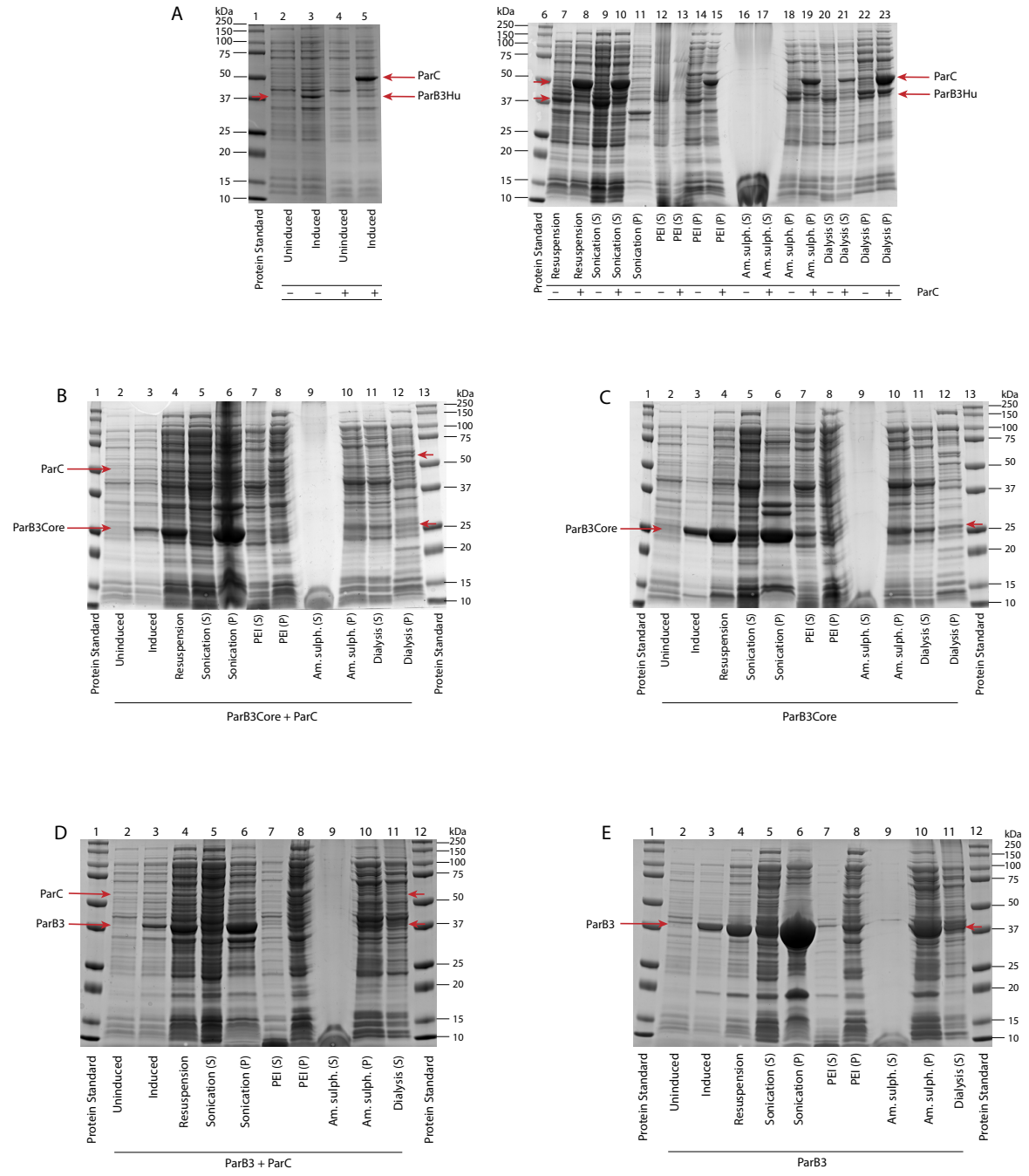


Figure 3.9. Purification of ParB3s +/- ParC. Fractions from each step of the purification process were analyzed by 12% SDS-PAGE. The fractions are labelled according to the protocol steps: (S) supernatant, (P) pellet. Induced and uninduced samples from the expression are shown for comparison. The band corresponding to the expected sizes of the proteins are indicated by red arrows. **(A)** ParB3Hu +/- ParC. In lanes 4, 5, 8, 10, 13, 15, 17, 19, 21, and 23 are + ParC samples, indicated by a (+) mark. Molecular weight protein standards in lanes 1 and 6. **(B)** ParB3Core + ParC. Molecular weight protein standards in lanes 1 and 13. **(C)** ParB3Core - ParC. Molecular weight protein standards in lanes 1 and 13. **(D)** ParB3 + ParC. Molecular weight protein standards in lanes 1 and 12. **(E)** ParB3 - ParC. Molecular weight protein standards in lanes 1 and 12. The gels were fixed and stained with PageBlue protein staining solution.

3.4.3.2 His-tag purification

The protein samples were purified with a HisPur™ Ni-NTA Purification Kit to separate the his-tagged ParB proteins from all other proteins in the samples. This experiment was performed with the help of Marit Ledsaak. Protein present in the eluted samples was checked by a filter paper dye binding assay. A droplet from each elution step was applied onto a 3mm Whatman Grade 17 Chr Cellulose Chromatography paper, which was then soaked in PageBlue protein staining solution and destained in 10% acetic acid. Protein was detected in all three fractions, which were pooled into one sample. The flow through of the sample loading and the two first washing fractions were analyzed by 12% SDS-PAGE together with a compiled sample of the three elution fractions **(Figure 3.10 A** and **B)**. The gels show presence of target ParB proteins together with other proteins present as impurities in all elution samples, especially in the control samples where no His-tagged proteins were present (lanes 5 and 9 in **Figure 3.10 A**). In the elution sample of ParB3Hu – ParC (lane 13), the ParB protein has higher concentration than the impurities. In the + ParC sample the ParB sample was weakly detected. In **Figure 3.10 B** the ParB band is equally clear in the ParB3 + and – samples, but in the ParB3Core samples the band is much more prominent in the – ParC sample (lane 13). In all lanes in both gels in **Figure 3.10** an unknown background protein of ~45 kDa is observed. The protein has likely an affinity for the resin due to its presence in all washing fractions and final eluates.

3.4.3.3 Nuclease Assay

A nuclease assay was performed to determine if the ParB proteins of interest have nuclease activity. The eluated proteins were tested on the pUC19 plasmid on which ParB from *E. coli* has earlier been observed to have unspecific nuclease activity (Johnson et al., 1999).

The concentrations of the protein samples were checked by a Bradford assay (section 2.3.3.3), where concentrations were determined to be in the ranges of 0.07-0.29 µg/µl which represents both ParBs and background proteins. An empty pUC19 plasmid was digested with EcoRI restriction enzyme in an EcoRI buffer (NEB) with added 5mM $CaCl_2$ to promote nuclease activity.

A total of 400 ng of each of the purified ParB +/- ParC proteins and pET16b +/- ParC. **Figure 3.10 C** revealed nuclease activity in all samples tested. In the pET16b control sample we observed DNA digestion visualized by a long smear (lanes 4 and 5). The pET16b – ParC sample (lane 4) had a shorter and less clear smear than in the + ParC sample (lane 5), indicating that the nuclease activity in this sample was higher. In the ParB3Hu - ParC and ParB3 – ParC

samples (lane 6 and 8, respectively) more nuclease activity was observed, indicated by a diffuse band of 100 bp and less which was formed by highly fragmented DNA that runs faster on the agarose gel. Samples ParB3Hu + ParC (lane 7), ParB3 + ParC (lane 9), ParB3Core - ParC (lane 10), and ParB3Core + ParC (lane 11) do not display the similar nuclease activity although some was present as a smear, as in the pET16b control samples.

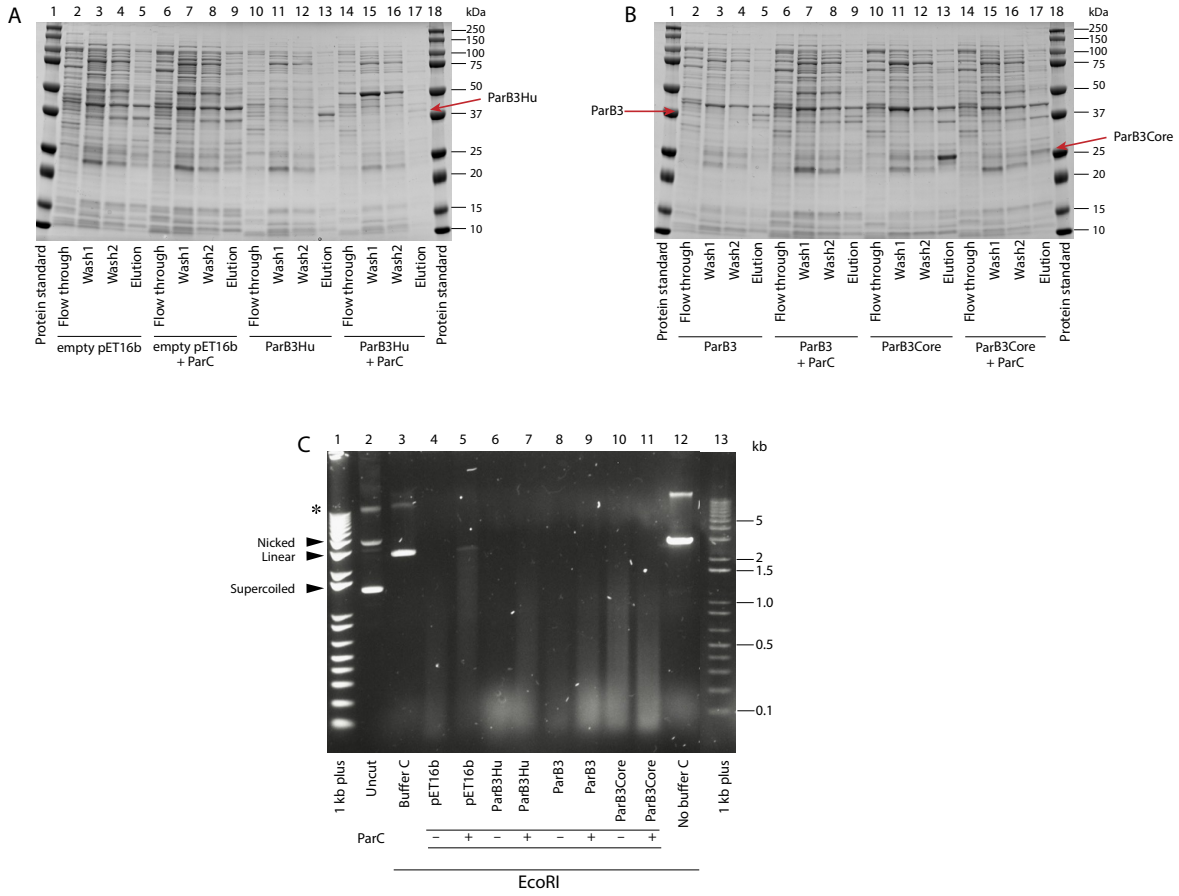


Figure 3.10. Purified his-tag ParB proteins and nuclease assay. Fractions from each step of the his-tag purification process were collected and analyzed by 12% SDS-PAGE: protein flow through, washing steps 1 and 2, and eluted sample. **(A)** Empty pET16b +/- ParC (lanes 2-9) and ParB3Hu +/- ParC (lanes 10-17). Molecular weight protein standards (lanes 1 and 18). **(B)** ParB3 +/- ParC (lanes 2-9) and ParB3Core +/- ParC (lanes 10-17). Molecular weight protein standards (lanes 1 and 18). The gels were fixed and stained with PageBlue protein staining solution. **(C)** Nuclease assay of ParB proteins. Plasmid pUC19 was digested by restriction enzyme EcoRI in EcoRI buffer (NEB) with added 5mM $CaCl_2$ (lane 12) and treated with purified proteins from BL21(DE) expressing pET16b +/- ParC, ParB3Hu +/- ParC, ParB3 +/- ParC, and ParB3Core +/- ParC in buffer C solution (lanes 4-11). pUC19 (lane 2), pUC19 cut with EcoRI in buffer C (lane 3), pUC19 cut in MQ H_2O (lane 12) were used as controls. Samples with and without ParC coexpression are marked with (+) and (-), respectively. Band marked with an asterisk is likely genomic DNA contamination in the pUC19 plasmid sample. 1X TAE 0.8% Agarose gel electrophoresis, 1 kb plus ladder (lanes 1 and 13).

3.5 Expression of ParB3Hu in OS25_INT3 mouse embryonic stem cells

OS25_INT mouse embryonic stem cells with a CRISPR/Cas9-integrated INT3 cassette (repeat *parS* sequences for ParB3) by the *Nanog* locus performed by Beata Nadratowska-Wesolowska, were used for transfection to test the function of the humanized ParB3 plasmid. Original plasmid pDasher_ParB3, construct pDasher_ParB3Hu, and control vector expressing GFP were transfected as described in section 2.4.3. The nuclear DNA was stained in blue with DAPI fluorescent staining which can be seen in contrast to the green fluorescence of the Dasher and GFP tag **(Figure 3.11)**. The ParB proteins do not have a nuclear localization signal (NLS) and is therefore not be actively transported to the nucleus. Because the protein expression is high, some of the ParB proteins will leak into the nucleus, bind to the *parS* sequence, and form a focus driven by the intrinsic spreading mechanism of ParB proteins. Due to high expression, a lot of background fluorescence can be seen in the cytosol where the ParB protein expresses. Midsections of z-stacks must be selected, as ParB aggregates can from in the cytosol. Images have to be analyzed carefully to not confuse nuclear foci with cytoplasmic foci. **Figure 3.11 A** shows the expression of pDasher_ParB3 in OS25_INT3. A nuclear focus is indicated by white arrows. By making a composite image of the DAPI and the Dasher staining, we can observe that the nuclear foci locates to where the blue color is overlaid. In **Figure 3.11 B**, where OS25_INT3 cells transfected with pDasher_ParB3Hu is visualized, two foci are observed which can be attributed to focus formation on both alleles. The expression and focus formation in ParB3 and ParB3Hu do not appear to differ. In **Figure 3.11 C**, a plasmid with a GFP tag was transfected. As expected, we see no foci formation in the nucleus in these images. Due to the SARS-CoV-2 pandemic restrictions, supervised training on instruments were prevented and thus all imaging was performed by Beata Nadratowska-Wesolowska.

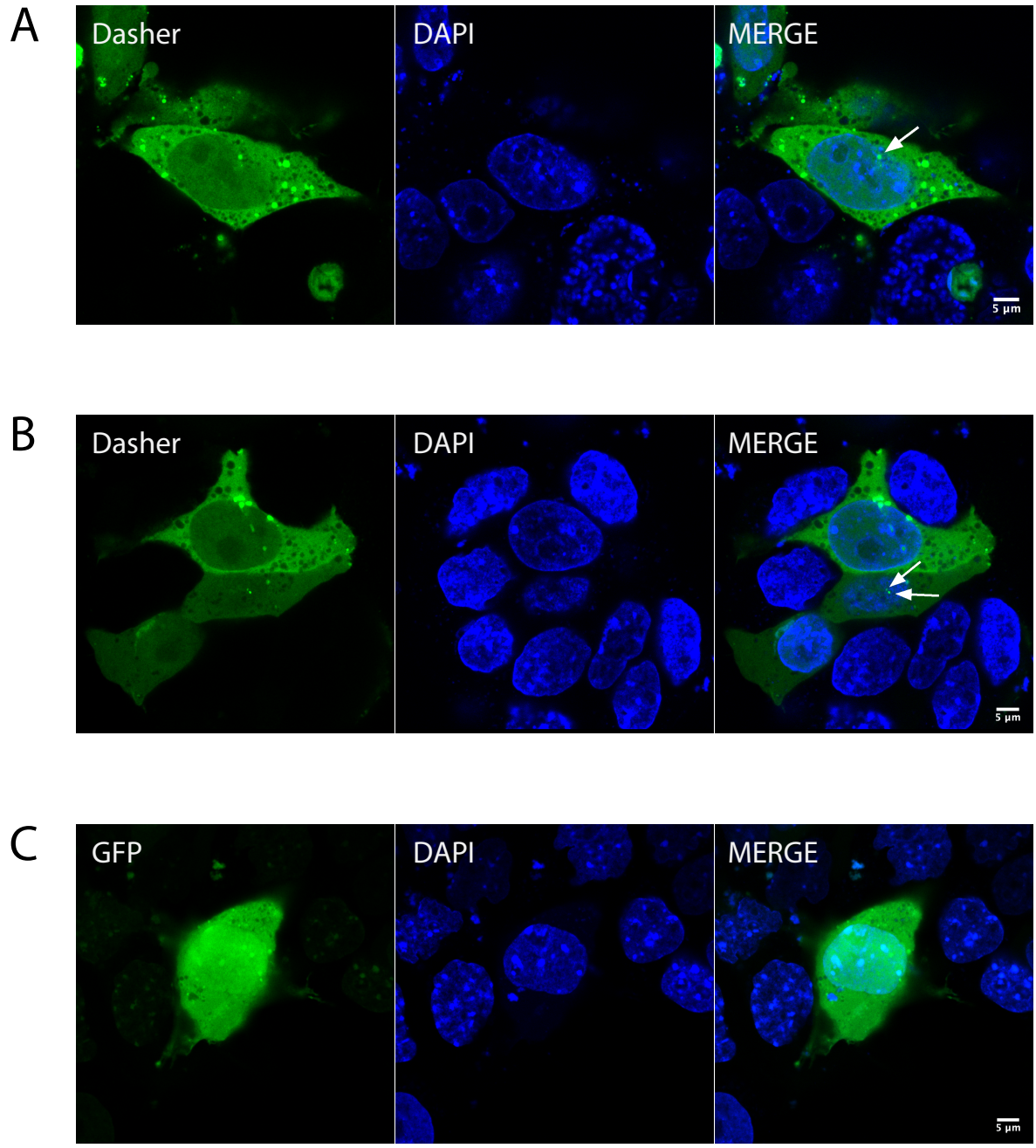


Figure 3.11. Expression of ParB3 and ParB3Hu OS25_INT3 mES cells. Transient transfection with Dasher-tagged ParB3 **(A),** ParB3Hu **(B)** proteins, and GFP-tagged control **(C)**. Foci are indicated with white arrows. Cells were fixed with 4% formaldehyde and DNA stained blue with DAPI (section 2.4.4). Scale bar: 5 µm. Images were captured by Beata Nadratowska-Wesolowska on the Zeiss710. Images were analyzed with Fiji ImageJ.

4 Discussion

The aim of this thesis was to investigate if ParB proteins repurposed for the ParB-INT imaging system have nuclease activity in order to understand and improve the system for live cell imaging of single genomic loci in embryonic stem cells. Moreover, to gain more insight into the protein structure, which is tightly connected to protein function, evolutionary relationship with homologous proteins were studied.

4.1 Three Predicted Models Indicate Potential Native Structure

Based on extensive structural analysis of predicted protein structures, three models were selected as the best representative structures: one for ParB1 and two for ParB3. The core of each model was predicted from the fold of *Bacillus subtilis*. On the constructed evolutionary tree, *B. subtilis* branches off only a few nodes before the outlined monophyletic group containing the ParB1 protein studied in this thesis (section 3.3.2, **Figure 3.5)**. Because protein structure is more conserved than sequence, and the same fold is found for ParB1 located on one of the final branches in this group, we expect that this fold is found in all proteins connected to the *B. subtilis* branching node.

The core of the predicted structures is highly stable and ordered, while the N-terminal regions (NTR) and C-terminal regions (CTR) of ParB1 and ParB3 were largely predicted to be intrinsically disordered. The CTR of ParB1 was not predicted to be disordered but had a peak of predicted ordered amino acids (aa) just below the cutoff threshold (section 3.2, **Figure 3.1 A)**. By comparing this result with the lack of structured NTR and CTR by SWISS-MODEL (SM) prediction (section 3.2, **Figure 3.2 A, C** and **D**), and then the Rampage analysis which placed many C-terminal aa in the outlier regions of the Ramachandran plot (section 3.2, **Figure 3.2 B)**, this peak should not be completely disregarded as it may be a false negative. Based on the multiple sequence alignment of homologous ParB proteins, we discover that the flexible NTR and CTR are highly variable between species and have a low degree of evolutionary conservation. This indicates that parts of these regions may have less relevance for protein function. However, there are a few aa conserved throughout these regions, which we believe to be responsible for the activity known to reside in these regions. New models were generated *ab initio* by I-TASSER, in an attempt to predict the structure of the flexible N- and C-terminal regions. The best models were selected based on subjective evaluations of the analysis results

and should therefore only be interpreted as suggestive representations of the native protein structure.

The modeled 3D structure of a protein adds additional information that linear sequences do not contain, for example enzymatic cavities and ligand-binding sites, topological information, exposed aa etc. In this project we have investigated the potential to use the structural information to locate the active site of the nuclease activity that earlier studies have demonstrated for ParB proteins (Grohmann et al., 1997; Johnson et al., 1999; Shaw et al., 2008).

4.2 Structure and Homology Analysis Reveal Possible Sites of Nuclease Activity

We investigated the possible location of the nuclease activity site in ParB1 through location of conserved motifs from a multiple sequence alignment and location in the predicted structural model. The location of the highly conserved GxxRxxA site in BOX II, which was recently found to exhibit CTP binding and/or hydrolysis activity (Jalal et al., 2020; Osorio-Valeriano et al., 2019; Soh et al., 2019), was identified in the predicted structures. This site is highly conserved among ParB homologs, including eukaryotic Sulfiredoxin (Srx) proteins (section 3.3.1, **Figure 3.4 A**). The GxxRxxA site was located in a large cavity in the predicted structure for ParB1. Previous mutational studies of the GxxRxxA motif in *B. subtilis* have shown that several mutations of these aa residues, such as R80A and R82A, lead to abolished foci formation (Song et al., 2017), and some, such as G77A and R80A, lead to faulty foci focus formation and defective subcellular localization (Breier and Grossman, 2007; Graham et al., 2014). However, the nuclease activity of these mutants was not studied. Chi et al. (2012) studied the nuclease activity of a Srx mutant where the catalytic Cysteine residue shown to be crucial for reductase activity, an evolutionary substitution of ParB Glycine in the GxxRxxA site, was mutated to Serine. This had no effect on the nuclease activity, and suggests that the cognate Glycine found in ParB homologs is likely not involved in nuclease activity.

As Srx proteins show high sequence similarity with ParB proteins, and have been demonstrated to retain their nuclease activity (Chi et al., 2012), it would be highly relevant to understand how mammalian cells containing cytosolic Srx proteins respond to coexpression of prokaryotic ParB proteins. Srx proteins are mainly localized in the cytosol and mitochondria (Chang et al., 2004), which is where the bulk of expressed ParB proteins are found. However, based on the multiple sequence alignment and phylogenetic tree, we observed that the Srx proteins have gained much evolutionary distance from ParB proteins, and would perhaps not

be perturbed by coexpression with its prokaryotic homolog. Neither the nuclease activity of ParB in bacteria nor Srx in eukaryotic cells has been well characterized and its purpose has not yet been uncovered.

The well-conserved ParB BOX I was also observed to localize in the same large protein cavity on the predicted models as the GxxRxxA site (section 3.3.1, **Figure 3.4 A**), and this sequence is also conserved in the nuclease retaining Srx proteins. Moreover, the well-conserved ParB BOX III was not found conserved in the Srx proteins. As it is very unlikely that the same function has appeared twice in close homologs throughout evolution, this suggests that BOX III is not involved in nuclease function. BOX I and BOX II should therefore be further investigated for potential mutational studies to abrogate nuclease activity and improve functionality of the ParB-INT system.

4.3 Nuclease Activity Observed in ParB3 and ParB3Hu

I purified six recombinant protein samples and all samples contained unspecific protein contamination some of which exerted nuclease activity (section 3.4.3.3, **Figure 3.10**). Therefore, the results of the nuclease assay will need further validation. However, the preliminary data obtained from the nuclease assay give an indication of nuclease activity in the tested proteins. The observation is especially relevant for ParB3Hu and ParB3Core expressed without ParC as these His-purified samples had a ParB protein band that was significantly more prominent than the contaminants. The ParB3 without ParC had a contaminating band of similar distinction.

The nuclease assay indicates that ParB3 and ParB3Hu both possess nuclease activity, and that this activity was higher than in the control (pET16b). The other proteins did not show the same extensive DNA fragmentation as these two proteins.

Generally, the ParB samples coexpressed with ParC did not show nuclease activity, indicating that either ParC results in lower yields of ParB or that ParC inhibits ParB activity by some unknown mechanism. A western blot with antibodies against the ParC HA-tag would confirm the presence of ParC. ParC is one of two subunits making up topoisomerase IV in bacteria. The topoisomerase is involved in the plasmid segregation process by relaxing and decatenating the interlinked plasmids (Kato et al., 1990; Zechiedrich et al., 1997). A potential mechanism for the possible ParC inhibition of ParB nuclease activity would be interesting to unveil as the nuclease activity of ParB is not well understood.

4.4 ParB3Core does not exhibit nuclease activity

The ParB3Core without ParC had the most prominent band of the proteins purified (section 3.4.3.3, **Figure 3.10 B**), which suggests that this is the protein that should have most effect, if any, on the pUC19 plasmid. Notably, the protein appeared to have little nuclease activity with similar cutting activity as the + ParC samples and the control pET16b samples. This data suggests that aa crucial to the nuclease activity can be found in the flexible NTD or CTD. The alignment between ParB1 and ParB of *B. subtilis* did not show many conserved aa in the terminal regions (section 3.3.1, **Figure 3.4 B**). The core domain contains the HTH domain necessary to bind DNA and the BOX II motif for CTP hydrolysis. Further investigations into the spreading activity of the ParB protein should be made to see if the core protein could be used in the ParB_INT imaging system. A recent study of a C-terminally truncated ParB homolog in *B. subtilis* showed that the C-terminal domain is not required for DNA-binding to the *parS* sequence, a function that has long been attributed to the HTH-domain (Fisher et al., 2017). However, the authors demonstrated that non-specific DNA (nsDNA) binding, and thus spreading, of ParB proteins was dependent of the dimerized C-terminal domain (Bartosik et al., 2009; Fisher et al., 2017). The ParB1 protein used in this project does not appear to have a well-conserved CTD in the alignment with *B. subtilis*, although both proteins have a Lysine-rich patch which was found to be important for dimerization (**Figure 3.4 B**). A core domain lacking a full CTD would likely be defective in spreading. The ParB3Core protein does also lack large portions of the NTD, which has been associated with protein-protein interactions with the ParA protein responsible for active segregation of chromosomes in prokaryotes. This function would not be relevant in the ParB_INT system. However, Soh et al. (2019) demonstrates that the N-terminal domain is also involved in the formation of dimers as shown for the CTD, leading to the assembly ParB dimers in clamp-like structures that close around the DNA strand. These studies all indicate that the core domain would not be able to form a focus around the parS site, as it likely can not properly bind, nor spread, on flanking DNA sequences. Nevertheless, the exact aa responsible for each of these activities has not been fully determined and the function of the core domain of the ParB3 protein should be investigated further.

4.5 ParC Likely Increases ParB Solubility but Lowers Total Yield

Johnson et al. (1999) showed that the coexpression of ParC topoisomerase with ParB proteins increased the solubility of ParB proteins and prevented formation of aggregates leading to precipitation during purification. In our protocol, this effect was only clearly observed when the resuspended pellet of ParB3 was centrifuged between step F and G (section 3.4.3.1, **Figure 3.8**), which resulted in a larger pellet of precipitated protein for samples without ParC than for samples coexpressed with ParC. This could suggest that more of the protein was in a soluble fraction in the presence of ParC, which supports the findings by Johnson (1999). However, the coexpressed ParB samples had lower expression efficiency in bacteria compared to ParB alone. His-tag purification protocol was selected because it is simple, requires no removal of fusion protein, and generally results in higher protein activity. To get a total higher yield of proteins, the solubility of ParB could be enhanced by other approaches, such as expression in vectors containing solubility enhancing tags such as an MBP-tag (pDEST566) or a GST-tag (pGEX2T1GW). These constructs are already available and were produced as a part of this thesis.

4.6 Weak Expression of Humanized Protein in Bacteria

A human codon optimized ParB3 was expressed in BL21 bacterial cells. The expression was very weak and not above background level compared to other ParB proteins, despite attempts to select well-expressing colonies (section 3.4.2, **Figure 3.7**). Studies show that translation of mRNA with codons that are not commonly found in the genome of the organism can result in delays in the elongation phase. This delay is due to a shortage of the specific tRNAs needed for transport of the correct aa to the ribosome (Hanson and Coller, 2018). Expression of a ParB3 protein optimized for mammalian expression may have resulted in reduced translation efficiency in *Escherichia coli*.

4.7 How can ParB Purification be Optimized?

A well-working protein purification protocol needs optimization for every new protein. Due to time limitations in this project, only one setup was performed for each sample. The purification protocol from Johnson et al. (1999) was valuable in setting up an in-lab protocol. Adjustments to the purification and precipitation protocol were made, however further

modifications are necessary to achieve better quality and yield (section 3.4.3). A few suggestions to improve the protocol are presented here.

I found that a prolonged sonication is necessary for cultures of larger volumes to release all proteins from the cell, as 6 minutes of sonication was adequate for the 50 ml expression cultures but not sufficient for the 250 ml expression cultures. Macromolecular precipitation by PEI appears to have worked well in precipitating large negatively charged macromolecules, like DNA and acidic proteins. To investigate if this step can be improved even further, lower salt concentrations in the buffer can be tested as high salt content could precipitate more proteins. A significant amount of protein was lost due to precipitation in the resuspension buffer after the ammonium sulphate precipitation step. To solubilize more protein after ammonium sulphate precipitation, buffer resuspension volume can be increased slightly to decrease protein concentration, and different salt concentrations in the buffer could be assessed. Ammonium sulphate can potentially be interchanged with other salts that perform the same "salting out" function, although ammonium sulphate is useful because of its additional ability to stabilize protein structure (Burgess, 2009). To increase yield from His-purification column and prevent weakly attached proteins to eluate, more protein needs to be solved after ammonium sulphate precipitation. Precipitated protein could possibly be solubilized by a longer incubation in resuspension buffer, and larger cultures for expression can be prepared. Moreover, the imidazole concentration in the wash buffers could be increased, while inspecting the wash fractions to make sure not too much ParB is lost.

4.8 Humanized ParB3 does not Improve the ParB_INT System

To compare the efficacy of the original ParB3 protein and the humanized ParB3Hu protein in the ParB-INT imaging system, these were transfected in OS25_INT3 cells. Because 64 possible codons encode only 20 different aa, each aa is coded by more than one codon. The tendency of some codons to be more frequently represented in the genomes of different species, is referred to as codon bias (Hanson and Coller, 2018). Codon optimization has been applied to recombinant proteins in order to interchange codons that are underrepresented in the genome for frequently selected codons. Codon optimization has been shown to increase elongation rates, expression efficiency, protein folding, and mRNA stability in some species, although this is a subject of debate (Hanson and Coller, 2018). Also, a recent study used codon-optimized ParB proteins in yeast, but did not perform a comparative study to look at efficiency of codon-optimization for ParB proteins (Audibert et al., 2020). In OS25_INT3 cells we observed that

the expression efficiency of the humanized version appeared to be similar to the prokaryotic version (section 3.5, **Figure 3.11**), and that the humanized ParB3 was not an improvement to the system. To further validate these results, more replicates should be made for quantitative observations.

4.9 Conclusion

The conclusions for this thesis will be divided into three parts, corresponding to each of the aims presented in section 1.7.

(1) The structural composition of ParB1 and ParB3 was predicted and analyzed, resulting in one ParB1 model and two ParB3 models. An evolutionary study of ParB homologs revealed that certain regions of ParB proteins are very well conserved, also in the eukaryotic homolog protein sulfiredoxin, while the predicted intrinsically disordered N- and C-terminal domains were highly variable between homologs. The well-conserved BOX I and GxxRxxA motif in BOX II were recognized as likely sites of nuclease activity, as they both reside in large pockets in the predicted model structure and should be further studied for mutagenesis experiments.

(2) I have in this project established a panel of clones to study the nuclease activity of two ParB proteins from the Burkholderiaceae family. Corresponding to findings by Johnson and colleagues, we saw a small increase in solubility of ParB proteins when coexpressed with ParC. Nonetheless, ParC appeared to have a negative impact on expression of ParB. ParB3 and ParB3Hu appeared to possess nuclease activity, and coexpression with ParC may inhibit nuclease activity. However, these experiments need to be repeated with ParBs of high quality and purity. The core domain of ParB3 did not show nuclease activity above controls and should be tested in the ParB-INT3 imaging system, although we can expect deficient spreading due to truncation of the N- and C-terminal regions which have been shown to be involved in the spreading mechanism of other ParB proteins.

(3) Preliminary qualitative data show that the humanized version of ParB3 was not an improvement to the ParB_INT3 imaging system as it appeared to express at equivalent levels to the prokaryotic version.

4.10 Further Research

Preliminary data suggests that ParB3 and ParB3Hu have nuclease activity. To validate these findings, the purification protocol must be improved. Proteins should be expressed with

a solubility-enhancing tag like GST and MBP, as the ParB proteins were only slightly more soluble with ParC coexpression. The purification can be improved by size-exclusion chromatography where proteins can be purified according to size, and ParB proteins can be separated from contaminating proteins.

The nuclease assay should be repeated to confirm that the core domain has lost its nuclease activity. The truncated protein should be fluorescently tagged and tested in stem cells to observe if the spreading mechanism is still intact and focus formation is observed. However, we do not expect a focus to be detectable above background levels as aa integral to focus formation has been found in the N- and C-terminal domains, resulting in a spreading deficient protein. A loss of nuclease activity in the ParB3Core protein indicates that aa involved in nuclease activity is also located in the terminal domains, which should be further investigated for a mutational study. Furthermore, a transfection can be performed to detect the nuclease activity of the ParB protein in mammalian cells. As a response to double stranded breaks (DBS), the H2A variant, H2AX, is phosphorylated (gamma-H2AX) and becomes a docking station for repair machinery recruited to the site (Rogakou et al., 1998; Rothkamm et al., 2015). Through immunostaining with antibodies against gamma-H2AX, focus formation around the DBS could confirm ParB nuclease activity.

To improve the ParB_INT imaging system, a mutational study of nuclease activity could lead to inactivation of the nuclease activity, while still retaining DNA-binding and spreading activity. A starting point would be to study the sequence and structure of other prokaryotic nucleases where the active sites are known, to compare with the ParB sequence. In this study we indicate that the location of the nuclease activity is in conserved BOX I or BOX II, which is recognized in ParB proteins and Sulfiredoxin proteins. In a recent mutational study of ParB in *B. subtilis*, Song et al. (2017) describes the ability of mutants to bind to *parS*, spread, and condense DNA. Mutants found to form fluorescent foci could be further tested for nuclease activity *in vitro*. This would form the basis of an improved ParB_INT system in the future.

5 References

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., 2015. Molecular biology of the cell, 6th ed. ed. Garland Science, New York.

Anchordoguy, T.J., Carpenter, J.F., Crowe, J.H., Crowe, L.M., 1992. Temperature-dependent perturbation of phospholipid bilayers by dimethylsulfoxide. Biochim. Biophys. Acta 1104, 117–122. https://doi.org/10.1016/0005-2736(92)90139-D

Anton, T., Bultmann, S., Leonhardt, H., Markaki, Y., 2014. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. Nucleus 5, 163–172. https://doi.org/10.4161/nucl.28488

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, 344–350. https://doi.org/10.1093/nar/gkw408

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, 529–533. https://doi.org/10.1093/nar/gkq399

Audibert, S., Tanguy-Le-Gac, N., Rech, J., Turlan, C., Bouet, J.Y., Bystricky, K., Lane, D., 2020. Addressing the role of centromere sites in activation of ParB proteins for partition complex assembly. PLoS One 15, 1–21. https://doi.org/10.1371/journal.pone.0226472

Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. Cell Res. 21, 381–395. https://doi.org/10.1038/cr.2011.22

Bartosik, A., Lasocki, K., Mierzejewska, J., Thomas, C., Jagura-Burdzy, G., 2004. ParB of Pseudomonas aeruginosa: Interactions with Its Partner ParA and Its Target parS and Specific Effects on Bacterial Growth. J. Bacteriol. 186, 6983–6998. https://doi.org/10.1128/JB.186.20.6983–6998.2004

Bartosik, A., Mierzejewska, J., Thomas, C.M., Jagura-Burdzy, G., 2009. ParB deficiency in Pseudomonas aeruginosa destabilizes the partner protein ParA and affects a variety of physiological parameters. Microbiology 155, 1080–1092. https://doi.org/10.1099/mic.0.024661-0

Basu, M.K., Koonin, E. V., 2005. Evolution of eukaryotic cysteine sulfinic acid reductase, sulfiredoxin (Srx), from bacterial chromosome partitioning protein ParB. Cell Cycle 4, 947–952. https://doi.org/10.4161/cc.4.7.1786

Baxter, J.C., Funnell, B.E., 2014. Plasmid Partition Mechanisms. Microbiol. Spectr. 2. https://doi.org/10.1128/microbiolspec.plas-0023-2014

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., De Santiago, I., Lavitas, L.M., Branco, M.R., Fraser, J., Dostie, J., Game, L., Dillon, N., Edwards, P.A.W., Nicodemi, M., Pombo, A., 2017. Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543, 519–524. https://doi.org/https://doi.org/10.1038/nature21411

Benkert, P., Biasini, M., Schwede, T., 2010. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350. https://doi.org/10.1093/bioinformatics/btq662

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T., 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci. Rep. 7, 1–15. https://doi.org/10.1038/s41598-017-09654-8

Bickmore, W.A., 2013. The Spatial Organization of the Human Genome. Annu. Rev. Genomics Hum. Genet. 14, 67–84. https://doi.org/10.1146/annurev-genom-091212-153515

Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., 2017. The SWISS-MODEL Repository—new features and functionality. Nucleic Acids Res. 45, 313–319. https://doi.org/10.1093/nar/gkw1132

Braitbard, M., Schneidman-Duhovny, D., Kalisman, N., 2019. Integrative Structure Modeling: Overview and Assessment. Annu. Rev. Biochem. 88, 113–135. https://doi.org/10.1146/annurev-biochem-013118-111429

Breier, A.M., Grossman, A.D., 2007. Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the Bacillus subtilis chromosome. Mol. Microbiol. 64, 703–718. https://doi.org/10.1111/j.1365-2958.2007.05690.x

Burgess, R.R., 2009. Chapter 20 Protein Precipitation Techniques, in: Methods in Enzymology. Elsevier Inc., pp. 331–342. https://doi.org/10.1016/S0076-6879(09)63020-2

Bystricky, K., 2015. Chromosome dynamics and folding in eukaryotes: Insights from live cell microscopy. FEBS Lett. 589, 3014–3022. https://doi.org/10.1016/j.febslet.2015.07.012

Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., Ben-Tal, N., 2013. ConSurf: Using evolutionary data to raise testable hypotheses about protein function. Isr. J. Chem. 53, 199–206. https://doi.org/10.1002/ijch.201200096

Chang, T.S., Jeong, W., Woo, H.A., Sun, M.L., Park, S., Sue, G.R., 2004. Characterization of mammalian sulfiredoxin and its reactivation of hyperoxidized peroxiredoxin through reduction of cysteine sulfinic acid in the active site to cysteine. J. Biol. Chem. 279, 50994–51001. https://doi.org/10.1074/jbc.M409482200

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., Huang, B., 2013. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. Cell 155, 1479–1491. https://doi.org/https://doi.org/10.1016/j.cell.2013.12.001

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., Gregor, T., 2018. Dynamic interplay between enhancer–promoter topology and gene activity. Nat. Genet. 50, 1296–1303. https://doi.org/10.1038/s41588-018-0175-z

Cheng, C.Y., Song, J., Pas, J., Meijer, L.H.H., Han, S., 2015. DMSO Induces Dehydration near Lipid Membrane Surfaces. Biophys. J. 109, 330–339. https://doi.org/10.1016/j.bpj.2015.06.011

Chi, Y.H., Kim, S.Y., Jung, I.J., Shin, M.R., Jung, Y.J., Park, Jin Ho, Lee, E.S., Maibam, P., Kim, K.S., Park, Joung Hun, Kim, M.J., Hwang, G.Y., Lee, S.Y., 2012. Dual functions of Arabidopsis sulfiredoxin: Acting as a redox-dependent sulfinic acid reductase and as a redox-independent nuclease enzyme. FEBS Lett. 586, 3493–3499. https://doi.org/10.1016/j.febslet.2012.08.002

Cremer, T., Cremer, C., 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. 2, 292–301. https://doi.org/10.1038/35066075

Davanloo, P., Rosenberg, A.H., Dunn, J.J., Studier, F.W., 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.81.7.2035

Dekker, J., Rippe, K., Dekker, M., Kleckner, N., 2002. Capturing chromosome conformation. Science (80-.). 295, 1306–1311. https://doi.org/10.1126/science.1067799

Dötsch, V., Wagner, G., 1998. New approaches to structure determination by NMR spectroscopy. Curr. Opin. Struct. Biol. 8, 619–623. https://doi.org/https://doi.org/10.1016/S0959-440X(98)80154-1

Dubarry, N., Pasta, F., Lane, D., 2006. ParABS Systems of the Four Replicons of Burkholderia cenocepacia: New Chromosome Centromeres Confer Partition Specificity. J. Bacteriol. 188, 1489–1496. https://doi.org/10.1128/JB.188.4.1489

Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V.A., Pieper, U., Stuart, A.C., Marti-

Renom, M.A., Madhusudhan, M.S., Yerkovich, B., Sali, A., 2003. Tools for comparative protein structure modeling and analysis. Nucleic Acids Res. 31, 3375–3380. https://doi.org/10.1093/nar/gkg543

Faure, G., Makarova, K.S., Koonin, E. V., 2019. CRISPR–Cas: Complex Functional Networks and Multiple Roles beyond Adaptive Immunity. J. Mol. Biol. 431, 3–20. https://doi.org/10.1016/j.jmb.2018.08.030

Fisher, G.L.M., Pastrana, C.L., Higman, V.A., Koh, A., Taylor, J.A., Butterer, A., Craggs, T., Sobott, F., Murray, H., Crump, M.P., Moreno-Herrero, F., Dillingham, M.S., 2017. The structural basis for dynamic DNA binding and bridging interactions which condense the bacterial centromere. Elife 6, 1–25. https://doi.org/10.7554/eLife.28086

Fraser, J., Williamson, I., Bickmore, W.A., Dostie, J., 2015. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. Microbiol. Mol. Biol. Rev. 79, 347–372. https://doi.org/10.1128/mmbr.00006-15

Funnell, B.E., 2016. ParB partition proteins: Complex formation and spreading at bacterial and plasmid centromeres. Front. Mol. Biosci. 3, 86–91. https://doi.org/10.3389/fmolb.2016.00044

Germier, T., Audibert, S., Kocanova, S., Lane, D., Bystricky, K., 2018. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. Methods 142, 16–23. https://doi.org/10.1016/j.ymeth.2018.04.008

Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politi, A.Z., Ellenberg, J., Gallardo, F., Bystricky, K., 2017. Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. Biophys. J. 113, 1383–1394. https://doi.org/10.1016/j.bpj.2017.08.014

Giancarlo, R., Rombo, S.E., Utro, F., 2019. DNA combinatorial messages and Epigenomics: The case of chromatin organization and nucleosome occupancy in eukaryotic genomes. Theor. Comput. Sci. 792, 117–130. https://doi.org/10.1016/j.tcs.2018.06.047

Gomez-Lamarca, M.J., Falo-Sanjuan, J., Stojnic, R., Abdul Rehman, S., Muresan, L., Jones, M.L., Pillidge, Z., Cerda-Moya, G., Yuan, Z., Baloul, S., Valenti, P., Bystricky, K., Payre, F., O'Holleran, K., Kovall, R., Bray, S.J., 2018. Activation of the Notch Signaling Pathway In Vivo Elicits Changes in CSL Nuclear Dynamics. Dev. Cell 44, 611-623. https://doi.org/https://doi.org/10.1016/j.devcel.2018.01.020

Graham, T.G.W., Wang, X., Song, D., Etson, C.M., van Oijen, A.M., Rudner, D.Z., Loparo, J.J., 2014. ParB spreading requires DNA bridging. Genes Dev. 28, 1228–1238. https://doi.org/10.1101/gad.242206.114

Grohmann, E., Stanzer, T., Schwab, H., 1997. The ParB protein encoded by the RP4 par region is a Ca2+-dependent nuclease linearizing circular DNA substrates. Microbiology 143, 3889–3898. https://doi.org/10.1099/00221287-143-12-3889

Grosveld, F., de Boer, E., Dillon, N., Gribnau, J., McMorrow, T., Milot, E., Trimborn, T., Wijgerde, M., Fraser, P., 1998. The dynamics of globin gene expression and position effects. Novartis Found. Symp. 214, 67–86. https://doi.org/10.1002/9780470515501.ch5

Guex, N., Peitsch, M.C., Schwede, T., 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30, 162–173. https://doi.org/10.1002/elps.200900140

Gurtovenko, A.A., Anwar, J., 2007. Modulating the structure and properties of cell membranes: The molecular mechanism of action of dimethyl sulfoxide. J. Phys. Chem. B 111, 10453–10460. https://doi.org/10.1021/jp073113e

Han, J., Zhang, Z., Wang, K., 2018. 3C and 3C-based techniques: The powerful tools for spatial genome organization deciphering. Mol. Cytogenet. 11, 1–10. https://doi.org/10.1186/s13039-018-0368-2

Hanson, G., Coller, J., 2018. Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. 19, 20–30. https://doi.org/10.1038/nrm.2017.91

Hoffbrand, A.V., Vyas, P., Campo, E., Haferlach, T., Gomez, K., 2019. Molecular Biology of the Cell, in: Color Atlas of Clinical Hematology: Molecular and Cellular Basis of Disease. John Wiley & Sons Ltd., pp. 1–26. https://doi.org/10.1002/9781119170655

Hoffman, E.A., Frey, B.L., Smith, L.M., Auble, D.T., 2015. Formaldehyde crosslinking: A tool for the study of chromatin complexes. J. Biol. Chem. 290, 26404–26411. https://doi.org/10.1074/jbc.R115.651679

Hong, Y., Lu, G., Duan, J., Liu, W., Zhang, Y., 2018. Comparison and optimization of CRISPR/dCas9/gRNA genome-labeling systems for live cell imaging. Genome Biol. 19, 1–10. https://doi.org/10.1186/s13059-018-1413-5

Isalan, M., 2013. Zinc Fingers: Structure and Design, in: Lennarz, W.J., Lane, M.D. (Eds.), Encyclopedia of Biological Chemistry. Academic Press, pp. 575–579. https://doi.org/10.1016/b978-0-12-809633-8.21266-1

Jalal, A.S., Tran, N.T., Le, T.B., 2020. ParB spreading on DNA requires cytidine triphosphate in vitro. Elife 9, 1–24. https://doi.org/10.7554/eLife.53515

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 (80-.). 337, 816–821. https://doi.org/10.1126/science.1225829

Johnson, E.P., Mincer, T., Schwab, H., Burgin, A.B., Helinski, D.R., 1999. Plasmid RK2 ParB protein: Purification and nuclease properties. J. Bacteriol. 181, 6010–6018. https://doi.org/10.1128/jb.181.19.6010-6018.1999

Jones, D.T., Cozzetto, D., 2015. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857–863. https://doi.org/10.1093/bioinformatics/btu744

Joyce, A.P., Zhang, C., Bradley, P., Havranek, J.J., 2015. Structure-based modeling of protein: DNA specificity. Brief. Funct. Genomics 14, 39–49. https://doi.org/10.1093/bfgp/elu044

Kato, J. ichi, Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., Suzuki, H., 1990. New topoisomerase essential for chromosome segregation in E. coli. Cell 63, 393–404. https://doi.org/10.1016/0092-8674(90)90172-B

Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166. https://doi.org/10.1093/bib/bbx108

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kawalek, A., Wawrzyniak, P., Bartosik, A.A., Jagura-Burdzy, G., 2020. Rules and exceptions: The role of chromosomal ParB in DNA segregation and other cellular processes. Microorganisms 8, 19–22. https://doi.org/10.3390/microorganisms8010105

Kempfer, R., Pombo, A., 2020. Methods for mapping 3D chromosome architecture. Nat. Rev. Genet. 21, 207–226. https://doi.org/10.1038/s41576-019-0195-2

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181, 662–666. https://doi.org/10.1038/181662a0

Khanna, V., Ranganathan, S., Petrovsky, N., 2019. Rational Structure-Based Drug Design. Encycl. Bioinforma. Comput. Biol. 2, 585–600. https://doi.org/https://doi.org/10.1016/B978-0-12-809633-8.20275-6

Kuraku, S., Zmasek, C.M., Nishimura, O., Katoh, K., 2013. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res. 41, W22–W28. https://doi.org/10.1093/nar/gkt389

Langer-Safer, P.R., Levine, M., Ward, D.C., 1982. Immunological method for mapping genes

on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. 79, 4381–4385. https://doi.org/10.1073/pnas.79.14.4381

Li, Jianzong, Feng, Y., Wang, X., Li, Jing, Liu, W., Rong, L., Bao, J., 2015. An overview of predictors for intrinsically disordered proteins over 2010–2014. Int. J. Mol. Sci. 16, 23446–23462. https://doi.org/10.3390/ijms161023446

Lieberman-Aiden, E., Berkum, N.L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science (80-.). 326, 289–293. https://doi.org/10.1126/science.1181369

Liu, Y., Wang, X., Liu, B., 2019. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. Brief. Bioinform. 20, 330–346. https://doi.org/10.1093/bib/bbx126

Lovell, S.C., Davis, I.W., Arendall, W.B., De Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by Cα geometry: φ,ψ and Cβ deviation. Proteins Struct. Funct. Genet. 50, 437–450. https://doi.org/10.1002/prot.10286

Ma, H., Reyes-Gutierrez, P., Pederson, T., 2013. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. Proc. Natl. Acad. Sci. 110, 21048 LP – 21053. https://doi.org/10.1073/pnas.1319097110

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-guided human genome engineering via Cas9. Science (80-.). 339, 823–826. https://doi.org/10.1126/science.1232033

Mariamé, B., Kappler-Gratias, S., Kappler, M., Balor, S., Gallardo, F., Bystricky, K., 2018. Real-Time Visualization and Quantification of Human Cytomegalovirus Replication in Living Cells Using the ANCHOR DNA Labeling Technology. J. Virol. 92, 1–22. https://doi.org/10.1128/jvi.00571-18

McGinty, R.K., Tan, S., 2015. Nucleosome structure and function. Chem. Rev. 115, 2255–2273. https://doi.org/10.1021/cr500373h

Miyanari, Y., Ziegler-Birling, C., Torres-Padilla, M.-E., 2013. Live visualization of chromatin dynamics with fluorescent TALEs. Nat. Struct. Mol. Biol. 20, 1321–1324. https://doi.org/10.1038/nsmb.2680

Moffatt, B.A., Studier, F.W., 1987. T7 lysozyme inhibits transcription by T7 RNA

polymerase. Cell 49, 221–227. https://doi.org/10.1016/0092-8674(87)90563-0

Murray, H., Ferreira, H., Errington, J., 2006. The bacterial chromosome segregation protein Spo0J spreads along DNA from parS nucleation sites. Mol. Microbiol. 61, 1352–1361. https://doi.org/10.1111/j.1365-2958.2006.05316.x

Onufriev, A. V, Schiessel, H., 2019. The nucleosome: from structure to function through physics. Curr. Opin. Struct. Biol. 56, 119–130. https://doi.org/10.1016/j.sbi.2018.11.003

Osorio-Valeriano, M., Altegoer, F., Steinchen, W., Urban, S., Liu, Y., Bange, G., Thanbichler, M., 2019. ParB-type DNA Segregation Proteins Are CTP-Dependent Molecular Switches. Cell 179, 1512–1524. https://doi.org/10.1016/j.cell.2019.11.015

Ovejero, S., Bueno, A., Sacristán, M.P., 2020. Working on genomic stability: From the S-phase to mitosis. Genes (Basel). 11, 1–28. https://doi.org/10.3390/genes11020225

Passot, F., Calderon, V., Fichant, G., Lane, D., Pasta, F., 2012. Centromere binding and evolution of chromosomal partition systems in the Burkholderiales. J. Bacteriol. 194, 3426–3436. https://doi.org/10.1128/JB.00041-12

Passot, F., Nguyen, H.H., Dard-Dascot, C., Thermes, C., Servant, P., Espéli, O., Sommer, S., 2015. Nucleoid organization in the radioresistant bacterium Deinococcus radiodurans. Mol. Microbiol. 97, 759–774. https://doi.org/10.1111/mmi.13064

Pernthaler, J., Glöckner, F.O., Schönhuber, W., Amann, R., 2001. Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. Methods Microbiol. 30, 207–210. https://doi.org/10.1016/s0580-9517(01)30046-6

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell 152, 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Cai, L., Mcdonel, P., Garber, M., Guttman, M., 2018. Organization in the Nucleus. Cell 174, 744–757. https://doi.org/10.1016/j.cell.2018.05.024

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, 95–99. https://doi.org/https://doi.org/10.1016/S0022-2836(63)80023-6

Récamier, V., Izeddin, I., Bosanac, L., Dahan, M., Proux, F., Darzacq, X., 2014. Single cell

correlation fractal dimension of chromatin: A framework to interpret 3D single molecule super-resolution. Nucleus 5, 75–84. https://doi.org/10.4161/nucl.28227

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., Bonner, W.M., 1998. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J. Biol. Chem. 273, 5858–5868. https://doi.org/10.1074/jbc.273.10.5858

Rothkamm, K., Barnard, S., Moquet, J., Ellender, M., Rana, Z., Burdak-Rothkamm, S., 2015. DNA Damage Foci: Meaning and Significance. Environ. Mol. Mutagen. 56, 491–504. https://doi.org/https://doi.org/10.1002/em.21944

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. https://doi.org/10.1038/nprot.2010.5

Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., Bystricky, K., 2014. DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living Cells. PLoS Genet. 10. https://doi.org/10.1371/journal.pgen.1004187

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Serber, Z., Keatinge-Clay, A.T., Ledwidge, R., Kelly, A.E., Miller, S.M., Dötsch, V., 2001. High-Resolution Macromolecular NMR Spectroscopy Inside Living Cells. J. Am. Chem. Soc. 123, 2446–2447. https://doi.org/10.1021/ja0057528

Shaw, N., Tempel, W., Chang, J., Yang, H., Cheng, C., Ng, J., Rose, J., Rao, Z., Wang, B.-C., Liu, Z.-J., 2008. Crystal structure solution of a ParB-like nuclease at atomic resolution. Proteins 70, 263–267. https://doi.org/10.1002/prot.21641

Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. Proteins Struct. Funct. Bioinforma. 17, 355–362. https://doi.org/10.1002/prot.340170404

Soh, Y.M., Davidson, I.F., Zamuner, S., Basquin, J., Bock, F.P., Taschner, M., Veening, J.W., de Los Rios, P., Peters, J.M., Gruber, S., 2019. Self-organization of parS centromeres by the ParB CTP hydrolase. Science (80-.). 366, 1129–1133. https://doi.org/10.1126/science.aay3965

Song, D., Rodrigues, K., Graham, T.G.W., Loparo, J.J., 2017. A network of cis and trans interactions is required for ParB spreading. Nucleic Acids Res. 45, 7106–7117. https://doi.org/10.1093/nar/gkx271

Souaid, C., Bloyer, S., Noordermeer, D., 2018. Promoter–Enhancer Looping and Regulatory Neighborhoods, in: Nuclear Architecture and Dynamics. Academic Press, Cambridge, pp. 435–456. https://doi.org/10.1016/b978-0-12-803480-4.00018-1

Sparks, T.M., Harabula, I., Pombo, A., 2020. Evolving methodologies and concepts in 4D nucleome research. Curr. Opin. Cell Biol. 64, 105–111. https://doi.org/10.1016/j.ceb.2020.04.005

Straight, A.F., Belmont, A.S., Robinett, C.C., Murray, A.W., 1996. GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. Curr. Biol. 6, 1599–1608. https://doi.org/https://doi.org/10.1016/S0960-9822(02)70783-5

Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, R., Haas, J., Schwede, T., 2020. QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics 36, 1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113–130. https://doi.org/10.1016/0022-2836(86)90385-2

Sullivan, N.L., Marquis, K.A., Rudner, D.Z., 2009. Recruitment of SMC by ParB-parS Organizes the Origin Region and Promotes Efficient Chromosome Segregation. Cell 137, 697–707. https://doi.org/10.1016/j.cell.2009.04.044

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., Vale, R.D., 2014. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell 159, 635–646. https://doi.org/https://doi.org/10.1016/j.cell.2014.09.039

Tessarz, P., Kouzarides, T., 2014. Histone core modifications regulating nucleosome structure and dynamics. Nat. Rev. Mol. Cell Biol. 15, 703–708. https://doi.org/10.1038/nrm3890

Wako, T., Yoshida, A., Kato, J., Otsuka, Y., Ogawa, S., Kaneyoshi, K., Takata, H., Fukui, K., 2020. Human metaphase chromosome consists of randomly arranged chromatin fibres with up to 30-nm diameter. Sci. Rep. 10, 1–6. https://doi.org/10.1038/s41598-020-65842-z

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, 296–303. https://doi.org/10.1093/nar/gky427

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410.

https://doi.org/10.1093/nar/gkm290

Wójcik, S., Birol, M., Rhoades, E., Miranker, A.D., Levine, Z.A., 2018. Targeting the Intrinsically Disordered Proteome Using Small-Molecule Ligands, in: Methods in Enzymology. pp. 703–734. https://doi.org/10.1016/bs.mie.2018.09.036

Wu, B., Chao, J.A., Singer, R.H., 2012. Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. Biophys. J. 102, 2936–2944. https://doi.org/10.1016/j.bpj.2012.05.017

Wu, X., Mao, S., Ying, Y., Krueger, C.J., Chen, A.K., 2019. Progress and Challenges for Live-cell Imaging of Genomic Loci Using CRISPR-based Platforms. Genomics, Proteomics Bioinforma. 17, 119–128. https://doi.org/10.1016/j.gpb.2018.10.001

Yamaichi, Y., Niki, H., 2000. Active segregation by the Bacillus subtilis partitioning system in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 97, 14656–14661. https://doi.org/10.1073/pnas.97.26.14656

Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43, W174–W181. https://doi.org/10.1093/nar/gkv342

Zechiedrich, E.L., Khodursky, A.B., Cozzarelli, N.R., 1997. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli. Genes Dev. 11, 2580–2592. https://doi.org/10.1101/gad.11.19.2580

Zhang, Y., 2009. I-TASSER: Fully automated protein structure prediction in CASP8. Proteins Struct. Funct. Bioinforma. 77, 100–113. https://doi.org/10.1002/prot.22588

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9. https://doi.org/10.1186/1471-2105-9-40

Zink, D., Sadoni, N., Stelzer, E., 2003. Visualizing chromatin and chromosomes in living cells. Methods 29, 42–50. https://doi.org/10.1016/S1046-2023(02)00289-X

6 Appendix

6.1 Appendix 1: ParB3 Confidentiality Agreement

Information that could potentially lead to disclosure of sequence or species identity for the ParB3 protein was withheld in this thesis, bound by proprietary and confidential agreements. The ParB_INT system and related proteins were developed in the lab of Prof. Kerstin Bystricky, Paul Sabatier University Tolouse III, France.

6.2 Appendix 2: Abbreviations

aa	Amino acids
bp	Base pair
BSA	Bovine serum albumine
CIP	Calf interstinal phosphatase
CMV	Cytomegalovirus
DAPI	4′,6-diamidino-2-phenylindole
DMEM	Dulbecco’s modified eagle medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotides
DPBS	Dulbecco’s phosphate-buffered saline
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESCs	Embryonic stem cells
EtOH	Ethanol
FBS	Fetal bovine serum
GFP	Green fluorescent protein
His-tag	Poly-histidine affinity tag
HTH	Helix-turn-helix
IPTG	Isopropyl-beta-D-thiogalactoside
IT	I-TASSER
Kb	Kilobases
kDa	Kilodalton

LB	Lysogeny broth
LIF	Leukemia inhibitor factor
MOD	ModWeb
OD	Optical density
ParB1	Partitioning Protein B1
ParB3	Partitioning Protein B3
ParC	Partitioning Protein C
PCR	Polymerase chain reaction
PEI	Polyminethyleneimine
P/S	Penecillin/Streptomycin
PTM	Post-translational modifications
RMSD	Root mean square deviation
RPM	Revolutions per minute
SDS	Sodium dodecyl sulphate
SM	SWISS-MODEL
SNP	Single nucleotide polymorphisms
SOB	Super optimal broth
Srx	Sulfiredoxin
TAE	Tris-acetate-EDTA
TE-buffer	Tris and EDTA buffer
UV	Ultraviolet

6.3 Appendix 3: Supplementary Results

Table S1. Overview of individual protocol specifications for each purification step in the three experiments performed.

Step	Purification							
	1		2		3			
Protein	ParB3 Hu	ParB3 Hu + ParC	ParB3 Core	ParB3 Core +ParC	ParB3	ParB3 +ParC	pET16b	pEt16b +ParC
Pelleted culture volume[a]	50 ml	50 ml	250 ml	250 ml	250 ml	250 ml	250 ml	250 ml
Resuspension buffer A[b] (pellet g; ml buffer)	0.45; 4.5	0.53; 5.3	0.70; 7.0	0.93; 9.3	0.69; 6.9	0.9; 9.0	0.92; 9.2	0.89; 8.9
Sonication (amplitude 50%, 30 sec on/off)	6 minutes	6 minutes	8 minutes due to viscosity	6 minutes	6 minutes	6 minutes	8 minutes due to viscosity	6 minutes
10% PEI	7.5% (v/v)	7.5% (v/v)	8.1% (v/v)	8.1% (v/v)	8.1% (v/v)	8.1% (v/v)	8.1% (v/v)	8.1% (v/v)
Ammonium sulphate (80% saturation) 4 x sample volume	12 ml	16.8 ml	24.8 ml	32.8 ml	20.8 ml	29.2 ml	33.2 ml	30.8 ml
Resuspension (1:1 buffer and sample volume)	Buffer B (100 mM KCl)	Buffer B (100 mM KCl)	Buffer B (100 mM KCl)	Buffer B (100 mM KCl)[c]	Buffer A (25 mM KCl)	Buffer A (25 mM KCl)	Buffer A (25 mM KCl)	Buffer A (25 mM KCl)
Centrifugation	-	-	-	-	23,000 x g, 10 min, 4°C	23,000 x g, 10 min, 4°C	23,000 x g, 10 min, 4°C	23,000 x g, 10 min, 4°C
Dialysis	Buffer B (100 mM KCl) 2x500 ml + 1L ON	Buffer B (100 mM KCl) 3x500 ml	Buffer B (100 mM KCl) 2x500 ml + 1L ON	Buffer B (100 mM KCl) 3x500 ml	Buffer B (150 mM KCl) 3x500 ml	Buffer B (150 mM KCl) 3x500 ml	Buffer B (150 mM KCl) 3x500 ml	Buffer B (150 mM KCl) 3x500 ml

[a] Induced cultures pelleted and frozen after expression experiment

[b] Refer to appendix 5 for buffer recipes

[c] Small spillage may reduce yield

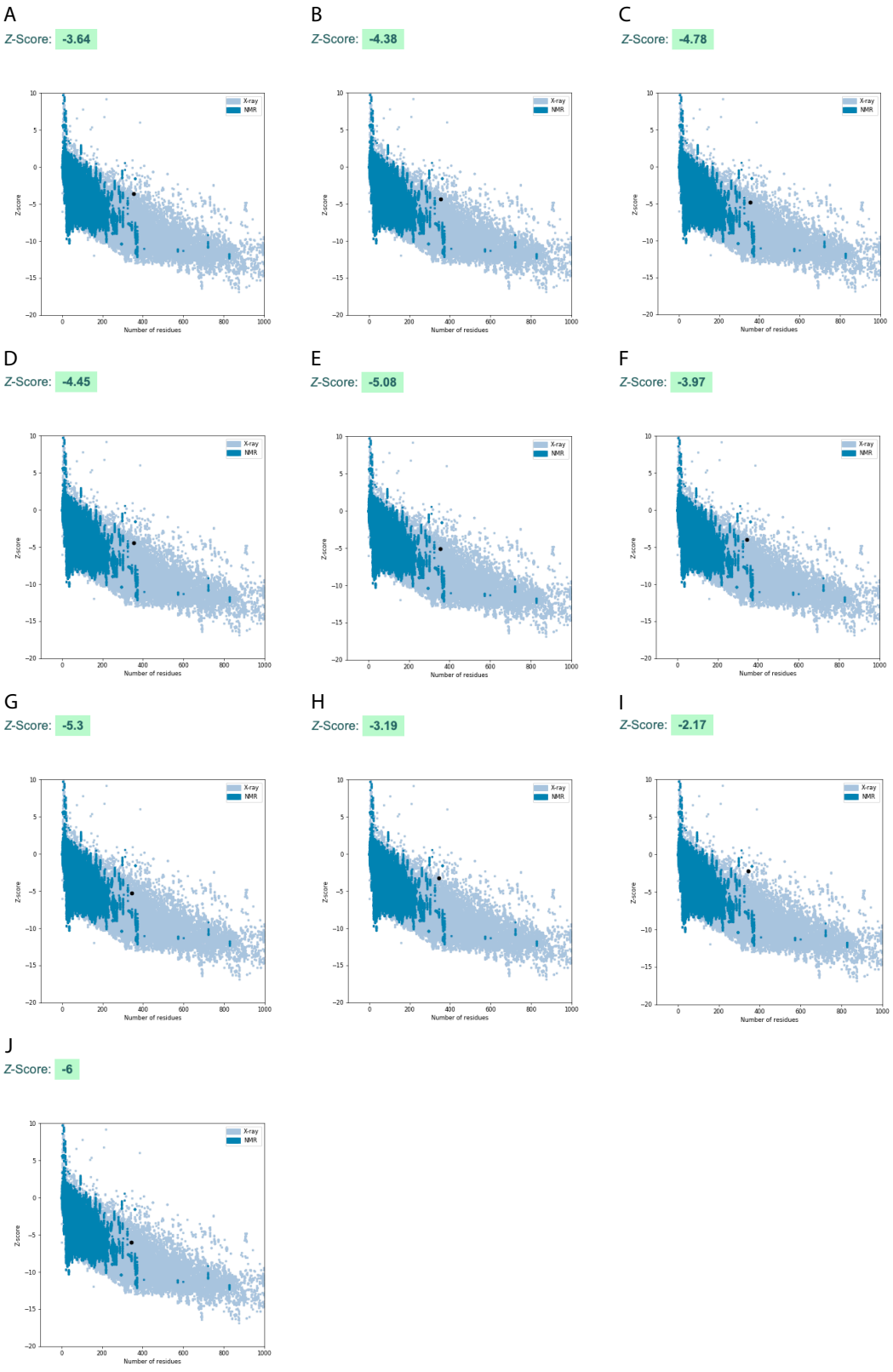


Figure S1. Energy Z-scores for IT models 1-5 for ParB1 (A-E) and ParB3 (F-J). The score in green is indicated by a black dot in the graph where of number of aa is plotted on the x-axis against Z-score on the y-axis. The dark blue field shows the Z-score of protein structures solved by NMR and the light blue field shows proteins solved by x-ray diffraction, sorted according to size. The analysis was generated by ProSA.

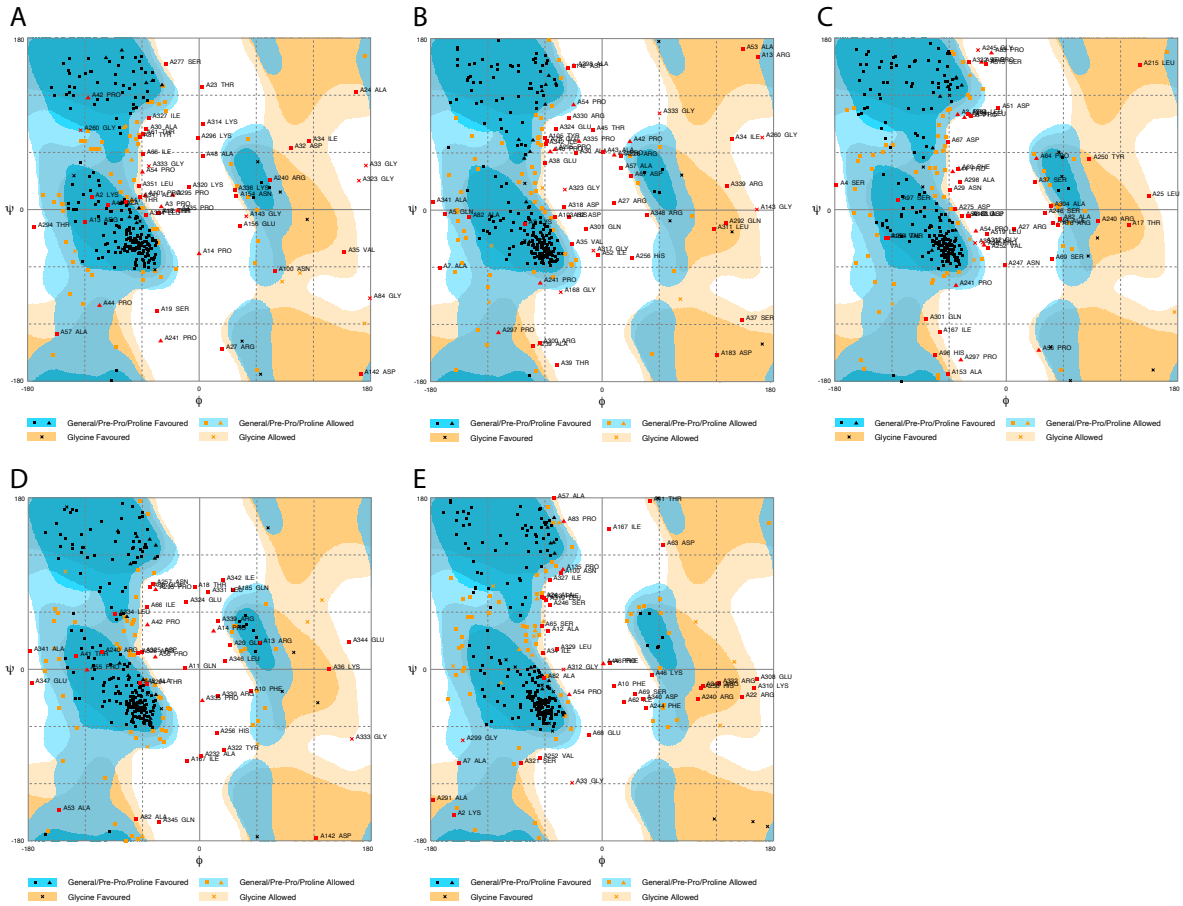


Figure S2. Ramachandran plots generated for IT models 1-5 (A-E) for ParB1. Aa residues can be located in either favored (dark blue and orange), allowed (light blue and orange), or outlier (white). Proline aa are indicated by triangles, glycine with x, and all other aa with a square. Aa located in favored region are colored black, in allowed region orange, and outlier aa are colored red. They are sorted based on how energetically favorable the dihedral angles are. The plots are generated by Rampage. Plots generated for ParB3 models can not be disclosed.

Table S2. Constructed clones verified with restriction enzyme cutting.

Clone	Restriction enzyme(s)	Band sizes (bp)
pDEST14_ParB1	EcoRI	1427 + 4263
pDEST17_ParB1	EcoRI	1427 + 4336
pDEST566_ParB1	SalI + XbaI	1864 + 5980
pGEX2T1GW_ParB1	BtgI	1387 + 4726
pDEST14_ParB3	PvuII	2572 + 3094
pDEST17_ParB3	PvuII	2645 + 3094
pDEST566_ParB3	PstI + XbaI	2002 + 5818
pGEX2T1GW_ParB3	PstI	1268 + 4821
pDEST14_ParB3Core	PvuII	2380 + 2845
pDEST17_ParB3Core	PvuII	2452 + 2876
pDEST566_ParB3Core	PstI + XbaI	1810 + 5599
pGEX2T1GW_ParB3Core	PstI	1049 + 4629
pET28a_ParC	EcoRV	1898 + 4841
pET15b_ParB3Hu	PstI	1343 + 5425
pDasher_ParB3Hu	BstXI	722 + 5005
pOriGFP_CMVmut_ParB1	EcoRI	1647 + 4294

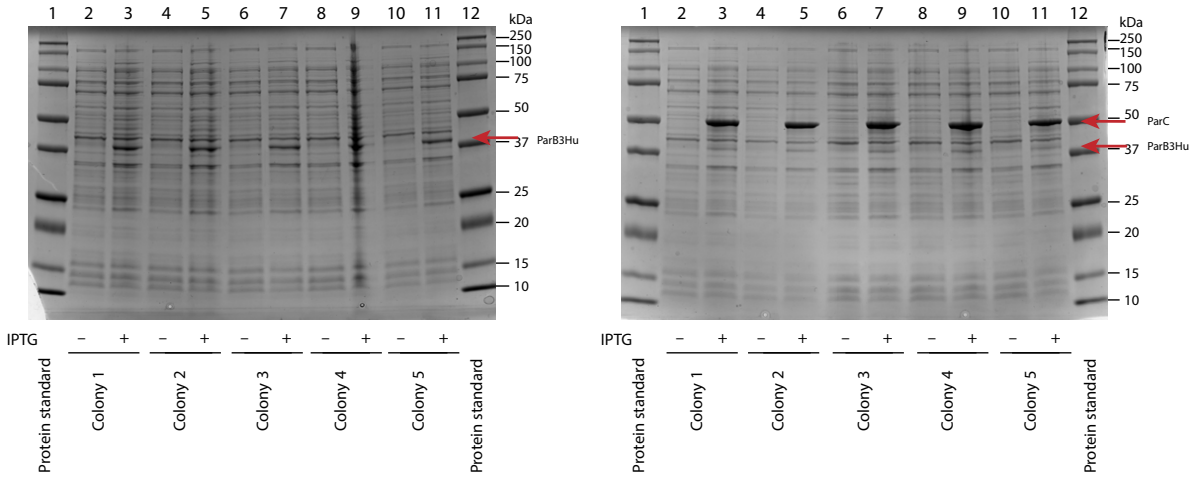


Figure S3. Overexpression of individual pET15b_ParB3Hu +/- ParC colonies in 50 ml cultures of BL21 cells. pET15b_ParB3Hu (~40 kDa) and pET28a_ParC (57 kDa). Samples collected from uninduced cultures (-) and induced cultures (+) analyzed by 12% SDS-PAGE. Bands of expected protein sizes are indicated by red arrows in samples – ParC **(A)** and samples + ParC **(B)**. Molecular weight proteins standards lanes 1 and 12.

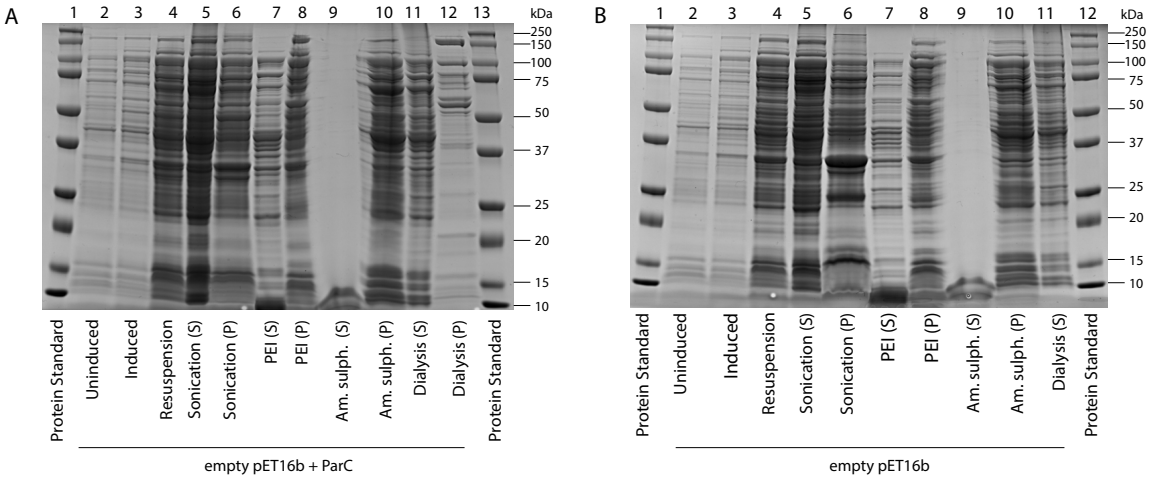


Figure S4. Purification of empty pET16b control +/- ParC. Fractions from each step of the purification process was analyzed by 12% SDS-PAGE. The fractions on the gel were labelled by the steps in the protocol. (S) indicates a supernatant sample and (P) a pellet sample. Induced and uninduced samples from the expression experiment are shown for comparison (lane 2 and 3). Molecular weight protein standards for gel A (lane 1 and 13) and B (lane 1 and 12). No pellet was present after dialysis for empty pET16b – C. PageBlue protein staining solution was used.

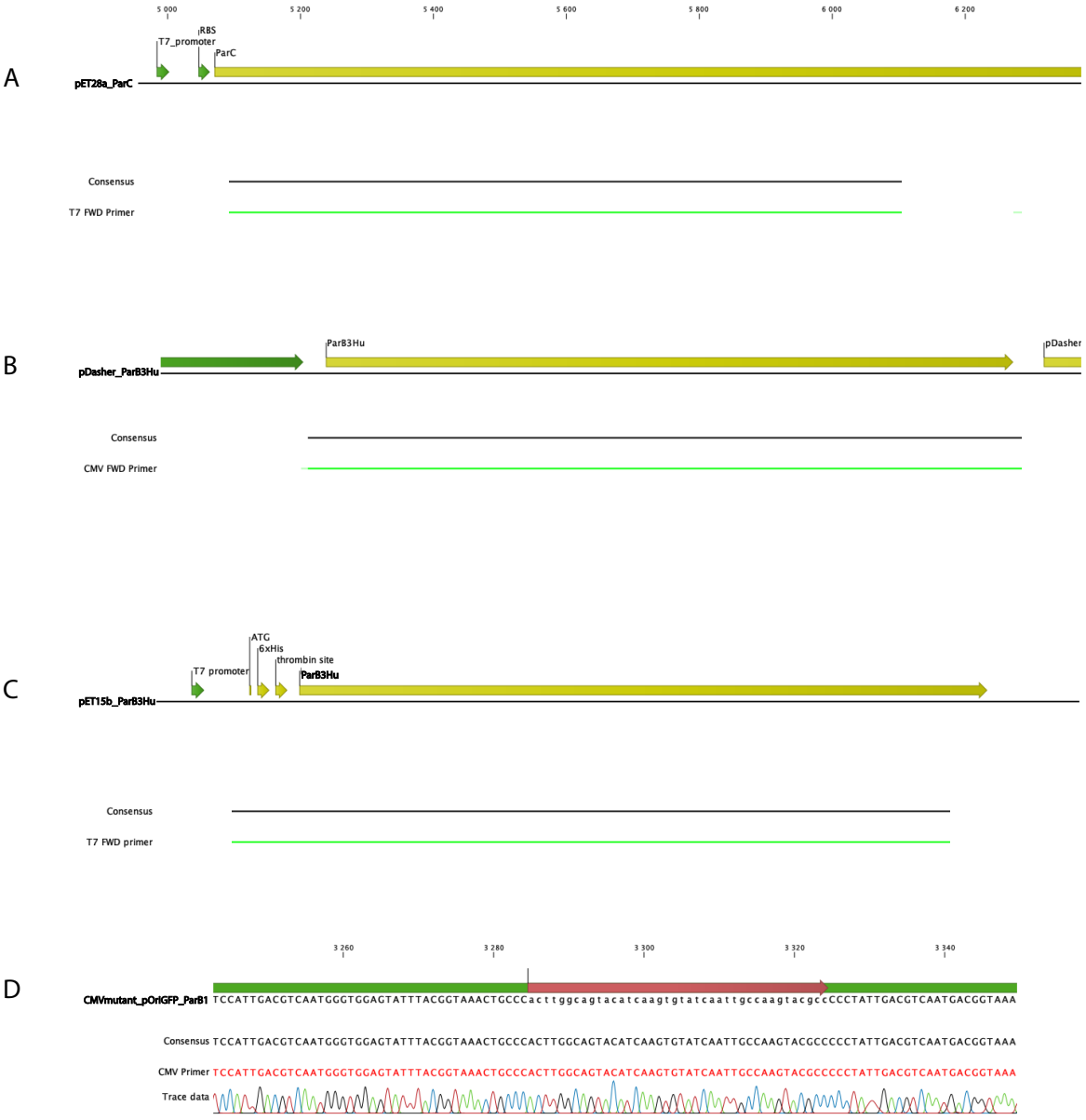


Figure S5. Validation of constructed clones and mutation sites by sequencing. Correctly integrated sequences are validated by a green line where the read aligns with the reference sequence. **(A)** pET28a_ParC checked with T7 FWD primer, **(B)** pDasher_ParB3Hu checked with T7 FWD primer, **(C)** pET15b_ParB3Hu checked with MBP FWD primer, and **(D)** pOriGFP_ParB1 mutation site checked with CMV REV primer. Further sequencing needs to be performed to validate the full plasmid.

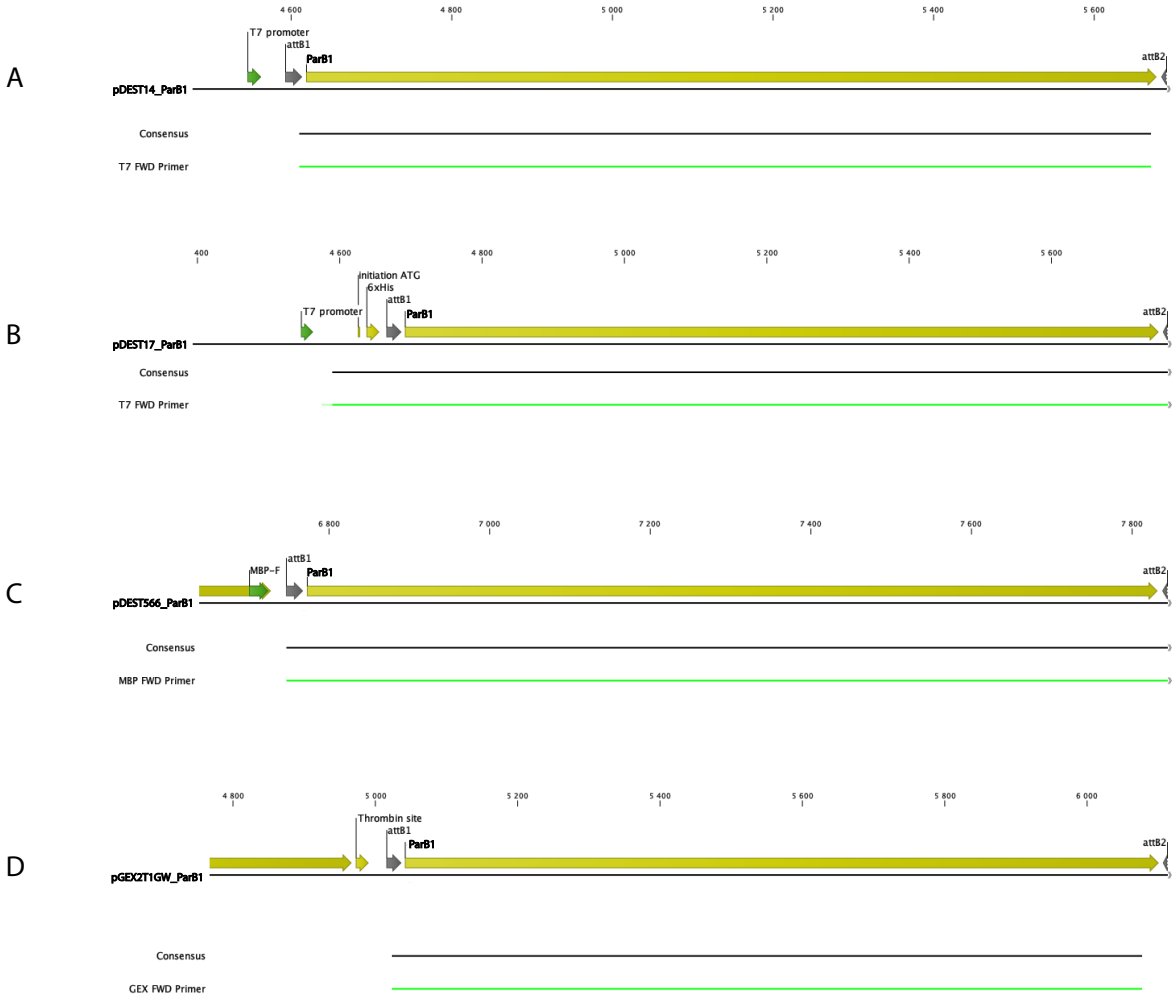


Figure S6. Validation of constructed ParB1 clones by sequencing. Correctly integrated sequences are validated by a green line where the read aligns with the reference sequence. **(A)** pDEST14_ParB1 checked with T7 FWD primer, **(B)** pDEST17_ParB1 checked with T7 FWD primer, **(C)** pDEST566_ParB1 checked with MBP FWD primer, and **(D)** pGEX2T1GW_ParB1 checked with GEX FWD primer.

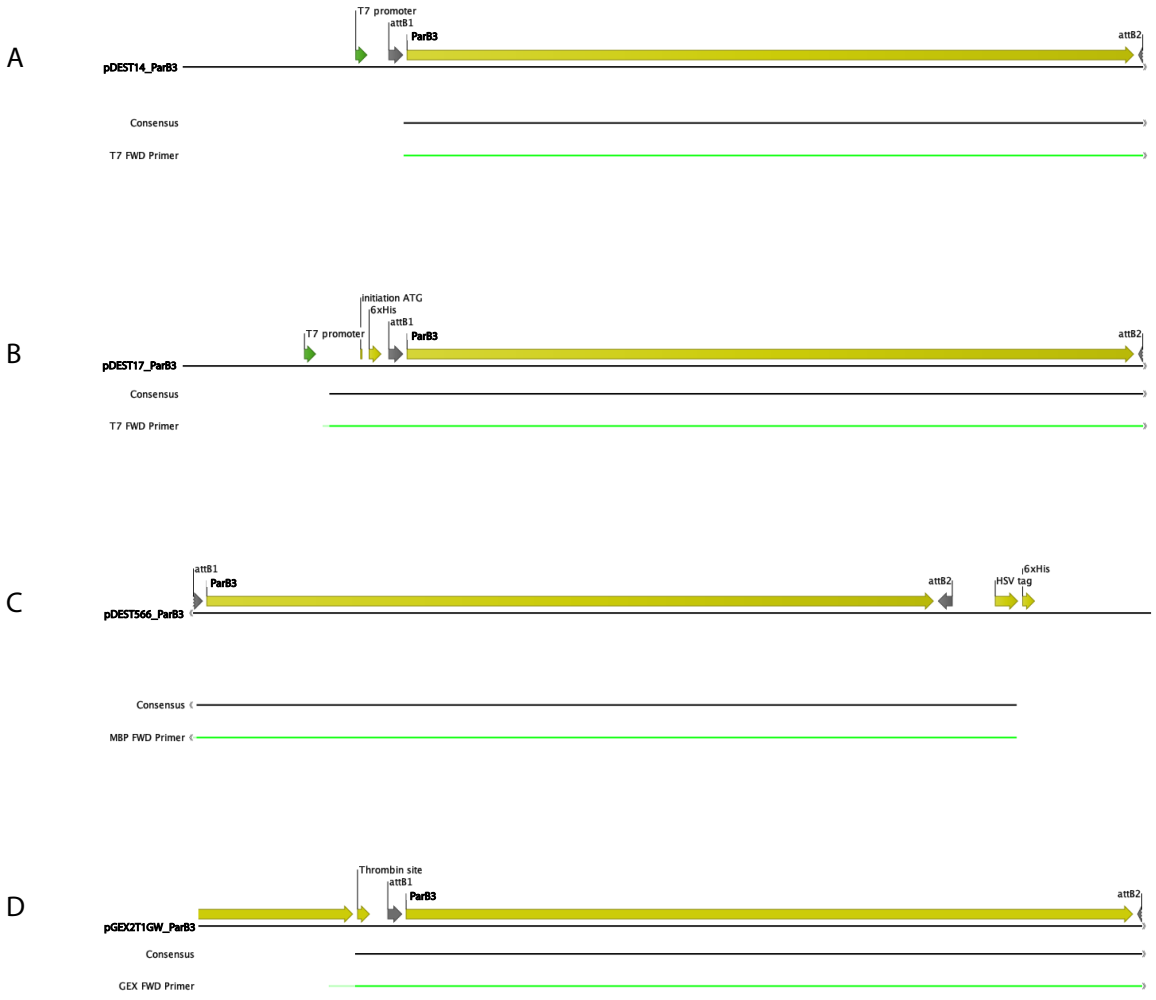


Figure S7. Validation of constructed ParB3 clones by sequencing. Correctly integrated sequences are validated by a green line where the read aligns with the reference sequence. **(A)** pDEST14_ParB3 checked with T7 FWD primer, **(B)** pDEST17_ParB3 checked with T7 FWD primer, **(C)** pDEST566_ParB3 checked with MBP FWD primer, and **(D)** pGEX2T1GW_ParB3 checked with GEX FWD primer.

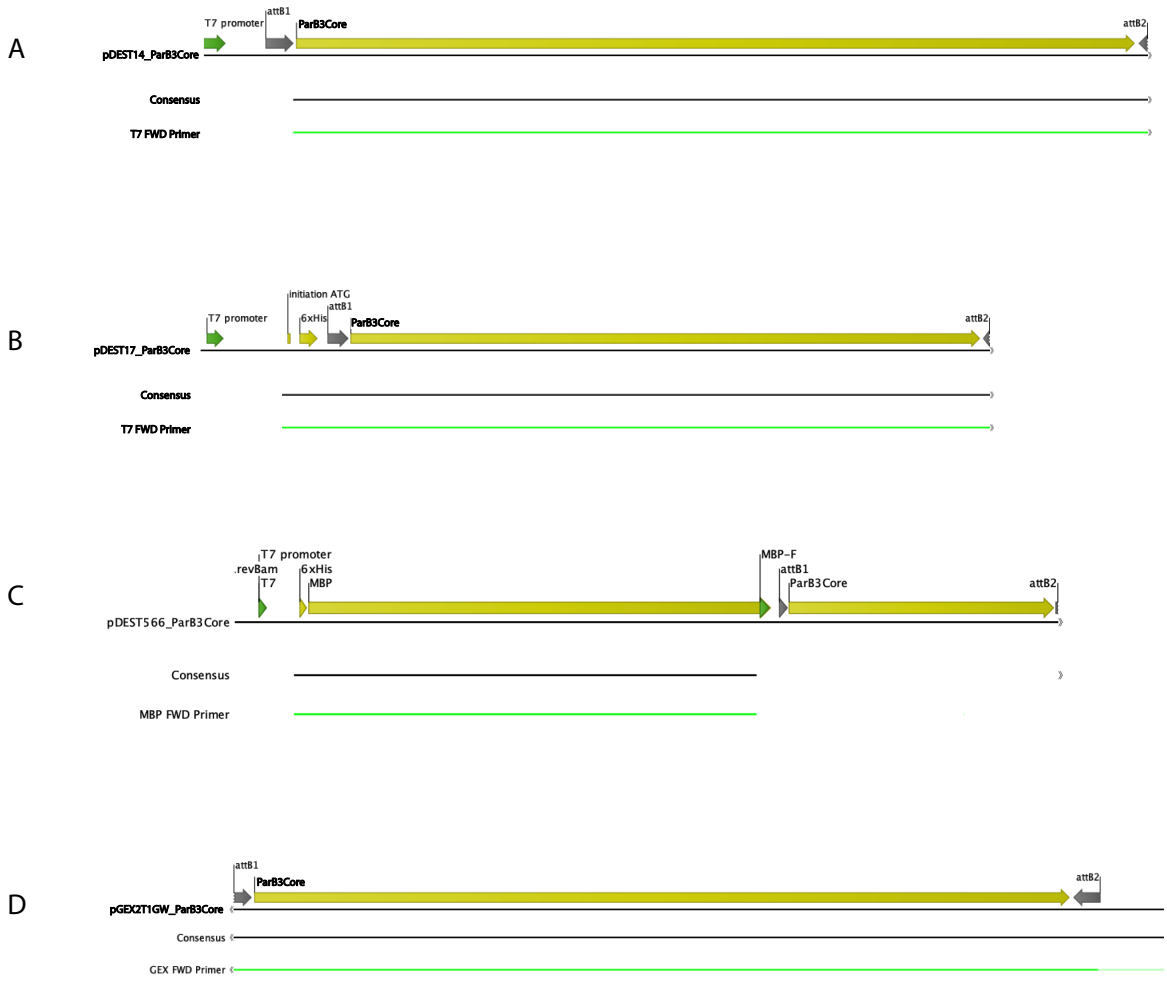


Figure S8. Validation of constructed ParB3Core clones by sequencing. Correctly integrated sequences are validated by a green line where the read aligns with the reference sequence. **(A)** pDEST14_ParB3Core checked with T7 FWD primer. **(B)** pDEST17_ParB3Core checked with T7 FWD primer. **(C)** pDEST566_ParB3Core checked with MBP FWD primer and needs further sequencing from other direction to span the gene sequence. **(D)** pGEX2T1GW_ParB3Core checked with GEX FWD primer, and needs further sequencing to confirm insert in right vector.

6.4 Appendix 4: Plasmid Maps

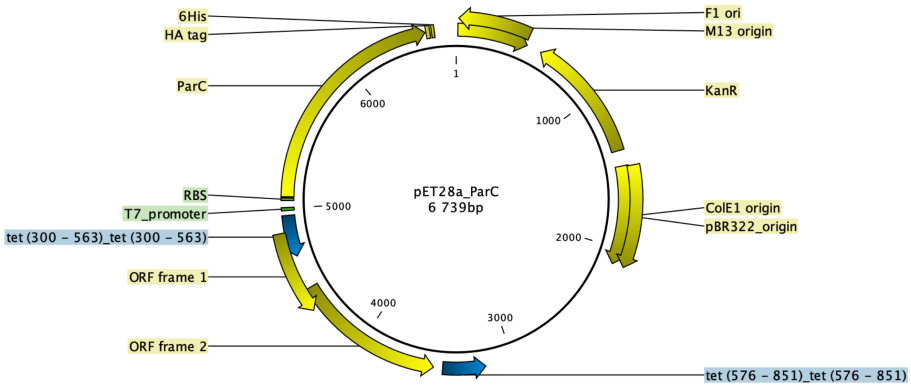


Figure S9. Constructed clone pET28a_ParC.

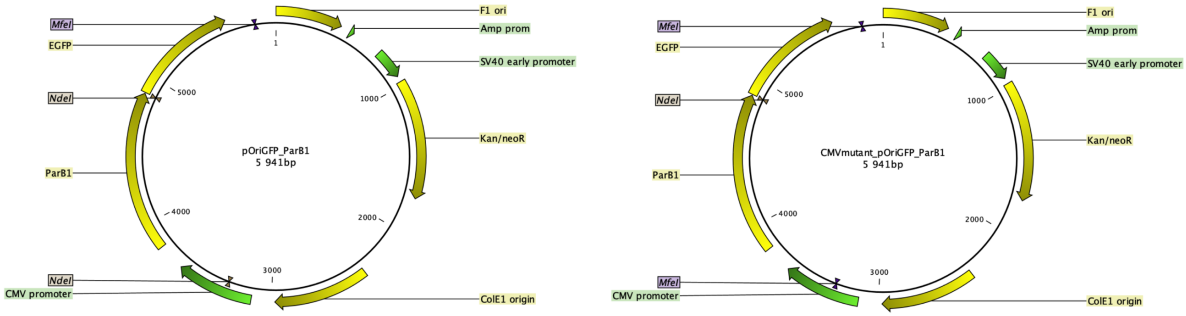


Figure S10. Plasmids from mutagenesis experiment. Original pOriGFP_ParB1 plasmid (left) with NdeI restriction cutting site in the CMV promoter and pOriGFP_ParB1 CMV mutant (right) with inserted MfeI site in the CMV promoter.

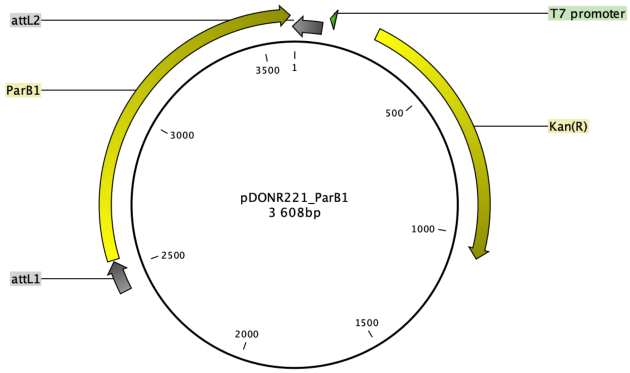


Figure S11. Constructed entry clone for gateway cloning of ParB1 in pDONR221.

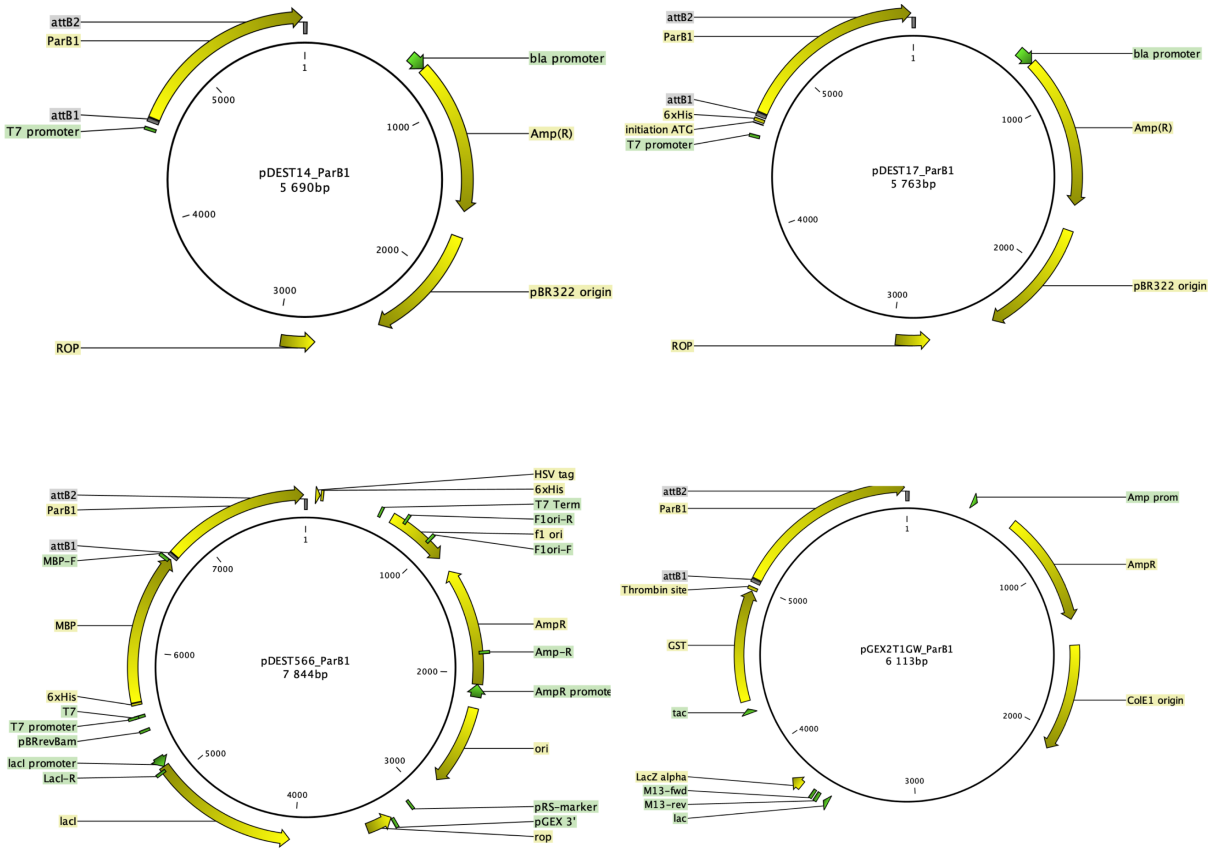


Figure S12. Constructed expression clones ParB1 in pDEST14, pDEST17, pDEST566, and pGEX2T1GW.

6.5 Appendix 5: Recipes

General:

0.5 M EDTA pH 8.0

93.06 g $Na_2EDTA•2H_2O$ (372.24 g/mol)

dH_2O to 400 ml for dissolving

pH was adjusted to 8.0 with 10 M NaOH

dH_2O to 500 ml

The solution was autoclaved

1 M Tris-HCl pH 8.0

60.55 g Tris base (121.14 g/mol)

dH_2O to 300 ml

pH was adjusted to 8.0 with 36% HCl (solution was cooled to room temperature before final adjustments)

dH_2O to 500 ml

The solution was autoclaved

0.5 M Tris-HCl pH 6.8

30.28 g Tris base (121.14 g/mol)

dH_2O to 300 ml

pH was adjusted to 6.8 with 36% HCl (solution was cooled to room temperature before final adjustments)

dH_2O to 500 ml

The solution was autoclaved

1X TE-buffer pH 8.0

10 mM 1 M Tris-HCl pH 8.0

1 mM 0.5 M EDTA pH 8.0

dH_2O to 250 ml

50% Glycerol

100 ml 100% glycerol

100 ml Milli Q
The solution was autoclaved

50X TAE
242 g Tris base (121.14 g/mol)
57 ml acetic acid (1M)
100 ml 0.5 M EDTA pH 8.0 (50 mM)
dH_2O to 1000 ml

10M NaOH
40 ml ddH_2O
40 g NaOH (39,997 g/mol) was added
The solution was cooled down on ice
MQ H_2O to 100 ml

5M NaCl
146.1g NaCl (58.44 g/mol)
MQ H_2O to 500 ml

70% Ethanol
70 ml ethanol
30 ml dH_2O

Solutions for bacterial culture:

LB medium
10 g tryptone
5 g yeast extract
10 g NaCl (58.44 g/mol)
dH_2O to 800 ml
pH was adjusted to 7.2 with 10 M NaOH
dH_2O to 1000 ml
The solution was autoclaved immediately

LB agar plates

400 ml LB medium

6.0 g agar

The solution was autoclaved immediately

The solution was cooled to 50-55°C in a water bath

If needed, antibiotics were added

Approximately 20 ml of solution was spread on one petri dish in a swirling motion to get bubbles to the edges of the plate and then was dried in a laminar flow hood and stored at 4°C

SOB medium

20 g tryptone

5 g yeast extract

0.5 g NaCl (58.44 g/mol)

dH_2O to 800 ml

1 ml 2.5 M KCl

pH was adjusted to 7.0

dH_2O to 1000 ml

The solution was autoclaved

10 ml of sterile filtered 2 M $MgSO_4$ was added prior to use

Transformation buffer

3.0 g PIPES (302.4 g/mol)

2.2 g $CaCl_2$ (147.02 g/mol)

18.6 g KCl (74.56 g/mol)

dH_2O to 800 ml

pH was adjusted to 6.7 with 10 M KOH

10.9 g $MnCl_2$ (197.9 g/mol)

dH_2O to 1000 ml

The solution was filter sterilized

Solutions for mammalian culture:

OS25 culturing medium

500 ml GMEM bhk12 (Gibco)

500 U P/S (Gibco)
50 ml FBS (Capricorn Scientific)
5 ml 100 mM sodium pyruvate (Gibco)
5 ml 100X NEAA (Gibco)
3.5 µl 14.3 M β-mercaptoethanol
1000 U LIF

0.1% gelatin
0.2 g gelatin
200 ml 1X D-PBS (Gibco)
The solution was autoclaved

Solutions for SDS-PAGE:

1 M DTT
7.72 g DTT (154.3 g/mol)
dH_2O to 50 ml

4X SDS loading buffer
200 mM 0.5 M Tris-HCl pH 6.8 (20 ml)
4 g SDS (228.37 g/mol)
20 ml 100% glycerol (40%)
0.3 mg bromphenol blue (669.96 g/mol)
dH_2O to 50 ml
10% 1 M DTT was freshly added prior to use

10X SDS running buffer
150.1 g glycine (75.07 g/mol)
30.28 g Tris base (121.14 g/mol)
10 g SDS (299.38 g/mol)
dH_2O to 1000 ml

Fixing solution (for SDS gel prior to staining)
125 ml 2-propanol (25%)

50 ml acetic acid (10%)

dH_2O to 500 ml

Solutions for fixation and mounting of slides:

4% paraformaldehyde

In fume hood

20 g paraformaldehyde (30.03 g/mol)

1X PBS to 500 ml

Added to a 50°C water bath and shook well to dissolve

The solution was cooled to room temperature, and checked for precipitate

1M NH_4Cl

5.35 g NH_4Cl (53,49 g/mol)

dH2O to 1000 ml

The solution was filter sterilized

Solutions for protein expression and purification:

LB Rich Medium for Protein Expression

16 g tryptone

10 g yeast extract

5 g NaCl (58.44 g/mol)

dH_2O to 800 ml

pH was adjusted to 7.2 with 10 M NaOH

dH_2O to 1000 ml

The solution was autoclaved immediately

2 M $MgSO_4$

123,25 g $MgSO_4$ (246.49 g/mol)

dH_2O to 250 ml

The solution was filter sterilized

1 M IPTG

1 g IPTG (238.30 g/mol)
Sterile Milli Q was added until total volume 4.2 ml

1M Tris-HCl pH 7.4
60.55g Tris base (121.14 g/mol)
dH_2O to 300 ml
pH was adjusted to 7.4 with 36% HCl (solution was cooled to room temperature before final adjustments)
dH_2O to 500 ml
The solution was autoclaved

1M HEPES pH 7.4
119.16g HEPES (238.32 g/mol)
MQ H_2O to 300 ml
5-6 pellets of NaOH (39,997 g/mol) were added
pH was adjusted to 7.4 with 10M NaOH
The solution was filter sterilized and stored at 4°C

2.5M KCl
74.55g KCl (74,5513 g/mol)
dH_2O to 300 ml
The solution was heated until homogenous
dH_2O to 400 ml
The solution was autoclaved

Buffer A
20 mM Tris-HCl pH 7.4
10% glycerol
25 mM KCl
MQ H_2O to 500 ml
1% complete protease inhibitor tablet (Roche) dissolved in 500 µl ddH_2O was added fresh before use

Buffer B

20 mM HEPES pH 7.4
10% glycerol
150 mM KCl
0.1% protease inhibitors were added fresh before use (1mg/ml aproptinin, pepstatin, leupeptin, PMSF)

Buffer C
20 mM HEPES pH 7.4
10% glycerol
150 mM NaCl
0.1% protease inhibitors were added fresh before use (1mg/ml aproptinin, pepstatin, leupeptin, PMSF)

20 ml 10% PEI pH 7.9
4g 50% PEI
MQ H_2O to 17g
The flask was shaken vigorously to dissolve PEI
pH was adjusted to 7.9 with HCl
MQ H_2O to 20g
The solution was filter sterilized

100% saturated ammonium sulphate at 25°C
100 ml dH_2O
76.68g $(NH4)_2SO4$ (132,14 g/mol) was added little by little while stirring with a magnet stirrer until salt was dissolved
pH was adjusted to 6.9 with 10 M NaOH
The solution was filter sterilized

His-tag equilibration buffer
10 mM imidazole
1 ml 10X dPBS
MQ H_2O to 10 ml
1% complete protease inhibitor tablet (Roche) dissolved in 500 µl ddH_2O was added fresh before use

His-tag wash buffer

25 mM imidazole

1 ml 10X dPBS

MQ H_2O to 10 ml

1% complete protease inhibitor tablet (Roche) dissolved in 500 µl ddH_2O was added fresh before use

His-tag elution buffer

250 mM imidazole

1 ml 10X dPBS

MQ H_2O to 10 ml

1% complete protease inhibitor tablet (Roche) dissolved in 500 µl ddH_2O was added fresh before use

6.6 Appendix 6: Materials

Table S3. Overview of materials used in this project.

Agarose Gel Electrophoresis		
1 Kb Plus DNA Ladder	Invitrogen	#10787026
Gel Loading Dye, Purple (6X)	New England Biolabs	#B70224S
GelRed® Nucleic Acid Gel Stain	Biotium	#41003
Antibiotics		
Ampicillin sodium salt	Sigma-Aldrich	#A9518
Chloramphenicol	Sigma-Aldrich	#C1919
Kanamycin sulfate	Sigma-Aldrich	#K1377
Cell Culturing		
2-Mercaptoethanol	Sigma-Aldrich	#M6250
10X Trypsin-EDTA solution	Sigma-Aldrich	#T4174
Dulbecco's Phosphate Buffered Saline (1X DPBS)	Gibco™	#14190094
EVE™ cell counting slides	NanoEnTek	#734-2676
Fetal Bovine Serum (FBS)	Capricorn Scientific	#FBS-11A
Gelatin from porcine skin	Sigma-Aldrich	#G1890
GMEM media	Gibco™	#21710082
Human Recombinant LIF	Homemade by B. Nadratowska Wesolowska	N/A
MEM Non-Essential Amino Acids Solution (100X)	Gibco™	#11140050
Penicillin-Streptomycin (10,000 U/mL)	Gibco™	#15140122
Trypan Blue Stain (0.4%)	Invitrogen	#T10282
Diverse		
Diluent A	New England Biolabs	#B8001S
HisTrap™ High Performance	Sigma-Aldrich	#GE17-5248-02
Enzymes		
Alkaline Phosphatase, Calf Intestinal (CIP)	New England Biolabs	#M0290
BamHI	New England Biolabs	#R0136S
BbsI	New England Biolabs	#R0539S
BstXI	New England Biolabs	#R0113
BtgI	New England Biolabs	#R0608S

DpnI	New England Biolabs	#R0176
EcoRI-HF	New England Biolabs	#R3101S
EcoRV	New England Biolabs	#R0195S
KpnI	New England Biolabs	#R0142S
MfeI	New England Biolabs	#R0589S
MspI-HF	New England Biolabs	#R0106S
NcoI	New England Biolabs	#R0193S
NdeI	New England Biolabs	#R0111S
NheI	New England Biolabs	#R0131S
PstI	New England Biolabs	#R0140
PvuII	New England Biolabs	#R0151S
SalI	New England Biolabs	#R0138
XbaI	New England Biolabs	#R0145
Gateway® BP Clonase® II enzyme mix	Invitrogen	#11789020
Gateway® LR Clonase® II enzyme mix	Invitrogen	#11791020
T4 DNA Ligase	New England Biolabs	# M0202S
Imaging		
Zeiss LSM 710 Confocal Microscope	Zeiss	N/A
Olympus CK40 Phase Contrast Tissue Culture Microscope	Olympus	N/A
VECTASHIELD® Antifade Mounting Medium with DAPI	Vector Labs	#H-1000
Instruments		
Bioruptor UCD-200	Diagenode	N/A
LABSONIC® M	B. Braun Biotech International	N/A
Centrifuge Avanti™ J-25	Beckman	N/A
Universal Refrigerated Centrifuge Model 5930	Kubota	N/A
ChemiDoc™ Touch Imaging System	Bio-Rad	N/A
Countess™ automated cell counter	Thermo Fischer Scientific	N/A
GelDoc XR+ System	Bio-Rad	N/A
NanoDrop™ 2000c Spectrophotometer	Thermo Scientific	N/A
SevenCompact pH meter S220	Mettler Toledo	N/A
Ultrospec 3100 Pro UV/Visible Spectrophotometer	GE Healthcare / Amersham Biosciences	N/A
Veriti™ 96-Well Thermal Cycler	Applied Biosystems	N/A

Kits		
NucleoBond Xtra Maxi	Macherey-Nagel	#740414
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	#740609
NucleoSpin Plasmid	Macherey-Nagel	#740588
PCR		
5x HOT FIREPol® Blend Master Mix Ready to Load	Solis BioDyne	#04-25-02015
5X Q5® High GC Enhancer	New England Biolabs	#B9028
5X Q5 Reaction Buffer	New England Biolabs	#B9027
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	#D2650
dNTP Set (100 mM)	Invitrogen	#10297018
PfuUltra II Fusion High-fidelity DNA Polymerase	Agilent Technologies	#600672
PfuUltra II Fusion Reaction Buffer	Agilent Technologies	#600672
Q5® High-Fidelity DNA Polymerase	New England Biolabs	#M0491L
Plasmids		
OriGFP_ParB1	Gifted by Kerstin Bystricky	-
pDasher_ParB3	Gifted by Kerstin Bystricky	-
pUC57-CAG	GenScript	-
pUC57-ParB1Hu	GenScript	-
pUC57-ParB3Hu	GenScript	-
pUC57-ParC	GenScript	-
Mammalian cells		
OS25_INT3 mESC	Generated by Beata Nadratowska-Wesolowska	-
Protein expression		
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma-Aldrich	#I5502
PageBlue™ Protein Staining Solution	Thermo Scientific	#24620
Semi-micro cuvette, PS	Sarstedt	#67.742
Protein purification		
Quick Start™ Bradford 1x Dye Reagent	Bio-Rad	#500-0205
Pierce™ Saturated Ammonium Sulfate Solution	Thermo Scientific	#45216

cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche	#4693132001
HisPur™ Ni-NTA Purification Kit, 0.2 ml	Thermo Scientific	#88227
Poly(ethyleneimine) solution ~50% in H_2O	Sigma-Aldrich	#03880
Spectrum™ Spectra/Por™ 1 RC Dialysis Membrane Tubing 6000 to 8000 Dalton MWCO, 20.4 mm	Thermo Fisher Scientific	#11415849
Spectrum™ Spectra/Por™ 1 RC Dialysis Membrane Tubing 6000 to 8000 Dalton MWCO, 6.4 mm	Thermo Fisher Scientific	#11495839
Whatman Grade 17 Chr Cellulose Chromatography Papers	VWR	#588-3010
Reaction Buffers		
10X NEB CutSmart® Buffer	New England Biolabs	#B7204S
10X NEBuffer™ 1.1	New England Biolabs	#B7201S
10X NEBuffer™ 2.1	New England Biolabs	#B7202S
10X NEBuffer™ 3.1	New England Biolabs	#B7203S
T4 DNA Ligase Reaction Buffer	New England Biolabs	#B0202S
SDS-PAGE		
12% Criterion™ TGX™ Precast Midi Protein Gel, 12+2 well, 45 µl	Bio-Rad	#5671043
12% Criterion™ TGX™ Precast Midi Protein Gel, 18 well, 30 µl	Bio-Rad	# 5671044
Precision Plus Protein™ Dual Color Standards	Bio-Rad	#1610374
Software		
Fiji/ImageJ version 2.0.0-rc-69/1.52p	(Schindelin et al., 2012)	N/A
Image Lab 6.0.1 For Windows	Bio-Rad	N/A
Image Lab Touch 2.3	Bio-Rad	N/A
NanoDrop 2000 / 2000c Operating Software, version 1.6	Thermo Scientific	N/A
Software - bioinformatics		

ConSurf	Ashkenazy H., Abadi S., Ben-Tal N, Celniker G., Chay O., Erez E., Glaser F., Martz E., Mayrose I., Nimrod G., Pupko T.	N/A
CLC Main Workbench	Qiagen	N/A
ITOL	Ivica Letunic, Peer Bork	N/A
I-TASSER	Yang Zhang Lab, University of Michigan	N/A
MAFFT	Kazutaka Katoh	N/A
Modeller	Andrej Sali Lab, University of California San Fransisco	N/A
ProSA-web	Manfred J. Sippl	N/A
PsiPred	Bioinformatics Group, University College London	N/A
Rampage	Crystallography and Bioinformatics Group, University of Cambridge	N/A
SCSF Bio	Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology	N/A
Swiss-Model	Biozentrum, University of Basel	N/A
UCSF Chimera	Resource for Biocomputing, Visualization, and Informatics, University of California, San Fransisco	N/A
Transfection and fixation		
DreamFect Stem 5000 Transfection Reagent	Nordic Diagnostica	#OZ-DF45000
Opti-MEM™ I Reduced Serum Medium	Gibco™	#31985062
Polyethylenimine (PEI), Branched	Polysciences	#23966-2

6.7 Appendix 7: Oligonucleotides

Table S4. Overview of nucleotides used in PCR reactions in this project.

Primer Label	Sequence (5'→3')	Purpose	Design
ParB1OrigattB_F	GGGGacaagtttgtacaaaaaagcaggct CCATGAAACCCTCCCAATTTGC	Primer for Gateway cloning	Ragnhild Eskeland
ParB1OrigattB_R	AGCGGATGCTGTTGTCGGAG taaacccagctttcttgtacaaagtggtCCCC	Primer for Gateway cloning	Ragnhild Eskeland
ParB3OrigattB_F*	----acaagtttgtacaaaaaagcaggct	Primer for Gateway cloning	Ragnhild Eskeland
ParB3OrigattB_R*	taaacccagctttcttgtacaaagtggt----	Primer for Gateway cloning	Ragnhild Eskeland
ParB3StruccoreattB_F*	----acaagtttgtacaaaaaagcaggct	Primer for Gateway cloning	Ragnhild Eskeland
ParB3StruccoreattB_R*	taaacccagctttcttgtacaaagtggt----	Primer for Gateway cloning	Ragnhild Eskeland
CMV-NdeI-to-MfeI-FWD	acttggcagtacatcaagtgtatcAATT gccaagtacgcc	Mutagenesis of NdeI site to MfeI site	Naima Azouzi
CMV-NdeI-to-MfeI-REV	acttggcagtacatcaagtgtatcAATT gccaagtacgcc	Mutagenesis of NdeI site to MfeI site	Naima Azouzi

*Protein specific primer sequences not shown

6.8 Appendix 8: PCR Cycling Conditions

Table S5. Cycling conditions for mutagenesis reaction.

Temperature (°C)	Time	Cycles	Step
95	15 min	1	1
95	20 sec	28	1
59	60 sec	28	2
72	5 min	28	3
72	5 min	1	1
4-12	∞	1	1

Table S6. Reaction mix for Gateway cloning of ParB1 (PfuUltraII).

Reagents	Volume
10X PfuUltra II Reaction Buffer	5 µl
5 mM dNTP	0.5 µl
10 µM Primer F	1 µl
10 µM Primer R	1 µl
DNA_template	1 µl (100 ng)
PfuUltra II fusion HS DNA Polymerase	1 µl
Sterile MQ H_2O	40.5 µl
Final volume	50 µl

Table S7. Cycling conditions for Gateway cloning of ParB1 (PfuUltraII).

Temperature (°C)	Time	Cycles	Step
95	2 min	1	1
95	20 sec	30	1
62	20 sec	30	2
72	20 sec	30	3
72	3 min	1	1
4	∞	1	1

Table S8. Reaction mix for Gateway cloning of ParB3 (Q5).

Reagents	Volume
5X Q5 reaction buffer	5 µl
5 mM dNTP	1 µl
10 µM Primer F	1.25 µl
10 µM Primer R	1.25 µl
DNA_template	1 µl (10 ng)
Q5 High-Fidelity DNA Polymerase	0.25 µl
5X Q5 High GC Enhancer	5 µl
100% DMSO	1.25 µl
Sterile MQ H_2O	9 µl
Final volume	25 µl

Table S9. Cycling conditions for Gateway cloning of ParB3 (Q5).

Temperature (°C)	Time	Cycles	Step
98	30 sec	1	1
98	10 sec	28	1
40	30 sec	28	2
72	40 sec	28	3
72	2 min	1	1
4	∞	1	1

Table S10. Reaction mix for Gateway cloning of ParB3Core (Q5).

Reagents	Volume
5X Q5 reaction buffer	5 µl
5 mM dNTP	1 µl
10 µM Primer F	1.25 µl
10 µM Primer R	1.25 µl
DNA_template	1 µl (10 ng)
Q5 High-Fidelity DNA Polymerase	0.25 µl
5X Q5 High GC Enhancer	5 µl
Sterile MQ H_2O	10.25 µl
Final volume	25 µl

Table S11. Cycling conditions for gateway cloning of ParB3Core (Q5).

Temperature (°C)	Time	Cycles	Step
98	30 sec	1	1
98	10 sec	28	1
60	30 sec	28	2
72	40 sec	28	3
72	2 min	1	1
4	∞	1	1

6.9 Appendix 9: Homologous Sequences

*sequence with available PDB entry, listed below

*>sp|Q8GY89|SRX_ARATH Sulfiredoxin, chloroplastic/mitochondrial OS=Arabidopsis thaliana OX=3702 GN=SRX PE=1 SV=1
MANLMMRLPISLRSFSVSASSSNGSPPVIGGSSGGVGPMIVELPLEKIRRPLMRTRSNDQ
NKVKELMDSIRQIGLQVPIDVIEVDGTYYGFSGCHRYEAHQKLGLPTIRCKIRKGTKETL
RHHLR
(PDB: 6KY4)

>tr|Q5WAG8|Q5WAG8_BACSK Stage 0 sporulation protein J OS=Bacillus clausii (strain KSM-K16) OX=66692 GN=spo0J PE=3 SV=1
MARTAGKKGLGKGLQAFFPEQEDKQEEQIVQVDLADVRPNPYQPRKTFSEEALKELSNSI
REHGILQPVTVRKAIKGYEIVMGERRVKAAKQAGLTQIPVIVQELDENKMMEIALIENLQ
REDLNPIEEAIAYEKLMEHTNSTQEQLAKRLGKSRPHIANHMRLLQLPKVVQEFISVGKL
TMGHGRALLGLQDKQKLSQLLEKVLQDKLSVREVEQLVQRLNEHVPRETKQVKVKLPPII
KEQQERLRDTLGTSVLIKPGKKKGKIEIDYFSEDDLERILSLLVHTEKE

*>sp|P26497|SP0J_BACSU Stage 0 sporulation protein J OS=Bacillus subtilis (strain 168) OX=224308 GN=spo0J PE=1 SV=2
MAKGLGKGINALFNQVDLSEETVEEIKIADLRPNPYQPRKHFDDEALAELKESVLQHGIL
QPLIVRKSLKGYDIVAGERRFRAAKLAGLDTVPAIVRELSEALMREIALLENLQREDLSP
LEEAQAYDSLLKHLDLTQEQLAKRLGKSRPHIANHLRLLTLPENIQQLIAEGTLSMGHGR
TLLGLKNKNKLEPLVQKVIAEQLNVRQLEQLIQQLNQNVPRETKKKEPVKDAVLKERESY
LQNYFGTTVNIKRQKKKGKIEIEFFSNEDLDRILELLSERES
(PDB: 6SDK)

>sp|O51395|PARB_BORBU Probable chromosome-partitioning protein ParB OS=Borrelia burgdorferi (strain ATCC 35210 / B31 / CIP 102532 / DSM 4680) OX=224326 GN=parB PE=3 SV=1
MDGVFKMIDIHLLDIDNDQPRKSVSLVELEELSISIKENGILQPLIVCKANDRYKIIVGE
RRFRAAKLIQLTNIPVIEVDIKESCKDFMPLVENIQRENFTPVEEAYAYKNIMNKYSLTQ
KDLSEKIGKSRTYISNLVRILDLEQEILNAIHRKEISFGHAKVILSLKDRQDRYNLYLII
LKKKFSVRDAEKYVKNFSKSIVKKRELEQDPFLNNIKEFLFDKIQTKIDIKGNQNKGKIE
IEYFTAGDLKRIVSLFGHSS

>tr|B4EFY9|B4EFY9_BURCJ Putative partitioning protein ParB OS=Burkholderia cenocepacia (strain ATCC BAA-245 / DSM 16553 / LMG 16656 / NCTC 13227 / J2315 / CF5610) OX=216591 GN=BCAM0004 PE=3 SV=1

MKPSQFAKGFQARPDITTSEKRTALDRLNAIDGIVKSETPTPAPTKSAKKDIAPPPAPEF
TIDPSIDESQQYRAWRLENRYAPGQVIELPLKAIKHSPFNPRHFYLKSSIAELAVNLAKQ
GQQQAIHVIPDYDNPGTYFVSDGGRRVRALKEANKESVKAIVIDVPIGIQSYKLGYDLNV
QRDSQTVFDNAVVWRRFLDDKHFQSQKELSEHLGLDESTVAVALSIGKLPEAIMQEMVAR
PDRFGSNMAYQVGRYHNARGTEATLRLINKIVSDDLSTRQVSDIVKGRVAAQETPKPAGR
QRYAQRLEIKLGGKSVGDLKSYGEDRIELRLRGLPKDKRDAILEQLERMLLSE

>tr|C5ANI5|C5ANI5_BURGB ParB-like partition protein OS=Burkholderia glumae (strain BGR1) OX=626418 GN=bglu_3p0020 PE=3 SV=1
MTKLDLRSRVRAGMDEQTRSATARLEAARVVETAHKEVAHALPSDVADRIPGTVSREGIL
ASERRPVKIRVAETIPNPYNPRVFYDEATIDELVNSFESQGQLEAIKVTRLNEYPGKWVI
IDGGRRTRAAKRRNDEFIDAEIIEEALEAKTLYLRAYHANKDREEQTDFDDAYAWKKLLD
DGIYRDQNELAQAVGRDPKHVSKVLQLTTLPETLLESMAKRADVVKLSHAYNLKLIFDRG
GETLASRWLNEVIAGTTSVRKLEQIAADAAEPKQAKRNKVHYQSKYPFQLEDGSEIGVLK
QFADGRTELSLKGVTGQDQVELGEKLKAVVLEWTKARQSSDPVADA

>tr|A0A3G3HJN9|A0A3G3HJN9_BURL3 Chromosome partitioning protein ParB OS=Burkholderia lata (strain ATCC 17760 / DSM 23089 / LMG 22485 / NCIMB 9086 / R18194 / 383) OX=482957 GN=CVS37_39475 PE=3 SV=1
MAKDTSKDKKPTGNLHLAAGLLRGLAQENAALETRLPEPAAAPNAVVDTAPASAPAAATP
AGTPDLGAPQKVLVKDCIPNPFNPRVFYSESSLHELALTLKREGQIEPIKVTRLPEFPGK
LVVIDGQRRLRATSINGDESINATFRTDHTPEQLYTIAYRANHDHERQTIFDDAVAWKRL
LDEKVFSDQNTLAEKIGKDKASISKTLSLNALPNTLLERMASANDVVGLQAAYFLKLIFE
RLGEPTADRLLTAVIDRKKSVRDLENFLRAQSDGSKKAGRTRYSVRHDFALESRAIGQLK
TYPDGRLDLQLKGVDASHQEALADKLKTVIDAYVAELAAATQQ

>tr|A0A1B4LIT4|A0A1B4LIT4_9BURK Chromosome partitioning protein ParB OS=Burkholderia ubonensis OX=101571 GN=WJ35_18840 PE=3 SV=1
MAKDTSKEKKATGNLHLAAGLLRGLAQENAALETRLPEPSVAPLTAVAAPAVKAAMTLPA
NAEDLGAPQKVAVKDCIPNPFNPRVFYSESSLHELALTLKREGQIEPIKVTRLPEFPGKL
VVIDGQRRLRATSINGDETINATFRSDHTPEQLYTIAYRANHDHERQTIFDDAVAWKRLL
DEKVFPDQNTLAEKIGKDKATISKTLSLNALSTTLLERMAAANDVVGLQAAYFLKLIFER
LGEPTADRLLTAVMERKKSVRDLENFLRTQGDGNKKTGRTRYSVRHDFALESRAVGQLKT
YPDGRLDLQLKGVDTSHQEALADQLKAVIDAYVAELAVAAQK

>tr|A0A158ARZ0|A0A158ARZ0_9BURK ParB-like partition protein OS=Caballeronia hypogeia OX=1777140 GN=AWB79_02692 PE=3 SV=1
MAKQLNLTSRVRAGMAVERESAEARLAKNDVAETARREAVNVIGKASEPIVRPSGRFKIK
ISEAVSNPYNPRSFYNAETIEGLARSFEQEGQLEPIKVTRLPEFPGKFVIIDGERRLRAG
KSRGDEFIDAEICEETLEKQTLYLRAYRANKERDEQTVFDDAVAWKRLLEEGIYREYSEL
AAAVGEAPSHVNKVILLNSLPTSFLTQMARFPKSVGLAHAYNLKLILERAGQPVAEHWLH
EILEGRASVRRLEQVAAADAPTKRGGPRRTHYQSRVQFKRGDGVALGEMKLFGDGRTELS
LVGVQGEAQQRLAERMKELIEVWSQELEG

>tr|A0A158DVN9|A0A158DVN9_9BURK ParB-like partition protein OS=Caballeronia temeraria OX=1777137 GN=AWB76_07564 PE=3 SV=1
MAKLNLTSRVRAGMALERQSAEARLQKNAVAETALREAATVVGVRPVHGDDVATMVPRQR
IKIADTVSNPYNPRTFYSAETIDGLARSLEEDGQLEPIKVTRLSEYPGKWVIIDGERRLR
AAKSRGAEFIDAELHEERLENKALYLRAYRANKERDAQTVFDDAIAWNRLLEKGIYRDYS
ELAAAVNEAPTHVNKVILLNSLPPSFLSRMAQSADTVGLAHAYNLKLILERGGLAVAEHW
LQEILEGRASVRRLEQAASADAPTRRGGSRRVHYQSRVQFKRRDGGSLGELKLFADGRTE
LSLQGVDGESQQMLAERMKALIESWTRELDGEASA

>sp|B8GW30|PARB_CAUVN Chromosome-partitioning protein ParB OS=Caulobacter vibrioides (strain NA1000 / CB15N) OX=565050 GN=parB PE=1 SV=2
MESVVVGEPGMSEGRRGLGRGLSALLGEVDAAPAQAPGEQLGGSREAPIEILQRNPDQPR
RTFREEDLEDLSNSIREKGVLQPILVRPSPDTAGEYQIVAGERRWRAAQRAGLKTVPIMV
RELDDLAVLEIGIIENVQRADLNVLEEALSYKVLMEKFERTQENIAQTIGKSRSHVANTM
RLLALPDEVQSYLVSGELTAGHARAIAAAADPVALAKQIIEGGLSVRETEALARKAPNLS
AGKSKGGRPPRVKDTDTQALESDLSSVLGLDVSIDHRGSTGTLTITYATLEQLDDLCNRL
TRGI

>tr|A0A1Y0I2T9|A0A1Y0I2T9_CELCE Chromosome partitioning protein ParB OS=Cellulosimicrobium cellulans OX=1710 GN=CBR64_07610 PE=3 SV=1
MSAGGRGPLGASDAEDERSGTQVTSPGGGDTAPPPADSPDDVVSRETAPQAQNGDGSAVS
GAKSTSESPADGDGAVTEEWAGSAAADLVPVPGARFAEVPTASIRPNPRQPRTVFDEGDL
DELIGSIREIGVLQPIVVRPVPGEDGSFELIMGERRWRATQAAGLDVIPAIIRETDDADL
LRDALLENLHRSALNPLEEAAAYRQLLDDFGCTHEELAERISRSRPQISNTLRLLRLPPL
VQRRVAAGVLSAGHARALLGLTDGADIERLAQRIVAEGLSVRATEEIVAMGGLDGERRAP
RAPRAGQRSAAIDELAHRLSDRFETRVKVDLGKNKGRLTVEFASVEDLNRILDVMAPEDP
GLLRK

>tr|A0A0P1MLP1|A0A0P1MLP1_9BACT Chromosome partitioning protein, ParB family OS=Chrysopegis kryptomonas OX=1633643 GN=JGI23_00133 PE=3 SV=1
MSKQRLGRGLDALIPKSATKEISIDSKDLRFDDGQSVGVIAKIEISKIKPNPYQPRENFD
RASLEELKQSIIEKGVIQPITVRRLPGGMYELITGERRVRASSEAGLTHIPAYIIEVESE
KELLELALIENIQREKLNPIEIAKAYQRLIEELGYTQEEIAKKIGKDRTTVANFIRLLKL
PEQIQESLKRGEITMGHARALINIPSRKLQIEIWNKIVKQGWSVRKVEKAVRDLLRSKDG
EEKPKKKTHSSPGVRDIESKLKRIFGTQVRIRQTGNSKGEIVIEFYSNDDFERLIELFEI
IEKHAR

>sp|Q9RYD8|PARB1_DEIRA Probable chromosome 1-partitioning protein ParB OS=Deinococcus radiodurans (strain ATCC 13939 / DSM 20539 / JCM 16871 / LMG 4051 / NBRC 15346 / NCIMB 9279 / R1 / VKM B-1422) OX=243230 GN=parB1 PE=3 SV=1
MSKKSSLGRGLDALLTKKGEPVAQAGTGTQVQTLKIERIAQAAYQPRQVFEPESLAELAQ
SIREKGVLQPLLVRPRGDAFEIVAGERRWRASQLAGLTELPVMIRDLGDREALEIAIVEN
LQREDLGPLEEARAYQALLDQGLNQEGVAQAVGKGRSTVTNALRLLTLPEPVLRALDEGS
ISASHARAVLTQPEADRLWAFEQIRSRGLNVREAEALKRERGGRDKGQGAPIKVNPPRAY
RQLELDLSRRTGTRVKITGEDKGRVELNYGSREELDRILQILGYEAEE

*>sp|P62558|SOPB_ECOLI Protein SopB OS=Escherichia coli (strain K12) OX=83333 GN=sopB PE=1 SV=1
MKRAPVIPKHTLNTQPVEDTSLSTPAAPMVDSLIARVGVMARGNAITLPVCGRDVKFTLE
VLRGDSVEKTSRVWSGNERDQELLTEDALDDLIPSFLLTGQQTPAFGRRVSGVIEIADGS
RRRKAAALTESDYRVLVGELDDEQMAALSRLGNDYRPTSAYERGQRYASRLQNEFAGNIS
ALADAENISRKIITRCINTAKLPKSVVALFSHPGELSARSGDALQKAFTDKEELLKQQAS
NLHEQKKAGVIFEAEEVITLLTSVLKTSSASRTSLSSRHQFAPGATVLYKGDKMVLNLDR
SRVPTECIEKIEAILKELEKPAP
(PDB: 3KZ5)

>tr|A0A3S8ZA03|A0A3S8ZA03_9ACTO ParB/RepB/Spo0J family partition protein OS=Flaviflexus salsibiostraticola OX=1282737 GN=EJO69_08230 PE=3 SV=1
MAERRKTPARKSLGKGISALFPAPQETDGDLPKRQAIDVFFGDGPLTAVESAPTPEALDQ
VETPEPYVPTPDEPAGLVPVPGAVLTEIPVDQIFANTKQPRTVFDEDELAELADSIAEVG
LLQPIVVRHVPDGGYELIMGERRWRASARAGLESIPAIVRDTADDDLLRDALLENLHRVE

LNPLEEASAYQQLLEDFGCTQEELSRRIARSRPQISNTLRLLKLPPLVQRRVAAGVLSAG
HARALLGLTDPAAMERLAQRIVAEGLSVRAVEEIVALGDEPRVRSAGRATRETAPQLMRL
SEDLSNRLDTRVKIRLGARKGTMTVDFGSIEDLNRILDVLSLPVERIPDGV

>tr|A0A126UXT0|A0A126UXT0_9RHOB Chromosome partitioning protein ParB OS=Halocynthiibacter arcticus OX=1579316 GN=RC74_05885 PE=3 SV=1
MVQKKNERRGLGRGLSALMADVNIPAPQDEKEAPVRAEQTIPIELIEPNPDQPRRYFDEE
ALQDLARSITEKGVIQPLIVRPNPRKQGLYEIVAGERRWRASQIAQLHKVPAIIRELDDT
EVLEVAIIENIQRADLNPVEEAIGYRQLMDRFGHTQEKVASALSKSRSYIANLLRLLQLP
EDVLALLRDGSLTTGHARALITAPNASELAKKIVAEGLSVRQTEKLVKIGDNSTKNQKSP
KASKDADTVAIEGDLSAALGMDTVIDHLAGKESGKVSIKYKSLEQLDEICRILSSTR

*>sp|O25758|PARB_HELPY Probable chromosome-partitioning protein ParB OS=Helicobacter pylori (strain ATCC 700392 / 26695) OX=85962 GN=parB PE=1 SV=1
MAKNKVLGRGLADIFPEINEVYEQGLYERANRVVELGIDEVMPNPYQPRKVFSEDSLEEL
AQSIKEHGLLQPVLVVSENGRYHLIAGERRLRASKLAKMPTIKAIVVDIEQEKMREVALI
ENIQREDLNPLELARSYKELLESYQMTQEELSKIVKKSRAHVANIMRLLTLSSKVQNALL
EEKITSGHAKVLVGLDGEKQELILNSIIGQKLSVRQTEDLARDFKINANFDNKKHGFKQT
QTLIAGDELERLNQSLWDHYKLKAALKGNKIVLRCYENSLLEAFMKKMMS
(PDB: 4MUK)

*>sp|Q9BYN0|SRXN1_HUMAN Sulfiredoxin-1 OS=Homo sapiens OX=9606 GN=SRXN1 PE=1 SV=2
MGLRAGGTLGRAGAGRGAPEGPGPSGGAQGGSIHSGRIAAVHNVPLSVLIRPLPSVLDPA
KVQSLVDTIREDPDSVPPIDVLWIKGAQGGDYFYSFGGCHRYAAYQQLQRETIPAKLVQS
TLSDLRVYLGASTPDLQ
(PDB: 1XW3)

>tr|A0A1M6C2W5|A0A1M6C2W5_9BACT Chromosome partitioning protein, ParB family OS=Hymenobacter daecheongensis DSM 21074 OX=1121955 GN=SAMN02745146_1009 PE=3 SV=1
MSETNEEKTTPAAAPAAAKRKIGGLGRGLNALIEGSYEKKGERLVSHPMNSVGFIPVAHI
EANPYQPRTHFDQEALQELAESIRVQGIIQPVTVRQMGTNSYQLISGERRLQASKLAGLD
TIPAYIRKADDQQMLEMALIENIQRENLNAIEIALSYQRLVSECNLKQEELGDRVGKNRS
TVTNYLRLLKLPPDIQIGLRDTAISMGHARALINIENADQQLALFHRILAEDLSVRKVEQ
LVRAGGAAPKAAEAPTAQQVQEQQVHIPVAELRRTERYLSDRFGSKVLVKPGPQGNGEIK
IAFDSVEDMQRILHILQPA

>tr|A0A1T5KCC2|A0A1T5KCC2_9MICO Chromosome partitioning protein, ParB family OS=Krasilnikoviella flava OX=526729 GN=SAMN04324258_2060 PE=3 SV=1
MNAEPAADQFATSSAAVEQEDEDEGVDDAAVSRETELKPVPGARFAELPVEAIRPNTWQP
RSVFDDGELDELVASIAEIGVLQPIVVRPDLSEPDRYELIMGERRWRATQAAGLETIPAI
VRETDDSDMLRDALLENLHRAALNPLEEAAAYRQLLDDFGCTHEELATRIARSRPQISNT
LRLLKLPPLVQRRVAAGVLSAGHARALLSLSDGAEIERLAQRIVSEGLSVRATEEIATLL
TSGDGDKPTRRAPRAGLRSEAMDELSTRLADRFETRVKVSIGKTKGRLTVEFASVEDLNR
ILGVMAPEDPGLLRS

>tr|A0A0H4KXE7|A0A0H4KXE7_9RHOB Chromosome partitioning protein parB OS=Marinovum algicola DG 898 OX=988812 GN=MALG_03579 PE=3 SV=1
MKQEKRRGLGRGLSALMADVEQTPGETGEAPRRADRMIPVERIFPNPDQPRRQFTQDALQ
ELANSIAEKGVIQPLIVRPQGEDRYEIVAGERRWRASQMAQLHEIPVIIRDFTDTEVLEV
AIIENIQRADLNPVEEAMGFRQLMDRFGHTQEKLAKALGKSRSHIANLLRLLTLPDEVRE
MLQQGELSAGHARALVTCDDPVGLAKRVVAQGLSVRETEKAAKQGDAPEPKAPTTGFLAE

PKGKDADTRALEGDLSAHLGLKVSVEHKPGSEGGKLVLNYSSLEQLDDLCRRLAGE

>tr|A0A2J9KNE0|A0A2J9KNE0_9ACTO Chromosome partitioning protein ParB OS=Mobiluncus mulieris OX=2052 GN=parB PE=3 SV=1
MAKKRALGSGLGALIPPAPASRGVDVIVPPKHGSQASPEDAKVHDLLNPRSPRTQTKAKG
NSKSKGEPELVAVPGMRMAELPLQSVVPNPNQPREVFDEEALRELAESIKSVGVLQPIVV
RPLEDSGGESRYELVMGERRWRASKLAGKRQIPAIIRETADEDMRRDALLENLQRVNLNP
LEEAAAYQQMIAEFGITQELLAKKLSRSRPQISNTLRLLKLPATVQVKVAAGVISAGHAR
ALLGIGNHEAMAALADRIVAEGLSVRATEEIVSLGEAEKPKRERKPRVKPALSEHLMESK
TKLEDLLQTRVNLQVGKHKGSISIEFADEADLNRIVEFISRH

>sp|P9WIJ9|PARB_MYCTU Probable chromosome-partitioning protein ParB OS=Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv) OX=83332 GN=parB PE=1 SV=1
MTQPSRRKGGLGRGLAALIPTGPADGESGPPTLGPRMGSATADVVIGGPVPDTSVMGAIY
REIPPSAIEANPRQPRQVFDEEALAELVHSIREFGLLQPIVVRSLAGSQTGVRYQIVMGE
RRWRAAQEAGLATIPAIVRETGDDNLLRDALLENIHRVQLNPLEEAAAYQQLLDEFGVTH
DELAARIGRSRPLITNMIRLLKLPIPVQRRVAAGVLSAGHARALLSLEAGPEAQEELASR
IVAEGLSVRATEETVTLANHEANRQAHHSDATTPAPPRRKPIQMPGLQDVAERLSTTFDT
RVTVSLGKRKGKIVVEFGSVDDLARIVGLMTTDGRDKGLHRDAL

>tr|A0A1G8YKS3|A0A1G8YKS3_9ACTN Chromosome partitioning protein, ParB family OS=Nonomuraea jiangxiensis OX=633440 GN=SAMN05421869_113305 PE=3 SV=1
MSQQRRGLGKGLGALIPTGPIVDGTGAVAPANGSTGDAGPRPIAGAYFKEVALEAIVPNP
RQPRDVFDGERLEELAASIREVGLLQPIVVRPVGGGRYELVMGERRWRACKQVGLDPVPT
IVRNTQDTDLLRDALIENLQREQLNAIEEAAAYQQLLDDFQATHEQLAEKVGRSRSHITN
TLRLLHLPPEVQRKVAASVISAGHARALLGLNSAEEQIELAKRIVAELLSVRAVEEIVAM
GSAKAAQKATRERQPPRPTAPGLTHLADRLSDHFETKVKVDLGRRKGRIVVEFATIDDLE
RIIGAMAPEAVRVMREAED

>tr|A0A172ZCH9|A0A172ZCH9_9BACL Stage 0 sporulation protein J OS=Paenibacillus bovis OX=1616788 GN=AR543_02815 PE=3 SV=1
MSKRLGKGLDALIPSLSVSEDDKVIEIPLKELRPNPFQPRKAFDDENIQELADSIAQHGV
IQPIIVRKVLKGYEIIAGERRFRASQVAGKKTIPAVVRTFSDQQVMEIALIENLQRENLN
AMEVATAYQALMEEFKLTQEDLSARVGKSRSHIANFLRLLSLPEEVKEYVSRGTLSMGHA
RALVGIKKEDLILMLAHRAIEEGWSVRQLEEAVQNLETKRKDNDKDKDKKTDPYLVDVEQ
TLQSRFQTTVKIRNSSKDKGKIELNYYSRQELERLLEMLNNM

>tr|A0A2K9MHG0|A0A2K9MHG0_9RHOB Chromosome partitioning protein ParB OS=Paracoccus jeotgali OX=2065379 GN=CYR75_12955 PE=3 SV=1
MTSTKRSSLGRGLSALMADLQPVDAPAGSPGATLVPIEQVTPNPDQPRRIFDPAALNELA
ESIRSRGIIQPLIVRRHPDDGSLYQIVAGERRWRASQIAQLHELPVIIREFSDAEMLEVA
IIENIQRADLNSIDEAASYRQLMTRFGHTQEKLAEALGKSRSHIANLLRLLTLPDQVQDW
VREEKLTAGHARALVTAENPTQIARRIIDKDLSVREVEALMREKTQKPGKAAPKAEKDAD
TRALEGDLSATLKMRVQIKHVGTEGGQMTITYRDLDQLDRLCQILAGTA

>tr|A0A248VZ14|A0A248VZ14_9BURK Chromosome partitioning protein ParB OS=Paraburkholderia aromaticivorans OX=2026199 GN=CJU94_36910 PE=3 SV=1
MSDQNQSAPQRRSSLVSSLRKGMDMERGAVDARLNAQEVGSAPMAGAGSTHSPSVPAYGG
RGNTASVKPRLTLPLSDLVSNPYNPRTFYRPEKVDELAVKMKRDGQLEAIKVTENARFPG
KHVIIDGEYRFRAKKSVADGFIDVEVFPALSDRDLFLIANSLNKDRTPQSLFDDALAYQK
LLDDGVFPGQEDLAASLGISESLMSKILKLKKLPETLLRRMAESDEPLGISHAYNIALMF
ERKGLHTAEEVLDKVLAGEMSSKHLQDLNSRNDNGEGSRKKRAHYLARVQFAGADGTEYG

QLKRFRDGRTELKLTGLSEEQQEALGERLEAVVREFMTGNAGSGA

>tr|A0A1H7G4V2|A0A1H7G4V2_9BURK Chromosome partitioning protein, ParB family OS=Paraburkholderia caballeronis OX=416943 GN=SAMN05192542_101581 PE=3 SV=1
MATKLNFTSKVRAGMAVERVSAGERLAESDVVGQALRDAAPVVVTDASVESIAPQRIRKI
PIGDVVSNPYNPRAFYSAETIDELANSFAQQGQITPIQVTRLPEFPDKYVIVDGERRVRA
AKSRGDKFIDADLKDGLDNQKLYLRAYHANKEREEQTVFDDALAWKKLLDDRVYVDQVEL
GVAVGQDPKHISKVIALTSLPPFLLQRMAQFSKAVGLGHAYNIKLILDRAGIKVAEHWLQ
QVIDGKASVRKLEAAASAEAGARSVGPRRTHYQSRVQFNRPDGVALGELKLFGDGRAELS
LKGISPADQQKLAERVKTVIDQWAAELHAPQSDVAS

>tr|A0A1H4GQ61|A0A1H4GQ61_9BURK ParB family protein OS=Paraburkholderia sartisoli OX=83784 GN=SAMN05192564_106164 PE=3 SV=1
MKPSQFAKGFQARPDTTSSEKRTALDRLNAIDGLVQRSDAKSADVPGRSSQPDAAVVSMV
EEPALSQAASESPEYHAWRVEHGYRSGQVIELALKTIKPSPFNPRHFYLKSSIAELAVNL
AKQGQQQAIHVIPDYDNPGSYFVSDGGRRVRALKEANKDSVKAIVIDLPIGIQSYKLGYD
LNVQRDSQTVFDNAVVWKRFLDDKHFQSQKELAEHLGLDESTVAVALSIAKLPEAVMHEM
VVRPDRFGSNMGYQVGRYHSARGTDATLRLINKILSDDLSTRQVADIVKGRATAQEIPKP
AGRQRYAQRLDIKLDGVTVGDLKSYGDDRLELRLRGLTREKRDDILKQIEQMLLSK

>tr|A0A1H1JN96|A0A1H1JN96_9BURK ParB family protein OS=Paraburkholderia tuberum OX=157910 GN=SAMN05445850_5257 PE=3 SV=1
MKPSQFAKGFQARPDTTSSEKRTALDRLNAIDGLVSNEAKTPTIAGRAIPAVLPSRLQDA
EIPDPDNESAAYRAWRLEHGYRPGQIIELALKTIKPSPFNPRYFYVKSSIAELAVNLAKQ
GQQQAIHVIPDYDNPGTYFVSDGGRRVRALKEANKESVKAIVIDLPVGIQSYKLGYDLNV
QRDSQTVFDNAVVWRRFLDERHFQSQKELAEHLGLDESTVAVALSIAKLPETVMQEMVAR
PDRFGSNMAYQVGRYHTARGAEATLRLINRILSDDLSTRQVADIVKGRATSQESAKPAGR
QRYAQRLEIKLGGVSVGDLKSYGDDRLELRLKGLTREKRDDILRQIEKMLK

>sp|P0A151|PARB_PSEPK Probable chromosome-partitioning protein ParB OS=Pseudomonas putida (strain ATCC 47054 / DSM 6125 / NCIMB 11950 / KT2440) OX=160488 GN=parB PE=3 SV=1
MAVKKRGLGRGLDALLSGPSVSALEEQAVKIDQKELQHLPVELVQRGKYQPRRDMDPEAL
EELAHSIRNHGVMQPIVVRPIGDNRYEIIAGERRWRATQQAGLDTIPAMVREVPDEAAIA
MALIENIQREDLNPLEEAMALQRLQQEFELTQQQVADAVGKSRVTVANLLRLITLPDAIK
TMLAHGDLEMGHARALLGLDENRQEEGARHVVARGLTVRQTEALVRQWLSDKPDPVEPSK
PDPDIARLEQRLAERLGSAVQIRHGNKGKGQLVIRYNSLDELQGVLAHIR

>sp|Q1RGT4|PARB_RICBR Probable chromosome-partitioning protein ParB OS=Rickettsia bellii (strain RML369-C) OX=336407 GN=parB PE=3 SV=2
MKNKGLGRGLSSLLGEEVISTEKESLEIVQIINIDRIKPNENQPRKHFEYDKIKELSDSI
LNNGLLQPIIIDNSFQIIAGERRWRACKLAKISEIPVIIKNLDAKESMEIALIENIQRSD
LTVMEEARGFKYLVENFNYTAEKLAERLGKSRSHIANLLRLNNLPQSIQNKLDENTLSMG
HARCLINHEHAEIIANYVIDNDLNVRQTEELVRQWSQNEYTKYPDNNRIGKQLFKEKTED
NDLQSLVKILSEKFNIKVTIENYSLGGKLIFHYKDLKELDKILLELS

>sp|P36077|SRX1_YEAST Sulfiredoxin OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=SRX1 PE=1 SV=1
MSLQSNSVKPTEIPLSEIRRPLAPVLDPQKIDAMVATMKGIPTASKTCSLEQAEAAASAG
ELPPVDVLGVRVKGQTLYYAFGGCHRLQAYDRRARETQNAAFPVRCRVLPATPRQIRMYL
GSSLDIE

>tr|D1BGG4|D1BGG4_SANKS Chromosome segregation DNA-binding protein OS=Sanguibacter keddieii (strain ATCC 51767 / DSM 10542 / NCFB 3025 / ST-74) OX=446469 GN=Sked_37980 PE=3 SV=1
MSEKRRGLGRGLGALIPTAPDLERPVDVFFPPKQQVEGEEPNSTAIADSAAVVAKLRAPR
EPKKTTGGSQRGSMAAALTSAGGSSATTARDRGTVRIDSVSRETYLAATGTEPESTLVPV
PGATFAHLPVESIRPNARQPRTVFDEGDLAELVGSIREIGVLQPIVVRPTDEPDRYELIM
GERRWRASQEAGLGAIPAIIRETEDGDLLRDALLENLHRSALNPLEEAAAYRQLLDDFGC
THEELAVRISRSRPQISNTLRLLKLPPLVQRRVAAGVLSAGHARALLGLPDGAQIERLAQ
RIVAEGLSVRATEEIVALGGDEAPTKPAQRPRAGTRIQAIDELATRLSDRFETRVKVDLG
KTKGKVTVEFASVEDLNRILNVMAPEEPGLFS

>sp|P55392|Y4CJ_SINFN Putative replication protein B OS=Sinorhizobium fredii (strain NBRC 101917 / NGR234) OX=394 GN=NGR_a00020 PE=3 SV=1
MARKNLLAGLVDTAEIPHADVAPAYPMRGASKSMVRSLDELSRQAEKFLEGETVVELDPE
TLDGSFVSDRMGDSSEQFEELKQAIAERGQDTPILVRPHPSAADRYQIVFGHRRARVARE
LGRKVKAVVKALDDRTHVIAQGQENSARANLSFIERANFASHLEKLGYDRTIIGSALAAN
AAAISKMIAVIDRIPEETIARIGPCPAVGRERWVELSLLVGKTANEAKVKAIVSDPSFNE
LSTDDRFNSLFSGLNSAAKPVRKTTPKILENWQPADKTVSAKYSNSAKAFALSMKSRNAG
PFGRYIADNLDRLYAEFLEQGNRKED

>tr|A0A2K8PJ62|A0A2K8PJ62_STRLA Putative chromosome-partitioning protein ParB OS=Streptomyces lavendulae subsp. lavendulae OX=58340 GN=parB PE=3 SV=1
MAGAGATSPSAVSTLASERGIAAAKIAVPAQEVAPSAVAADVSRETSAAGVPAPANAEAD
PAAGARFAELPLDAIKPNPRQPREVFDEDALAELVTSIQEVGLLQPVVVRQTAPGRYELI
MGERRWRACREAGLERIPSIIRATDDDKLLLDALLENLHRAQLNPLEEAAAYDQLLQDFK
CTHDQLADRIGRSRPQVSNTLRLLRLSGPVQRRVAAGVLSAGHARALLSVEDSEQQDRLA
HRIVAEGLSVRAVEEIVTLMGSEPASTVKPKGPRAGTRVAPALSQLATRLSDRFETRVKV
DLGQKKGKITVEFASMEDLERILGTLAPGEGRVLEQGLSGE

>tr|A0A1V0UCG5|A0A1V0UCG5_STRVN Chromosome partitioning protein ParB OS=Streptomyces violaceoruber OX=1935 GN=B1H20_17200 PE=3 SV=1
MSERRRGLGRGLAALIPAAPLEKKGPETDDGAASFGAGAVLTADRGVPAAKLAGLPSSPL
VPEPASAPEERTEQGVGASGAYFAELPVGSITPNPRQPREVFDEDALAELVVSIKEVGLL
QPVVVRKTGPDSYELIMGERRWRACQEAGLERIPTIVRATDDEKLLLDALLENLHRAQLN
PLEEAAAYDQLLKDFQCTQDQLADRIGRSRPQVTNTLRLLRLSPPVQRRVAAGVLSAGHA
RALLSLEDAEEQDQLAHRIVAEGMSVRAVEEIVKLMDSHPSRTRKPKGPRAGGRVSPALT
DLATRLSDRFETRVKVDLGQKKGKIVVEFASMEDLDRILGSLAPGEGRVLERRLADSSEE
DEG

>tr|A0A0N7M294|A0A0N7M294_9RHOB Chromosome-partitioning protein ParB OS=Thalassobius mediterraneus OX=340021 GN=parB_1 PE=3 SV=1
MVDRPKTRGLGRGLSALMADVTPPTAAADTGQGASARRPDMMVPIEKVHPNPDQPRRTFT
EEQLQELAASIKEKGIIQPLVVRERPGDKGAYEIVAGERRWRAAQIAMQHTIPVIVRDFD
DTEVLEVAIIENIQRADLNPVEEAAGYKALMDKFGHTQEKLSEVLGKSRSYIANLVRLLQ
LPDEVQEFLRDGKLSAGHARALITSDDAVSLAKKVISSGLSVRETERLAKKTVSELSEMD
QPRPKAPRKSVEKDADTRALEADLSAALRMKVSIDHDAGGESGKLIISYEDFDGLDELCR
KLSVVS

>tr|A0A1H1CIV9|A0A1H1CIV9_9ACTN Chromosome segregation DNA-binding protein OS=Thermostaphylospora chromogena OX=35622 GN=SAMN04489764_1496 PE=3 SV=1
MSQQRRGLGKGLGALIPTVPTENGVGVVAPTVRAEPEPGPAPVPGAHFQEIPVNAIDPNP
RQPRTIFDEAALEELATSIREVGLLQPIVVRPSGRDRYELIMGERRWRASQIAGLTEIPA

IVRKTQDDELLRDALIENLQREQLNPLEEAAAYQQLLDDFGATHEQLAQRIGRSRPHITN
TLRLLQLPPEVQKRVAAGLISAGHARALLSLDDPAAQEHLAKRIVAEGLSVRSVEEIVAL
GDVKAGRAPRERTTKKAPSPVLVDLADRLSDRFETKVKVDLGQRKGRIVVEFASMEDLER
IIATMAPEALPSMQEKG

*>sp|Q72H91|SP0J_THET2 Chromosome-partitioning protein Spo0J OS=Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039) OX=262724 GN=spo0C PE=1 SV=1
MSRKPSGLGRGLEALLPKTGAGVVRLPLASIRPNPRQPRKRFAEESLKELADSIREKGLL
QPLLVRPQGDGYELVAGERRYRAALMAGLQEVPAVVKDLTDREALELALVENLQREDLSP
VEEARGYQALLEMGLTQEEVARRVGKARSTVANALRLLQLPPEALEALERGEITAGHARA
LLMLEPEDRLWGLKEILEKGLSVRQAEALRERLAMAPKRSAEPSPLSLELSRHLGLPVRV
VGGKKGKVVIQYRSLEELEALLRRLGYQA
(PDB:1VZ0)

>sp|Q9KNG7|PARB_VIBCH Probable chromosome-partitioning protein ParB OS=Vibrio cholerae serotype O1 (strain ATCC 39315 / El Tor Inaba N16961) OX=243277 GN=parB PE=3 SV=1
MTKRGLGKGLDALLATSSLAREKQQVASLSQSMSAEGELADLSISNLKPGIYQPRKDLSP
EALEELAASIQSQGIIQPIVVRHLPTGGYEIIAGERRWRAAKQAGLKQVPCLIKQVEDRG
AIAMALIENIQREDLNAMEEAQALERLQNEFNLTHQQVAEVIGKSRTTVTNLLRLNQLSD
DVKRLLETKQLEMGHARALLMLEGEQQVEIAQQVAKKQLTVRQTEQLVKKCLSDPSDAKN
VSEDLEIQQLSQNLSEKLAAKVSIVRTPNGKSKVTISLDEPHKLELLIAKLQN

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway