Structural and Biophysical Investigations of Two Negative Regulators of DNA Replication Initiation in *Bacillus subtilis*: YabA and SirA

Katie H. Jameson

PhD

University of York

Chemistry

September 2015

Abstract

DNA replication initiation is strictly controlled to maintain chromosome copy number. Over- or under-replication of an organism's genome is detrimental to survival, thus initiation events are tightly regulated to ensure only one round of DNA replication occurs per cell cycle. In prokaryotes, DNA replication commences when a protein initiator, DnaA, forms a helical filament at the origin of replication, *oriC,* inducing localised DNA unwinding and the recruitment of the replication machinery. The work described here sought to elucidate how DNA replication initiation is regulated in the Gram-positive model organism *Bacillus subtilis*. In particular, this work aimed to offer insight into the mechanisms of two negative regulators of DNA replication initiation, SirA and YabA; proteins that play significant roles in regulating replication initiation in sporulating and vegetatively growing cells, respectively. Both SirA and YabA interact directly with the initiator DnaA, and YabA has additionally been shown to interact with the DNA polymerase β-clamp, DnaN (an essential component of the replication machinery).

Detailed here are structural and biophysical studies of SirA and YabA. This includes the X-ray crystal structure of SirA bound to the N-terminal domain of DnaA in an inhibitory complex, and characterisation of the SirA-DnaA^{DI} interface using an *in vitro* assay. When coupled with *in vivo* localisation studies carried out by our collaborators, the work on SirA suggests a mechanism for its inhibition of DNA replication. Also detailed are biophysical studies used to characterise the architecture of YabA, guided by the use of *in silico* models, and the X-ray crystal structure of YabA's N-terminal domain. These results are discussed alongside structural and mutational studies of YabA carried out by our collaborators; collectively the results delineate the full-length structure of YabA and offer insight into its interactions with DnaA and DnaN.

Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisor Tony Wilkinson, for his continuous support and trust throughout my PhD; our discussions have been invaluable to the progression of the work described here. I would also like to thank Jean Whittingham and Elena Blagova for providing me with training in all of the techniques I required to carry out this work, and for offering advice throughout my PhD. Thanks are also due to Gideon Davies for his useful advice during 'Thesis Advisory Panel' meetings, and to Mark Fogg and Anne Grahl who laid the foundations of this work. Mentions must also be given to our collaborators, Heath Murray, Marie-Françoise Noirot-Gros and Laurent Terradot; my discussions with them have been valuable in shaping the direction of work detailed here.

I also want to thank our lab technicians, Sally, Simon and Jules for all the work they did in keeping the lab running smoothly throughout my PhD. I must also thank Johan Turkenburg and Sam Hart for their assistance with crystal testing, data collection and structure solution throughout my time in YSBL.

I want to thank everyone at YSBL for making my time here so enjoyable; your friendship and generosity (in giving time, advice and cake) is what makes YSBL a truly lovely place to work - I will sincerely miss being here. Special mentions must be given to: Jenni, Sophie, James, Dan and Ben, who made my time here particularly special - I (will) miss you all.

Finally, I must also thank all of my friends and family for their support over the course of my PhD. In particular, I want to thank my parents and my partner, Gareth, who I know are always there for me, and have offered so much love and support in the last few years; I love you all.

Author's Declaration

All work presented within this thesis was conducted by the author, with the exception of collaborative works outlined below and stated explicitly throughout the body of the text. All sources are acknowledged as references.

The Author declares that none of the original work presented in this thesis has previously been used for the achievement of a research degree at the University of York or any other university.

The original work presented in this thesis has been included in the following publications (see Bibliography):

> Jameson, K. H., Rostami, N., Fogg, M. J., Turkenburg, J. P., Grahl, A., Murray, H. & Wilkinson, A. J. Structure and interactions of the *Bacillus subtilis* sporulation inhibitor of DNA replication, SirA, with domain I of DnaA. *Mol. Microbiol.* **93,** 975–91 (2014).
>
> Felicori, L., Jameson, K. H., Roblin, P., Fogg, M. J. Garcia, T., Ventroux, M., Cherrier, M.V., Bazin, A., Noirot, P., Wilkinson, A. J., Molina, F., Terradot, L., Noirot-Gros, M-F. Tetramerization and interdomain flexibility of the replication initiation controller YabA enables simultaneous binding to multiple partners. *Nucleic Acids Research*, **44**, 449–463 (2016).

The following pieces of work discussed in this thesis, and the people who conducted them, are acknowledged below:

- Cloning of pET-YSBLIC3C-DnaA^{DI}SirA was carried out by Dr. Mark Fogg and Anne Grahl prior to the commencement of the work described herein.
- Cloning of pET-YSBLIC3C-YabA was carried out by Dr. Mark Fogg prior to the commencement of the work described herein.
- Two putative full-length YabA X-ray diffraction datasets were obtained by Dr. Mark Fogg prior to the commencement of the work described herein.

- SirA *in vivo* localization studies in *Bacillus subtilis* were performed by Nadia Rostami and Heath Murray at the University of Newcastle.

- *In silico* modelling of YabA was carried out by Liza Felicori and Frank Molina.

- Structure solution of YabA$^{1-62}$ was carried out by Mickaël Cherrier, Alaxandre Bazin and Laurent Terradot at CNRS (Centre national de la recherche scientifique), Lyon.

- Small angle X-ray scattering analysis of YabA was performed by Pierre Roblin at SOLEIL Synchrotron.

- Yeast 2-hybrid and 3-hybrid screening and YabA *in vivo* localisation studies were carried out by Transito Garcia, Magali Ventroux and Marie-Françoise Noirot-Gros at INRA (French National Institute for Agricultural Research), Jouy-en-Josas.

- Mass spectrometry analysis and data processing was carried out by Simon Grist and Andrew Leech (stated in methods). Analysis of the resultant spectra was carried out by the author.

- SEC-MALLS and AUC data processing was conducted by Dr. Andrew Leech of the York Bioscience Technology Facility.

Chapter 1 : Introduction

1.1: DNA Replication

For all organisms, replication of genomic DNA is a complex task. The scale of the genome, even in a bacterial cell (at 1-10 Mbp[1]), already presents a considerable challenge - the cell must be able to replicate its DNA very quickly, but also highly accurately to ensure the correct passage of information from one generation to the next. In order to initiate the replication process, a nucleoprotein complex must form on DNA and localised DNA unwinding must occur to allow the loading of a processive DNA helicase. A large multi-component molecular machine known as the replisome must then be assembled, and this molecular machine carries out the replication of the DNA - using different mechanisms to replicate each stand of DNA due to their different polarities. Furthermore, the need to coordinate DNA replication with cell cycle, ensuring the genome is replicated once per cell division, adds an additional level of complexity to the process. Below, these processes will be discussed in detail, with particular emphasis on how DNA replication is regulated. The exact proteins and regulatory mechanisms involved in DNA replication vary, both between and within each domain of life. Although a number of parallels can be drawn between the domains, this introduction will focus only on the replication of DNA in prokaryotes and differences observed within this domain.

1.2: Semi-conserved Replication

Prokaryotic DNA replication occurs from a single origin of replication and involves the progressive separation of the two parental strands of DNA with concurrent synthesis of new DNA strands complementary to each of the parental strands (i.e. each replicated strand uses one parental strand as a template). This results in two double-stranded semi-conservatively replicated daughter chromosomes. Replicating chromosomes have been described to have replication 'eyes' or 'bubbles' where the branch points of these structures are the locations of 'replication forks' - i.e. where new strand synthesis occurs (Fig. 1.1, A). Strand synthesis is usually bidirectional (contains two replication forks) but can be unidirectional[2].

1.3: Semi-discontinuous Replication

New strands of DNA are synthesised by the action of DNA polymerase, which extends DNA strands in a 5' to 3' direction. This means that as both strands of the DNA double helix are replicated simultaneously at the replication fork, they require different mechanisms of synthesis to allow the DNA polymerase to travel in only a 5' to 3' direction (Fig. 1.1, B). The strand which extends with the replication fork in the 5' to 3' direction will be continuously synthesised with replication fork movement - this strand is known as the "leading strand". However, the strand extending in the 3' to 5' direction is synthesised as a series of 5' to 3' segments (containing around 1000-2000 bps each) known as Okazaki fragments. This is known as the "lagging strand"; these lagging strand fragments are later covalently linked by the enzyme DNA ligase. This process of leading and lagging strand synthesis is given the term 'semi-discontinuous replication' because only one strand is replicated continuously[2].

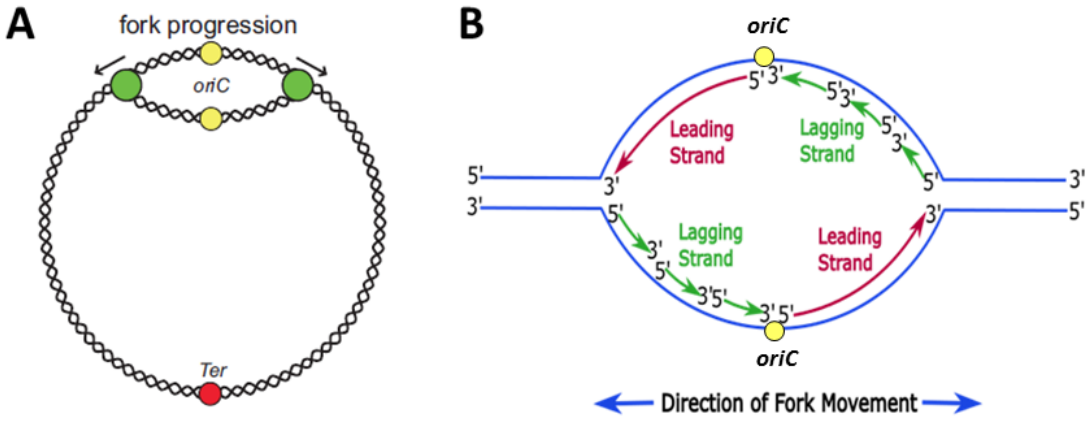


Figure 1.1: Bidirectional and semi-discontinuous DNA replication: **A**) Replication 'eye' or 'bubble'; yellow circles indicate the location of *oriC*, green circles indicate the replication forks during bidirectional replication and the red circle shows the point of replication termination. Taken from Robinson *et al*. *Current Drug Targets* 13, 352–372 (2012). **B**) Semi-discontinuous DNA replication diverging from the replication origin (yellow circles indicate location of *oriC*); all DNA polymerases synthesise DNA in a 5'→ 3' direction. Leading strand synthesis (red) occurs continuously whilst lagging strands (green) are synthesised as a series of Okazaki fragments.

1.4: Phases of DNA Replication

The process of DNA replication can be separated into 3 distinct phases: initiation, elongation and termination. During the initiation phase, a nucleoprotein complex induces localised DNA unwinding and facilitates helicase loading at the origin of replication. In the elongation phase, the replication machinery assembles and lagging and leading strands are prossessively synthesised by DNA polymerase to form new strands of DNA. Finally, during termination, DNA polymerisation is halted at a specific termination site. A number of proteins are required to collectively orchestrate each replication phase[2]. The role of these proteins and the events of each stage will be described in some detail below. Much of our current understanding of the DNA replication process in bacteria has been derived from work carried out in *E. coli* (model organism for Gram-negative bacteria). This knowledge is thought to be directly transferable to other bacterial species given the high conservation of DNA replication proteins (e.g. DnaA, DnaB-like helicase etc.)[3]. However differences between species have been observed, and these will be discussed where relevant (and known). In particular, differences observed in *B. subtilis* (a model organism for Gram-positive bacteria) will be presented as these are pertinent to this study.

1.5: Initiation of DNA Replication

1.5.1: DNA Replication Initiation in *E. coli*

DNA replication is initiated by the binding of a protein initiator, DnaA, to the origin of replication, *oriC.* DnaA forms a right-handed helical oligomer on the DNA[4,5] by binding a series of recognition sites within the origin termed DnaA-boxes[6]. The formation of this oligomer induces a localised unwinding of DNA within the origin at an AT-rich region termed the DUE (for 'DNA Unwinding Element')[7,8]. DnaA then plays a role in recruiting the prossessive DNA helicase, DnaB,[9] which is loaded onto the unwound ssDNA by a helicase loader, DnaC[10]. DnaB subsequently recruits the primase, DnaG, and the polymerase β-clamp DnaN, which will recruit other components of the replication machinery ready for strand synthesis[11].

DnaA is is a member of the AAA+ ATPase family (ATPases associated with diverse cellular activities), able to bind and hydrolyse ATP, it exists in the cell as both an ATP- and an ADP-bound form[12]. However, DnaA-ATP is considered to be the 'active' form of the protein as

this is required for oligomerisation at the origin[13,14], triggering DNA unwinding and ultimately, assembly of the replisome.

The DnaA-boxes within the origin of replication vary in their affinity for DnaA, based on their similarity to a consensus binding sequence and on the adenosine nucleotide bound state of DnaA[15,16]. The DnaA-boxes at the origin can be either 'strong' or 'weak'; where strong boxes bind both DnaA-ATP and DnaA-ADP with equal affinity and 'weak' boxes have a much greater affinity for DnaA-ATP (over DnaA-ADP)[17]. In order for the helical DnaA oligomer to form at the origin and induce DNA unwinding, both strong- and weak- DnaA boxes need to bind DnaA[18,19]. In the *E. coli* origin (Fig. 1.2), DnaA-boxes are distributed such that the 3 strong boxes lie at the extreme ends of the origin and in the centre. As DnaA-ATP recruitment to the origin has been shown to be co-operative, these 'strong'-boxes are thought to form anchoring points from which the DnaA oligomer can grow[18,20]. In this model, DnaA-ATP is recruited to weak-sites via co-operative interactions with DnaA-ATP molecules already bound to neighbouring sites[20].

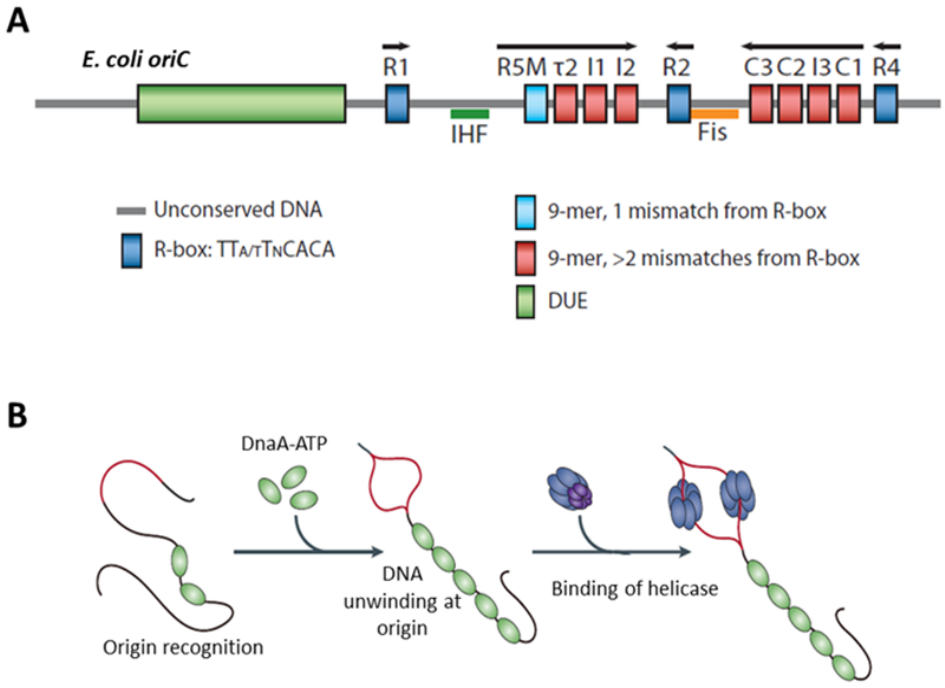


Figure 1.2: DNA replication initiation at *oriC* A) The *E. coli* origin of replication (*oriC*) contains a series of DnaA-boxes with differing affinity for DnaA (R-boxes have the consensus sequence, I and C boxes contain mismatches), a DNA unwinding element (DUE) and binding sites for accessory proteins (IHF/Fis). Figure A adapted from Leonard and Grimwade, *Front. Microbiol.* 6 1-13 (2015). **B)** DnaA recognises binding sites on *oriC,* forming a nucleoprotein complex which induces unwinding at the DUE. The DNA helicase and other components of the replication machinery are subsequently recruited. Figure B adapted from Mott and Berger, *Nat. Rev. Microbiol*. 5 343-354 (2007).

When the oligomer has formed at the origin, localised strand unwinding occurs at the DUE (Fig 1.3). Based on structural work carried out with *Aquifex aeolicus* DnaA, unwinding is thought to be directly mediated by the DnaA-oligomer, which introduces positive writhe in the bound DNA[4]; compensatory negative writhe at the DUE could subsequently cause DNA unwinding[5,21]. Unwound DNA is then stabilised by binding to DnaA at a ssDNA binding site in its ATPase domain[22,23].

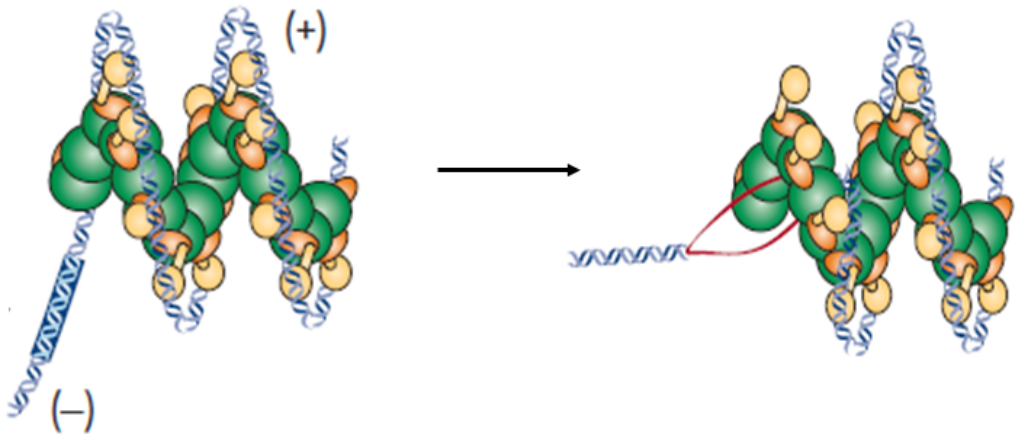


Figure 1.3: Model of DNA replication initiation at *oriC*: DnaA-ATP forms a right-handed helical oligomer at *oriC* introducing positive writhe in the bound DNA. Compensatory negative writhe destabilises the DUE causing DNA unwinding. The DnaA ATPase domain binds to single-stranded DNA, stabilising the unwound DUE. Figure adapted from from Mott and Berger, *Nat. Rev. Microbiol*. **5** 343-354 (2007).

Following the unwinding of the DUE, a homohexameric DNA helicase, DnaB, is loaded onto single stranded DNA at the replication origin by the action of a helicase loader protein, DnaC. Loading of the helicase onto ssDNA is understood to occur via a ring-breaking mechanism whereby DnaC forms a spiral oligomer which remodels the homo-hexameric DnaB ring, to produce a break in the ring which is large enough to allow loading onto ssDNA[24]. The DnaB-DnaC complex is recruited to the origin by DnaA. An interaction between the N-terminal domain of DnaA and the helicase, DnaB[9,25,26] is thought to create the correct orientation for loading of DnaB onto the bottom strand, whilst an interaction between the AAA+ domains of DnaA and DnaC is thought to recruit the complex in the correct orientation for DnaB loading on the upper strand (Fig. 1.4).[24,27] An interaction with the N-terminal domain of DnaB subsequently recruits the primase, DnaG. Active primer formation then appears to induce the dissociation of DnaC, a step which is necessary for

DnaB to function as an active helicase. Release of DnaC appears to be dependant on the ATPase activity of DnaC and is suggested to be induced by a confromational change in DnaB during primer formation[28] (DnaG interacts with the N-terminal domain of DnaB, whilst DnaC interacts with the C-terminal domain)[29].

The loading of the helicase is also important for the recruitment of the DNA polymerase clamp, DnaN. The clamp will, in turn, recruit the DNA polymerase, ready for elongation[29,30].

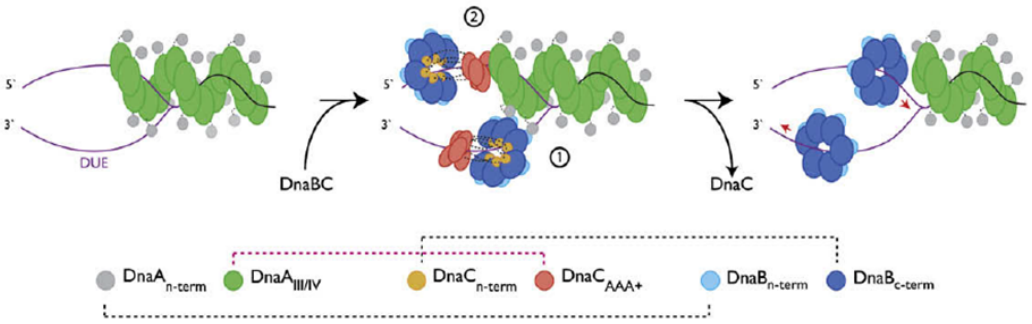


Figure 1.4: DnaC loads DnaB via a ring breaking mechanism: The initiator DnaA forms a helical oligomer during initiation which associates with the upper strand of DNA. Following unwinding of the DUE, interactions between DnaA and DnaB or DnaC in the DnaC-DnaB complex are thought to correctly orientate DnaB for loading onto the bottom and top strands of DNA, respectively. Figure taken from Mott *et al. Cell* 135, 623-34 (2008).

1.5.2: DNA Replication Initiation in *Bacillus*

Despite the high conservation of replication initiation proteins between bacterial organisms, significant differences between bacterial origins have been observed[31], as well as differences in the accessory proteins involved in initiation[3]. *Bacillus subtilis,* like *E. coli,* utilises the initiation proteins DnaA, a DnaB-like helicase known as DnaC, a helicase loader, DnaI, and a primase, DnaG[32]. *B. subtilis* DnaA has also been shown to form helical oligomers on both double and single stranded DNA[33] and requires DnaA-ATP for co-operative binding to the origin[34]. *Bacillus anthracis* DnaA displays an ATP-dependent variable affinity for DnaA-box sequences[35]. Together these findings suggest *Bacillus* DnaA functions at *oriC* in a similar manner to *E. coli* DnaA. However, in contrast to the single origin found in *E. coli*, *B. subtilis* has a bipartite origin whereby DnaA boxes are located on either side of the *dnaA* gene (see section 1.9). *B. subtilis* initiation also requires the roles of two additional essential proteins, DnaD and DnaB[32,36], both of which exhibit DNA remodelling properties[37] and bind to the origin upstream of helicase loading[38]; their role

and architecture is discussed below in section 1.11. For clarity, essential components of the *B. subtilis* DNA replication machinery and their *E. coli* equivalents are listed below in Table 1.1.

Table 1.1: The essential DNA replication initiation machinery of *B. subtilis* and *E. coli*

Role in DNA Replication Initiation	***B.subtilis***	***E. coli***
Initiator	DnaA	DnaA
Helicase	DnaC	DnaB
DNA Remodelling	DnaB, DnaD	_
Helicase Loader	DnaI	DnaC
Primase	DnaG	DnaG

Helicase loading in *B. subtilis* is thought to occur via an alternative mechanism to *E. coli* known as ring assembly[30]. In this model, DnaI facilitates the assembly of DnaC onto ssDNA[39,40]. This is supported by the fact that, in the presence of DnaI, hexameric DnaC shows no helicase or translocase activity whilst monomeric DnaC shows both helicase and translocase activities[39]. The helicase loader DnaI, like *E. coli*'s loader protein, contains an N-terminal helicase interaction domain and a C-terminal AAA+ domain[41]. The ATPase activity of DnaI's C-terminal domain is stimulated in the presence of ssDNA, but only in the absence of DnaI's N-terminal domain. Thus, it seems that binding of the N-terminal domain of DnaI to the helicase DnaC reveals a cryptic ssDNA binding site on the C-terminal domain[40]. It is thought that this can then stimulate helicase loading onto ssDNA and the ATPase activity of the C-terminal domain may stimulate the release of DnaI once loading has occured[40].

1.6: The Elongation Phase

During the elongation phase of DNA replication, DNA is processively synthesised by the action of a large multi-subunit complex known as the replisome (Fig. 1.4, A). This consists of three DNA polymerase complexes, the helicase (homohexamer), primase (assumed from structural studies to comprise of 3 subunits[42]), three processivity clamps, DnaN (only two of which are associated with the core replisome) and a pentameric clamp loader complex[43].

In E. coli, following initiation recruitment of the primase, DnaG, occurs via an interaction with the helicase DnaB; this causes the helicase loader, DnaC, to dissociate. A short RNA primer is synthesised by DnaG, and the processivity clamp DnaN is loaded by the clamp loader complex, followed by DNA pol III. These RNA fragments are extended by DNA pol III, but later replaced with DNA by DNA polymerase I and fused with adjacent DNA fragments by DNA ligase[2,44,45].

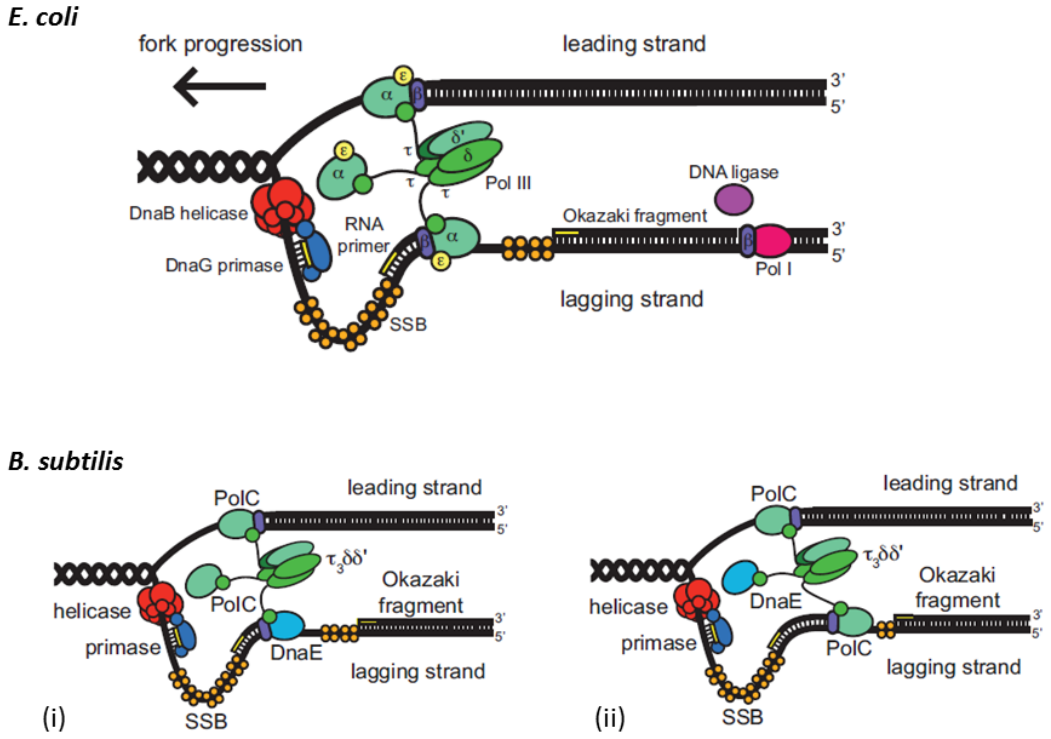


Figure 1.5: Replisome components in *E. coli* and *B. subtilis*: A) Schematic representation of the *E. coli* replisome showing locations of the helicase, primase, DnaA pol III, the-β clamp and clamp loader ($\tau_3\delta\delta'$) at the replication fork. **B)** Schematic representation of the *B. subtilis* replisome showing locations of the helicase, primase, DnaA pol III, the-β clamp and clamp loader ($\tau_3\delta\delta'$) at the replication fork. In (i) DnaE is shown extending the initial RNA primer with DNA and in (ii) PolC is shown extending that DNA further along the Okazaki fragment. Figure adapted from Robinson *et al*. *Current Drug Targets* **13,** 352–372 (2012).

The helicase forms a homohexameric ring which is understood to sit at the head of the replication fork on the lagging strand of the template DNA. The helicase mechanically separates dsDNA by translocating along the lagging strand template – a process driven by ATP-hydrolysis. Separated DNA strands are coated in single-stranded binding protein, SSB, which prevents strands from re-annealing and offers protection of the ssDNA from nucleases[2,44,45].

The primase, DnaG, contains three functional domains; an N-terminal Zn-binding domain (ZBD), a central RNA polymerase domain (RPD) and a C-terminal helicase binding domain. Three DnaG molecules associate with the N-terminal domain of the helicase, positioned such that the primase captures ssDNA which has been newly unwound by the helicase, ready for primer formation[46]. The primase contains a ssDNA binding groove which is thought to interact non-specifically with ssDNA, allowing the primase to track along the ssDNA and orientate it correctly for entry into the active site in the RPD (where primers are synthesised from available rNTPs); this also allows extrusion of the newly synthsised primer on the outside of the DnaB-DnaG complex, ready for handoff to SSB and DNA polymerase[46]. Whilst the RPD contains the catalytic site for RNA primer synthesis, the ZBD is responsible for modulating the activity of the RPD. Interestingly, the ZBD does not regulate the activity of the RPD intramolecularly on the same DnaG molecule in *cis*, but rather, regulates the RPD of an adjacent primase in *trans*[47]. A RPD and ZBD from separate molecules recognise a single ssDNA template to intitate primer synethesis at specific trinuceotide regognition sites; with the ZBD increasing the catalytic activity of the *trans* RPD, but also restricting processivity and primer length[47].

Strand extension of the primer is carried out by DNA pol III, which has a subunit structure of αεθ (where α is the catalytic subunit, ε is responsible for its proofreading function and θ is a non-essential subunit thought to stimulate the activity of ε). It extends the primer with the assistance of the processivity clamp, DnaN (also known as the β-clamp). DnaN is thought to sit directly behind DNA pol III, and forms a closed ring structure on the DNA which is formed of a homodimer of C-shaped monomers. When binding dsDNA, DnaN is thought to minimize its interactions by binding across the major and minor grooves of DNA (and not within the DNA groove), allowing the protein to essentially slide along the DNA. The β-clamp allows the polymerase to synthesise up to 1000 bases a second[2,44,45].

Each new lagging strand requires the loading of a new β-clamp; thus the clamp loader complex forms part of the replisome machinery. The clamp loader is a pentameric complex of $\tau_3\delta\delta'$. The τ subunit, the product of the gene DnaX, forms interactions with both the DNA helicase and DNA polymerase III – this is thought to play an architectural role at the replisome and couple DNA unwinding with DNA extension[2,44,45].

Elongation in *B. subtilis* occurs by a similar mechanism; however it uses two different (but related) DNA polymerases, PolC and DnaE for DNA synthesis (Fig. 1.4, B). (DnaE is more closely related to *E. coli* DNA pol III than PolC)[48]. Both polymerases have been shown to be essential for lagging strand synthesis, whilst only PolC is required for leading strand synthesis[49]. Both polymerases are able to extend DNA primers, but only DnaE is able to extend RNA primers produced by primase, DnaG[49]. During the handover of RNA primers, DnaE appears to form a ternary complex with DnaG and the helicase, DnaC[50]. In reconstituted in vitro assays, DnaC acts to stimulate the activity of DnaG, and DnaE is able to extend very short primers of only 2 or 4 nulecotides produced by DnaG; the abunance of which are increased by the presence of DnaC. Both DnaC and DnaG were shown to increase the fidelity of DnaE, in a nucelotide-dependant manner[50]. DnaE is thought to subsequently pass strand synthesis over to PolC, which is more procesive[49]. This is analogous to a system in eukaryotes where the DNA polymerase α extends RNA primers with DNA, before the lagging strand polymerase δ continues DNA synthesis[44].

1.7: Termination of DNA replication

1.7.1: Termination of DNA replication in *E. coli*

The termination of DNA replication occurs at the replication terminus – a locus positioned directly opposite *oriC.* In *E. coli*, this region is flanked on either side by five non-palindromic 23-bp sites, *Ter*A-J, which bind the monomeric protein Tus (Terminator utilisation substance)[51–53]. The orientation of these *Ter* sites dictates whether or not a travelling replication fork is able to pass the site or is halted in DNA replication[51,53]. Thus, a replication fork can bypass a *Ter* site unimpeded when travelling in the permissive direction, but is blocked when travelling in the non-permissive direction. For example, in Fig. 1.5, A, a replication fork travelling clockwise would bypass *TerH*, *TerI*, *TerE*, *TerD* and *TerA*, but would be halted at *TerC* (or failing that, at *TerB*, *TerF*, *TerG* or *TerJ*). Tus is a 36 kDa protein which specifically binds *Ter* sites in an asymmetric manner[54] (Fig. 1.5, B); collision with DnaB (*E. coli* helicase) causes Tus to rapidly dissociate in the permissive direction, however in the non-permissive direction Tus-*Ter* forms a roadblock which prevents the translocation of DnaB (and thus the replication fork)[55]. Tus is has been shown to function like a 'molecular mousetrap' at *Ter*; the trap is set by asymmetric binding of Tus to dsDNA in the non-permissive orientation, and subsequent strand unwinding by the oncoming machinery 'triggers' the trap causing a specific cytosine base at position 6 of the *Ter* site to flip into a binding site on Tus, forming a 'locked' Tus-*Ter* complex (Figure 1.5, B) and thus formation of a roadblock which prevents the progression of the replication fork[55,56].

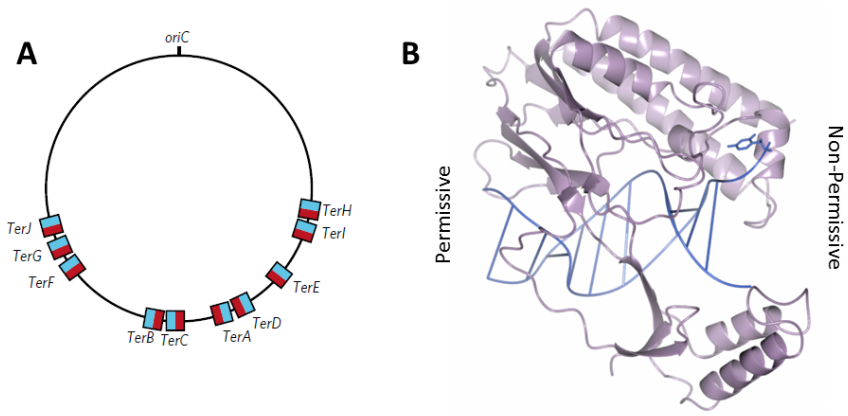


Figure 1.6: *Ter* sites on the *E. coli* chromosome and the Tus-*Ter* complex: A) Location and orientation of *Ter* sites in *E. coli*: permissive face shown in blue, non-permissive face shown in red. Figure adapted from Berghuis *et al. Nat. Chem. Biol.* **11** 579-585 (2015). **B**) Structure of the Tus-*Ter* complex showing permissive face (left) and non-permissive face (right). On the non-permissive face a specific cytosine base flips into Tus when dsDNA is unwound by the oncoming replication fork, creating a 'locked' complex.

1.7.2: Termination of Replication in *B. subtilis*

Replication termination in *B. subtilis* also occurs by a polar mechanism in which the *Ter* site can be approached from a 'permissive' or 'non-permissive' direction. In this case, the binding of two homodimers of the replication termination protein (RTP) at 'A' and 'B' sites within the *Ter* region are required to arrest replication (Figure 1.6, A)[57,58]. The approach of the replication machinery from the 'B' site results in termination of replication (non-permissive direction) whilst approach from the 'A' site allows replication to continue (permissive direction). The crystal structure of a single dimer bound to the native 'B' site has been shown to display asymmetry in the 'wing' region of the winged-helix domain[59] (Figure 1.6, B); the protomer which would lie adjacent to the A-site shows a 'wing-up' conformation, whilst the other protomer displays a 'wing down' conformation, making different contacts with the dsDNA. This asymmetry alone may provide a mechanism to allow 'permissive' and 'non-permissive' directions. However, binding of the A site is also required to block replication fork progression and binding to the A-site has been shown to be co-operative following B-site binding[60], thus the molecular mechanism RTP uses to block replication progression in *B. subtilis* remains unknown. Although the molecular details of *E. coli* and *B. subtilis* replication termination mechanisms vary, they appear to have evolved conceptually similar mechanisms for terminating replication in a direction specific manner.

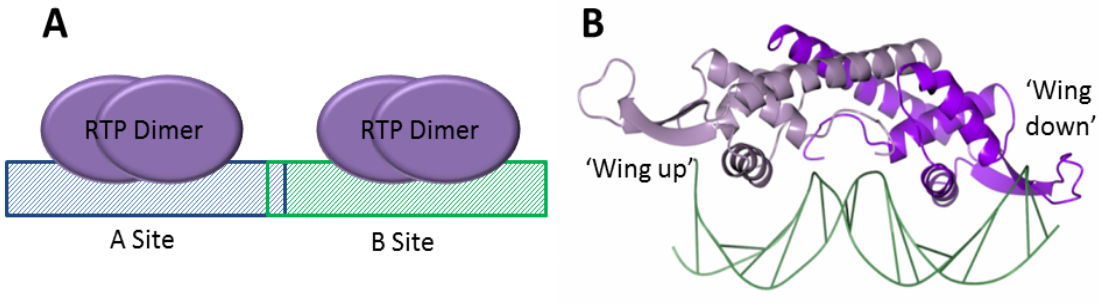


Figure 1.7: RTP binds as a dimer to A and B sites in *B. subtilis* replication terminus : A) Schematic image of two RTP dimers binding at the A and B sites of the terminus region. **B)** Structure of an RTP dimer bound to dsDNA with the sequence of a B-site region; one molecule displays a 'wing up' conformation (adjacent to the A site) and the other a 'wing down' conformation.

1.8: Regulation of DNA Replication

DNA replication is tightly regulated to ensure that it occurs in an accurate and timely manner, only once per cell cycle. Over- or under-replication of the organism's genome is detrimental to the organism's survival. Under-replication leads to cells which may be missing vital genetic information, whilst over-replication can lead to wasted resources and dysregulation of other cellular processes. The regulation of DNA replication occurs mainly at the initiation phase, during or prior to the recruitment of the replication machinery (once elongation commences it will progress in a largely uncontrolled manner). Regulatory mechanisms typically affect the ability of DnaA to form a nucleoprotein complex at the replication origin. Thus, an understanding of replication origins and the proteins involved in replication initiation is important for understanding regulatory mechanisms of DNA replication; these will first be covered in more detail below.

1.9: Replication Origins

Regulation of DNA replication often involves the prevention of DnaA or other initiation proteins from binding the origin (i.e. prevention of orisome formation)[61,62]. Thus, understanding of replication origins is of significance to our understanding of how replication is regulated[31]. Despite this, our knowledge of replication origins is relatively limited. The replication origins of many species have been predicted by computational programs, however only a few of these have been functionally confirmed *in vivo*; of these, the origins vary significantly in length, sequence and organisation (e.g. number and spacing of DnaA-boxes and the location of unwinding elements)[31,63]. Conserved features of replication origins include the presence of DnaA-box clusters, a DNA unwinding element (DUE) and locations for the binding of regulatory proteins[31].

The location of the replication origin and its gene context is well conserved across bacterial species, with most flanked by (or containing) the *dnaA* gene[32,64]. The genes surrounding *oriC* and *dnaA* are typically also well conserved with *oriC* residing in one or two intergenic regions[32]. Origins are described as either continuous or bipartite depending on whether all of their functional elements are found within one or two intergenic regions. For example, the origins of *B. subtilis* and *H. pylori* (Fig. 1.7) are bipartate, containing two DnaA-box clusters, separated by the *dnaA* gene[65,66]. *E. coli* is representative of a continuous origin,

but unusually, the *E. coli* origin has undergone a major rearrangement resulting in a translocation of the origin 44 kb away from the *dnaA* gene (and surrounding context)[64]. This suggests that *B. subtilis* may provide a better model for bacterial replication origins in general; representing a more primordial origin with conserved gene context[32]. The bipartate origin structure in *B. subtilis* has been shown to be important for proper replication initiation, however, the region could be shortened considerably[65]. Notably, during replication initiation, *B. subtilis* forms looped structures which are thought to be a consequence of the bipartite nature of its origin[67].

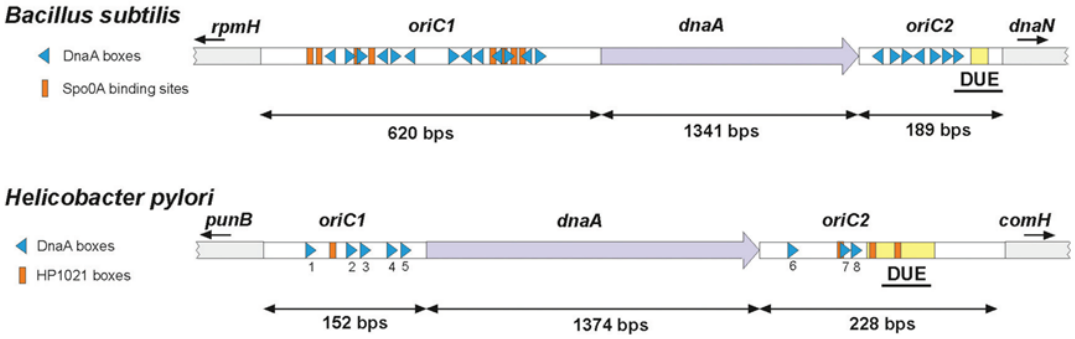


Figure 1.8: The bipartite replication origins of *B. subtilis* and *H. pylori*: In these origins, two clusters of DnaA-boxes are separated by the *dnaA* gene and the DUE lies at the end of one of the DnaA-box clusters. Figure adapted from Wolanski *et al. Front. Microbiol.* **5** 1-14 (2015).

One conserved feature of replication origins is the existence of a series of recognition sequences (or 'affinity boxes') for DnaA; although the number and spacing of DnaA-boxes varies between organisms and in some more distantly related species DnaA-box consensus binding sequences are different[31]. In organisms where they have been characterised, DnaA-boxes are 9 base pair elements, with the exception of a 12 base pair box identified in *Thermotoga maritima*[31]. All boxes share a conserved core sequence, and most deviate from the *E. coli* consensus by only one or two bases[31] (NB: *B. subtilis* and *E. coli* share the same consensus box sequence of 5'-TTATCCACA-3'[12]).

DnaA-box spacing and orientation have been shown to be important for replication initiation in *E.coli*[62] (e.g. altering box-spacing by part of a helical DNA turn is highly detrimental to initiation, whilst altering by a single turn is without effect[68]). DnaA-box spacing preferences have also been observed for other bacteria (e.g. preferring two closely spaced boxes over a single box)[63,69]. The affinity of DnaA for DnaA-boxes within the origin may vary, as is seen in *E. coli* and *C. crescentus*[18]. In *E. coli* the presence of high- and low-

affinity boxes is an essential feature for orisome formation[18,20]. It is not known if this is a conserved mechanism for orisome formation across other bacterial species.

Another conserved feature of replication origins is the presence of DNA unwinding elements. These are AT-rich sequences which typically contain 13-nucleotide repeat sequences[70]. These regions are noticeably AT-rich compared to neighbouring sequences, giving them reduced thermo-stability. DUEs are always located up- or down- stream of a DnaA box cluster and proper spacing between the DUE and its neighbouring DnaA-box cluster is essential for DNA unwinding[70].

Although bacterial origins differ in their composition, all origins harbour sequences which direct the formation of replication complexes, DNA unwinding and regulatory activities in their own species. Cognate DnaA proteins and replication origins have likely co-evolved to optimise their interactions, creating genus and species specific instructions for the initiation of DNA replication[31].

1.10: DnaA: Structure and Function

The DNA replication initiator protein, DnaA, is responsible for the recognition of DnaA-boxes at the replication origin, forming a nucleoprotein complex whereby a helical oligomer of DnaA self-associates and induces a localised unwinding at the origin DUE, which leads to the recruitment of the DNA helicase and further replication machinery.

DnaA is a member of the AAA+ ATPase family, containing four distinct domains[71,72] (Fig. 1.8, A). The N-terminal domain I is an 'interaction domain' which has been shown to interact with various protein regulators of DnaA across different organisms and is thought to play a role in the self-assembly of DnaA at the origin. In *E. coli, it* also interacts with the helicase, DnaB. Domain II is a poorly conserved flexible linker of variable length, whilst domain III contains Walker A and B type ATPase motifs capable of binding and hydrolysing ATP. Domain III also plays a role in self-interaction/oligomer formation and in ssDNA binding. The C-terminal domain IV is a dsDNA binding domain which is responsible for DnaA-box recognition at the origin. Since its discovery, DnaA has been the subject of mutational and structural studies which have led to our current understanding of its function. The outcome of these studies, focusing on each of DnaA's domains in turn, will be described below.

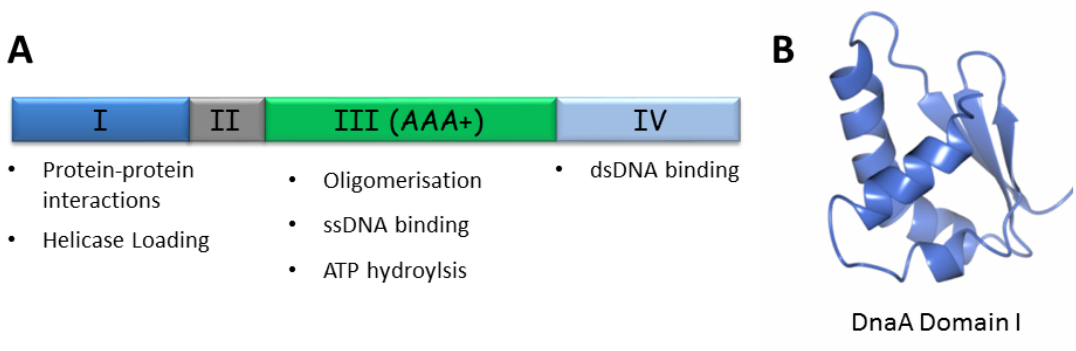


Figure 1.9: DnaA has four functional domains : A) Domain structure of DnaA. **B)** DnaA Domain I has a KH-domain fold similar to that found in some ssDNA binding proteins.

1.10.1: DnaA Domain I

DnaA Domain I has a K homology (KH)-domain fold typically found in ssDNA binding proteins[26,73,74] (Fig 1.8, B). Domain I is reported to play a role in the oligomerisation of DnaA at *oriC* [75,76] and in *E. coli*, it interacts with the helicase DnaB[77–79] where it is thought to help correctly orientate the loading of DnaB at the origin. Domain I has been recently termed an 'interaction hub' because in addition to the helicase, various protein regulators from different organisms have been shown to interact with the domain. These include *E. coli* DiaA[80] and *H. pylori* HobA[81] - structural homologues and promoters of initiation in their respective organisms and two negative regulators of initiation: *E. coli* Hda[82] and *B. subtilis* SirA[83] (see sections 1.13 and 1.14). DnaA Domain I has been shown to have a weak affinity for single-stranded *oriC* DNA but not poly(dT) DNA suggesting some sequence specificity to its binding[26]. However, no ssDNA binding role has been shown for DnaA *in vivo*.

1.10.2: DnaA Domain II

NMR studies suggest that domain II is unstructured and acts as a flexible tether between domains I and III. It is not completely dispensable for DnaA function, but it is poorly conserved and varies significantly in length between organisms[72]. One study in *E. coli* showed that deletions of 30-36 residues could be made at several positions along domain II (defined here as residues 79-135) without loss of cell viability, indicating that 21-27 residues is the minimum domain II length needed for DnaA functionality, and that no specific residues are required[84]. The authors did however note that deletion of 30 base pairs or more increases the doubling time of cells compared to wildtype, and it is only shorter deletions (17-19 residues) which show a phenotype indistinguishable from

wildtype. Consistent with this observation, another study suggests that domain II is important for replication efficiency, as spontaneous mutations causing a deletion within DnaA domain II were able to suppress a strain which showed poor growth due to an over-initiation phenotype[85]. In this study, *E. coli* cells lacking a regulator of DNA replication (SeqA) display slow growth and small colony formation, as well as severe over-initiation. A deletion mutation within DnaA domain II restored near-wildtype growth and DNA content. This behaviour mimics that of a deletion mutant of a stimulator of initiation (DiaA), suggesting that domain II is playing a role in replication efficiency[85].

It seems likely that the spacer is required to allow DnaA to adopt the correct conformation for initiation[84]; this would naturally play a role in replication efficiency. Although the spacer must be 21-27 residues long in *E. coli*, this number is likely to differ in other species where domain II has a highly variable length. One interesting theory is that domain II size may be related to the spacing of DnaA-boxes within the origin, with greater spacing between DnaA-boxes requiring a longer domain II [63]. If this is the case, perturbations in domain II size could affect the oligomerisation efficiency of DnaA – and hence the overall replication efficiency. However, further understanding of both the minimum domain II size of other species' DnaA proteins and of DnaA-box spacing at replication origins is required to draw any meaningful conclusions about this relationship.

1.10.3: DnaA Domains III-IV

Domain IV of DnaA is responsible for the recognition of DnaA-boxes at the origin[86]. In *E. coli*, DnaA has the highest affinity for the 9 base pair consensus sequence 5'-TTATNCACA-3' [87]. An X-ray structure of domain IV bound to a consensus sequence DnaA-box revealed that this interaction is mediated by a helix-turn-helix motif, which interacts primarily with the major groove of the dsDNA. Additional contacts are also made in the minor groove[88] (Fig 1.9, A). Base specific interactions could be observed at 8 of the 9 base pairs in the DnaA-box; the exception being the base pair at position 5, which has no specific base associated with the consensus binding motif. In the major groove, residues in helices 4 and 5 of Domain IV make base specific interactions with T4, C6, A7, G8* and T9* (where * indicates a base on the 3' to 5' strand). In the minor groove, a specific arginine residue (Arg399) makes several base-specific interactions; A2* forms a direct H-bond with the Arg399 side chain, and interactions with A1*, T2 and T3* and Arg399 are mediated via water molecules. Many interactions are also formed with backbone phosphate groups in both

major and minor grooves[88]. Mutations at residues involved in base-specific interactions result in loss of DnaA-box binding specificity, or eliminate DNA-binding completely[89]. Binding of domain IV introduces a local kink of 28° into the DNA, which may be significant for assisting DNA-unwinding[88].

Domain III of DnaA contains two distinct structural regions which are referred to as domain IIIa and IIIb. Domain IIIa has an αβ-structure and contributes to nucleotide binding via Walker A, Walker B and sensor I sequence motifs. Domain IIIb has a helical structure and contains the sensor II motif, which is also important in nucleotide binding. This structure is typical of AAA+ ATPases[90,91]. Unusually however, a crystal structure of ATP- bound DnaA domains III-IV (using a non-hydrolysable ATP analogue) shows that DnaA adopts an open spiral conformation[5] contrary to the planar ring conformations adopted by many AAA+ family ATPases[90] (Fig. 1.9, B). DnaA's spiral conformation is a right-handed helical filament, the formation of which is ATP-dependent and likely mimics the oligomer DnaA forms during initiation at the origin[4]. Adjacent protomers interact with one another via two clusters of conserved residues located on either side of the nucleotide binding pocket[5].

Crucially for DnaA's function, DnaA-ADP cannot form the right-handed oligomer displayed by DnaA-ATP[4], and this can be accounted for by a comparison of their structures. Whilst DnaA-ATP shows an open spiral conformation[5], DnaA-ADP appears to be a monomer[92]. A small but significant conformational change occurs upon ATP binding: repositioning of the 'lid' in domain IIIb. This appears to be mediated by interactions which are formed between the sensor II arginine and the γ-phosphate of the nucleotide. This opens the ATPase domain to allow a helix from an adjacent promoter to interact with the bound ATP via a conserved arginine residue, which seems to be an important factor for stabilising the DnaA helical filament[5]. This 'arginine finger' interaction is conserved across other AAA+ ATPases[93].

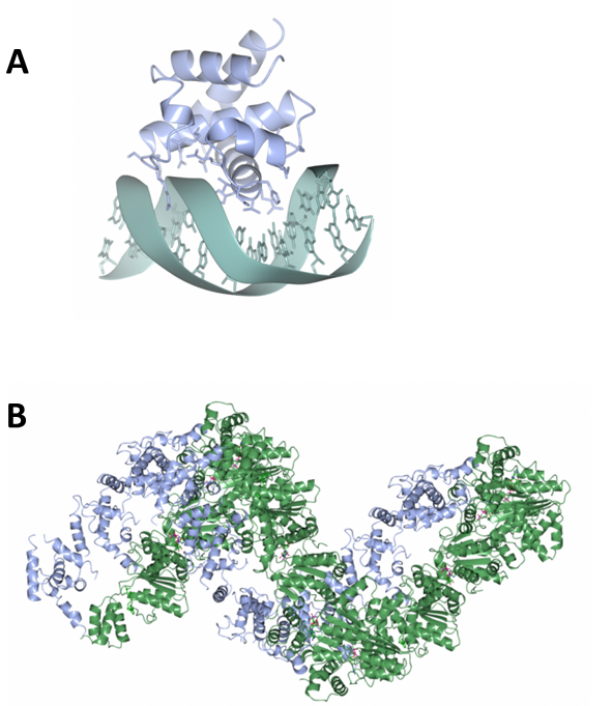


Figure 1.10: DnaA domains III and IV: DnaA-box recognition and self-oligomerisation: **A)** DnaA Domain IV bound to a DnaA-box motif; DnaA makes base specific interactions with 8 of the 9 DnaA-box bases. **B)** DnaA Domains III-IV bound to a non-hydrolysable ATP analogue, form a spiral structure which is thought to mimic DnaA oligomerisation at the origin. Domain III is displayed in green; Domain IV is in light blue.

Comparison of the Domain IV-DnaA-box binding modes and the domain III-IV structure suggests that a conformational change must occur at the linker helix between domains III and IV for dsDNA binding to be compatible with helical filament formation[88,94]. A significant kink in the linker helix is observed in the ATP-bound structure compared to the ADP bound form (where the helix is straight) suggesting that the two domains are 'conformationally uncoupled' and must rotate with respect to one another to allow filament formation at the origin[5].

A structure showing DnaA domain III's ssDNA-binding mode (also ATP-bound) displays the same spiral conformation as that seen for ATP-bound DnaA[23] (Fig 1.10, A). In this case, each DnaA protomer binds 3 nucleotides, making many interactions with the DNA phosphodiester backbone. Each nucleotide triplet displays a normal B-form DNA conformation, but the triplets are separated by gaps of approximately 10 Å creating an overall extended form of DNA[23] (Fig. 1.10, B). This strand extension was been shown to be ATP dependent in solution and is highly reminiscent of ssDNA binding displayed by the

homologous recombination protein, RecA. The third base of each triplet is rotated however, making bases in the DnaA-bound strand discontiguous; this presumably prevents re-annealing of the strand at the origin[23] (Fig. 1.10, C). DnaA is able to directly melt DNAs of moderate stability, enabling localised unwinding at the DUE. It is likely to be assisted in unwinding at the origin by DNA remodelling proteins (known to play a role in initiation). It is as yet unclear how DnaA co-coordinates both double and single-stranded DnaA binding modes for nucleoprotein filament formation and dsDNA unwinding at the DUE[95].

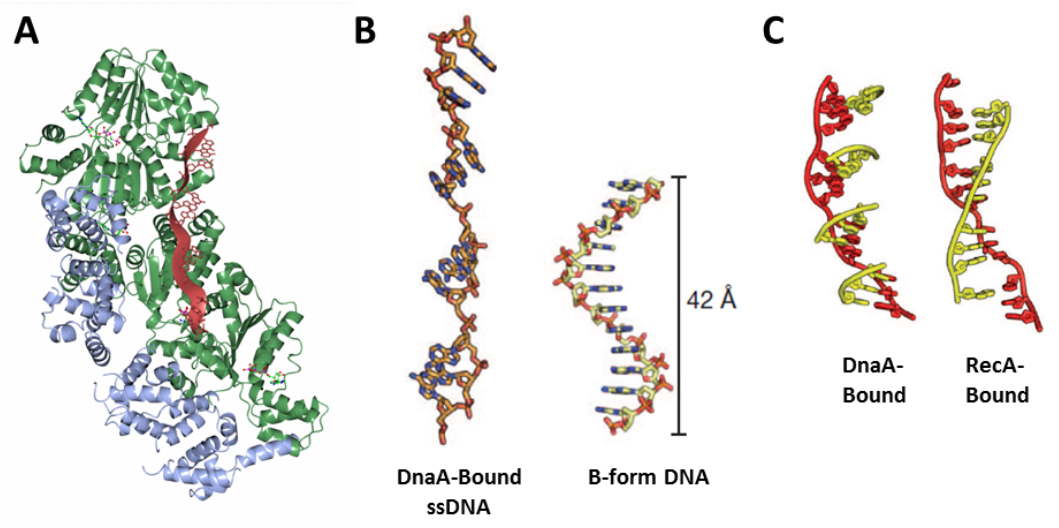


Figure 1.11: ssDNA binding by DnaA domains III-IV: **A)** ssDNA-binding mode of DnaA domains III-IV. Domain III mediates ssDNA DNA-binding. Domain III is in green, domain IV in light blue, and ssDNA in red. **B)** ssDNA binding by DnaA stretches the strand into an extended form compared to B-form DNA. **C)** Diagram showing how bound ssDNA (yellow) would pair with B-form DNA (red) when bound to DnaA and RecA. DnaA binds ssDNA so that triplets of bases are non-contiguous, assumed to prevent re-annealing of the strand at the origin. (B) and (C) were adapted from Duderstadt *et al*., *Nature* **478,** 209–213 (2011).

1.11: *B. subtilis* Initiation Proteins DnaD and DnaB

Besides DnaA, DnaC (equivalent to *E. coli* DnaB), DnaI and DnaG, DNA replication initiation in *B. subtilis* requires the presence of two additional essential proteins, DnaD and DnaB[32]. Both DnaD and DnaB are components of the replication initiation machinery at *oriC*[36] and of the replication restart machinery (DnaA-independent)[96]. Both proteins exhibit DNA remodelling activities[37] and share structural similarity[97]. *B. subtilis*' initiation machinery assembles in a specific hierarchical manner, with DnaD and DnaB required between DnaA binding at *oriC* and the loading of the helicase, DnaC[38]. DnaD is recruited after DnaA, forming a direct interaction with DnaA (and itself)[98]. DnaD is in turn required for the recruitment of DnaB, which then recruits DnaC-DnaI[38] and is thought to function as a dual helicase loader[39].

The exact roles of DnaD and DnaB in replication initiation remain somewhat unclear. DnaD displays DNA remodelling properties, able to untwist supercoiled DNA into an open looped form[99]. It forms tetramers which can assemble into large protein scaffolds that appear to mediate DNA loop formation and enhance melting of dsDNA[100]. Although the N-terminal domain (DDBH1) of DnaD is implicated in tetramer formation[101,102] and the C-terminal domain (DDBH2) in both double- and single-stranded DNA binding[100,102]. Full-length protein is required for DnaD to exhibit DNA remodelling activity (i.e. DNA looping and melting)[100,102]. Furthermore, the abundance of DnaD in the cell (estimated at 3000-5000 molecules[103]) has led to the suggestion that DnaD plays a global role in DNA remodelling (beyond that required for DNA replication initiation)[37]. This is supported by a recent study showing that DNA remodelling by DnaD stimulates DNA repair by Nth endonucleases in response to DNA damage by treatment with H_2O_2[104].

It is generally thought that DnaB acts together with DnaI to enable that loading of DnaC onto forked DNA[39]. However, a study by Ioannou *et al.*[40] suggests that DnaI alone is sufficient to load the helicase onto DNA. DnaB is however required to recruit DnaC-DnaI to the origin[38] and has been shown to stimulate the helicase and translocase activities of DnaC in the presence of DnaI[39]; thus it certainly assists with helicase function. It has also been implicated in association of the DNA replication machinery with the cell membrane[105,106] and additionally, has been shown to laterally compact DNA - although it is unknown how this contributes to its function[37].

Although DnaD and DnaB show little sequence similarity, analysis using a hidden markov model technique identified two shared domains known as DDBH1 and DDBH2. DnaD displays a DDBH1-DDBH2 architecture, whilst DnaB has a DDBH1-DDBH2-DDBH2 architecture[97] (Fig. 1.11, A). The structure of the DDBH1 domain of DnaD revealed a winged helix domain with two additional structural elements: an N-terminal helix-strand-helix and a C-terminal helix[101] (Fig. 1.11, C). The β-strand of the helix-strand-helix was found to mediate interactions between DnaD molecules in both dimer and tetramer interactions (Fig 1.11, C). The C-terminal helix has been shown to be important in higher-order oligomerisation (between tetramers)[101]. These structural elements appear to be present in DnaB DDBH1[97] and DnaB has also been shown to form tetramers mediated by its N-terminus[37], suggesting that DnaB and DnaD share similar oligomerisation properties.

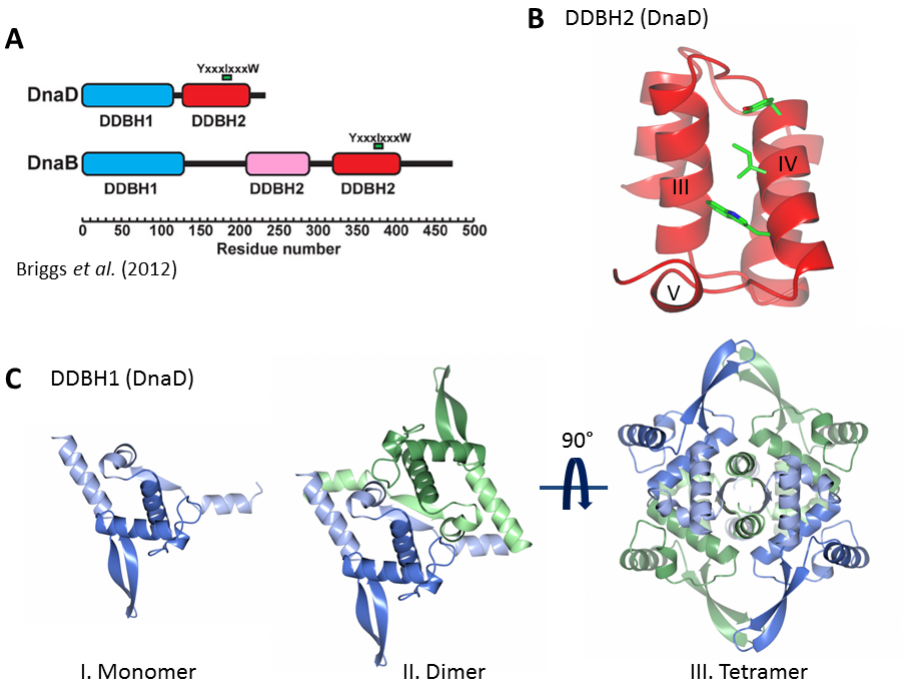


Figure 1.12: Architecture of DnaD and DnaB: A) Diagram showing architecture of DnaD and DnaB. Conserved DNA binding motif YxxxIxxxW is marked on the relevant DDBH2 domain. Image adapted from Briggs *et al.* (2012)[32]. **B**) Ribbon diagram of the DnaD DDBH2 domain from *Streptococcus mutans* (PDB code: 2ZC2). Tyrosine, Isoleucine and Tryptophan residues of the YxxxIxxxW motif are coloured by atom (carbon in green, nitrogen in blue and oxygen in red). **C**) Ribbon diagram of DnaD DDBH1 domain from *Bacillus subtilis* (PDB code: 2V79) showing a winged helix with additional structural elements; typlical winged helix architecture is shown in darker colours (blue, green) and additional structural elements are shown in lighter colours. I) Monomer of DDBH1; II) Dimer of DDBH1 and III) Tetramer of DDBH1. Dimer and tetramer interactions are mediated by the β-strand of the additional helix-strand-helix.

The DDBH2 domain of DnaD has been shown to be involved in DNA-binding and in DNA-dependent higher-order oligomerisation[102]. Two structures deposited by the Midwest Center for Structural Genomics of the DDBH2 domain of DnaD homologues from *Streptococcus mutans* (PDB code: 2ZC2) and *Enterococcus faecalis* (PDB code: 2I5U) show a compact helical structure with four longer helices I-IV and a short fifth helix (V) of only 4 residues (Fig. 1.11, B). Although residues after this point are poorly conserved across DnaD homologs, secondary structure prediction and analysis of the *Bacillus subtilis* DnaD DDBH2 domain by NMR suggests that helix V is longer than that shown in either of the X-ray structures, extending it by another 7 residues[97]. This final helix (V) is followed by a disordered region at the C-terminus. Structural elements which have been found to be important for ssDNA binding are a YxxxIxxxW motif residing in helix IV, the poorly conserved helix V and the C-terminal unstructured region[97]. These structural elements appear to be conserved in the second of the DDBH2 domains of DnaB[97], which has been implicated in both ds- and ss- DNA binding and in higher-order oligomerisation[107]; again suggesting that the domains play similar roles.

A fragment containing DnaB DDBH1-DDBH2 (residues 1-300 without the C-terminal DDBH2 domain) mediates tetramer formation, and is also able to bind ssDNA, suggesting the first DDBH2 domain may have evolved to bind only ssDNA[97,107]. Interestingly, the function of DnaB appears to be mediated by proteolysis from its C-terminus as truncated versions of the protein are seen in the cytosol during the mid-late growth phase (i.e. not membrane associated). Only full length DnaB is observed at *oriC*, thus proteolysis appears to be regulating DnaB's function.[107] It is unclear however whether the truncated version of the protein has its own function, or if its purpose is solely in self-regulation.[107] Nevertheless, the different DNA binding capabilities of the DDBH2 domains of DnaB may be important in differentiating the functions of the full-length and truncated versions of DnaB.

It has been previously noted that the requirement of two additional proteins in *Bacillus subtilis'* initiation of DNA replication, and their sequential recruitment at *oriC,* provides additional opportunities for the regulation of replication initiation – with opportunity for regulators to inhibit protein-protein interactions at each stage of initition[38]. Recent studies suggest that DnaD itself may indeed play a role in regulating DNA replication initiation as it has been found to affect the formation of the DnaA filament at *oriC.*[108,109] (see section 1.14.3).

1.12: Regulation of DNA Replication Initiation

It has long been recognised that DNA replication initiation is coupled to the cell cycle, ensuring that the frequency of genome duplication coordinates with the rate of cellular growth[110]. In rapidly dividing cells, DNA replication increases in frequency to support rapid growth[110]. Although the link between replication rate and cell cycle is well recognised, nutrient mediated regulatory control is not well understood[111]. It was previously thought that the cellular concentration of DnaA acted as a limiting factor for the rate of DNA replication initiation[112,113]. However, recent studies in both *B. subtilis* and *E. coli* have suggested that increasing DnaA concentration alone is not enough to trigger DNA replication initiation[113,114]. Nutrient mediated growth rate regulation was found to be dependent on both DnaA and *oriC* in *B. subtilis*[114] and there appear to be multiple regulatory mechanisms for coupling growth rate with replication; links with central carbon metabolism, respiration, fatty acid synthesis, phospholipid synthesis and protein synthesis were all reported[114].

1.12.1: DnaA Acts as a Transcription Factor

DnaA is known to act as a transcription regulator in *E. coli*[115], *B. subtilis*[116] and *C. cresentus*[117]. It has been shown to act in the auto-regulation of the *dnaA* gene promoter, reducing expression of DnaA when it is highly abundant in the cell[118]. It has also been shown to regulate the expression of many other genes in *E. coli* and *B. subtilis*[115,116] and a study into *B. subtilis* transcription regulation suggested that many DnaA-binding sites utilised for transcription regulation in *B. subtilis* are conserved in other organisms[116]. Furthermore, transcription regulation in *B. subtilis* responds to replication status to maintain cell viability (i.e. perturbations in replication elongation or initiation resulted in transcription regulation to inhibit cell division and maintain cell integrity)[116,119]. This provides the cell with a mechanism to couple DNA replication and the cell cycle. DnaA transcription mediated coupling of replication with the cell cycle has also been observed in *C. cresentus*[117,120]. These mechanisms are currently poorly understood, but are likely to play an important role in the coupling of DNA replication to the cell cycle.

1.12.2: Species Specific DNA Replication Regulation

A range of protein regulators and cis-acting DNA elements have been identified across various organisms which regulate the timing and frequency of DNA replication initiation[95,121]. However, all of the identified regulators are genus or species specific and the same regulators are only observed in very closely related bacteria. Notably, Gram-negative and Gram-positive bacteria appear to employ separate systems of regulation of replication initiation, with the model organisms *E. coli* and *B. subtilis* providing our most thorough understanding of the control of replication initiation in Gram-negative and Gram-positive bacteria, respectively[122]. Both organisms use a range of protein regulators and cis-acting DNA elements that have no known homolog in the other species. A comprehensive overview of these mechanisms is detailed below.

1.13: Regulation: *E. coli*

1.13.1: Regulatory Inactivation of DnaA (RIDA)

In *E. coli,* the concentration of available 'initiation-active' DnaA-ATP is considered an important limiting factor in the initiation of DNA replication from *oriC*[123], thus regulation of the available DnaA-ATP can be employed to control initiation events. The 'regulatory inactivation of DnaA' is a term given to the process of converting 'active' DnaA-ATP into 'inactive' DnaA-ADP by ATP hydrolysis. Hydrolysis is promoted by the protein Hda, and requires a complex between DnaA, Hda and the DNA-bound polymerase β-clamp, DnaN[124–126]. In this way, the regulation appears to couple initiation and elongation of DNA replication, as ATP hydrolysis is activated following the start of DNA synthesis[125].

Hda is homologous to DnaA domain III, with a 48% sequence similarity between the AAA+ ATPase domains[125]. Hda binds DnaN via an N-terminal clamp binding motif (sequence QL[SP]LPL)[127], whilst Hda:DnaA interactions are mediated via their respective ATPase domains[128]. Strains carrying a *hda* deletion, and inactivated Hda mutants or DnaA mutants unable to hydrolyse ATP, exhibit overinitiation of DNA replication and growth inhibition[13,123,124,129]. Hda mediated hydrolysis of DnaA-ATP to DnaA-ADP requires ADP-bound Hda[130] which has been shown to be monomeric. In contrast, apo-Hda appears to form homomultimers (mainly homodimers)[130]. Together this suggests that Hda's oligomeriation state plays a role in its ability to promote DnaA-ATP hydrolysis[130]. A structural study of Hda also showed the formation of dimers (where Hda is bound to the

nucleotide CDP). However because the DnaN binding motif is buried in this conformation, it is assumed to be an inactive form of the protein[131].

Mutations in both Hda and DnaA suggest that the proteins interact via their respective AAA+ domains, with Hda's arginine finger residue playing an important role in DnaA-ATP hydrolysis[128,132]. Models of the DnaA-Hda interaction suggest that it may form a similar interaction to that seen between molecules of DnaA[131,132], however Hda may make fewer contacts with DnaA than DnaA protomers[132]. Hda's interaction with the β-clamp is important for Hda-DnaA binding, suggesting that the β-clamp may adjust the conformation of the Hda-DnaA interaction to promote ATP hydrolysis[132]. Recently interactions of Hda with DnaA domains I and IV have also been shown to be important for RIDA as mutations at specific DnaA Domain I and IV residues leads to higher cellular concentrations of DnaA-ATP than seen in wildtype cells[82,133]. The Hda-DnaA model suggests domain IV makes contacts with Hda at a nucleotide interaction surface towards the C-terminus of Hda[133]. It has therefore been proposed that DnaA domain I interacts with the N-terminal portion of Hda's ATPase domain, to stabilise the DnaA-Hda interaction from both sides[82](Fig. 1.12). Further understanding of the interactions between Hda, DnaA and the β-clamp are required to fully elucidate Hda's molecular mechanism.

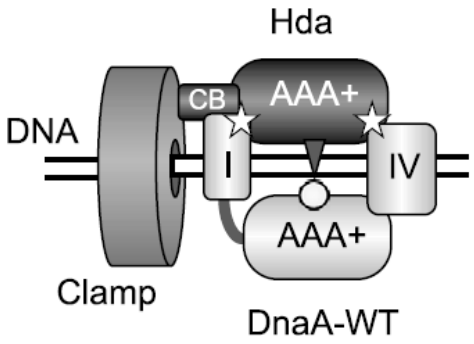


Figure 1.13: Schematic representation of the interaction of Hda with DnaA and the β-clamp:Hda makes contacts with Domains I, III and IV of DnaA, and the DNA-bound β-clamp during RIDA. Image taken from Su'etsugu, *et al*., *Environ. Microbiol.* **15,** 3183–95 (2013).

1.13.2: IHF and Fis

The DNA-bending proteins IHF and Fis are thought to play important roles in regulating the binding of DnaA at the origin of replication[134]. Further to this, they have been shown to play roles in the binding of DnaA to two cis-acting regulatory sites on the chromosome,

datA and DARS[135,136] (See below). Loss of IHF disrupts synchronous DNA replication. Curiously, Fis has been reported to play both inhibitory and stimulatory roles in DNA replication initiation[137–139].

Both proteins are able to bind *oriC* and have been shown to act in an antagonistic manner[134,140]. Binding of IHF to a specific site in the *E. coli* replication origin (see Fig. 1.2) promotes binding of DnaA at DnaA-boxes within the origin[141], contributing to DnaA oligomer formation. Fis has been reported to inhibit DNA unwinding at *oriC* by blocking binding of both DnaA and IHF[134]. Increasing concentrations of DnaA were found to relieve Fis inhibition, and IHF was found to redistribute DnaA molecules at *oriC*[134,141].

IHF binds to the origin between DnaA-boxes R1 and R5 during replication initiation (Fig. 1.2), causing a bend in the DNA; this DNA bending has been proposed to bring the R1 and R5 DnaA-boxes into close proximity with one another, allowing them to extend the helical DnaA oligomer[20]. Strains with a deletion of either of the high-affinity DnaA-boxes, R1 or R4, at the ends of *oriC* (Fig. 1.2), required both Fis and IHF for initiation; leading to the suggestion that the DNA bending exhibited by both proteins creates the correct conformation for an initiation complex[20] which is proposed to form a loop structure stabilised by DnaA molecules on high affinity DnaA boxes R1, R2 and R4 (Fig 1.13, A). Thus the authors suggest that Fis and IHF help to mimic this conformation in the absence of the R1 or R4 boxes (Fig. 1. 13, B) but that simultaneous binding of Fis and IHF may not be possible in the normal initiation complex – providing a mechanism for the dynamic behaviour of Fis and IHF[20].

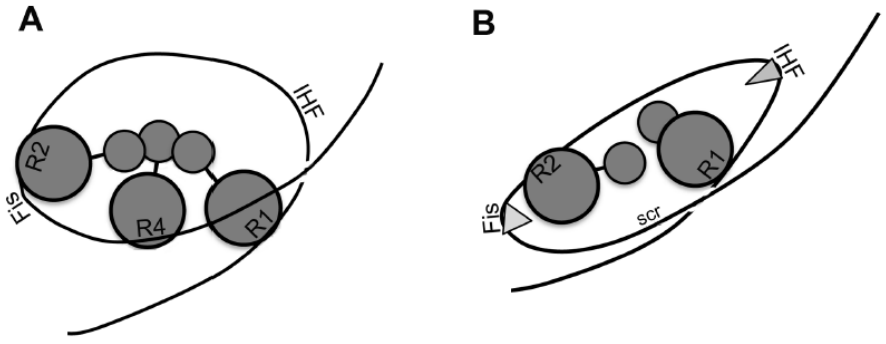


Figure 1.14: Schematic representation of the initiation complex: A) DnaA binding at DnaA-boxes R1, R2 and R4 on the *E. coli* replication origin are proposed to form a looped DNA structure. **B)** Binding of both IHF and Fis may act to mimic the looping of DNA at *oriC* in strains lacking the R4 (or R1) DnaA-box. Figure taken from Kaur *et al., Mol. Microbiol.* **91**, 1148–1163 (2014).

1.13.3: *datA*

datA is a chromosomal locus in *E. coli* around ~1 kb in length, located at 94.7 minutes on the chromosome[142]. It contains five DnaA-boxes and exhibits a high affinity for DnaA molecules. It acts as a negative regulator of DNA replication initiation; deletion of *datA* or mutations in DnaA-boxes within *datA* cause overinitiation of DNA replication[142–144]. *datA* is unique in this behaviour as other chromosomal DnaA-box sites (other than those within *oriC*) do not affect replication initiation events[144]. Binding of IHF to *datA* promotes DnaA-binding and is essential for the regulatory action of *datA* [145]. *datA* has mainly been suggested to act as a sink for DnaA molecules, titrating them away from the replication origin[142–145], however a recent study has revealed that *datA* promotes the hydrolysis of DnaA-ATP to DnaA-ADP [136]. This mechanism is dependent on both IHF binding to *datA* and the DnaA arginine finger residue (Arg285) suggesting that *datA* promotes the formation of a nucleoprotein complex, somewhat reminiscent of *oriC*, and stimulates the hydrolysis of DnaA-ATP at this site[136]. The binding of IHF was shown to take place immediately after initiation, providing a mechanism for the timing of *datA* mediated DnaA-ATP hydroylsis[136].

1.13.4: DARS

Two loci in *E. coli,* termed DARS1 and DARS2 for 'DnaA reactivating site' 1 and 2, respectively, have also been implicated in the reactivation of DnaA by promoting nucleotide exchange, generating DnaA-ATP from DnaA-ADP[146]. Deletion of DARS sequences causes inhibition of DNA replication due to a decrease in the cellular DnaA-ATP concentration[146]. DnaA-ADP molecules have been shown to assemble on DARS1 promoting the regeneration of DnaA-ATP[146]. The simultaneous binding of the DNA bending proteins IHF and Fis to DARS2 has been shown to facilitate DnaA-ATP regeneration *in vivo*[135], providing a mechanism for the timing of DnaA reactivation. The binding of IHF to DARS2 appears to be cell-cycle regulated and not dependent on DNA replication, whilst the binding of Fis is linked to growth phase: occurring during exponential growth but not stationary phase[135]. Fis' role in DARS2 stimulation is consistent with a report that Fis is required for the stimulation of replication initiation in rapidly growing cells[139].

1.13.5: Dynamics of IHF and Fis Binding

There is an increasing appreciation of the importance of the roles that IHF and Fis play in regulating DNA replication initiation, both through the regulation of available DnaA-ATP (*datA*/DARS2)[135,136] and via direct binding to the replication origin[134]. It will be interesting to understand how the timing of IHF and Fis binding at *oriC*, *datA* and DARS is controlled and how these dynamics link to replication, cell and growth cycles.

1.13.6: SeqA

SeqA negatively regulates DNA replication initiation by preventing re-initiation of replication for an intermediate period just after replication has occurred. It binds to hemi-methylated GATC sites in *oriC* which act to sequester the origin, preventing DnaA oligomer formation and transcription of the *dnaA* gene by blocking binding of the *dnaA* promotor[147–150]. The *E. coli* replication origin contains 11 GATC sites which will be hemi-methylated immediately after DNA replication. This is due to the semi-conserved nature of DNA replication: the newly synthesised strand will be unmethylated whilst the parental DNA strand will be methylated. SeqA binds with strongest affinity to hemi-methylated GATC sites[151]. The SeqA molecules bound at GATC sites on *oriC* are able to recruit further SeqA proteins to sequester the origin, as long as six or more of the GATC sites are hemi-methylated[152]. Sequestration of the origin persists for around 1/3 of the cell cycle, although it is not fully understood how SeqA is released after this time[149,153]. Dam methyl transferase methylates adenosine bases in GATC sites, and it seems likely that spontaneous dissociation of SeqA coupled with Dam methylation relieves sequestration[154]. SeqA has also been implicated in faithful chromosome segregation[155].

1.13.7: DiaA /HobA

DiaA is a positive regulator of DNA replication initiation in *E. coli*. It influences the frequency and timing of DNA replication initiation[156]. It is thought to function by promoting the oligomerisation of DnaA at the origin, by binding to domain I of DnaA[73,157,158]. HobA, a structural homolog of DiaA in *H. pylori*[159], has been shown to be an essential regulator of DNA replication[81]. Both proteins form tetramers, and a HobA-DnaA domain I structure revealed a 4:4 HobA:DnaA domain I complex, where each HobA protomer binds to one DnaA domain I[73]. DiaA has been shown to bind an equivalent site on DnaA[80]. Despite this,

HobA and DiaA are not interchangeable *in vivo* due to differences in their cognate DnaA domain I sequences, which appear to have co-evolved with the regulators[157]. Formation of heterologous complexes can be achieved however with hybrid DnaA molecules in which the regulators' cognate DnaA domain I was exchanged with the heterologous species' DnaA domain I (i.e. *E. coli* DiaA can interact with a chimeric DnaA protein containing *E. coli* domain I fused to *H. pylori* domains II-IV). This confirmed that DiaA and HobA are functional homologs, each promoting DnaA binding at the origin, but with different dynamics (HobA accelerates DnaA binding, whilst DiaA decelerates DnaA binding)[157]. Structural and mutational studies with HobA have led to the suggestion that the tetramers function as molecular scaffolds at *oriC* which promote the formation of DnaA oligomers, however direct experimental evidence is still required[160]. DiaA may play a role in regulating the timing of helicase loading in *E. coli* as both DiaA and the helicase appear to bind an overlapping site on DnaA, and DiaA can inhibit helicase loading *in vitro*[80].

1.14: Regulation: *B. subtilis*

1.14.1: YabA

YabA is the principle regulator of DNA replication initiation in vegetatively growing *B. subtilis* cells[161]. YabA is a negative regulator of DNA replication; a deletion of yabA (Δ*yabA*) causes an increased frequency of initiation events and asynchronous DNA replication[161] as well as a growth phenotype associated with increased initiation events[162,163]. YabA interacts with both the replication initiator, DnaA, and the DNA polymerase clamp, DnaN, as revealed by yeast two hybrid studies and a Zn^{2+} precipitation assay[162,163]. Mutations which affect the interaction of YabA with either DnaA or DnaN in a yeast two hybrid assay have been shown to act like Δ*yabA* mutants in *B. subtilis*: resulting in increased initiation and asynchronous DNA replication[163]. Furthermore, these mutations have been shown to disrupt the formation of GFP-YabA foci, which typically form at mid-cell (suggestive of localisation with the replisome)[164]. Together this suggests that the interactions between YabA, DnaA and DnaN are physiologically relevant.

As a regulator interacting with DnaA and DnaN, YabA is reminiscent of the *E. coli* replication regulator Hda, which acts in the "regulatory inactivation of DnaA" by promoting the hydrolysis of ATP in DnaA (and thus 'inactivating' DnaA)[125]. However, YabA is a much smaller protein than Hda, shares no apparent homology with Hda (or DnaA), and does not promote DnaA-ATP hydrolysis *in vitro*[109]. Thus, it appears to employ an alternative

regulatory mechanism for the inhibition of DNA replication, which also couples regulation of replication initiation with elongation through interactions with DnaA and DnaN.

The nature of YabA's interaction with DnaA and DnaN is debated in the literature[34,164]. YabA is able to form a simultaneous interaction with both DnaA and DnaN in a yeast 3 hybrid experiment[163], however it remains unclear whether or not this ternary complex is physiologically relevant. In particular, the role DnaN plays in this system is debated. One model (Fig. 1.14, A) suggests that DnaN tethers DnaA to the replisome via an intermediary interaction with YabA (also giving physiological significance to a ternary complex)[164]. Another model however, suggests that DnaN removes YabA from *oriC*, allowing cooperative binding of DnaA at *oriC*[34] (Fig. 1.14, B). The first model suggests that DnaN enables YabA's ability to inhibit DnaA, whilst the latter model suggests DnaN antagonises YabA's inhibition of DnaA.

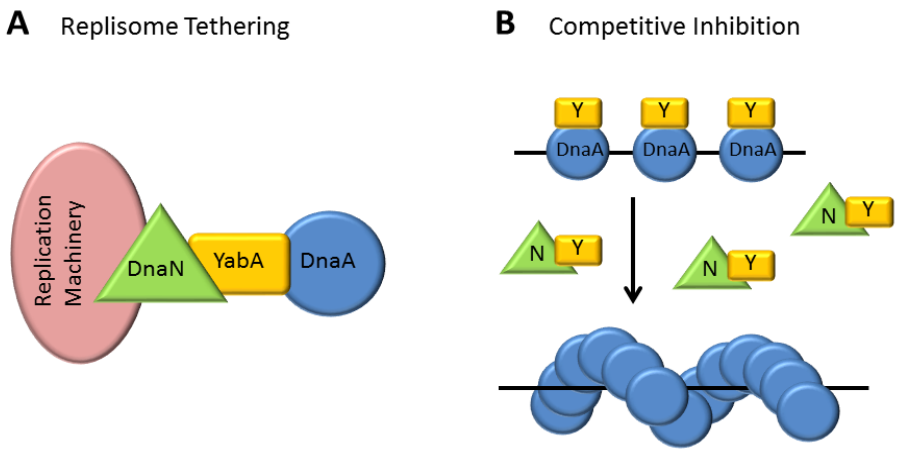


Figure 1.15: Alternative models for YabA's inhibition of DNA replication: A) YabA tethers DnaA to the replisome via an interaction with both DnaA and DnaN, titrating DnaA away from the replication origin. **B)** YabA inhibits the cooperative binding of DnaA at the origin during replication. When DnaN is released after replication, YabA binds DnaN, releasing DnaA.

The first model (Fig. 1.14, A) is based on the localisation of DnaA in wildtype and Δ*yabA* cells. In wildtype cells, DnaA localises at the origin in small cells (early cell cycle) and at the cell centre (co-incident with DnaX - i.e. replisome localised) in larger cells (remainder of the cell cycle)[164]. However, in Δ*yabA* cells, DnaA localised with the origin throughout the cell cycle. This led to the proposal of a tethering model, whereby YabA tethers DnaA to the replisome via DnaN for most of the cell cycle (thereby sequestering DnaA away from the

origin and preventing re-initiation)[164]. However, an earlier study suggested that YabA co-localises with DnaN only at late stages of the cell cycle[161] and recent studies have shown that increased levels of DnaN in the cell cause an increase in the number of DNA replication initiation events, and conversely decreases in the level of DnaN decreases initiation events[34]. Intriguingly, deletion of YabA and increasing DnaN did not produce an additive affect, suggesting that they were acting at the same rate-limiting step. In addition, in a strain replicating from *oriN* (i.e. DnaA-independent), YabA was shown to affect the cooperativity of DnaA binding at *oriC*, and increased levels of DnaN remove YabA from *oriC*. This evidence led to the proposal of the second model (Fig. 1.14, B) whereby YabA prevents the co-operative binding of DnaA at *oriC* until the concentration of free DnaN increases in the cell, removing YabA from DnaA and enabling the co-operative binding of DnaA at *oriC*[34]. This is consistent with observations that YabA is able to disrupt DnaA-oligomer formation on DNA *in vitro*[109].

1.14.2: Soj/Spo0J

Soj is an ATPase which has been found to both negatively and positively regulate DNA replication in *B. subtilis*[165]. This activity is dependent on its oligomeric state[166], which is controlled by nucleotide binding; ATP-bound Soj forms dimers which co-operatively interact with DNA in an unspecific manner, whilst ADP-bound Soj is monomeric[167]. Dimeric (ATP-bound) Soj appears to stimulate overinitiation of replication, whilst monomeric Soj inhibits replication[33,166]. Spo0J regulates Soj activity by stimulating the ATPase activity of dimeric Soj, converting it back into its monomeric form[166].

Soj appears to interact with the ATPase domain (III) of DnaA, but it does not affect the ability of DnaA to bind or hydrolyse ATP[33]. Instead, it acts by inhibiting DnaA oligomer formation at *oriC*; a Soj mutant trapped in a monomeric state has been shown to inhibit DnaA oligomer formation both *in vitro* and *in vivo*[33]. Curiously, Soj trapped in a dimeric state is also able to interact with DnaA on a similar DnaA surface to monomeric Soj, without inhibiting DnaA oligomerisation. Thus, it has been suggested that monomeric Soj may be acting to prevent conformational changes in DnaA required to form an active initiation complex[33] (e.g. the proposed conformational change required between domains III and IV for open complex formation[5]), whilst dimeric Soj may stabilise DnaA oligomer formation[33]. This is consistent with evidence that suggests dimeric Soj plays an active role in stimulating DNA replication[74,75].

Soj and Spo0J are orthologues of the proteins ParA and ParB, respectively. ParA, and ParB, along with a cis-acting DNA sequence *parS*, form a plasmid partitioning system found in many prokaryotic species. The system ensures the partitioning of low copy number plasmids into daughter cells. In these partitioning systems, ParB binds to *parS* sequences on the plasmid, and ParA forms filaments on chromosomal DNA. An interaction between ParA and ParB simulates ParA's ATPase activity, which is thought to cause dissociation of the terminal ParA molecule from the filament; the plasmid can then either dissociate or translocate along the chromosomal DNA by binding to the next ParA molecule. Continuous cycles of ParA assembly and disassembly are thought to ensure equidistribution of the plasmids within the cell[169], ensuring that in cases were more than one copy of the plasmid is present - they will be located on either side of the division plane[170]

Chromosomal orthologues of ParA and ParB and *parS* sites are common throughout bacteria, and in some species they are known to directly affect chromosome segregation; it is thus attractive to assume that the role that the ParA/ParB chromosomal orthologues play in chromosome segregation is similar to that of the plasmid partitioning proteins. However, the role of *parABS* in many bacteria is still unclear. In *B. subtilis,* Spo0J-*parS* contributes to accurate chromosome segregation, but they are not essential for this function[171]. Instead they play a role in the recruitment of the SMC complex to the origin, and it is the SMC proteins that are responsible for proper DNA segregation and condensation of the chromosome[172,173]. These combined roles of Spo0J may provide a mechanism by which *B. subtilis* is able to co-ordinate DNA replication with chromosome segregation[173].

1.14.3: DnaD

DnaD has also been reported to play a role in the regulation of DNA replication initiation in *B. subtilis*. Like YabA, DnaD has been shown to inhibit the ATP-dependent cooperative binding of DnaA to *oriC* DNA[108] and is able to affect the formation of DnaA helical filaments in an *in vitro* DnaA-crosslinking assay[109]. This suggests that DnaD may play a role in regulating *B. subtilis* DNA replication initiation by altering the formation of the helical DnaA filament at the origin of replication.

1.14.4: DnaA-Box Clusters

Six DnaA-box clusters (DBCs) in *B. subtilis* found outside of the replication origin have been shown to play a role in regulating DNA replication. A deletion strain in which all six DBC boxes were removed displayed an early initiation of DNA replication phenotype. This deletion was only significant when all six clusters were removed and could be partially relieved by the re-introduction of a single DBC at various locations[168]. Nevertheless, this indicates that *B. subtilis* may be utilising chromosomal locations to titrate DnaA molecules away from the origin[163].

1.14.5: Sporulation and DNA Replication

A significant characteristic of *B. subtilis* cells is their ability to sporulate under nutrient limiting conditions. This process of cellular differentiation allows the cell form a dormant spore under conditions of nutrient limitation. The endospore is metabolically inactive and highly resistant to extreme conditions (such as high temperatures, desiccation and ionizing radiation). When nutrient availability improves, the endospore can germinate, returning the cell to vegetative growth, even after thousands of years[174,175]. The process of sporulation, contrary to vegetative growth (where cells divide symmetrically at their centre), involves the asymmetric division of the cell into a larger mother cell compartment and smaller forespore compartment (Fig. 1.15). These two compartments must each contain a copy of the organism's genome; differential gene expression subsequently gives them dramatically different cell fates. The forespore is engulfed and nurtured by the mother cell, transforming it into a resistant endospore, whilst the mother cell ultimately lyses to release the endospore[176] (Fig. 1.15). Entry into the sporulation pathway is under the control of a complex phosphorelay which culminates in the phosphorylation of the protein Spo0A, the master regulator of sporulation[177]. Spo0A~P acts as a transcription regulator, controlling the expression of over 500 genes[178].

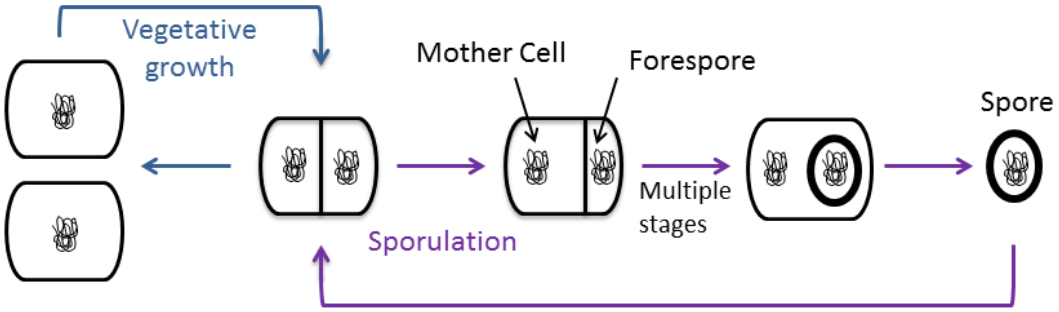


Figure 1.16: Vegetative growth and sporulation in *B. subtilis* : In normal vegetative growth (blue arrows) cells divide symmetrically, producing identical daughter cells. During sporulation (purple arrows) cells divide asymmetrically forming a mother cell and forespore; each receive a copy of the organisms genome, and through differential gene regulation they endure different fates. The mother cell engulfs the forespore, nurturing it into a mature spore before ultimately lysing to facilitate the spores release.

The process of DNA replication must be coordinated with entry into sporulation to ensure that the cell contains only two copies of the chromosome – one destined for the mother cell and the other for the forespore. Trapping of two (or more) chromosomes in the forespore compartment can be lethal to the cell, or otherwise, lead to a reduced viability of the spore (i.e. an increased number of spores fail to germinate in environments of high nutrient availability)[179].

1.14.5.1: Spo0A~P Pulsing is controlled by Sda and *B. subtlis*' Chromosomal Arrangement

It has long been recognised that there is a 'sensitive period' in the cell cycle when the cell can enter sporulation; if the cell progresses beyond this point it is committed to a new round of vegetative division[180,181]. This 'sensitive period' is associated with fluctuating Spo0A~P levels, which pulse throughout the cell cycle: at their highest concentration immediately after rounds of DNA replication are complete[179,182]. As a threshold level of Spo0A~P must be reached for sporulation to be triggered, this gradual increase within the cell cycle is thought to allow the turning on of low-threshold, and then, high threshold genes which trigger the sporulation process[183,184].

Although it is clear that Spo0A~P pulsing must be linked to DNA replication, until recently it was not known how this is achieved. The cellular Spo0A~P concentration has been directly linked to the expression of the sporulation inhibitor protein, Sda[179]. Sda is a histidine kinase

inhibitor which inhibits the action of KinA and KinB, important members of the phosphorelay responsible for the phosphorylation of Spo0A[185–187](Fig. 1.16). Sda expression levels are directly controlled by DnaA via transcription regulation[179,185] and this expression correlates with the presence of replication active DnaA in the cell. Thus, Sda levels spike at the same time as, or just after, the replisome forms[179]. Sda is subsequently rapidly proteolysed[188]. This provides a feedback mechanism whereby Sda blocks phosphorylation of Spo0A and hence entry into sporulation during ongoing rounds of DNA replication[179]. However, Sda alone does not appear to control Spo0A~P pulsing, as deletion of Sda does not stop the pulsing of Spo0F levels (a 'low-threshold' gene under the control of Spo0A~P); [189] suggesting that Spo0A~P pulsing still occurs.

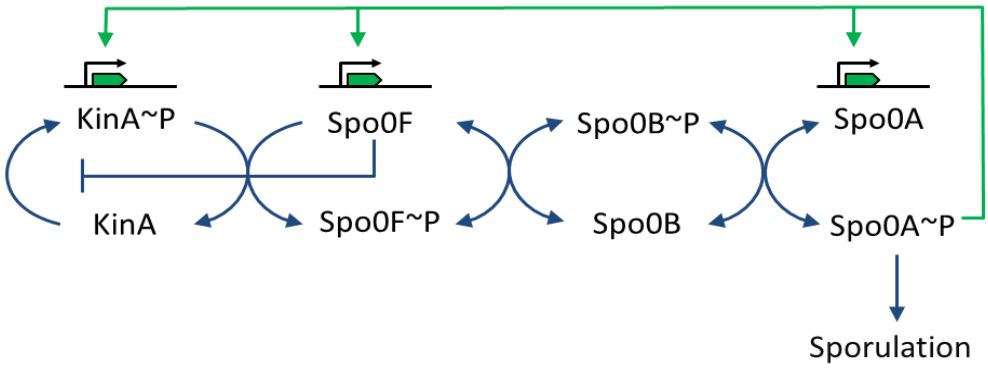


Figure 1.17: Phosphorelay leading to the induction of sporulation : A series of phosphate transfer reactions are required to obtain the threshold level of Spo0A~P needed for entry into the sporulation pathway.

The chromosomal arrangement of the phosphorelay genes s*po0F* and *kinA* has recently been shown to be important for Spo0A~P pulsing[179]. *spo0F* is located close to the replication origin whilst *kinA* is located near the terminus, leading to a 'gene dosage' imbalance during ongoing replication rounds[182]. The expression of high levels of Spo0F has been reported to block the phosphorylation of KinA, blocking entry into sporulation[190,191]. Due to the chromosomal arrangement of *spo0F* and *kinA*, two copies of *spo0F* will be present in the cell during most of the period of DNA replication (as it is replicated early), alongside only one copy of *kinA*[182]. Alterations in the chromosomal positioning of *spo0F* and induction of Spo0F production from an inducible promoter, directly affect Spo0A~P pulsing, and hence entry into sporulation[182]. Thus, it seems that the high Spo0F: KinA ratio inhibits KinA phosphorylation, and hence entry into sporulation. As rapidly growing cells can undergo multiple rounds of DNA replication at once, the relative Spo0F: KinA ratio also provides a mechanism for inhibiting sporulation in high nutrient conditions (where

replication cycles will be initiated during ongoing replication and thus more copies of the Spo0F gene will be present in the cell)[182]. Collectively, the chromosomal arrangement of *Bacillus subtilis* and direct inhibition of KinA activity by Sda serve to coordinate the entry into sporulation with DNA replication – where Sda serves as a control in the dynamic range of Spo0A activity[189].

1.14.5.2: SirA

SirA, a protein produced under the direct control of Spo0A~P, has been shown to inhibit DNA replication in sporulating cells. It is responsible for the control of chromosomal copy number in cells which are artificially induced to sporulate under conditions of rapid growth – a situation where cells will be undergoing multifork replication[192]. Inhibition of DNA replication by SirA has been shown to be mediated via a direct interaction with DnaA domain I[83,193]. SirA overexpression mimics DnaA depletion from the origin, suggesting it is able to inhibit the co-operative binding of DnaA to *oriC* [193]. The role of SirA will be covered more thoroughly in section 2.2.

1.14.5.3: A Direct Role of Spo0A~P in Replication Control

The *Bacillus subtilis* replication origin contains a number of (mainly imperfect) Spo0A-boxes which partially overlap with the location of DnaA-boxes[194] (Fig. 1.17). Notably, Spo0A-boxes have the consensus sequence 5'- TGTCGAA -3' which partially overlaps with the DnaA-box consensus sequence of 5'- TGTGNATAA – 3'[195]. Spo0A~P has been shown to bind these Spo0A-boxes *in vitro*[195], and mutations of these boxes that specifically alter consensus away from Spo0A-boxes (while preserving DnaA-boxes) affects Spo0A~P binding (but not DnaA binding) to the origin[194]. The binding of Spo0A~P to *oriC* appears to play a role in chromosome copy number control, however the effects of the Spo0A-box mutations are small compared to the loss of copy number control in a ΔSda strain (and to a lesser extent, ΔSirA strain).[194] This will be discussed further in section 2.2.

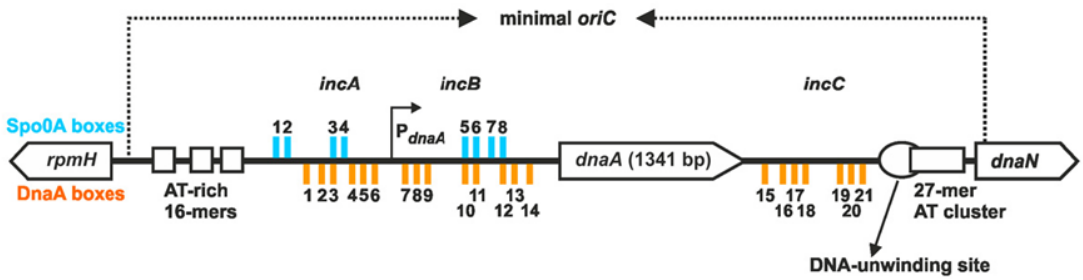


Figure 1.18: Overlapping Spo0A-boxes and DnaA-boxes in the replication origin: A number of Spo0A~P binding sites (Spo0A-boxes – blue) overlap with DnaA-boxes (orange) in the B. subtilis replication origin. These have been shown to play a direct role in inhibiting DNA replication.

1.15: ATP Hydrolysis vs. Inhibition of the DnaA-*oriC* Interaction

In recent years, opposing themes have emerged in the regulation of DNA replication initiation in *E. coli* and *B. subtilis*. It has long been recognised that the cellular DnaA-ATP concentration plays an important role in the regulation of *E. coli* replication initiation[21,61,196]. Hda, along with the DNA polymerase clamp, DnaN, acts to promote the hydrolysis of DnaA-ATP after initiation has occurred[125]. *datA*, a site which until recently was thought to be involved in DnaA titration, also acts to promote DnaA-ATP hydrolysis following initiation[136], and the DARS sites reactivate DnaA by promoting nucleotide exchange from ADP to ATP[146]. The DNA-binding proteins IHF and Fis play a role coordinating DnaA-ATP availability by regulating activity at *oriC*, *datA* and DARS; coupling

DNA replication initiation with the cell cycle and growth rate by an, as yet, poorly understood mechanism[134,135,145]. Thus, many of *E. coli*'s regulators are acting to control availability of DnaA-ATP as a method of initiation control.

In *B. subtilis*, no regulator has been identified which affects the conversion of DnaA-ATP to DnaA-ADP. Instead, several of *B. subtilis'* regulators appear to act by directly altering DnaA binding at *oriC*. The regulators YabA and DnaD have both been found to affect the cooperativity of DnaA-ATP binding at *oriC*[34,108], and both appear to be able to affect DnaA oligomer formation *in vitro*[109]. Soj also appears to act by directly inhibiting DnaA oligomer formation on DNA[33]. Furthermore, even though Soj and YabA interact with the ATPase domain of DnaA, neither protein has an effect on the ATP hydrolysis in DnaA-ATP[33,109]. Finally, SirA has been shown to affect DnaA localisation in the cell, mimicking the effects of DnaA depletion and leading to suggestions that it also acts by preventing DnaA binding at *oriC*[193]. Together, these results suggest that the primary method *B. subtilis* uses to inhibit DNA replication initiation is to directly affect DnaA binding to the origin.

DnaA-ATP is required for the co-operative binding of DnaA to the origin in both *E. coli* and *B. subtilis*[34]. It is therefore curious that the organisms have evolved divergent regulatory systems. One simple explaination for this may be that the different origin architectures of each species has led to the evolution of differing regulatory mechanisms. Another theory is that the spontaneous rate of *E. coli* and *B. subtilis* DnaA-ATP hydrolysis is substantially different[108]. *In vitro*, the exchange of DnaA-ADP for ATP has been found to be ~9 x faster in *B. subtilis* compared to *E. coli*[197]. Thus, if this is true *in vivo*, stimulation of ATP hydrolysis and nucleotide exchange in *E. coli* DnaA would be expected to have a much greater effect on replication initiation than in *B. subtilis* where turnover of ATP is already rapid[108]. Hence, the two organisms may have evolved different mechanisms of replication regulation due to differences in their intrinsic rates of DnaA-ATP hydrolysis.

1.16: DNA Replication as a Potential Antibiotic Target

Much of our understanding of bacterial DNA replication initiation has been derived from studies in *E. coli.* However the bacterial initiation machinery and the regulation of DNA replication varies between prokaryotes, making studies of replication in other organisms desirable for a fuller understanding of bacterial replication. Furthermore, it has been suggested that DNA replication may be an underexploited antibiotic target given the high

conservation of replication proteins[3,44]. However in order to create effective drugs, it is important to understand which aspects are conserved between bacteria (and other domains of life) and which are not[44]. Gram-positive pathogens such as *C. difficile* and *B. anthracis* are important targets of new antibiotics. The organism *B. subtilis* provides a benign model for the study of DNA replication initiation in Gram-positive bacteria; perhaps revealing important nuances of DNA replication which can be targeted in the development of antibiotics against Gram-positive pathogens.

1.17: Aims of This Project

The aim of this project was to improve our understanding of the regulatory mechanisms used by *B. subtilis* to control DNA replication initiation. More specifically, the project aimed to characterise protein regulators of DNA replication and their interaction partners using structural and biophysical methods, to determine the molecular mechanism each protein uses to regulate DNA replication initiation. The focus of the project later became the characterisation of two negative regulators of *B. subtilis* DNA replication: SirA and YabA.

The specific aims of the project were to express and purify the DNA replication regulator SirA in complex with DnaA domain I (SirA-DnaA^{DI}) in both native and SeMet derivative forms for structure determination; to crystallise and solve the structure of SirA-DnaA^{DI} using single anomolous dispersion (SAD) phasing methods and to characterise the SirA-DnaA^{DI} complex; and to produce mutants to assist in the characterisation of SirA-DnaA^{DI} binding.

For YabA; to express and purify YabA in native and SeMet derivative forms for structure determination; to crystallise YabA and solve the structure of YabA to facilitate understanding of how it inhibits DNA replication initiation; and to carry out biophysical characterisation of YabA to experimentally validate an *in silico* model of YabA and gain understanding of YabA's quaternary architecture.

Chapter 2 : Biophysical and Structural Characterisation of SirA-DnaA^{DI}

2.1: Abstract

Cells of *B. subtilis* can form endospores in response to nutrient starvation. This process of differentiation requires an asymmetric cell division creating two cellular compartments, a forespore and a mother cell, each of which require a single copy of the organism's genome to ensure the fitness of the resultant spore. DNA replication must therefore be coupled to sporulation to ensure that the sporulating cell has the correct chromosome copy number. SirA, a protein produced under the control of the master sporulation regulator Spo0A~P, directly inhibits DNA replication in diploid cells committed to sporulation, by binding to the replication initiator protein, DnaA. Here we have solved the structure of SirA in complex with domain I of DnaA (DnaA^{DI}) revealing a heterodimer. The interacting surfaces of both proteins are α-helical in character with predominantly apolar side chains contributing to the packing of the interface. Site-directed mutagenesis experiments have been used to confirm the significance of the interface for the interaction of SirA and DnaA^{DI} *in vitro*. These results are further supported by collaborative work on site-directed mutants *in vivo*, which suggest that the SirA-DnaA^{DI} interface identified in the crystal structure is significant for DnaA-mediated localisation of SirA at the replisome in sporulating cells.

2.2: Introduction

In response to nutrient starvation, *B. subtilis* cells enter the sporulation pathway, creating an endospore which lies dormant until nutrient conditions improve[175]. The process of sporulation involves the asymmetric division of the cell into a larger mother cell and a smaller forespore compartment, each of which require one copy of the organism's genome to form the resultant spore[175]. Sporulation entry typically follows the completion of DNA replication due to a feedback regulatory mechanism which creates a pulsing of Spo0A~P levels with the cell cycle. The peak in Spo0A~P concentration occurs immediately after the completion of DNA replication (see section 1.14.5). Although this feedback provides a mechanism for preventing replicating cells from entering sporulation, cells which are artificially induced to sporulate under conditions of rapid growth can still maintain correct chomosome copy number[192]. This implies that the cell also carries an active mechanism for blocking DNA replication in cells already committed to sporulation.

Deletion of a gene under the direct control of Spo0A~P, *yneE* (*sirA*), resulted in cells which lose the ability to control chromosome number when induced to sporulate from conditions of rapid growth[192]. Overproduction of the encoded protein, SirA, in cells growing on agar plates results in an inability to produce colonies and expression of SirA in liquid culture produces elongated and anucleate cells, with some cells containing nucleoids which have been severed by division septa – a phenotype reminiscent of DnaA depletion [192,193]. These experiments defined SirA as a negative regulator of DNA replication during spore formation. Yeast two-hybrid and localisation studies show that SirA acts to inhibit DNA replication via a direct interaction with DnaA[193]. A genetic assay was used to isolate mutations in DnaA which would allow the cell to overcome the no-growth phenotype displayed by cells overproducing SirA on agar plates. This identified four strains resistant to the overproduction of SirA, which contained mutations in *dnaA* that mapped to a region on DnaA domain I ($DnaA^{DI}$), strongly indicating that this cluster of amino acid residues is located at the SirA-interaction site on DnaA[83].

The major sporulation regulator, Spo0A~P has been shown to play a role in regulating DNA replication initiation by binding Spo0A-boxes within *oriC* which overlap with DnaA-binding sites[195] (see section 1.14.5.3). Loss of binding to these Spo0A boxes affects chromosome copy number control during sporulation[194]. However, these effects are only significant in a Δ*sda*/Δ*sirA* mutant strain when sporulation is induced from low nutrient conditions or in rapidly growing cells suddenly induced to sporulate[194]. Otherwise the most important

factors for copy number control were; in low nutrient conditions, Sda, a factor which inhibits sporulation entry in response to transcription regulation by DnaA (see section 1.14.5.1)[194]; and in rapidly growing cells induced to sporulate, SirA[194]. This suggests that Sda and SirA are the main contributors to chromosome copy number control during sporulation, with Sda acting to prevent replicating cells from entering sporulation, and SirA preventing the initiation of new rounds of DNA replication in cells already committed to sporulation.

Here, we were interested in studying the structure of SirA, to gain insight into its mechanism of action. SirA has no homologues other than those found in *bacilli*, which adds to its interest as a structural target.

Preliminary work done at York Structural Biology Laboratory by M. Fogg and A. Grahl found that the production of recombinant SirA in *E. coli* yielded only insoluble protein in all of the tested *E. coli* cell strains and growth conditions[198]. However, the creation of a duet expression vector, pYSBLICDnaA^{DI}SirA, which directs the expression of MQAGPA-SirA and histidine tagged domain I of DnaA (His_6DnaA^{DI}) from separate T7 promotors, led to the production of SirA in a soluble form (where SirA is assumed to be in complex with DnaA^{DI}).

2.3: Aims

SirA interacts with Domain I of DnaA to inhibit DNA replication during the onset of sporulation. There is no structural information available about SirA to infer its molecular mechanism. The aim of the work detailed in this chapter was to purify SirA in complex with domain I of DnaA (DnaA^{DI}) and solve the structure of the SirA-DnaA^{DI} inhibitory complex. Ultimately to use this structure to gain insight into the mechanism SirA employs to inhibit the initiation of DNA replication.

2.4: Methods

2.4.1: Expression

The plasmid pET-YSBLIC3C-DnaA^{DI}SirA was introduced into *E. coli* BL21 (DE3) cells for the co-overproduction of SirA and His-tagged DnaA^{DI}. Overnight cell cultures were used to inoculate 500 ml of LB media supplemented with 30 µg/ml kanamycin. Cultures were grown to an OD_{600} of 0.6-0.9 at 37 °C with shaking at 180 rpm before protein production was induced with 1 mM Isopropyl-β-D-1-thiogalactoside (IPTG). Following induction, cultures were grown at 37 °C (180 rpm shaking) for a further 4 hours before cells were harvested by centrifugation. SirA/DnaA^{DI} proteins harbouring site-directed mutations were produced in an analogous manner.

For SeMet substituted protein production, overnight cultures of *E. coli* BL21 (DE3) harbouring pET-YSBLIC3C-DnaA^{DI}SirA were used to inoculate 500 ml minimal media supplemented with 30 µg/ml kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (180 rpm shaking) prior to the addition of an amino acid mixture (50 mg lysine, 50 mg phenylalanine, 50 mg threonine, 25 mg isoleucine, 25 mg leucine, 25 mg valine) to suppress methionine production [199], and 30 mg selenomethionine. Cultures were grown at 37 °C (180 rpm shaking) for a further 15 minutes prior to induction of recombinant protein production with 1 mM IPTG. Cultures were subsequently grown at 30 °C (180 rpm shaking) overnight (16-20 hours) before cells were harvested by centrifugation.

2.4.2: Purification

The protein purification procedure was identical for native and SeMet substituted proteins. Harvested cells were resuspended in 50 mM Tris pH 8.5, 200 mM KCl, 20 mM imidazole and 10 mM 2-mercaptoethanol (βME) (Buffer A), and an EDTA-free protease inhibitor cocktail tablet (Roche) was added. Resuspended cells were lysed by sonication and the lysate clarified by centrifugation. The cell lysate was loaded on to a His Trap FF crude Ni-affinity column (GE Healthcare) and bound protein eluted with an increasing imidazole concentration gradient (20 – 500 mM) by addition of buffer B (50 mM Tris pH 8.5, 200 mM KCl, 500 mM imidazole and 10 mM βME). This step was followed by size-exclusion chromatography on a HiLoad 16/60 Superdex S75 column (GE Healthcare) equilibrated with 50 mM Tris pH 8.5, 200 mM KCl, 10 mM βME (Buffer C). The chromatogram exhibited two protein peaks. SDS-PAGE analysis of the peak fractions showed the earlier eluting peak

corresponded to the SirA-His_6DnaA^{DI} complex with the later eluting peak containing His_6DnaA^{DI} which is produced in excess in the duet expression system. The protein complex and DnaA^{DI} fractions were combined separately and the N-terminal histidine tag was removed from DnaA^{DI} in both cases by incubation with 3C protease overnight (protease : protein ratio of 1:50). Passage through a second Ni-affinity column to remove the histidine tag and the tagged protease yielded pure protein in a buffer of 50 mM Tris pH 8.5, 200 mM KCl, 10 mM BME. The proteins were judged to be pure according to Coomassie staining of SDS-polyacrylamide gels.

2.4.3: Mass Spectrometry

Purified SirA-DnaA^{DI} and DnaA^{DI} were buffer exchanged into 2mM Tris for analysis by Electrospray Ionisation-Mass Spectrometry (ESI-MS). Samples were provided at a concentration of 5 mg ml^{-1}. Spectra were obtained and processed by Simon Grist.

2.4.4: SEC-MALLS

SEC-MALLS analysis of DnaA^{DI} and SirA-DnaA^{DI} was carried out at a range of protein concentrations: DnaA^{DI} was analysed at 1.0 mg ml^{-1}, 2.5 mg ml^{-1} and 5.0 mg ml^{-1} and SirA-DnaA^{DI} was analysed at 0.5 mg ml^{-1}, 1.0 mg ml^{-1} and 2.5 mg ml^{-1}. For each run, 100 µl of sample was loaded onto a Superdex 75 HR 10/30 size-exclusion column equilibrated with 50 mM Tris pH 8.5, 200 mM KCl at a flow rate of 0.5 ml/min. The eluate was analysed successively by a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor as described previously [200]. Data were analysed with Astra software (Wyatt).

2.4.5: Crystallisation and Structure Solution: SirA-DnaA^{DI}

Native crystals of SirA-DnaA^{DI} were grown in hanging-drops containing a 1:1 ratio of concentrated protein solution and reservoir solution. Native crystals suitable for data collection were obtained following mixing of a protein solution of 6.4 mg ml^{-1} and a reservoir solution of 100 mM HEPES pH 7.5, 200 mM ammonium acetate, 25 % (w/v) PEG 3350, 1% (v/v) DMF. Crystals were transferred to a cryoprotectant solution consisting of

the reservoir solution containing 20% (v/v) glycerol before being cryocooled in liquid nitrogen. X-ray diffraction data were collected to 1.7 Å resolution on beamline I03 at the Diamond Light Source (DLS), Harwell. The crystal belongs to space group $P2_1$ with unit cell dimensions a = 77.3 Å, b = 34.7 Å, c = 84.7 Å and $\alpha = \gamma$ = 90 °, β = 102.1 °. SeMet crystals were grown in hanging-drops containing a 2:1 ratio of concentrated protein solution: reservoir solution. SeMet-substituted crystals suitable for data collection were obtained using a protein concentration of 1.9 mg ml^{-1} and a reservoir solution of 100 mM MMT pH 6.0 (DL-malic acid, MES and Tris buffers in a molar ratio of 1:2:2), 20 % (w/v) PEG 3350, 2% (v/v) DMF. Crystals were soaked in a cryoprotectant solution consisting of the reservoir solution containing 20% (v/v) glycerol before being cryo-cooled in liquid nitrogen. X-ray diffraction data were collected to 2.1 Å resolution on beamline I24 at DLS. The crystal belongs to space group $P2_1$ with cell dimensions a = 51.4 Å, b = 35.6 Å, c = 63.3 Å and $\alpha = \gamma$ = 90 °, β = 92.8 °.

Diffraction datasets obtained from the SeMet derivative and native crystals were processed using the automated data processing pipeline Xia2 [201] with options that run XDS [202]. Data were merged using AIMLESS [203]. The structure of SirA-$DnaA^{DI}$ was solved by single-wavelength anomalous dispersion (SAD) phasing. Heavy atom sub-structure determination, density modification and model building were carried out using the CRANK [204] pipeline available within the Collaborative Computational Project No. 4 (CCP4) program suite [205]. Nine selenium atom sites were identified using SHELX C/D [206] and their positions refined using BP3. The correct hand for the phases was identified using SOLOMON [207] and density modification carried out in PARROT [208] before atomic model building in BUCCANEER [209]. The SeMet-SirA-$DnaA^{DI}$ model was partially refined using maximum-likelihood methods in REFMAC [210] and manual model building in COOT [211]. The partially refined SeMet-SirA-$DnaA^{DI}$ model was used as a model for the solution of native SirA-$DnaA^{DI}$ by molecular replacement with MOLREP [212], the search for SirA molecules preceding that for $DnaA^{DI}$ domains. The native SirA-$DnaA^{DI}$ model was refined through iterative cycles of refinement in REFMAC and manual model building in COOT to an R-factor of 13.1 (R_{free} = 19.7). Refinement statistics are shown in Table 2.8. The SirA-$DnaA^{DI}$ atomic coordinates and crystallographic structure factors have been deposited in the Protein Data Bank with the accession code 4TPS.

2.4.6: Crystallisation & Structure Solution: DnaA^{DI}

SeMet derivative crystals of His_6DnaA^{DI} were grown in hanging-drops containing a 1:1 ratio of concentrated protein solution and reservoir solution. Crystals suitable for data collection were obtained after 3 months following mixing of a protein solution of 4 mg ml^{-1} and a reservoir solution of 100 mM Bis-Tris Propane pH 7.5, 0.5% dioxane, 25 % (w/v) PEG 3350. Crystals were transferred to a cryoprotectant solution consisting of the reservoir solution containing 20% (v/v) glycerol before being cryo-cooled in liquid nitrogen. X-ray diffraction data were collected to 2.4 Å resolution on beamline I03 at the Diamond Light Source (DLS), Harwell. The crystal belongs to space group $P4_32_12$ with unit cell dimensions a = 41.27 Å, b = 41.27 Å, c = 111.46 Å and $\alpha = \beta = \gamma$ = 90 °. A diffraction dataset obtained from the SeMet derivative was processed using MOSFLM[213] and data were scaled and merged using AIMLESS (Evans, 2006). The structure of DnaA^{DI} was initially solved by single-wavelength anomalous dispersion (SAD) phasing, however improved refinement statistics were obtained when the structure was re-solved by molecular replacement using *B. subtilis* DnaA^{DI} structure from the SirA-DnaA^{DI} complex as a molecular replacement model. One DnaA^{DI} molecule was found to occupy the asymmetric unit, with a solvent content of 48%. The DnaA^{DI} model was refined through iterative cycles of refinement in REFMAC and manual model building in COOT to an R-factor of 20.9 (R_{free} = 28.6). Refinement statistics are shown in Table 2.9.

2.4.7: Mutagenesis

Six individual alanine substitution mutations were introduced into pET-YSBLIC3C-DnaA^{DI} SirA by site-directed mutagenesis. On *dnaA^{DI},* mutations were introduced to direct substitutions at T26, W27 and F49. On s*irA*; at F14, Y18 and Q48. Oligonucleotides used for the mutagenesis reactions are shown in Table 2.1. PCR reactants and cycling parameters are shown in Tables 2.2 and 2.3, respectively. PCR products (50 µl) were incubated with 15 units of DpnI at 37°C for 4-5 hours. 50µl of Top10 cells were transformed with 2µl of DpnI-digested product and cells plated onto kanamycin-supplemented agar to select for the mutated plasmids. Individual colonies were selected and grown overnight in 5 ml LB supplemented with 30 µg/ml kanamycin. Overnight cultures were harvested by centrifugation and DNA extracted and purified from the harvested cells using a Mini-prep kit (Qiagen). Introduction of mutations at the desired sites was confirmed by sequencing (data not shown).

Table 2.1: Oligomers used for the mutagenesis of pET-YSBLIC3C-DnaA^{DI} SirA

Construct Name	Mutagenesis Oligomers
pET-YSBLIC3C-DnaA^{DI}T26A SirA	5'– CCGAGTTTTGAGGCTTGGATGAAGTCAACC –3'
	5'– GGTTGACTTCATCCAAGCCTCAAAACTCGG –3'
pET-YSBLIC3C-DnaA^{DI}W27A SirA	5'– CGAGTTTTGAGACTGCGATGAAGTCAACC –3'
	5'– GGTTGACTTCATCGCAGTCTCAAAACTCG –3'
pET-YSBLIC3C-DnaA^{DI}F49A SirA	5'– CGGCTCCCAATGAAGCTGCCAGAGACTGGC – 3'
	5' – GCCAGTCTCTGGGCAGCTTCATTGGGAGCCG – 3'
pET-YSBLIC3C-DnaA^{DI}SirAF14A	5' – CCTGATCAAAGAGGAAGCTGCCAATCACTATTTCGG – 3'
	5' – CCGAAATAGTGATTGGCAGCTTCCTCTTTGATCAGG – 3'
pET-YSBLIC3C-DnaA^{DI}SirAY18A	5' – GGAATTTGCCAATCACGCTTTCGGCCGGGAATCGG – 3'
	5' – CCGATTCCCGGCCGAAAGCGTGATTGGCAAATTCC – 3'
pET-YSBLIC3C-DnaA^{DI}SirAQ48A	5' – GAAATGACAGAGAAAGCGATTCAATATATTACAC – 3'
	5' – GTGTAATATATTGAATCGCTTTCTCTGTCATTTC – 3'

Table 2.2: PCR components for the site-directed mutagenesis of pET-YSBLIC3C-DnaA^{DI} SirA

Reactant	Volume/ µl (50µl reaction)
pET-YSBLIC3C-DnaA^{DI}SirA (100 ng µl^{-1})	1
Mutagenesis oligo I (20 pmol µl^{-1})	0.5
Mutagenesis oligo II (20 pmol µl^{-1})	0.5
Pfu Turbo Buffer	5
dNTPs (2mM)	5
Pfu Turbo DNA Polymerase (10,000 U/ ml)	1
DMSO	1
H_2O	36

Table 2.3: PCR cycling parameters for the site-directed mutagenesis of pET-YSBLIC3C-DnaA^{DI} SirA

PCR Step	Temperature (°C)	Time / min	Cycles
Initial denature	95	5	1
Denature	95	1	30
Anneal	50	1	
Extend	72	10	
Final denature	95	1.25	1
Final extend	72	5	1

2.4.8: Solubility Assay

E. coli BL21 (DE3), harbouring wild type and mutated pET-YSBLIC3C-DnaA^{DI}SirA plasmids were grown in 200 ml LB-kanamycin until the OD_{600} reached ~0.6. A portion of cells was set aside and harvested (uninduced aliquot (U)) while IPTG was added to the remaining cells. After a further 4 hours growth, aliquots of the induced (I) cells were taken and used to prepare total cell samples. The remaining cells from the induced culture were harvested by centrifugation and the cell pellets were re-suspended in 20 ml of 50 mM Tris pH 8.5, 200 mM KCl, 20 mM imidazole, 10 mM 2-mercaptoethanol. Cells were lysed by sonication and the lysate clarified by centrifugation. An aliquot of this soluble fraction (S) was retained. The remaining lysate was loaded onto a 1ml HisTrap FF crude Ni-affinity column (GE Healthcare), washed with 6 ml re-suspension buffer, and bound protein eluted with 4 ml of 50 mM Tris pH 8.5, 200 ml KCl, 500 mM imidazole, 10 mM 2-mercaptoethanol and the eluate (E) was collected. For each of the wild type and alanine variants, samples of the total fractions from uninduced (U) and induced (I) cells together with the soluble lysis (S) and high imidazole column eluate (E) fractions were analysed by SDS polyacrylamide gel electrophoresis followed by Coomassie blue staining.

2.5: Results

2.5.1: Production of SirA-DnaA^{DI}

Cloning and expression testing carried out by Mark Fogg and Anne Grahl prior to the start of this project revealed that production of recombinant His-SirA in *E. coli* yielded only insoluble protein. A number of expression strains and solubilisation buffers were tested and although high levels of SirA were consistently produced, the protein always partitioned into the insoluble fraction upon cell lysis[198]. Following a report that the SirA binding determinants of DnaA reside in domain I [83], a co-expression construct was created which directs expression of *dnaA*^{DI} (fused to a sequence encoding an N-terminal cleavable polyhistadine tag) and *sirA* from separate T7 promoters on the same vector. Remarkably, this co-expression strategy resulted in the expression of high levels of SirA in the soluble cell fraction after cell lysis - a result attributed to an assumed interaction with DnaA^{DI}. The work described here utilises this co-expression vector for protein production.

Native and SeMet-substituted SirA-DnaA^{DI} were expressed as described in section 2.4.1. Fig. 2.1 shows expression tests of both Native and SeMet forms of the protein. Expression of SirA and DnaA^{DI} is clearly induced in both SeMet and Native cultures (comparison of lanes 2 & 3 and 5 & 6, respectively). Both derivatives appear to have high solubility, although the SeMet derivative is slightly reduced compared to wildtype (lanes 4 and 7 respectively).

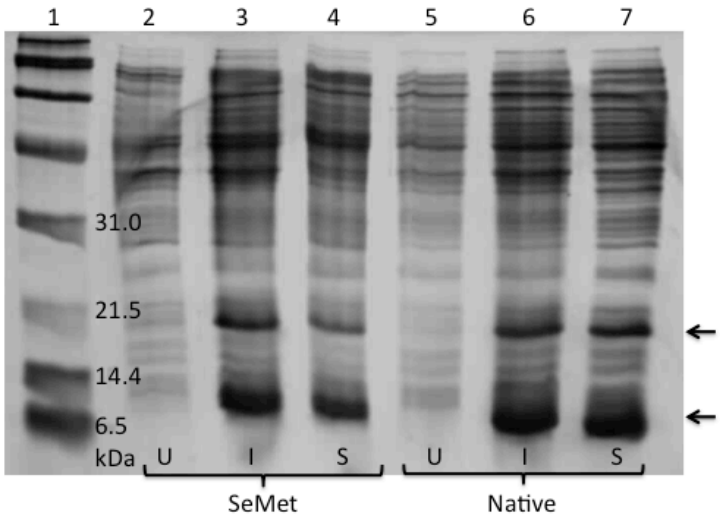


Figure 2.1: Expression of wild-type and SeMet-substituted SirA-DnaA^{DI}; lane 1: broad-range markers; lane 2: lysate from uninduced SeMet cell culture; lane 3: total lysate from induced SeMet cell culture; lane 4: soluble lysate from SeMet cell culture; lane 5: lysate from uninduced native cell culture; lane 6: lysate from induced native cell culture; lane 7: soluble lysate from native cell culture.

2.5.2: Purification of SirA-DnaA^{DI} and DnaA^{DI}

Native and SeMet-derivative SirA-DnaA^{DI} was purified by the procedure outlined in section 2.4.2. Clarified lysate was loaded onto a Ni-affinity column and eluted over a gradient of increasing imidazole concentration. SirA-DnaA^{DI} is retained on the column through a His-tag located at the N-terminus of DnaA^{DI}. The protein complex eluted as a single peak at an imidazole concentration of appoimately 125 mM. Fig. 2.2, A, shows a typical chromatogram. Fractions containing protein eluted from the Ni column were combined (and concentrated when necessary) to a total volume of less than 5ml. For the chromatogram in Fig. 2.2, A, fractions 9-20 were combined (and concentrated to approx. 5ml) following analysis by SDS-PAGE (Fig. 2.2, B) which showed high levels of SirA and DnaA^{DI} in fractions within this range.

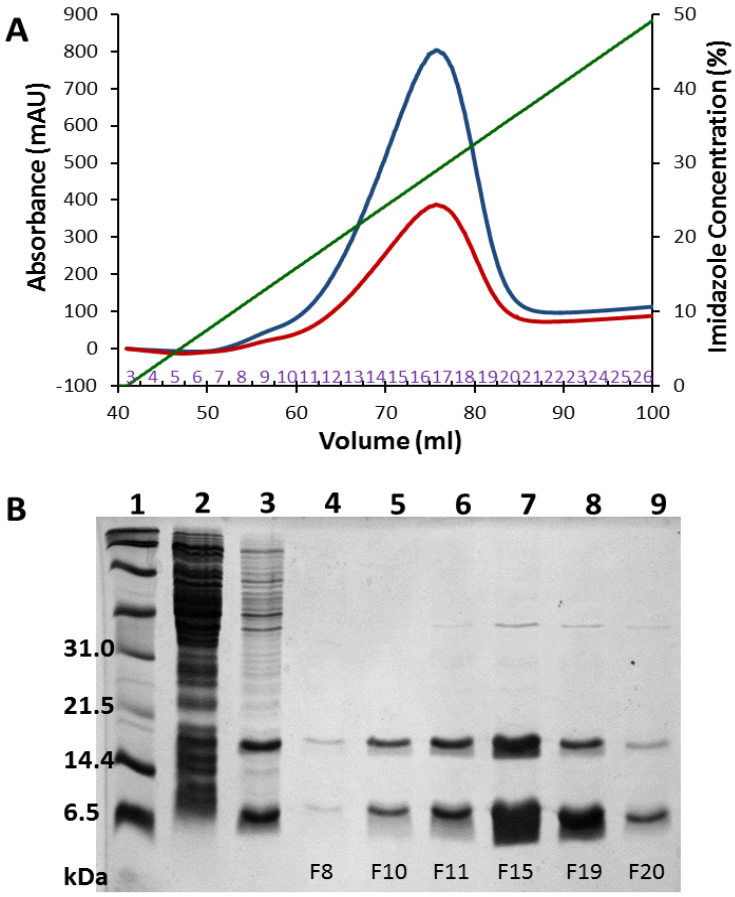


Figure 2.2: Ni-affinity chromatography for SirA-DnaA^{DI}: A) Chromatogram showing elution of protein from Ni-affinity column; In blue, absorbance at 280 nm; in red, absorbance at 254 nm; in green; percentage of buffer B (buffer B concentration = 500 mM imidazole); numbers in purple denote fractions collected. **B**) SDS-PAGE analysis of fractions eluted from Ni-affinity column; Lane 1, broad-range markers; lane 2, Flow-through during Ni-affinity column loading; lane 3, wash with buffer A; lane 4, 10 µl Fraction 8; lane 5, 10 µl Fraction 10; lane 6, 10 µl Fraction 11; lane 7, 10 µl Fraction 15; lane 8, 10 µl Fraction 19; lane 9, 10 µl Fraction 20.

This SirA-DnaA^{DI} complex was further purified using size-exclusion chromatography. The protein elutes in two peaks – a chromatogram is shown in Fig. 2.3, A. Analysis by SDS-PAGE (Fig. 2.3, B) reveals that the first peak contains SirA and DnaA^{DI} (assumed to be in complex) and the second peak, DnaA^{DI}. In the example shown in Figure 2.3, A, fractions 9-11 were combined as analysis by SDS-PAGE indicates that these fractions contain the SirA-DnaA^{DI} complex. Fractions 14-18 were also combined as these are assumed to contain only DnaA^{DI}.

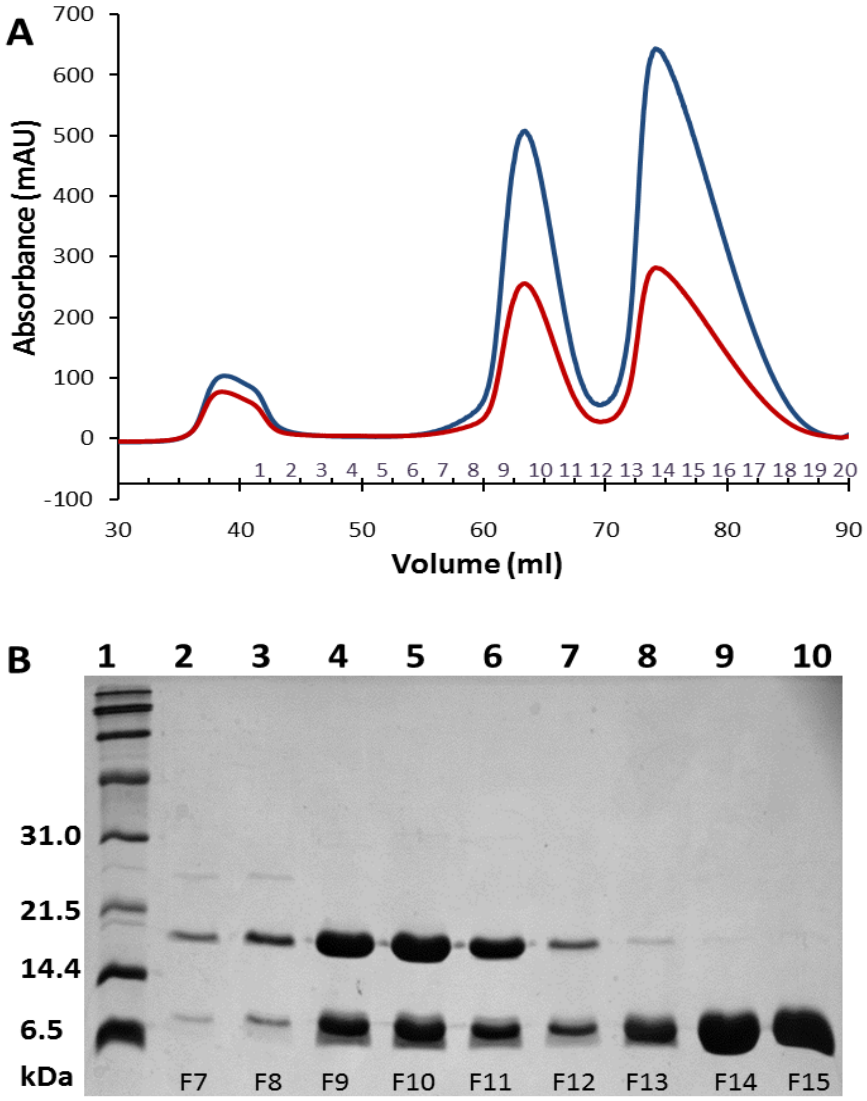


Figure 2.3: Size-exclusion chromatography for SirA-DnaA^{DI} : A) Chromatogram showing elution of protein from size-exclusion column; In blue, absorbance at 280 nm; in red, absorbance at 254 nm; numbers in purple denote fractions collected. **B)** SDS-PAGE analysis of fractions eluted from size-exclusion column; Lane 1, broad-range markers; lane 2, 10µl fraction 7; lane 3, 10µl fraction 8; lane 4, 10 µl Fraction 9; lane 5, 10 µl Fraction 10; lane 6, 10 µl Fraction 11; lane 7, 10 µl Fraction 12; lane 8, 10 µl Fraction 13; lane 9, 10 µl Fraction 14; lane 10, 10 µl Fraction 15.

The hexa-histidine-tag can be cleaved from His_6DnaA^{DI} in both the SirA-His_6DnaA^{DI} complex or His_6DnaA^{DI} alone by incubation with 3C protease overnight at a ratio of approximately 1:100 of 3C protease: protein (Fig. 2.4). Pure protein is then obtained by passing the cleaved protein through a second Ni column. His-tagged protein and (His-tagged) 3C protease are retained on the column whilst pure SirA-$DnaA^{DI}$ (or $DnaA^{DI}$) emerge in the flowthrough. Protein collected in the flowthrough is concentrated and shown to be pure according to SDS-PAGE. (Although $DnaA^{DI}$ bands appear diffuse, this appears to be an artifact of the protein size and not due to impurity, as the lowest molecular weight marker also runs with a diffuse band).

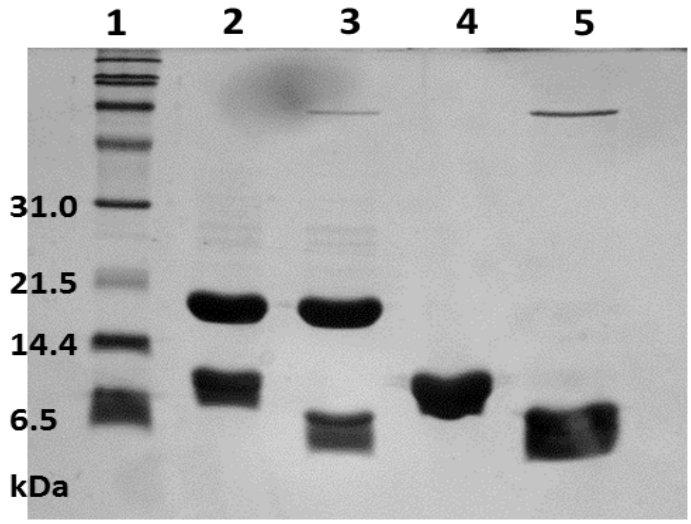


Figure 2.4: SDS-PAGE analysis of His-tag cleavage from $DnaA^{DI}$: Lane 1, broad-range markers; lane 2, SirA-His-$DnaA^{DI}$; lane 3, SirA-$DnaA^{DI}$ after overnight incubation with 3C protease; lane 4, His-$DnaA^{DI}$; lane 5, $DnaA^{DI}$ after incubation overnight with 3C protease.

2.5.3: Analysis of Purified SirA-$DnaA^{DI}$ and $DnaA^{DI}$ by SDS and Native-PAGE

Purified SirA-$DnaA^{DI}$ and $DnaA^{DI}$ were analysed by SDS-PAGE to assess their purity. Fig. 2.5, A, shows SDS-PAGE analysis of native and SeMet derivatives of SirA-$DnaA^{DI}$ and $DnaA^{DI}$ both histidine-tagged and un-tagged following purification. All species appear to be pure. For the SirA-$DnaA^{DI}$ complex (lanes 1-4), the relative staining of SirA and $DnaA^{DI}$ indicates that the two exist as a 1:1 complex; i.e. SirA would be expected to show a band almost twice as large as $DnaA^{DI}$ for a 1:1 ratio on account of their expected masses - 18, 698 Da for SirA and 11,749 (His-tagged) or 9, 734 Da (untagged) for $DnaA^{DI}$.

Native-PAGE analysis was also carried out for native and SeMet forms of SirA-DnaA^{DI} and DnaA^{DI} to analyse the homogeneity of the complex and DnaA^{DI} species in both their histidine-tagged and untagged forms (Fig. 2.5, B). The gel indicates that native and SeMet His$_6$DnaA^{DI} and DnaA^{DI} form a single homogenous species (Fig. 2.5 B, lanes 5-8). Analysis of native and SeMet SirA-His$_6$DnaA^{DI} each reveal 3 bands (Fig. 2.5, B, lanes 1-4); the lowest band in each case corresponds to (His$_6$)DnaA^{DI}; present in the sample either as excess protein remaining with the complex during purification, or as protein that has dissociated from the complex. A band of similar intensity to DnaA^{DI} which appears just below the predominant species in these samples, is likely to correspond to SirA, consistent with the notion that a proportion of the SirA-DnaA^{DI} complex has undergone dissociation. The predominant species in lanes 1-4 is thought to correspond to the SirA-His$_6$DnaA^{DI} complex.

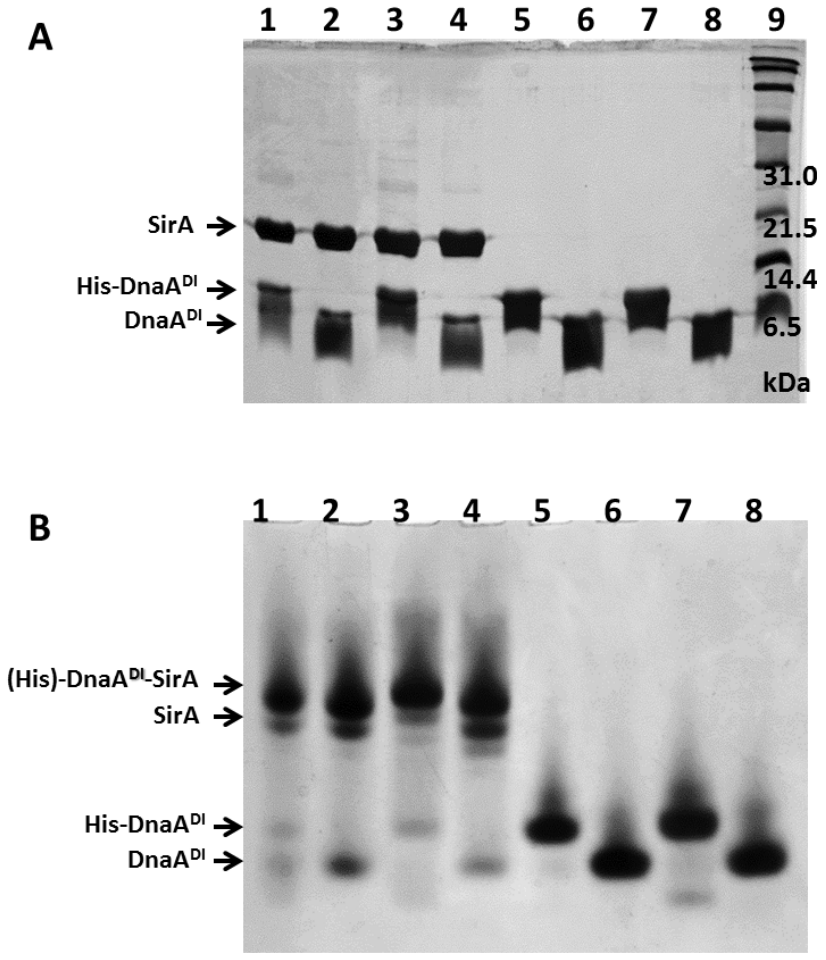


Figure 2.5: SDS- and native-PAGE analysis of pure SirA-DnaA^{DI} and DnaA^{DI}: : **A)** SDS-PAGE, and **B)** Native-PAGE. For A and B: Lane 1, 1µg Native SirA-His$_6$DnaA^{DI}; lane 2, 1µg Native SirA-DnaA^{DI}; lane 3, 1µg SeMet SirA-His$_6$DnaA^{DI}; lane 4, 1µg SeMet SirA-DnaA^{DI}; lane 5, 1µg Native His$_6$DnaA^{DI}; lane 6, 1µg Native DnaA^{DI}; lane 7, 1µg SeMet His$_6$DnaA^{DI}; lane 8, 1µg SeMet DnaA^{DI}. For A only: Lane 9, broad-range markers.

2.5.4: Mass Spectrometry

Electrospray-ionisation mass spectrometry was used to analyse purified SirA-His_6DnaA^{DI} (i.e. prior to His-tag cleavage) and DnaA^{DI} samples with and without the histidine tag.

The spectrum of SirA-His_6DnaA^{DI} revealed 4 mass peaks (Table 2.4). The two higher molecular mass peaks of 18,776 Da and 18,698 Da correspond to SirA; the latter corresponds exactly to the expected mass of the recombinant SirA construct (MQAGPA-SirA = 18,698 Da) and the former has a mass 78 Da greater than this (18,776 Da). This peak arises from the covalent attachment of β-mercaptoethanol to SirA which presumably occurred during purification when addition of βME was found to increase the solubility of the complex significantly (βME has a mass of 78 Da). This is the larger peak in the spectrum and is thus considered to be the predominant species. This βME adduct was later observed in the crystal structure (see section 2.5.8.1). The two lower mass peaks of 11,750 Da and 9,734 Da correspond to molecular weights for His_6DnaA^{DI} without the initiator methionine (11,747 Da) and DnaA^{DI} plus 3 amino acids, Gly-Pro-Ala (9,734 Da) – this is equivalent to a DnaA^{DI} molecule where the His-tag has been removed.

Table 2.4: Peaks observed in mass spectra of the SirA-His_6DnaA^{DI} complex

Sample	Peak	Observed Mass (Da)	Calculated Mass (Da)	Species
SirA-His_6DnaA^{DI}	1	18,775	18,776	SirA-βME
	2	18,698	18,698	SirA
	3	11,750	11,880 – 131 (Met) = 11,749	His_6DnaA^{DI}
	4	9,735	9,734	DnaA^{DI}

The mass spectra of His_6DnaA^{DI} and DnaA^{DI} produced single peaks at 11,747 Da and 9,731 Da, respectively (Table 2.5). These values are in good agreement with their theoretical molecular masses. His_6DnaA^{DI} has a theoretical molecular mass of 11,880 Da and as above, the observed mass corresponds to a loss of the start methionine (11,749 Da). DnaA^{DI} has a theoretical mass of 9,734, in good agreement with the empirical mass of 9, 731 Da.

Samples of native and SeMet His_6DnaA^{DI} which had been stored at 4 °C for approximately 4 months were also analysed (Table 2.5). This protein had been used for crystal optimisation,

however the crystals were not reproducible with a fresh preparation of His_6DnaA^{DI}, and mass spectrometry was used to analyse the crystallising sample. Analysis confirmed the incorporation of 2 SeMet residues in the case of SeMet His_6DnaA^{DI}, and both native and SeMet proteins displayed 3 peaks in their spectra. In both cases, these peaks corresponded to 3 different N-terminal truncations of the protein; 2 within the His-tag, leaving 4 and 5 residues attached, respectively, and one missing the His-tag and an additional 2 amino acids at the N-terminus. Crystallisation experiments were subsequently switched to $DnaA^{DI}$ following His-tag cleavage, however, SeMet $DnaA^{DI}$ diffracting crystals were obtained from an earlier crystal tray and the structure subsequently solved. This analysis would suggest that the crystallising species did not contain the full his-tag (the predominant species in the case of SeMet $DnaA^{DI}$ had 4 His-tag residues).

Table 2.5: Peaks observed by mass spectrometry of $DnaA^{DI}$ proteins

Sample	Peak	Observed Mass (Da)	Calculated Mass (Da)	Species
His_6DnaA^{DI}	1	11747	11,880 – 131 (Met) = 11,749	His_6DnaA^{DI}
$DnaA^{DI}$	1	9731	9,734	$DnaA^{DI}$
His_6DnaA^{DI} (storage at 4°C for ~4 months)	1 2 3	10,045 9861 9247	10,009 9862 9248	FQGPA-$DnaA^{DI}$ QGPA-$DnaA^{DI}$ $DnaA^{DI}$ (-ME)
SeMet His_6DnaA^{DI} (storage at 4°C for ~4 months)	1 2 3	10,139 9955 9295	10,103 9956 9294	SeMet FQGPA-$DnaA^{DI}$ SeMet QGPA-$DnaA^{DI}$ SeMet $DnaA^{DI}$ (-ME)

2.5.5: SEC-MALLS

Samples of $DnaA^{DI}$ and of the SirA-Dna^{DI} complex were analysed by SEC-MALLS (Size Exclusion Chromatography - Multi-Angle Laser Light Scattering). This technique allows the weight average molecular weight of eluting species from a size exclusion column to be calculated, thus the oligomeric states of proteins, and the stoichiometry of protein complexes can be inferred.

2.5.5.1: Analysis of DnaA^{DI}

Samples of DnaA^{DI} were analysed at concentrations of 1.0 mg ml^{-1}, 2.5 mg ml^{-1} and 5.0 mg ml^{-1}. (Fig. 2.6). Table 2.6 shows the elution times and experimentally determined molecular masses for each DnaA^{DI} concentration.

Some discrepancy is observable between samples, notably the elution time decreases slightly and the measured weight average molecular mass increases slightly with increasing DnaA^{DI} concentration. This suggests some self-association may be occurring at higher concentrations. However, on average, the samples correspond to a weight average molecular mass of ~11 kDa, strongly suggesting that DnaA^{DI} is a monomer in solution (calculated molecular mass of 9.7 kDa).

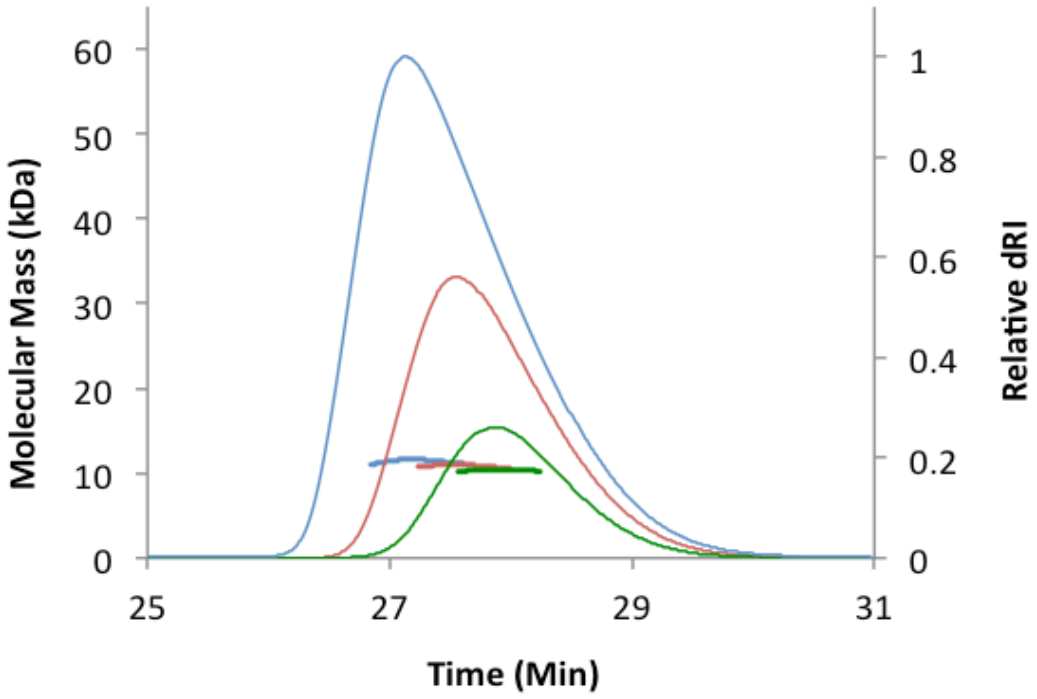


Figure 2.6: Overlay of SEC-MALLS analysis for DnaA^{DI} at varying concentrations: the thinner lines display the differential refractive index of the DnaA^{DI}-containing elute from a Superdex 10/300 S75 column as a function of time. Thicker lines show the weight average molecular mass of the eluting species, calculated from refractive index and light scattering measurements. Blue shows analysis of DnaA^{DI} at 5 mg ml^{-1}; red at 2.5 mg ml^{-1} and green at 1 mg ml^{-1}, revealing species of approximately 11 kDa, indicating that DnaA^{DI} is a monomer in solution.

Table 2.6: DnaA^{DI} SEC-MALLS results

Concentration (mg ml^{-1})	Elution Time (min)	Molecular Weight (kDa)
1.0	27.9	10.4
2.5	27.6	10.9
5.0	27.2	11.4

2.5.5.2: Analysis of SirA-DnaA^{DI}

Samples of SirA-DnaA^{DI} were analysed at concentrations of 0.5 mg ml^{-1}, 1.0 mg ml^{-1} and 2.5 mg ml^{-1} (Fig. 2.7, A). Results are shown in Table 2.7. The elution profile for SirA-DnaA^{DI} has two peaks, a major peak at ~24 minutes and a minor peak at ~28 minutes. The minor peak comprises 8-12% of the total protein content in the samples analysed, and the analysis above suggests that this corresponds to monomeric DnaA^{DI}. The major peak comprises 88-92% of the total protein content, and corresponds to a weight average molecular mass of between 25 and 28 kDa (depending on concentration). This correlates well with a 1:1 heterodimer of SirA-DnaA^{DI}, which would have an expected molecular mass of 28.7 kDa.

Table 2.7: SirA-DnaA^{DI} SEC-MALLS results

Concentration (mg ml^{-1})	Peak	Contribution (%)	Elution Time (Min)	Molecular Weight (kDa)
0.5	Major	87.7	24.1	25.4
1.0		91.5	23.8	26.8
2.5		91.7	23.7	28.0
0.5	Minor	12.3	28.1	10.0
1.0		9.5	28.1	11.8
2.5		8.3	20.0	14.7

Again, some discrepancy in elution times and determined molecular weights occurs at different SirA-DnaA^{DI} sample concentrations. Notably, a shift occurs in the ratio of eluting species as the concentration changes. The relative area under the minor peak increases with respect to the major peak as the concentration decreases (in Fig. 2.7, B major peaks have been scaled to the same height to show the relative shift in minor peak size). This change in the ratio of eluting species is accompanied by a small shift in the elution time of the major peak (to a later time) and a decrease in its associated molecular mass. Together this suggests that some complex dissociation is occurring at lower SirA-DnaA^{DI}

concentrations. Nevertheless, it is clear that under these experimental conditions, a 1:1 SirA-DnaA^{DI} heterodimer is the predominant species.

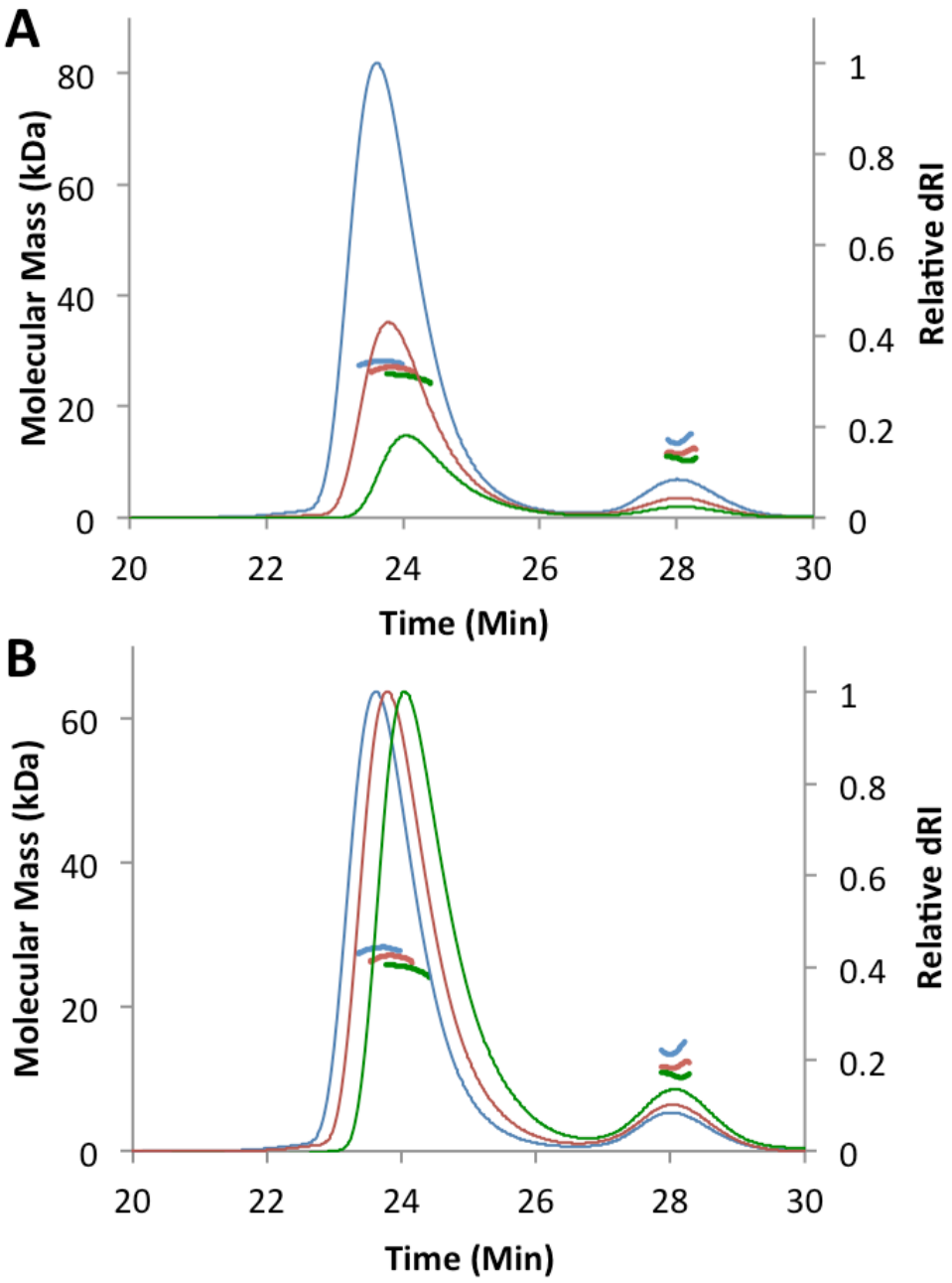


Figure 2.7: Overlay of SEC-MALLS analysis for SirA-DnaA^{DI} at varying concentrations : the thinner lines display the differential refractive index of the Sir-DnaA^{DI}-containing elute from a Superdex 10/300 S75 column as a function of time. Thicker lines show the weight average molecular mass of the eluting species, calculated from refractive index and light scattering measurements. Blue shows analysis of SirA-DnaA^{DI} at 2.5 mg ml^{-1}; red at 1.0 mg ml^{-1} and green at 0.5 mg ml^{-1}. The major species elutes at a mass of 25-28 kDa indicating SirA-DnaA^{DI} is a homodimer in solution. A peak at ~28 minutes is also visible, deriving from excess/dissociating DnaA^{DI}. **A)** Overlay of SEC-MALLS traces for SirA-DnaA^{DI} **B)** Overlay of SEC-MALLS traces for SirA-DnaA^{DI} with major peaks scaled to show the change in ratio of eluting species with concentration.

2.5.6: Crystallisation of SirA-DnaA^{DI}

Initial crystallisation conditions for native and SeMet derivatives of SirA-DnaA^{DI} were obtained from commercially available 96 well screens. Native crystals went through multiple rounds of optimisation in 24-well format using hanging drop vapour diffusion methods. SeMet crystals were initially optimised in 96-well format using sitting drop vapour diffusion before further optimisation in 24-well format using hanging drop vapour diffusion methods. Fig. 2.8 displays snapshots at stages of the optimisation process for both Native (Fig. 2.8, A) and SeMet (Fig. 2.8, B) crystals. In each case, picture I shows the best condition obtained in the 96 well screen, and respectively, Figure 2.8, A, IV and 2.8 B, V show crystals used for data collection for native and SeMet crystals. (In the case of A, IV, a single crystal was obtained from the cluster for data collection).

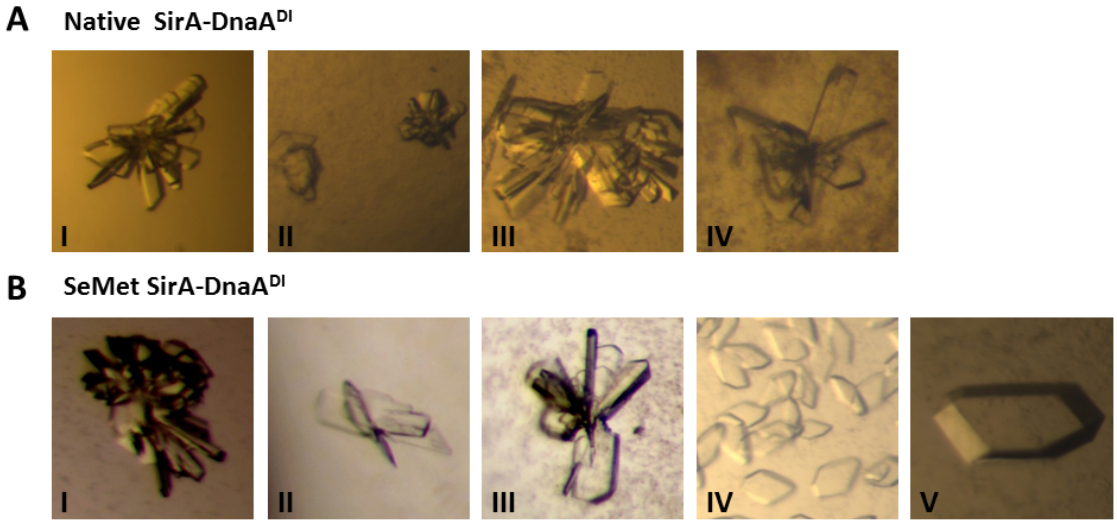


Figure 2.8: Snapshots of the crystal optimisation process for native and SeMet SirA-DnaA^{DI}: A) Native SirA-DnaA^{DI} crystals. Images I-IV display snapshots of the crystal optimisation process with image I showing the best crystals obtained from 96-well commercial screens, and image IV showing a crystal of diffraction quality. **B)** SeMet SirA-DnaA^{DI} crystals. Images I-V display snapshots of the crystal optimisation process with image I showing the best crystals obtained from 96-well commercial screens, and image V showing a crystal of diffraction quality.

Both native and SeMet SirA-DnaA^{DI} crystals diffracted in house to a resolution beyond 2.7Å (Figure 2.9). Full 3D data were collected at Diamond Light Source, Harwell.

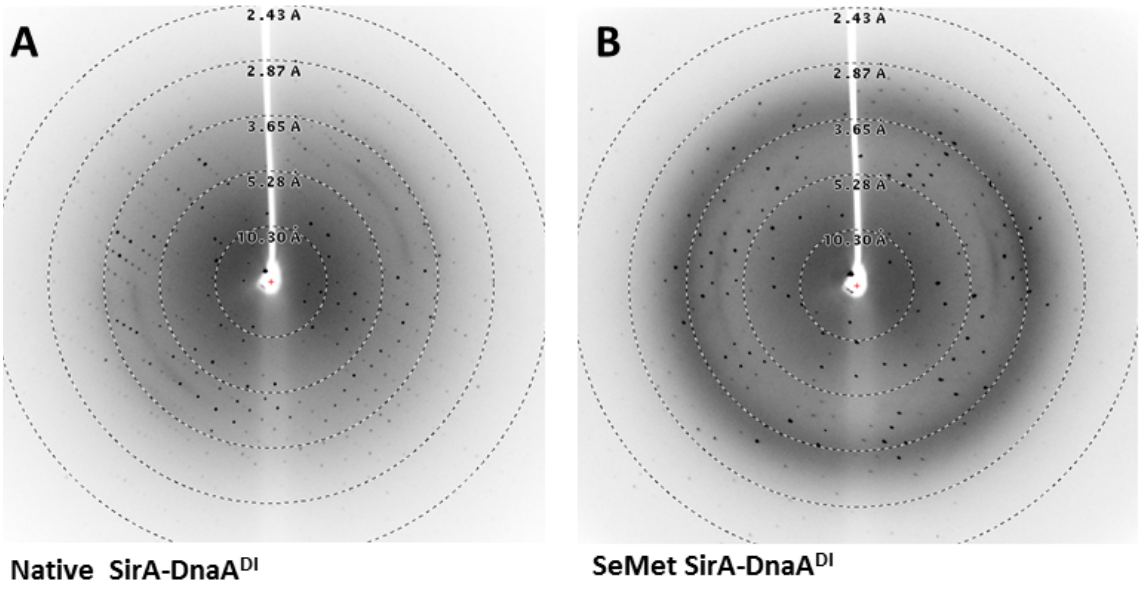


Figure 2.9: In-house testing of diffraction quality SirA-DnaA^{DI} crystals: Diffraction patterns from in-house testing of crystals of **A)** native SirA-DnaA^{DI} and **B)** SeMet-derivative SirA-DnaA^{DI}. Data were collected from each crystal at Diamond Light Source, Harwell.

2.5.7: Crystallisation of DnaA^{DI}

Initial crystallisation conditions for native and SeMet derivatives of DnaA^{DI} were obtained from commercially available 96-well screens. Native and SeMet crystals went through multiple rounds of optimisation in 24-well format using hanging drop vapour diffusion methods; crystals of suitable diffraction quality were obtained for SeMet DnaA^{DI} (Fig. 2.10), however native crystals of DnaA^{DI} diffracted poorly despite multiple attempts at optimisation. A SAD data set was collected from SeMet DnaA^{DI} at Diamond Light Source, Harwell.

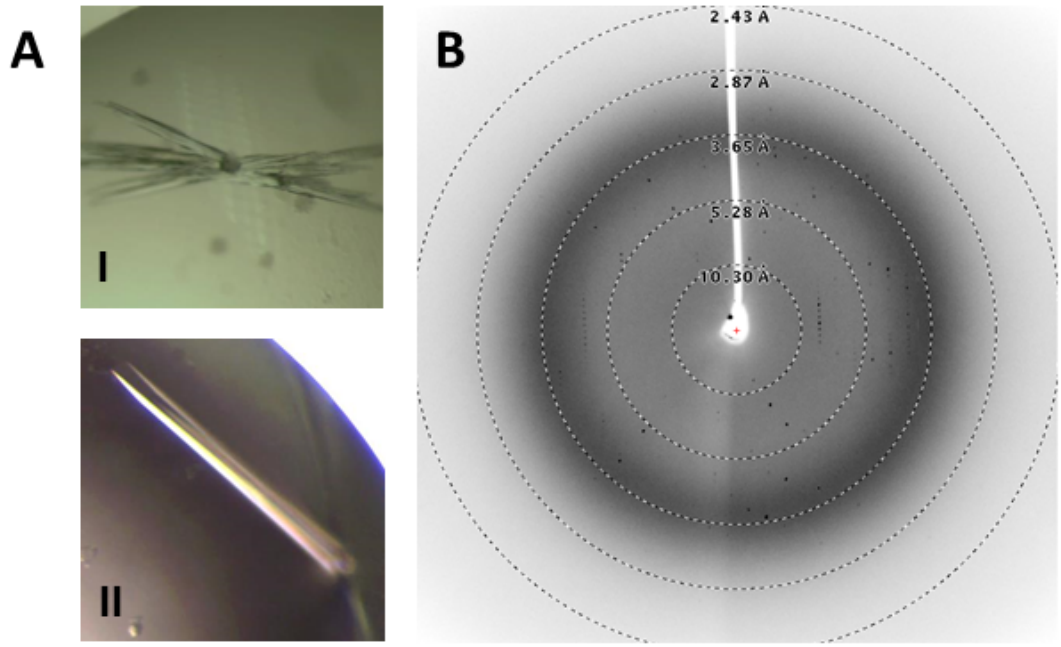


Figure 2.10: SeMet DnaA^{DI} crystals and in-house diffraction: A) SeMet DnaA^{DI} crystals; Image I – best crystals formed in 96-well commercial screen . Image II - crystal of diffraction quality. **B)** Diffraction pattern from in-house testing of SeMet DnaA^{DI} crystal.

2.5.8: SirA-DnaA^{DI} Structure

The SirA-DnaA^{DI} crystal structure was solved to 1.7 Å resolution using SAD (Single Anomalous Dispersion) phasing methods (Table 2.8). Native and SeMet derivative crystals were grown from similar conditions producing crystals in the space group $P2_1$ with different unit cells. SeMet-derivative crystals contain one complex per asymmetric unit (i.e. one molecule of SirA and one molecule of DnaA^{DI}), whilst native crystals contain two complexes per asymmetric unit. The SeMet-derivative structure was solved and partially refined to allow the native structure to be solved using molecular replacement. The refined model encompasses residues 2-141 of SirA in both chains A and C (Met^{1} and residues 142-148 being disordered), and residues 1-81 of DnaA^{DI} in molecules B and D (Gln^{82} being disordered). A vestigial N-terminal Gly-Pro-Ala sequence inherited from the DnaA^{DI} polyhistidine-tag can be seen in molecule D, and an additional N-terminal Ala is visible in molecule B. SirA-DnaA^{DI} is seen as a heterodimer, with molecules A (SirA) and B (DnaA^{DI}) forming one heterodimer and molecules C and D forming the other.

2.5.8.1: SirA

SirA consists of a single globular domain comprising seven β-strands and five α-helices in the order β1-α1-α2-α3-α4-β2-β3-β4-β5-α5-β6-β7 (Fig. 2.11). The SirA fold consists of a central seven-stranded twisted β-sheet with strand order β2-β3-β4-β5-β1-β6-β7, flanked on either side by two α-helical regions, one comprising helices α1, α2 and α3 and the other, helices α4 and α5. Comparison to other structures in the PDB using the online server PDBeFold identified similarity to several structures across the β1-α4-β2-β3-β4-β5-α5-β6-β7 region (residues 2-9, and 54-139). The search did not identify a protein which also encompassed structural similarity to the DnaA-interacting helices α1-α2-α3 (described below), however a search with these helices alone found homologous regions in many proteins of varying function.

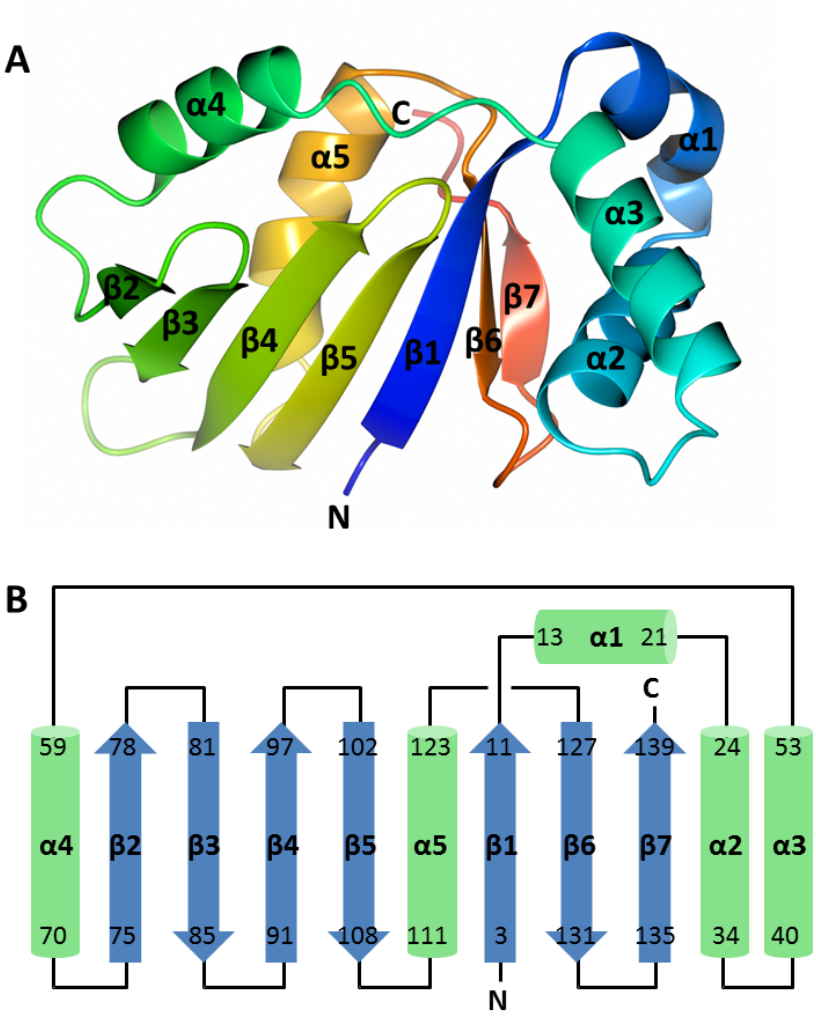


Figure 2.11: SirA structure and topology : A) Ribbon diagram of SirA from the SirA-DnaA^{DI} complex. The chain is colour-ramped from the N-terminus (blue) to the C-terminus (red). **B)** SirA chain topology diagram. Numbers denote residues at the start and end of each chain.

Table 2.8: Data and refinement statistics for SirA-DnaA^{DI}

	SirA-DnaA^{DI} SeMet	SirA-DnaA^{DI} Native
Data Collection		
X-ray source	DLS, i24	DLS, i03
Wavelength (Å)	0.9789	0.9763
Resolution Range (Å)	40.8-2.09	62.76-1.65
Space Group	P1 2_1 1	$P2_1$
Unit Cell Parameters		
a b c (Å)	51.35, 35.63, 63.27	77.29, 34.69, 84.74
α β γ (°)	90, 92.77, 90	90, 102.09, 90
No. of reflections [a]	13549/989	52893/2598
Completeness (%) [a]	98.7/99.1	98.9/99.6
Redundancy [a]	3.2/3.3	2.8/2.8
I/σ(I) [a]	11.9/1.9	12.5/1.9
R_{merge} [b] (%) [a]	7.4/79.9	3.9/45.8
Refinement and Model Statistics		
Resolution Range (Å)		62.84-1.65
R-factor[c] (R_{free}[d])		13.1 (19.7)
Reflections (working/R_{free})		50172/2706
Outer-shell/High Res range		1.69-1.65
Outer-shell[e]/High Res R-factor[c] (R_{free})[d]		19.0 (27.9)
Outer-shell/High Res Reflections (working/free)		3677/214
Molecules per asymmetric unit		4
rmsd from ideal geometry[ff]		
Bond Lengths (Å)		0.017
Bond Angles (°)		1.8
Average B-factor (Å^2)		27.8
Ramachandran Plot[g]		98.16/0.92/0.92

a . The first number refers to the overall data set, the second refers the outer resolution shell; 1.68-1.65 Å;

b . Rmerge=$\sum_{hkl}\sum_i |I_i - \langle I \rangle| / \sum_{hkl}\sum_i \langle I \rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes *hkl* and $\langle I \rangle$ is the statistically weighted average reflection intensity.

c . R-factor =$\sum||F_o| - |F_c|| / \sum|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively.

d . R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.

e . Outer shell for refinement corresponds to 1.69–1.65 Å.

f . Root-mean-square deviation of bond lengths and bond angles from ideal geometry.

g . Percentage of residues in most-favoured/disallowed/allowed regions of the Ramachandran plot.

The electron density maps reveal a 2-mercaptoethanol (BME) molecule linked through a disulphide bond to Cys125 of both SirA chains in the asymmetric unit (Fig. 2.12). The presence of this adduct explains species observed in the electrospray ionization mass spectrum of SirA of 18 776 Da, 79 Da larger than that of SirA; 18 697 Da (See above). BME was present during the purification steps as it was found to improve the solubility of the SirA-DnaA^{DI} complex.

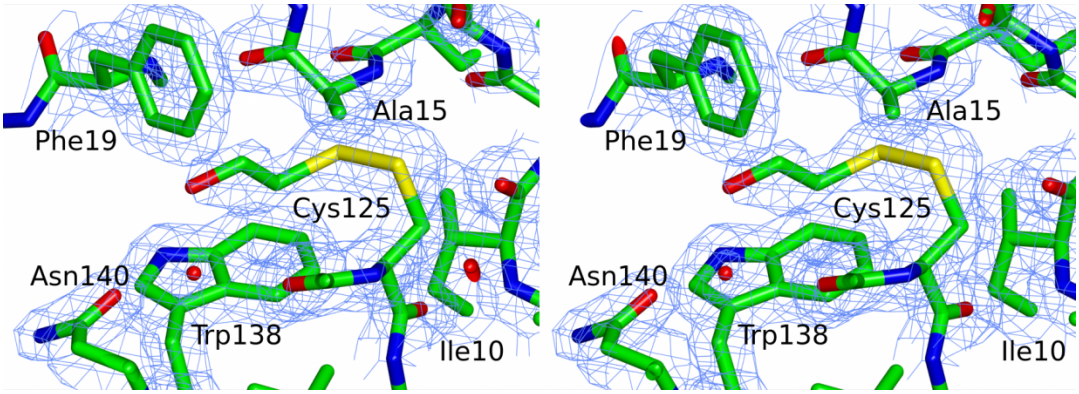


Figure 2.12: 2-Mercaptoenthanol adduct on SirA : Stereoview of residues in the neighbourhood of Cys^{125} of SirA, showing 2-mercaptoenthanol connected to the thiol of the side chain of Cys^{125} through a disulphide bond. Electron density is shown in blue contoured at 1σ. The model is coloured by atom with carbon in green, oxygen in red, nitrogen in blue and sulphur in yellow.

2.5.8.1: DnaA^{DI}

DnaA^{DI} from *B. subtilis* (Fig. 2.13) consists of four alpha helices and three beta strands in the order α1-α2-β1-β2-α3-α4-β3 with a β-sheet topology of β1-β2-β3. It has a similar topology to the previously determined structures of the corresponding domains of DnaA from *E. coli* (Abe et al., 2007), *M. genitalium* (Lowery et al., 2007) and *H. pylori* (Natrajan et al., 2009), sharing the K-homology domain motif that is widespread in single- stranded nucleic acid binding proteins.

2.5.9: DnaA^{DI} Structure

The structure of DnaA^{DI} was solved to 2.5 Å resolution using molecular replacement methods (Table 2.9). An initial solution was obtained by SAD phasing, however the model could not be refined past R/R-free of 0.23/0.34. Solution by molecular replacement using *Bsu* DnaA^{DI} from SirA-DnaA^{DI} allowed a modest improvement in the refinement statistics, to an R/Rfree of 0.21/0.29. The 2 solutions superpose well, with little modification to the overall structure (the N-terminal β-sheet is better defined in the molecular replacement solution). The model encompasses residues 1 to 81 of DnaA (Gln^{82} is disordered) and has a topology as described above (Fig. 2.13). The domain displays the same conformation in both its free and SirA-bound states.

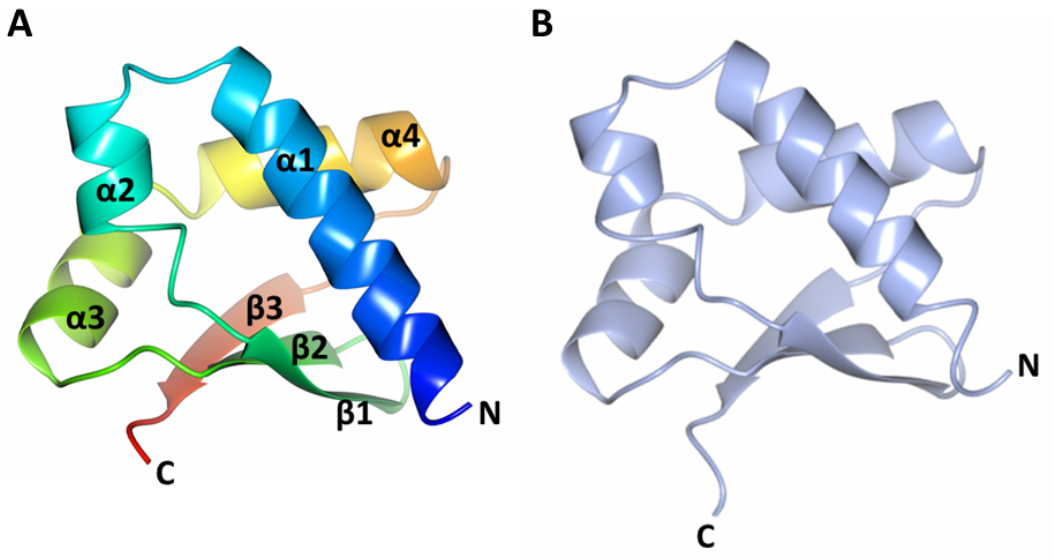


Figure 2.13: DnaA^{DI} structures:A**)** Ribbon diagram of DnaA^{DI} from SirA-DnaA^{DI} colour-ramped from N-terminus (blue) to C-terminus (red). **B**) Structure of *B. subtilis* DnaA^{DI} : Ribbon diagram of DnaA^{DI} (crystallised without SirA).

Table 2.9: Data and refinement statistics for DnaA^{DI}

	DnaA^{DI}
Data Collection	
X-ray source	DLS, i03
Wavelength (Å)	0.9686
Resolution Range (Å)	41.3-2.38
Space Group	$P4_3\,2_1\,2$
Unit Cell Parameters	
a b c (Å)	41.27, 41.27, 111.46
α β γ (°)	90, 90, 90
No. of reflections [h]	4113/411
Completeness (%) [h]	97.4/96.4
Redundancy [h]	9.5/9.1
I/σ(I) [h]	13.1/1.8
R_{merge} [i] (%) [h]	11.5/153
Refinement and Model Statistics	
Resolution Range (Å)	38.7-2.38
R-factor[j] (R_{free}[k])	20.9 (28.6)
Reflections (working/R_{free})	3904/199
Outer-shell/High Res range	2.44-2.38
Outer-shell[l]/High Res R-factor[j] (R_{free})[k]	36.2 (63.8)
Outer-shell/High Res Reflections (working/free)	289/10
Molecules per asymmetric unit	1
rmsd from ideal geometry[m]	
Bond Lengths (Å)	0.018
Bond Angles (°)	1.94
Average B-factor (Å^2)	57.2
Ramachandran Plot[n]	96.3/2.5/1.2

h . The first number refers to the overall data set, the second refers the outer resolution shell; 2.47-2.38 Å.

i . Rmerge=$\sum_{hkl}\sum_i |I_i - \langle I \rangle| / \sum_{hkl}\sum_i \langle I \rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes *hkl* and $\langle I \rangle$ is the statistically weighted average reflection intensity.

j . R-factor =$\sum ||F_o| - |F_c|| / \sum |F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively.

k . R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.

l . Outer shell for refinement corresponds to 2.44-2.38 Å.

m . Root-mean-square deviation of bond lengths and bond angles from ideal geometry.

n . Percentage of residues in most-favoured/disallowed/allowed regions of the Ramachandran plot.

2.5.10: The SirA-DnaA^{DI} Interface

The SirA-DnaA^{DI} interface is extensively α-helical in character, with the binding of SirA to DnaA^{DI} mediated by helices α1-α2-α3 on SirA and helices α2 and α3 on DnaA^{DI} (Fig. 2.14 and 2.15, A). The interaction between the two proteins is dominated by side-chain contacts, with a significant degree of hydrophobic character.

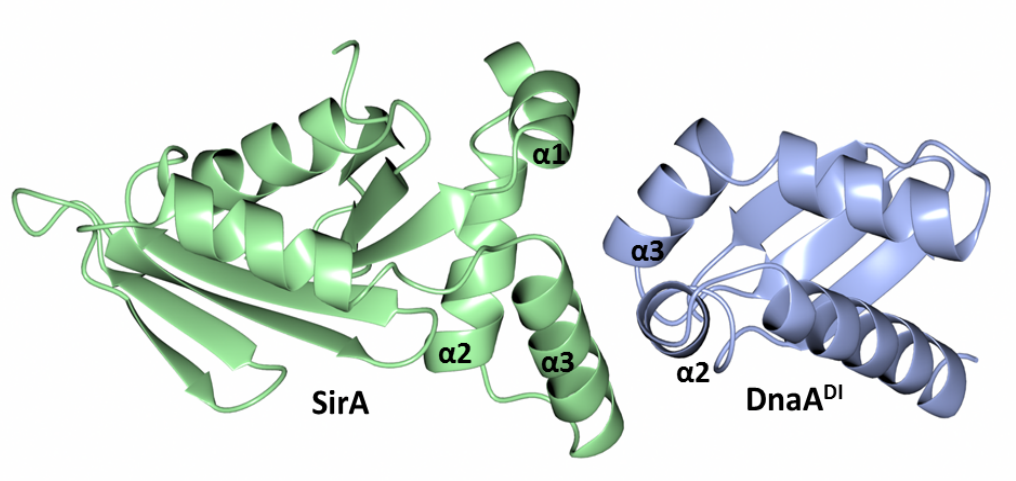


Figure 2.14: SirA-DnaA^{DI} complex: Ribbon diagram of SirA-DnaA^{DI} complex with SirA shown in light green and DnaA^{DI} shown in ice blue.

The complex buries 1240 Å^2 of otherwise accessible surface area, which comprises 8% and 12% of the surface areas of DnaA^{DI} and SirA, respectively (with 17 contributing residues from each protein). In the SirA binding surface of DnaA^{DI}, residues Thr26, Trp27 and Phe49 contribute to the core of the interface with residues Pro22, Ser23, Glu25, Ser30, Pro46, Asn47, Glu48, Asp52, Ser56 and Trp53 contributing to the rim. As shown in Fig. 2.15, B, these residues are very strongly conserved in a set of DnaA orthologues from endospore-forming bacteria. On the corresponding DnaA^{DI} binding surface of SirA, Phe14, Tyr18, Gln48 and Ile52 contribute to the core and Glu13, His17, Val24, Leu28, Gln41, Met44, Lys47 and Tyr51 contribute to the rim. Again core residues are well conserved in SirA orthologues with some variation observed in the residues constituting the rim (Fig. 2.15, C)

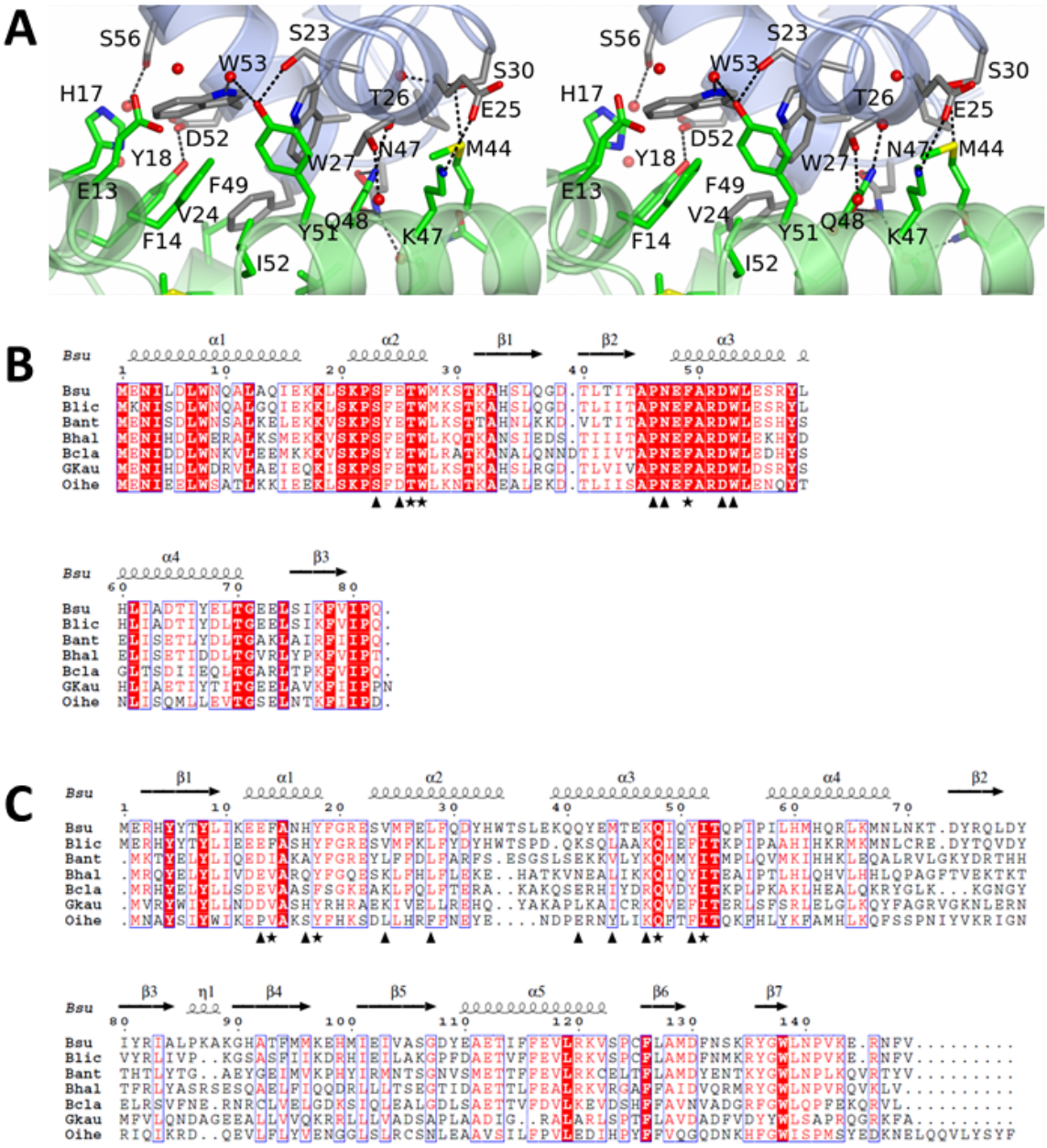


Figure 2.15: Stereoview of SirA-DnaA^{DI} interface and conservation of key residues A) Stereoview of the complex between DnaA^{DI} (chain D) and SirA (chain C) represented as light blue and light green ribbons, respectively. Side-chains of labelled residues are displayed in cylinder format and coloured by atom type with nitrogen (blue), oxygen (red), sulphur (yellow) and carbons coloured in grey for DnaADI and green for SirA. Water molecules are represented as red spheres, and polar interactions are denoted by dashed lines. **B)** and **C)** Alignment of the sequences of orthologues of DnaA^{DI} (B) and SirA (C) from selected *Bacillus* species; Bsu, *B. subtilis*; Blic, *B.licheniformis*; Bant, *B. anthracis*; Bhal, *B. halodurans*; Bcla, *B. clausii*; Gkau, *Geobacillus klaustophilus*; Oihe, *Oceanobacillus iheyensis*. Symbols below the alignments indicate interfacial residues in the respective molecules contributing to the core (asterisks) and the rim (triangles). Secondary structure elements and residue numberings are displayed above the alignment. The images were created using ESPript (Gouet, 2003).

2.5.10.1: The SirA interaction site on DnaA^{DI} maps to residues previously implicated in interaction with SirA

SirA interacting residues on DnaA have been previously identified using a genetic approach. Induction of SirA production in vegetatively growing cells inhibits growth, thus *B. subtilis* cells with SirA under the control of an IPTG-inducible promoter are unable to produce colonies on IPTG-supplemented LB agar (i.e. have a "no-growth" phenotype). A screen for DnaA mutations which could suppress this no-growth phenotype identified four strains which could grow under conditions of SirA expression, each harbouring a different *dnaA* mutation. Analysis of the sequence of these *dnaA* alleles revealed point mutations which would produce Asn47Asp, Phe49Tyr, Ala50Val and Ala50Thr substitutions. Yeast two hybrid analysis was used to confirm that these mutants were unable to interact with SirA suggesting that they affect SirA-DnaA complex formation[83].

The SirA–DnaA^{DI} structure reveals that Asn47 and Phe49 make direct interactions with SirA (Fig. 2.16). Asn47 of DnaA forms a pair of hydrogen bonds with Gln48 on helix α3 of SirA, and the side-chain of Phe49 of DnaA resides in a hydrophobic pocket created by helices α1, α2 and α3 of SirA. In contrast, Ala50 does not contact SirA in the complex but its side chain is buried within DnaA^{DI} between helices α2 and α3, in such a way that it determines the structure of the interface (Fig. 2.16). It is expected that mutations at this position introducing bulkier side-chains, such as valine or threonine, would alter the structure of the interaction surface leading to a lower affinity for SirA. In summary, the structure of SirA–DnaA^{DI} is consistent with the previous interpretation of the genetic data.

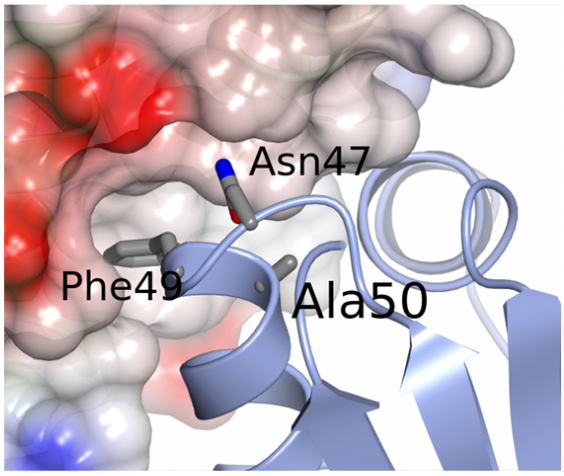


Figure 2.16: Mapping onto the structure of DnaA^{DI} the sites corresponding to mutations in *dnaA* that allow growth of *B. subtilis* when *sirA* is being overexpressed : SirA is rendered as a partially transparent electrostatic surface and DnaA^{DI} as a ribbon with the side-chains of residues Asn47, Phe49 and Ala50 in cylinder format.

2.5.10.2: Verifying the SirA-DnaA^{DI} Interface *in vitro*

To confirm that the protein-protein interface observed in the SirA-DnaA^{DI} structure is physiologically relevant, we designed an assay to investigate the interaction between SirA and DnaA^{DI} *in vitro.* As we were unable to express SirA in a soluble form in the absence of DnaA^{DI}, classical mixing experiments were not possible. Instead, we took advantage of the fact that SirA's solubility was dependent on its co-expression with and binding to DnaA^{DI}, developing a qualitative binding assay. We hypothesised that disrupting the interaction between SirA and DnaA^{DI} during co-expression would reduce or abolish SirA's solubility, which we could observe following cell lysis/protein purification.

Using site-directed mutagenesis, we introduced alanine mutations into the pET-YSBLIC3C-DnaA^{DI}SirA co-expression vector at codons which specify residues located in the structurally determined SirA-DnaA^{DI} interface. Three substitutions were made on SirA: Phe14Ala, Tyr18Ala and Gln48Ala; residues contributing respectively, 45 Å^2, 50 Å^2 and 75 Å^2 to the buried surface area of the interface (Fig. 2.17, A). Additionally, the hydroxyl group of Tyr18 forms a hydrogen bond with the carboxyl group on Asp52 of DnaA^{DI} and Gln48 forms a pair of hydrogen bonds with Asn47 of DnaA^{DI} (via side chain-side chain interactions). Three mutations were also made on DnaA^{DI}: Thr26Ala, Trp27A and Phe49Ala; contributing respectively, 75 Å^2, 50 Å^2, and 135 Å^2 of buried surface area to the interface. (Fig. 2.17, B) After confirmation of these mutations by sequencing, expression trials were conducted in *E. coli* BL21 (DE3) cells. The solubility of the recombinant proteins was compared to wild type proteins by SDS-PAGE analysis of cell fractions following lysis. The effect of each mutation on the interaction between SirA and DnaA^{DI} was further probed using a Ni pull-down assay. Soluble cell lysate was loaded onto a Ni-affinity column, washed extensively with loading buffer and eluted with a high imidazole concentration. Expression from pET-YSBLIC3C-DnaA^{DI}SirA directs the expression of DnaA^{DI} with a hexahistidine-tag and untagged SirA. Thus, histidine-tagged DnaA^{DI} is expected to bind to the column, but any retention of SirA is expected to be mediated by its interaction with histidine-tagged DnaA^{DI}.

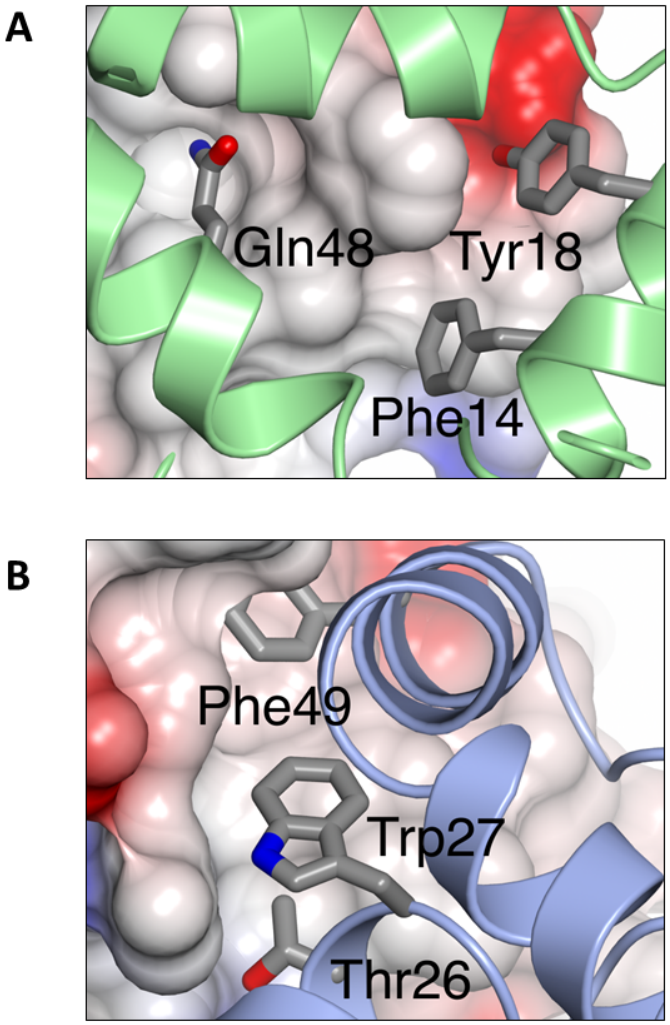


Figure 2.17: Residues within the SirA-DnaA^{DI} interface chosen as sites for site-directed mutagenesis: A) Core residues from the DnaA^{DI}-interacting surface of SirA which were chosen as sites for alanine substitution. DnaA^{DI} is shown as an electrostatic surface with SirA represented as a green ribbon with the side-chains of F14, Y18 and Q48 displayed as cylinders. **B**) Core residues from the SirA interacting surface of DnaA^{DI} which were chosen as sites for alanine substitution. SirA is shown as an electrostatic surface with DnaA^{DI} represented as a blue ribbon with the side-chains of T26, W27 and F49 displayed as cylinders.

The expression experiments indicate a reduced solubility of SirA in soluble lysate fractions from cells producing SirA with the mutations Phe14Ala, Tyr18Ala and Gln48Ala (Fig. 2.18, lanes 4, 8, 12 and 17). Furthermore, a reduced amount of the modified SirA protein is eluted in the high imidazole fraction during the Ni-pull down experiments, compared to that observed for wildtype SirA (Fig. 2.18, lanes 5, 9, 13 and 18). The DnaA mutants Trp27Ala and Phe49Ala appear to show only a negligible level of soluble SirA in the soluble lysis fractions (Fig. 2.18, lanes 27 and 31) and show no or very little SirA eluting in the high

imidazole fraction after the Ni-pull down experiment (Fig. 2.18, lanes 28 and 32). DnaA Thr26Ala shows near-wild type solubility and SirA recovery after the Ni-pull down (Fig. 2.18, lanes 22 and 23). This implies that all mutations on SirA and the DnaA Trp27Ala and DnaA Phe49Ala mutations elicit weaker binding of SirA to $DnaA^{DI}$.

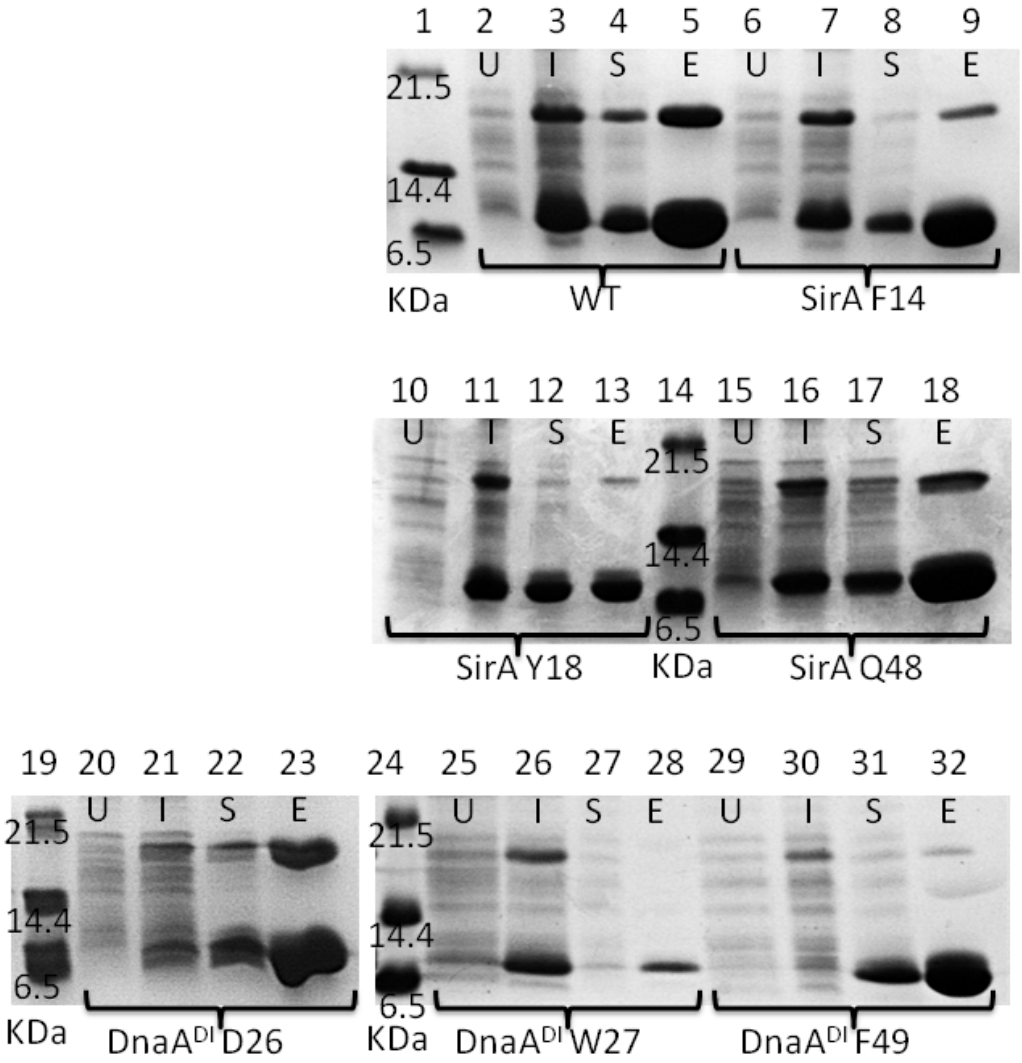


Figure 2.18: SDS-PAGE analysis for SirA solubility assay: Cultures of cells induced to express wild type and alanine-substituted variants of His_6DnaA^{DI} and SirA were grown and soluble cell lysate passed over a Ni-NTA column. Samples of cell lysate of fractions from uninduced cells (U), induced cells (I) and soluble lysate (S) were analysed as well as samples of the flowthrough from the Ni-column load (F) and a high-imidazole elution fraction (E). Lanes 1, 14 and 23 contain molecular weight markers. Uninduced (U), Total (T), Soluble (S)induced (I), Flowthrough (F) and Eluate (E) fractions were loaded for the wild type proteins in Lanes 2-5, and for wild type $DnaA^{DI}$ produced together with SirA mutants as follows; SirA(F14A) in Lanes 6-9: SirA(F18A) in Lanes 10-13: SirA(Q48A) in Lanes 15-18. Uninduced (U), Total (T)Induced, FlowthroughSoluble (S) and Eluate (E) samples of SirA and the $DnaA^{DI}$ variants were loaded as follows; $DnaA^{DI}$(T26A) in Lanes 19-22: $DnaA^{DI}$(W26A) in Lanes 24-27; $DnaA^{DI}$(F49A) in Lanes 28-31.

However, interpretation of these results is complicated by the variability in the level of $DnaA^{DI}$ present in these fractions. Thus, the elution fractions shown in Fig. 2.18, were diluted to normalise $DnaA^{DI}$ quantities to an approximately equivalent amount, and samples again resolved by SDS-PAGE. (Fig. 2.19) This clearly demonstrates that mutations in $DnaA^{DI}$ of Trp27Ala and Phe49Ala have the most significant effect on SirA solubility, with little discernable SirA present in the eluate fractions. In contrast, DnaA Thr26Ala shows similar levels of SirA in the Ni pull-down as observed for wildtype. The SirA mutations Phe14Ala, Tyr18Ala and Gln48Ala display a more modest effect, with an obvious reduced quantity of SirA present in each elute compared to wildtype. Quantification of the SirA band intensities in Fig. 2.19 by the software ImageJ suggests that, relative to the wild type SirA, there is a 1.5-fold lower recovery of SirA Phe14Ala and SirA Gln48Ala and a 2.5-fold lower recovery of SirA Tyr18Ala.

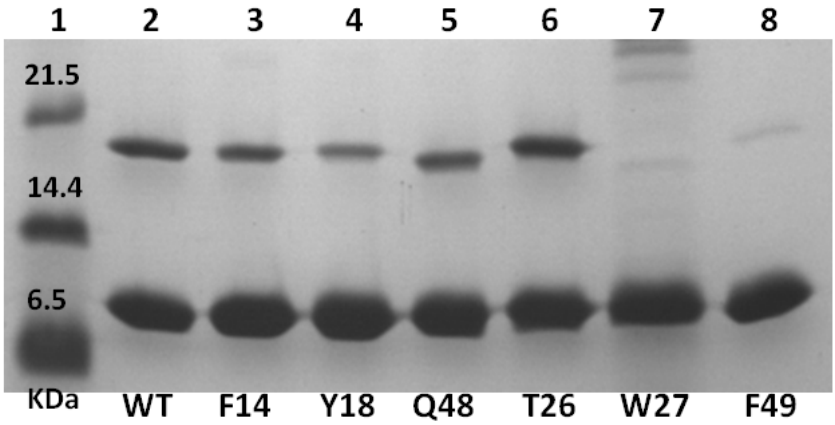


Figure 2.19: SDS-PAGE analysis showing normalised levels of $DnaA^{DI}$: Samples of the eluate fractions containing approximately normalized levels of $DnaA^{DI}$ were loaded so that the efficiency of SirA pull-down could be compared. Lane 1: molecular weight markers; Lane 2: wild type $DnaA^{DI}$–SirA; Lanes 3–5: Native $DnaA^{DI}$ and the SirA variants, loaded as follows; Lane 3: SirA(F14A), Lane 4: SirA(Y18A), Lane 5: SirA(Q48A). Lanes 6–8: Samples of native SirA and the $DnaA^{DI}$ variants, loaded as follows; Lane 6: $DnaA^{DI}$ (T26A), Lane 7: $DnaA^{DI}$ (W27A), Lane 8: $DnaA^{DI}$ (F49A).

Together these results correlate well with the SirA-DnaA interface inferred from the crystal structure. Residues Phe14, Tyr18 and Gln48 of SirA all form side chain contacts with $DnaA^{DI}$ which would be weakened by a truncation of these side chains to alanine (Fig. 2.17, A). The residues Trp27 and Phe49 on DnaA have side chains which project into a hydrophobic pocket on the surface of SirA, contributing to extensive van der Waals contacts across the SirA-DnaA interface (Fig. 2.17, B). Truncation of either of these residues to alanine has a

clear effect on the SirA-DnaA interaction, presumably due to a loss of van der Waals contacts. By contrast, Thr26 of DnaA sits at the edge of the hydrophobic pocket, and although largely buried, the hydroxyl group remains solvent exposed (it is able to hydrogen bond with a water molecule on the protein surface). Crucially, the substitution of threonine to alanine is less dramatic than the other mutations, allowing DnaA Thr26Ala to maintain its interaction with SirA (Fig. 2.17, B).

2.5.11: Comparison of SirA-DnaA^{DI} structure with HobA-DnaA^{DI} structure

The structure of the SirA-DnaA^{DI} complex and that of HobA-DnaA^{DI} [73] (a structure of *H. pylori* DnaA^{DI} bound to a regulator of this organism, HobA) were compared to one another. Strikingly, this revealed that HobA and SirA bind to the same structural site on DnaA^{DI}, burying equivalent surface residues (Fig. 2.20, A and B). This is particularly surprising given the divergent effects on DnaA exerted by SirA and HobA. HobA is a stimulator of replication initiation in *H. pylori*, essential to the organism's survival. Thus, HobA and SirA are able to achieve opposing regulatory functions by binding to the same structural site on DnaA^{DI}. Despite a shared structural topology however, residues contributing to this binding site are not conserved between *H.pylori* and *B. subtilis* (Fig. 2.20, C), perhaps reflecting a divergence in their respective regulatory mechanisms. Nevertheless, this indicates an important structural site on DnaA^{DI} for the regulation of replication initiation.

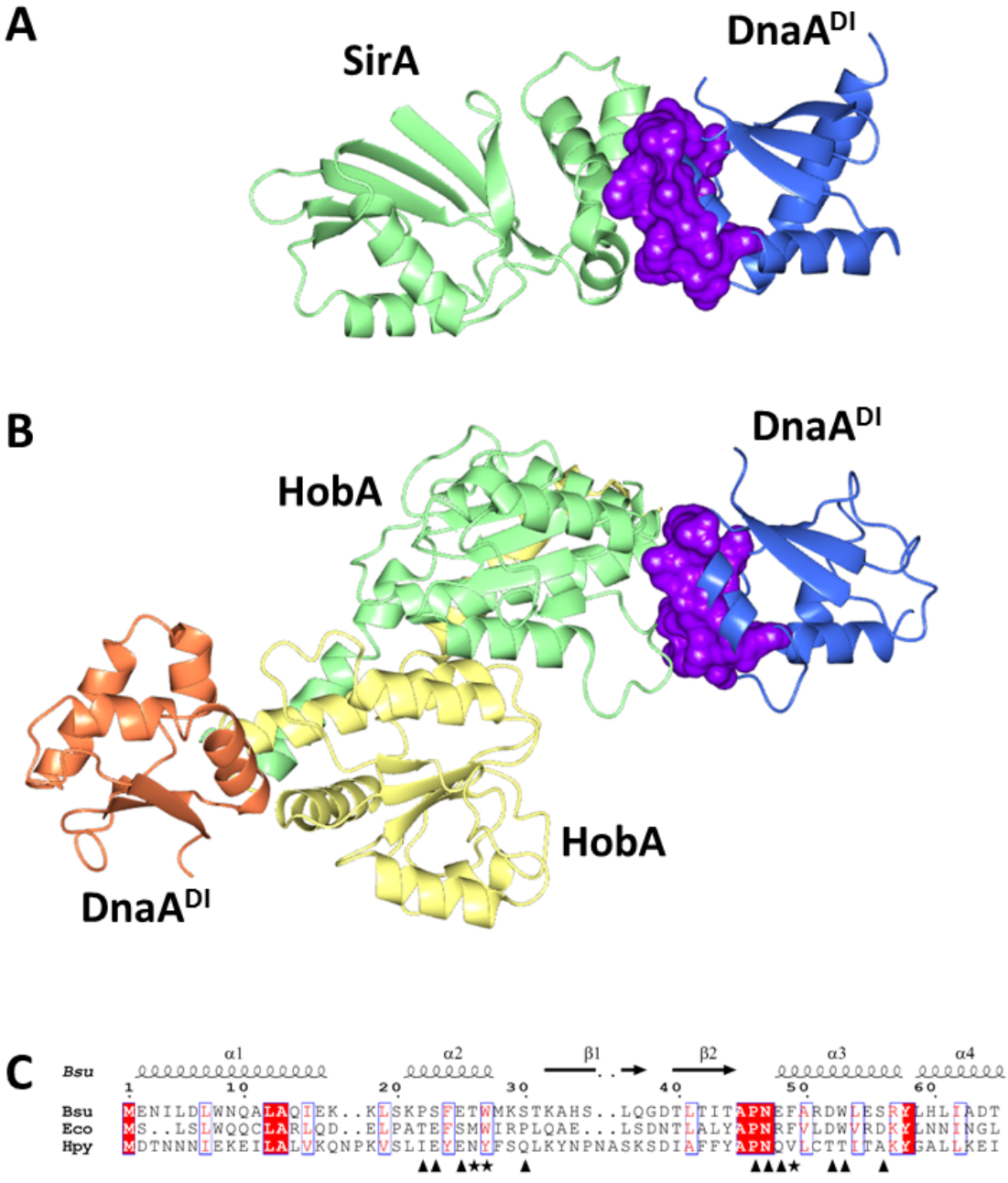


Figure 2.20: Comparison of *B. subtilis* SirA-DnaA^{DI} with *H. pylori* HobA-DnaA^{DI}: A) Ribbon diagram of the *B. subtilis* SirA-DnaA^{DI} complex with SirA shown in light green and DnaA^{DI} shown in blue. **B)** Ribbon diagram of the *H. pylori* HobA-DnaA^{DI} complex (PDB id code: 2wp0) with two molecules of HobA shown in yellow and light green, and two molecules of DnaA^{DI} shown in blue and coral. The binding surfaces of SirA or HobA on DnaA^{DI} in (**A**) and (**B**) are shown in purple. **C)** Alignment of DnaA^{DI} from *B. subtilis*, *E. coli* and *H. pylori*. Symbols below the alignments indicate interfacial residues on DnaA^{DI} in the SirA–DnaA^{DI} structure; core residues are indicated by asterisks and residues in the rim by triangles. Secondary structure elements and residue numberings are displayed above the alignment. The images were created using ESPript (Gouet, 2003).

2.6: Discussion

DNA replication initiation is strictly regulated to ensure chromosome copy number is coordinated with cell division and differentiation. The prokaryotic initiator protein, DnaA, has a number of identified regulators which vary across bacterial species[122]. In *B. subtilis,* several regulators of DNA replication initiation have been identified, two of which act at the entry to sporulation. Spo0A~P is thought to bind to Spo0A boxes located at the replication origin, occluding DnaA box regions and thus, preventing DnaA binding to *oriC* [194]. Additionally, the regulator SirA is known to assist with the maintenance of chromosome copy number on entry into sporulation[192,193]. Here, we have elucidated SirA's mode of binding to $DnaA^{DI}$ by solving the crystal structure of SirA in an inhibitory complex with $DnaA^{DI}$. Collaborative studies with Nadia Rostami and Heath Murray at the University of Newcastle have confirmed the physiological relevance of the SirA-$DnaA^{DI}$ complex and these are described below.

SirA has recently been shown to form foci at the mid-cell during sporulation in approximately 20% of cells, 150 mins after sporulation is induced (Fig. 2.21, A). These foci were shown not to coincide with replication origins which are located at the cell poles (Fig. 2.21, B), and moreover, SirA was found to co-localise with DnaN (the polymerase β-clamp (Fig. 2.21, C), suggesting that SirA is associated with the replisome (and away from the origins) during the early stages of sporulation[214].

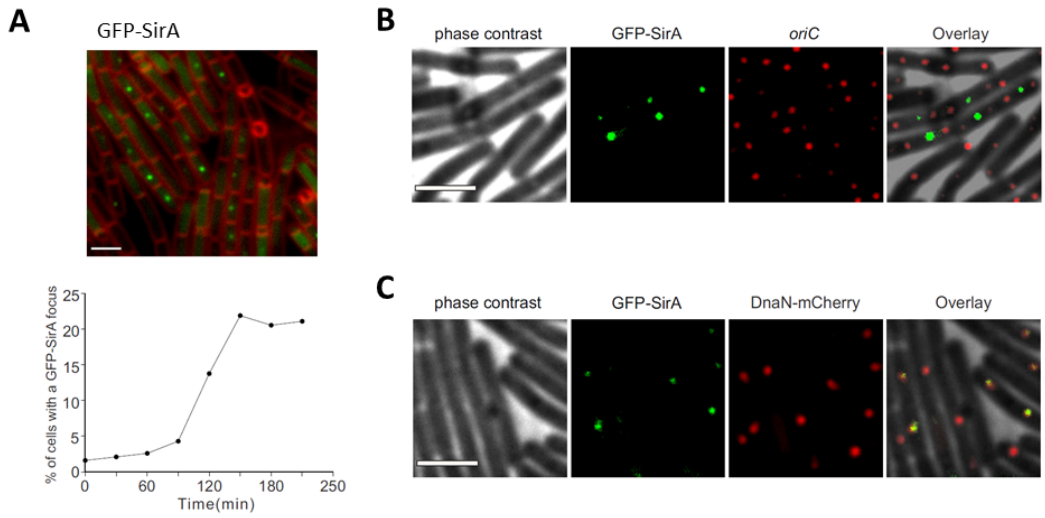


Figure 2.21: Localisation of SirA *in vivo:* Figure to show GFP-SirA localisation in vivo; adapted from figure produced by N. Rostami and H. Murray. All images are 150 mins post starvation in sporulating *B. subtilis* cells. Scale bars = 3µm **A)** GFP–SirA foci formation (at maximum 150 mins post starvation). Membrane dye FM5-95 was used to highlight cell outlines. **B)** Localisation of GFP–SirA foci compared to replication origins. **C)** Co-localisation of GFP–SirA foci and DnaN-mCherry foci.

Furthermore, the mutations on SirA which were found to disrupt the SirA-DnaA^{DI} interaction *in vitro* (above); Phe14Ala, Tyr18Ala and Gln48Ala, significantly reduce SirA foci-formation (Phe14Ala, Tyr18Ala to background levels and Gln48Ala 2.5-fold) (Fig. 2.22, A), demonstrating the necessity of these residues for localisation of SirA at the replisome. The mutations on DnaA; Thr26Ala, Trp27Ala and Phe49Ala could not be introduced into *B. subtilis* cells, suggesting that they may also play a role in an essential interaction between DnaA and the replication machinery[214]. By contrast, the previously identified mutations Asn47Ala and Ala50Val [83] could be readily introduced into *dnaA* and were also found to have a significant effect on SirA foci formation - Asn47Ala reduced foci formation 2.4-fold, whilst Ala50Val reduced them to background levels[214] (Fig. 2.22, B). Together, these results indicate that the SirA-DnaA^{DI} interface identified in the crystal structure is required for SirA foci formation, and that SirA localisation at the replisome is mediated by an interaction with DnaA.

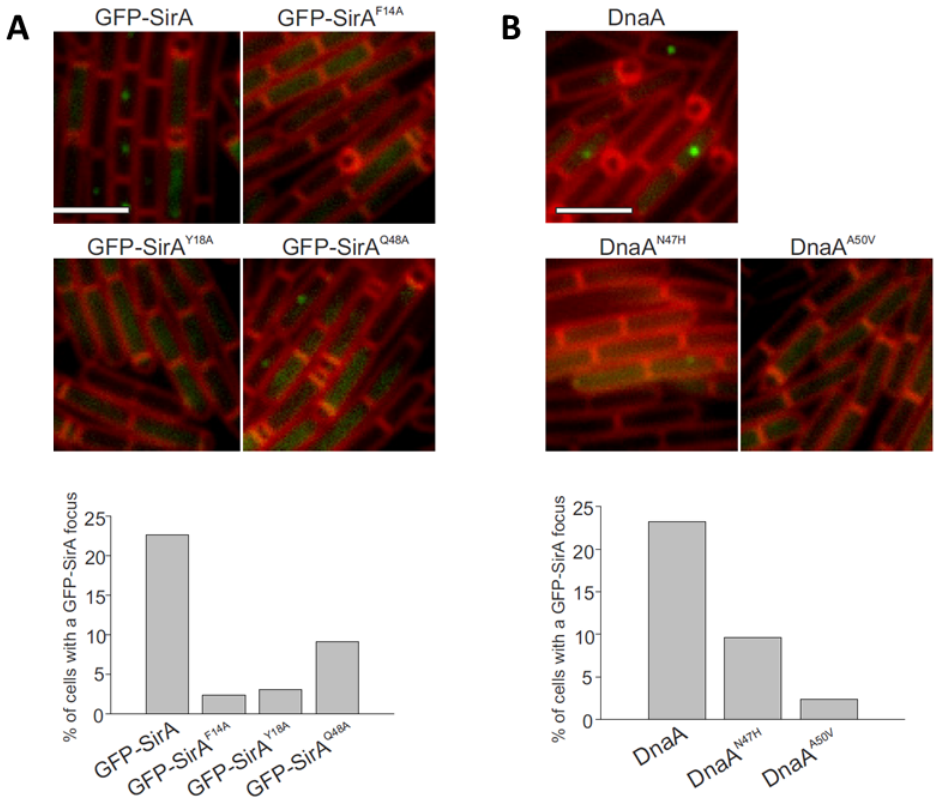


Figure 2.22: Effect of mutations in the SirA-DnaA interface on GFP-SirA foci formation *in vivo*: Figure adapted from a figure originally produced by N. Rostami and H. Murray. All images are 150 mins post starvation in sporulating *B. subtilis* cells. Scale bars = 3µm **A)** Images to show amino acid substitutions F14A, Y18A and Q48A in SirA inhibit GFP-SirA foci formation *in vivo.* Quantification of a representative dataset is shown below. **B)** Images to show amino acid substitutions N47H, and A50V in DnaA inhibit GFP-SirA foci formation *in vivo.* Quantification of a representative dataset is shown below.

In addition, the above mutations in SirA were able to overcome the phenotypes shown by induction of wildtype SirA in vegetatively growing cells on/in both solid and liquid media. As described earlier, induction of wild type SirA from an inducible promotor inhibits vegetative cell growth on solid media[192,193]. It also produces elongated cells with a single nucleoid in liquid media. However, the introduction of mutations directing the Phe14Ala, Try18Ala and Gln48Ala substitutions into *sirA* allowed cell growth on solid media (Fig. 2.23, A) and produced cells which had a normal DNA distribution and morphology in liquid media (Fig. 2.23, B)[214]. This suggests that the amino acids necessary for SirA localisation are also essential for inhibiting DnaA activity in vivo. Altogether, these results provide a strong physiological relevance to the SirA-DnaA^{DI} interface identified in the crystal structure.

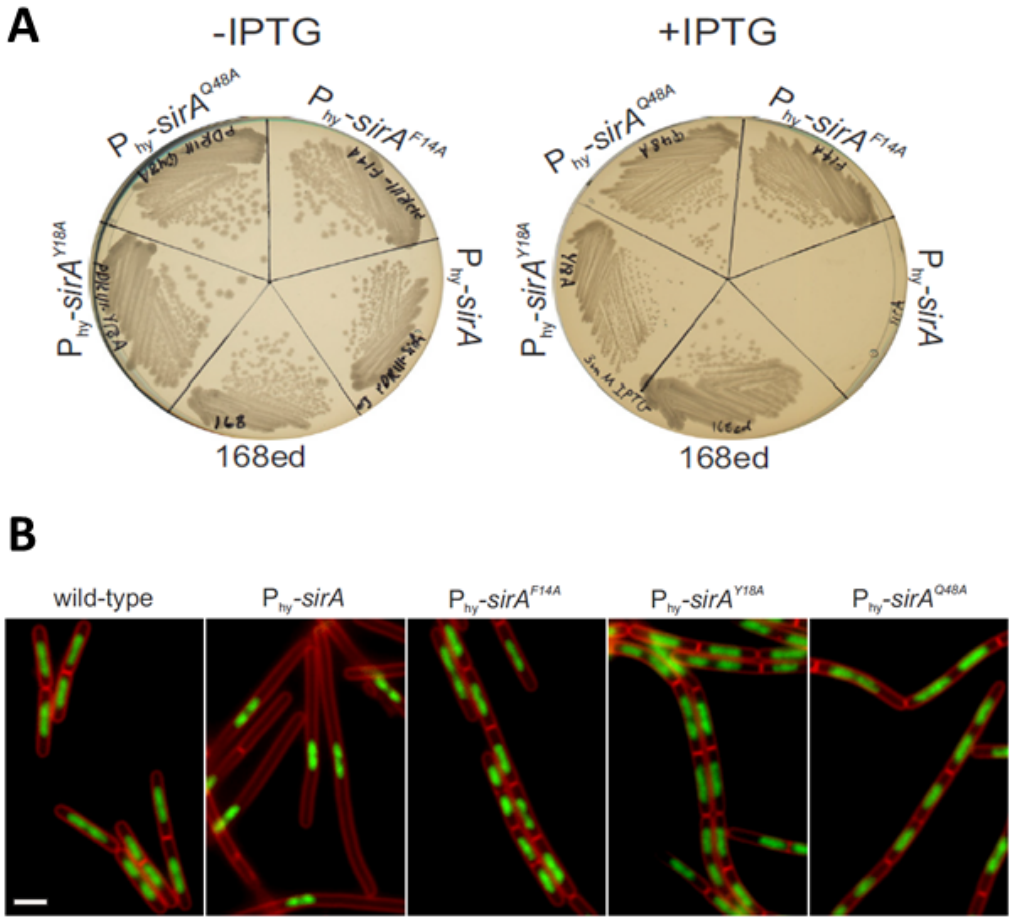


Figure 2.23: Growth phenotypes of wild-type SirA vs. SirA variants: Figure adapted from a figure originally produced by N. Rostami and H. Murray. **A)** Wild-type and mutant sirA were placed under control of an IPTG-inducible promoter and streaked on nutrient agar plates in the presence and absence of IPTG (3 mM). **B).** Comparison of overexpression of wild-type and mutant SirA in cells grown in liquid media. Images taken 180 min post-induction of gene expression (with 3mM IPTG). Membrane dye FM5-95 was used to highlight cell outlines and DAPI was used to stain the DNA. Scale bar = 3 µm.

The structure of SirA-DnaA^{DI}; an inhibitory complex of DnaA^{DI}, complements that of HobA-DnaA^{DI} from *H. pylori*[73] where DnaA^{DI} is in a complex which leads to activation of DNA replication initiation. As previously inferred, we have found that SirA binds a structurally equivalent site on DnaA^{DI} as the regulators HobA[73] and DiaA[80]; structural homologues from the organisms *H. pylori* and *E. coli*, respectively. HobA and DiaA are activators of DNA replication initiation, known to form tetramers which promote the oligomerisation of DnaA at *oriC*[157]. Each HobA/DiaA tetramer binds to four DnaA^{DI} molecules in a way that is thought to facilitate DnaA-binding to the array of DnaA-boxes distributed at *oriC*[160]. Conversely, SirA binds only one molecule of DnaA^{DI} in a manner which inhibits DNA replication initiation. Although SirA and HobA/DiaA have quite different 3D folds, each masks a structurally equivalent surface on DnaA. It is intriguing therefore, that they elicit such different regulatory outcomes.

It has previously been suggested that SirA inhibits DNA replication initiation by preventing the binding of DnaA at *oriC* based on the fact that induction of SirA reduces the number of DnaA-GFP foci at *oriC* during vegetative growth, and that the amount of DnaA at *oriC* decreases in a SirA-dependent manner following the induction of sporulation[83,193]. However, the finding that SirA forms foci at the mid-cell during sporulation, at sites distal from the replication origins suggests that SirA cannot be directly inhibiting DnaA oligomerisation at *oriC*. In *E. coli,* DnaA^{DI} is thought to play a role in oligomerisation, and is suggested to form dimers mediated on one surface of DnaA^{DI} [75]. The corresponding 'dimerisation surface' in *B. subtilis* DnaA^{DI} is located opposite the SirA binding surface, suggesting that SirA could not inhibit this mode of dimerisation. However, in *B. subtilis*, we did not observe DnaA^{DI} mediating dimer or oligomer formation *in vitro*, and the equivalent tryptophan residue to that essential in *E. coli* is buried within the DnaA^{DI} structure in both its 'free' and 'SirA-bound' forms, suggesting that *B. subtilis* DnaA^{DI} does not form dimers in an analogous manner to the *E. coli* orthologue. There is no published evidence to suggest that *B. subtilis* DnaA^{DI} mediates dimer formation. Collectively, this evidence suggests SirA is not likely to act by inhibiting DnaA^{DI}-DnaA^{DI} interactions.

Based on the observation that SirA accumulates at the replisome during sporulation, and that this interaction appears to be mediated by DnaA^{DI}, one hypothesis for SirA's inhibition mechanism is that SirA interacts with replisome-bound DnaA in a manner which stabilises the replisome-DnaA complex, and thus inhibits DnaA from re-binding to *oriC*. This is similar to the DnaA-tethering model proposed for YabA[215]. However, the observation that the

mutations Thr26Ala, Trp27Ala and Phe49Ala could not be introduced to *dnaA* suggests that these residues may be involved in an essential interaction with another component of the initiation complex, and thus SirA may instead act by inhibiting this interaction. In *E. coli*, $DnaA^{DI}$ is directly implicated in the recruitment of the DNA helicase[74]. However, in *B. subtilis* there is no evidence for a direct interaction between DnaA and the helicase. *B. subtilis* helicase recruitment requires two additional essential DNA remodelling proteins DnaD and DnaB[37]; DnaD interacts directly with DnaA, and forms multimeric scaffolds on DNA[100]. DnaD subsequently recruits DnaB, which is thought to bridge an interaction with the helicase-helicase loader[38]. Thus, SirA may serve to inhibit a DnaA-DnaD interaction which arrests the assembly of the initiation complex.

In summary, the SirA-$DnaA^{DI}$ complex presented here is the first of $DnaA^{DI}$ in an inhibitory complex. It defines the interaction surfaces of SirA and $DnaA^{DI}$ and the stoichiometry of their interaction. The complex is physiologically relevant, as sporulating cells in *B. subtilis* containing mutations in the SirA-$DnaA^{DI}$ interface have been shown to affect GFP-SirA foci formation at the replisome. These results will assist the elucidation of SirA's mechanism of action, which is limited by our current knowledge of the role of $DnaA^{DI}$ in DNA replication initiation within *B. subtilis*.

Chapter 3 : YabA Biophysical Characterisation

3.1: Abstract

YabA is a key regulator of DNA replication in *B. subtilis*; it controls the timing and synchrony of DNA replication in vegetatively growing cells by a currently unknown mechanism of action. In an attempt to provide structural insight into YabA's mechanism, an *in silico* model of YabA has been produced based on regions of homology within YabA for a leucine zipper domain and a Zn-binding domain. The initial model is a dimer, resembling the shape of a hammer; however YabA has been experimentally determined to be a tetramer. Therefore, three potential 'dimer of dimer' models have been proposed to account for the tetramer structure. Here, we have used biophysical methods to characterise the shape of YabA, to determine which of the proposed tetramer models is the most consistent with YabA's experimental behaviour. Initial characterisation of YabA by SEC-MALLS and AUC suggested that (as expected) our recombinant YabA forms a tetramer, and that this tetramer has an elongated structure. Individual domains of YabA were subsequently cloned, produced and characterised to further analyse the overall quaternary structure. These results suggest that the model which is the most consistent with our *in vitro* knowledge of YabA resembles a 'doggy bone': in this model, two coiled-coil domains form an elongated helical 'core' with two Zn-cluster domains residing at either end of the structure.

3.2: Introduction

YabA is a negative regulator of DNA replication in *B. subtilis*, affecting the synchrony of DNA replication and frequency of DNA replication initiation[161]. YabA has been shown to form physiologically relevant interactions with both the DNA replication initiator, DnaA, and the DNA polymerase clamp, DnaN[162,163]. Although this is reminiscent of the *E. coli* regulator Hda, which hydrolyses ATP within DnaA as a method for down-regulating DNA replication initiation[125,126,128], it is clear that YabA must employ a different mechanism for regulating DNA replication initiation frequency. YabA shares no homology with Hda (or DnaA) and does not affect the turnover of ATP in DnaA[109] . YabA has been shown to affect the cooperativity of DnaA binding at *oriC* [34], and over-expression of DnaN has been shown to increase DNA replication events in a manner which mimics a YabA deletion phenotype[34,216]. YabA is conserved in low-GC Gram positive bacteria, representing a distinct family of replication regulator proteins. This makes its molecular mechanism of significant interest. Models have been proposed for YabA's activity, but an understanding of the structure of YabA (and its interactions with DnaA and DnaN) would help to better understand the mechanism YabA uses to regulate the initiation of DNA replication.

Homology based sequence analysis suggests that YabA consists of an N-terminal leucine zipper motif, and a C-terminal zinc-binding domain separated by an area of poor sequence conservation[163]. Collaborators L. Felicori and F. Molina have proposed an *in silico* dimer model of YabA (Fig. 3.1) based on structural templates which were identified due to their homology to either the leucine zipper or zinc-binding portion of YabA. In the dimer model, the N-terminal region forms an α-helical coiled-coil domain connected to two adjacent Zn-binding domains via a flexible 'hinge' region. The overall shape of the dimer resembles that of a hammer, with the coiled-coil domain forming the handle and the Zn-binding domains forming the hammerhead.

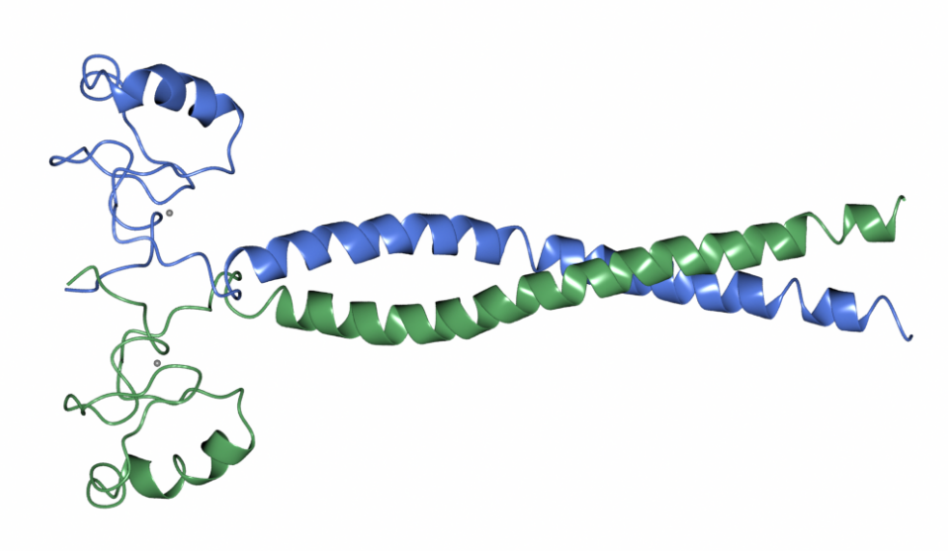


Figure 3.1: *In silico* model of YabA: Model of YabA created using constraint-based homology modelling in the program MODELLER (9v8). YabA forms a dimer which resembles a hammer. The N-terminal leucine zipper residues from a coiled-coil domain (the hammer's handle) which is linked to two Zn-binding domains via a 'flexible hinge'. The Zn-binding domains lie adjacent to one another, forming the hammerhead.

YabA has been previously shown to form tetramers in solution[163]; an observation which is not directly addressed by the dimer model of YabA. However, this discrepancy can be reconciled by assuming that the model is able to form a tetramer from a dimer of dimers. It is not immediately obvious how this dimer of dimers would form however, and thus, 3 alternative conformations have been proposed (Fig 3.2). In the first, (Fig. 3.2, A) the hammerhead regions of the YabA dimers come together to form a group of 4 Zn-binding domains with the coiled-coil domains protruding out in opposite directions. Overall this produces an elongated structure which represents 'beads on a rail'. In the second model (Fig 3.2, B), 'doggy bones', the dimers associate via their N-terminal coiled-coil domains, and the Zn-binding domains are separated at either end of the model - again this displays an overall elongated structure. In the final model, 'bunch of flowers', the dimers lie parallel, with both coiled-coil domains together and all Zn-binding domains associated with one another, forming a more compact structure than the first two models.

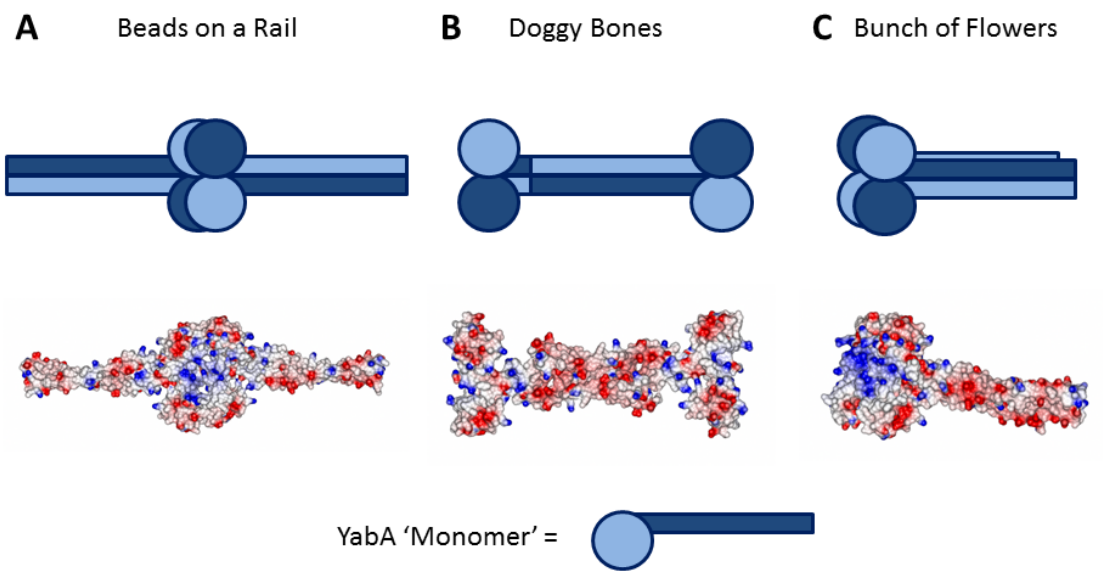


Figure 3.2: YabA tetramer models : proposed dimers of dimers of YabA (models of the tetramer). **A)** The dimers come together at their Zn-binding domains to form a structure resembling 'beads on a rail'. **B)** The dimers come together at their coiled-coil domains, resembling 'doggy bones'. **C)** The dimers run parallel to one another, with both coiled-coil and Zn-binding domains together, resembling a 'bunch of flowers'.

Here we are interested in using biophysical characterisation to identify which of the tetrameric models of YabA is most consistent with YabA's behaviour *in vitro*. Identifying YabA's shape will aid our understanding of how YabA interacts with DnaA and DnaN.

A construct which directs the expression of full length YabA with an N-terminal hexa-histidine tag was previously cloned at YSBL by M. Fogg. This construct has been utilised here for the production of YabA.

3.3: Aims

Understanding the quaternary structure of YabA may provide insight into its interactions with both DnaA and DnaN, and thus the potential mechanisms YabA employs to regulate the initiation of DNA replication. The purpose of this work was to determine which of several proposed *in silico* tetramer models are most consistent with YabA's behaviour *in vitro*. The initial aim of the work was to produce YabA by expression in *E. coli* and to purify it by chromatography methods. Subsequently, to carry out SEC-MALLS and AUC experiments with purified YabA to characterise its quaternary structure. Using these results, we planned to develop further hypotheses to investigate the quaternary structure of YabA.

3.4: Methods

3.4.1: YabA Expression

The plasmid pET-YSBLIC3C-YabA was introduced into *E. coli* BL21 (DE3) cells for the over-expression of *yabA*. Overnight LB cultures were used to inoculate 500 ml LB cultures supplemented with 30 μg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (shaking at 180 rpm) before protein production was induced with 1 mM IPTG. Following induction, cultures were grown at 16 °C (180 rpm shaking) overnight (~20 hours) before cells were harvested by centrifugation.

3.4.2: YabA Purification

Harvested cells were re-suspended in a buffer of 50 mM Tris, 500 mM NaCl, 20 mM Imidazole, containing an EDTA-free protease inhibitor cocktail tablet (Roche). Re-suspended cells were lysed by sonication and lysate clarified by centrifugation. The cell lysate was loaded on a HisTrap FF crude Ni-affinity column (GE Healthcare) at approx. 2-2.5 mL/min and bound protein eluted over an increasing imidazole concentration gradient (0 – 500 mM). Fractions containing YabA were concentrated to less than 2 mL for size-exclusion chromatography. Protein was loaded onto a HiLoad 16/60 Superdex S200 column (GE Healthcare) equilibrated with 50mM Tris, 150 mM NaCl and run at 0.75 ml/min. Fractions containing YabA were combined according to analysis by SDS-PAGE, and the N-terminal histidine tag was removed by incubation with 3C protease overnight at a protease:protein ratio of 1:50. Passage through a second Ni-affinity column to remove the histidine tag and tagged protease yielded pure protein in a buffer of 50 mM Tris, 150 mM NaCl, 20 mM Imidazole. Protein was assessed as pure according to analysis by SDS-PAGE.

3.4.3: Analytical Ultracentrifugation: YabA

A sedimentation velocity (SV) experiment was carried out across a range of YabA concentrations (0.07 mg ml^{-1}, 0.15 mg ml^{-1}, 0.31 mg ml^{-1}, 0.63 mg ml^{-1}, 1.25 mg ml^{-1}) in 50 mM Tris, 150 mM NaCl. The SV experiment was run at 50,000 rpm at 20 °C using an Optima XL/I (Beckman) instrument and an AN50Ti rotor, with scans measured at 275 nm. The results were analysed to provide a frictional ratio and sedimentation coefficient specific to YabA (analysis carried out by A. Leech). These results were then compared to theoretical values calculated for various tetrameric YabA models.

3.4.4: Cloning: YabA$^{1-58}$ and YabA$^{70-119}$

Constructs encoding YabA N-terminal domain (YabA$^{1-58}$) and YabA C-terminal domain (YabA$^{70-119}$) were cloned using a deletion mutagenesis method. A whole vector amplification of pET-YSBLIC-YabA was performed by PCR using primers which would produce a desired deletion in the YabA coding region (Table 3.1). PCR reactants and conditions are listed in Tables 3.2 and 3.3. 45µl of PCR product was incubated with 1.5 µl DpnI (15 units) at 37 °C for 4-5 hours, before transformation into Top10 competent cells. 1µl of digested product was added to 50 µl Top10 competent cells before incubation for 30 minutes on ice. Cells were heat-shocked at 42 °C for 50 seconds before re-incubation on ice for 10 minutes. 150 µl LB media was added and cells were incubated at 37°C for 60 minutes prior to plating of 100 µl of the culture onto LB agar supplemented with 30 µg ml^{-1} kanamycin. Plates were incubated for 18-20 hours at 37°C until the production of colonies around 1.0-1.5 mm in diameter. Cultures containing 5 ml LB supplemented with 30 µg ml^{-1} kanamycin and a single colony were incubated at 37 °C overnight, before isolation of plasmid DNA using a mini-prep kit (Qiagen). Resultant plasmids were analysed by double digest and sequences confirmed to be correct by sequencing (GATC). The recombinant plasmids (pET-YSBLIC3C-YabA$^{1-58}$ and pET-YSBLIC3C-YabA$^{70-119}$) direct the expression of YabA residues 1 to 58 or 70-119 fused to a 3C cleavable N-terminal His-tag.

Table 3.1: YabA construct primers and coding regions: : listed are the primers used for cloning of YabA constructs and their resultant coding regions. Bases written in bold have been appended for annealing or insertion into the YSBLIC3C vector.

Construct	Residues of YabA Encoded	Cloning Primers		Template
pET-YSBLIC3C-YabA$^{1-58}$	1-58	Fwd	5'-**CGACGCAGCAG**TAACGCGCCTTC TCCTCACATATGGC-3'	pET-YSBLIC3C-YabA
		Rev	5'-**AAGGCGCGTTA**CTGCTGCGTCGT ATCGTCCAGC-3'	
pET-YSBLIC3C-YabA$^{70-119}$	70-119	Fwd	5'-**AGGGACCAGCAA**AAACACAGAA GACAGAGCAAACTGATATAGGG-3	pET-YSBLIC3C-YabA
		Rev	5'-**GTCTTCTGTGTTTT**TGCTGGTCC CTGGAACAGAACTTCC-3'	

Table 3.2: PCR components for cloning pET-YSBLIC3C-YabA$^{1-58}$ and pET-YSBLIC3C-YabA$^{70-119}$

Reactant	Volume/ µl (50µl reaction)
pET-YSBLIC3C-YabA (100 ng µl^{-1})	1
Oligo I (20 pmol µl^{-1})	0.5
Oligo II (20 pmol µl^{-1})	0.5
KOD Hot Start DNA Polymerase Buffer (10x)	5
dNTPs (2mM)	5
KOD Hot Start DNA Polymerase (1,000 U/ ml)	1
$MgSO_4$ (25mM)	2
H_2O	35

Table 3.3: PCR Conditions for cloning of pET-YSBLIC3C-YabA$^{1-58}$ and pET-YSBLIC3C-YabA$^{70-119}$

Step	Temperature /°C	Duration /min	Cycles
Initial Denature	94	5	1
Denature	94	1	30
Annealing	55	1	
Extension	72	7	
Final Extention	72	3	1

3.4.5: Double Digest

50 µg DNA was digested with 10 units of NcoI and 10 units of NdeI in NEB buffer 2 at a total volume of 25 µl. The mixture was incubated for 1 hour at 37 °C before a 20µl sample was combined with 5µl DNA loading dye and analysed on a 1% agarose gel.

3.4.6: Expression: YabA$^{1-58}$ and YabA$^{70-119}$

The recombinant plasmid pET-YSBLIC3C-YabA$^{1-58}$ or pET-YSBLIC3C-YabA$^{70-119}$ was introduced into *E. coli* BL21 (DE3) cells for over-expression.

For small scale expression tests, 200 µl of overnight LB culture was used to inoculate 10 ml LB cultures supplemented with 30 µg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (shaking at 180 rpm) before cells from 1 ml of culture were havested by centrifugation, and protein production was induced in the remaining culture with 1 mM IPTG. Following induction, cultures were grown at 37°C for 4 hours, or 16°C for 18-20 hours

with shaking at 180 rpm. Cells from 1 ml of culture were harvested by centrifugation for analysis of protein production.

For large scale expression, overnight LB cultures were used to inoculate 500 ml LB cultures supplemented with 30 μg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (shaking at 180 rpm) before protein production was induced with 1 mM IPTG. Following induction, cultures were grown at 37°C for 4 hours, with shaking at 180 rpm. Cells were subsequently harvested by centrifugation (Sorvall RC 5B Plus, 15 min at 15 K (SS34 rotor)). The pelleted cells were re-suspended in 50 mM Tris pH 8.0, 500 mM NaCl, 20 mM Imidazole (purification lysis buffer) prior to re-pelleting by centrifugation (10 min, 5K), for storage at -20°C.

3.4.7: Purification: $YabA^{1-58}$ and $YabA^{70-119}$

$YabA^{1-58}$ and $YabA^{70-119}$ were purified by the procedure outlined for YabA in section 3.4.2. with a modified size-exclusion step; protein fractions were combined and concentrated to ≤ 1 mL and loaded onto a HiLoad 16/60 Superdex S75 column (GE Healthcare).

3.4.8: Mass Spectrometry

Purified $YabA^{1-58}$ and $YabA^{70-119}$ were buffer exchanged into 2 mM Tris for analysis by Electrospray Ionisation-Mass Spectrometry (ESI-MS). Samples were provided at a concentration of 5 mg ml^{-1}. Spectra were obtained and processed by A. Leech.

3.4.9: Circular Dichroism

Circular dichroism spectra were recorded at 20 °C on a Jasco J-810 CD spectrophotometer using a quartz cell with a 0.1 cm path length. Experiments were carried out at a concentration of 0.2 mg mL^{-1} in a buffer of 20 mM K Phosphate. Spectra were recorded across the wavelength range of 260 nm - 185 nm. A scan of the buffer (20 mM K Phosphate) was also measured and subtracted from the protein spectra to remove any contribution from the buffer.

3.4.10: SEC-MALLS

To determine the oligomeric state of YabA, YabA$^{1-58}$ and YabA$^{70-119}$ analysis was carried out by SEC-MALLS. 100 µl samples of protein were loaded onto a size-exclusion column equilibrated with 50 mM TRIS pH 8.0, 150 mM NaCl. YabA was loaded at a concentration of 2.5 mg ml^{-1} onto a Superdex 200 HR 10/30 column. YabA$^{1-58}$ and YabA$^{70-119}$ were loaded at a concentration of 5 mg mL^{-1} onto a Superdex 75 HR 10/30 column. The eluate was successively analysed by a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor. Data was analysed with Astra software (Wyatt).

3.5: Results

3.5.1: YabA Protein Production

The vector pET-YSBLIC3C-YabA directs the expression of YabA with an N-terminal cleavable hexa-histidine tag. *YabA* was over-expressed from this vector yielding soluble protein following the procedure outlined in methods (section 3.4.1). Analysis of protein production and solubility by SDS-PAGE (Fig. 3.3) showed that high levels of YabA protein production were induced (Fig. 3.3, lane 3) and that this protein was soluble in a re-suspension buffer of 50 mM Tris pH 8.0, 500 mM NaCl and 20 mM Imidazole (Fig. 3.3, lane 4).

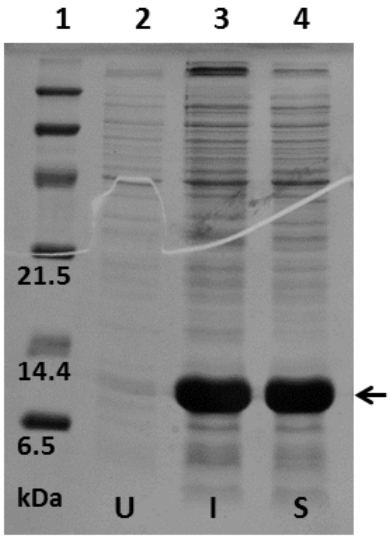


Figure 3.3: Production of YabA: 1) Low molecular weight markers; 2) Lysate from un-induced cell culture; 3) Lysate from cell culture induced to produce YabA; 4) soluble cell lysate from cell culture induced to produce YabA.

3.5.2: YabA Purification

YabA can be purified by a combination of Ni-affinity and size-exclusion chromatography as outlined in section 3.4.2. Clarified cell lysate containing His-tagged YabA is loaded onto a Ni-affinity column before elution over an increasing imidazole gradient (20-500 mM). YabA elutes from the column in a single peak at an imidazole concentration of approx. 200mM (a chromatogram is shown in Fig. 3.4, A). A shoulder is visible on the left-hand side of the peak indicating elution of some protein at a lower imidazole concentration. Analysis by SDS-PAGE shows that although YabA elutes across the entirety of the peak, the protein eluting in the shoulder region at a lower imidazole concentration contains more contaminants (Fig. 3.5, lane 4). Therefore, only protein contributing to the main peak is

combined (in this case, fractions 19 to 35). The combined fractions are concentrated to approx. 2ml for size-exclusion chromatography. The chromatogram from the size-exclusion run contains two peaks (3.4, B); a minor peak at approx. 48 ml and a major peak approx. 70 ml. Analysis by SDS-PAGE indicates that the minor peak contains a small quantity of YabA (suspected aggregate) and some contaminants (Fig. 3.5, lane 8) whilst the major peak contains the majority of YabA protein (Fig. 3.5, lanes 9-11).

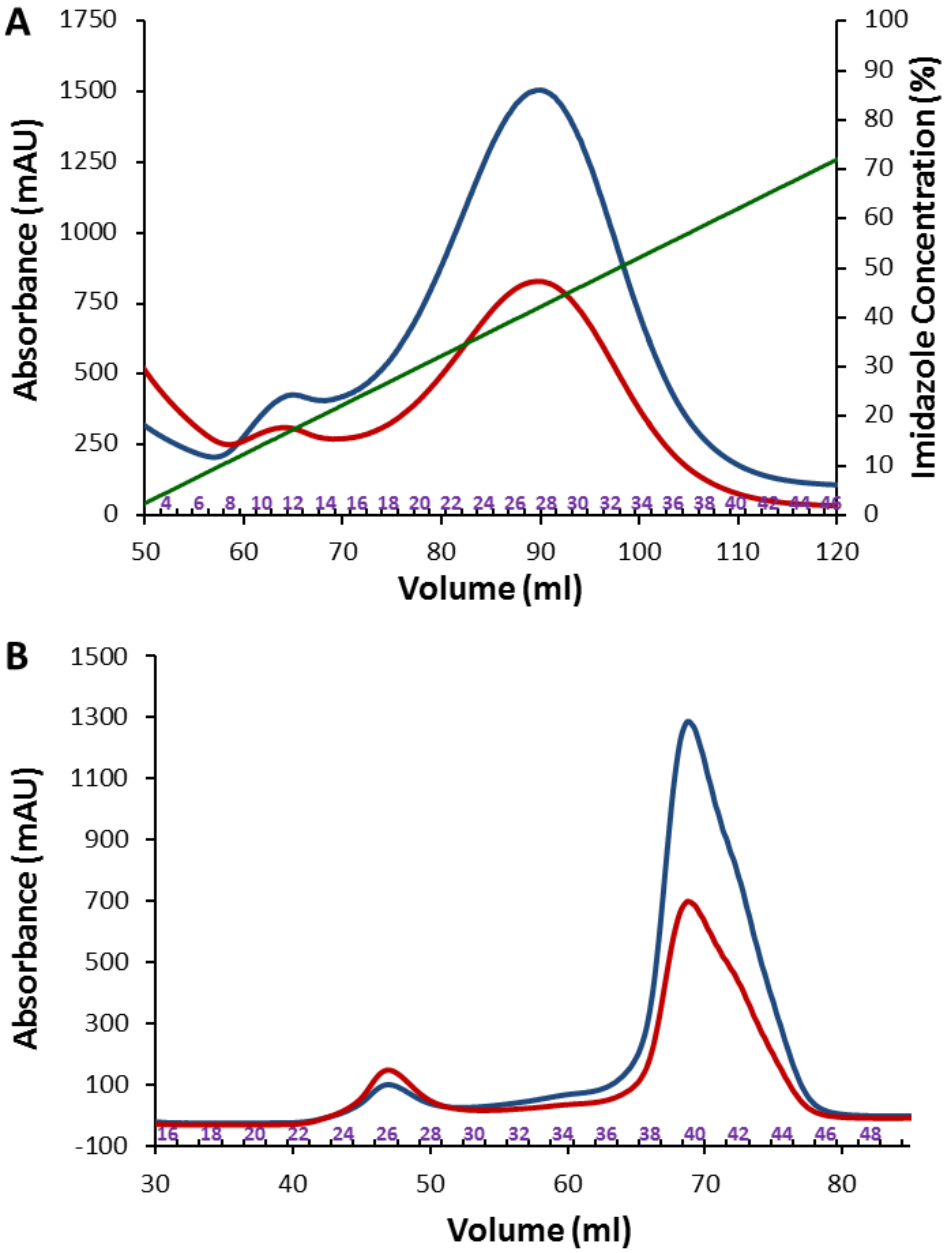


Figure 3.4: YabA purification by Ni-affinity and size-exclusion chromatography: Chromatograms showing elution of YabA protein from; **A)** a Ni-affinity column; and **B)** a size-exclusion column. In blue, absorbance at 280 nm; in red, absorbance at 254 nm; numbers in purple denote fractions collected.

Fractions containing YabA are combined and the histidine tag is removed by incubation overnight with 3C protease (Fig. 3.5, lanes 14 and 15). The protein is then passed over a second Ni-column where tagged-protease and un-cleaved proteins are retained on the column, whilst pure YabA emerges in the flowthrough (Fig. 3.5, lane 15).

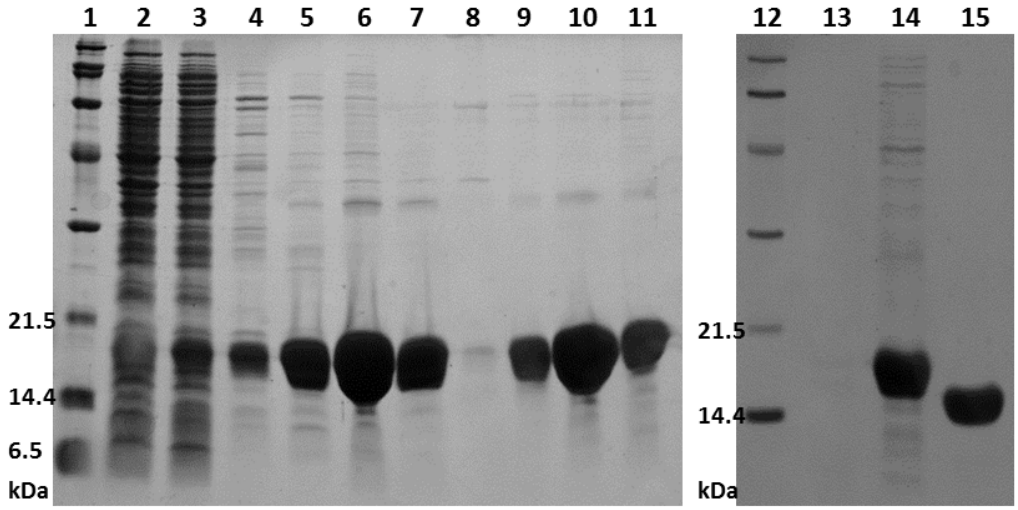


Figure 3.5: YabA purification shown by SDS-PAGE: lane 1) Broad-range markers. Lanes 2-7) fractions from Ni-affinity chromatography: lane 2) Flowthrough during column loading; lane 3) wash with buffer A; lane 4) 10 µl fraction 11; lane 5) 10 µl fraction 18; lane 6) 10µl fraction 27; lane 7) 10µl fraction 35. Lanes 8-11) fractions from size-exclusion chromatography: lane 8) 10µl fraction 26; lane 9) 10µl fraction 36; lane 10) 10µl fraction 41; lane 11) 10µl fraction 44. Lane 12) Low molecular weight markers; lane 13) blank; lane 14) His-YabA (combined fractions after size exclusion); lane 15) YabA in flow-through from 2nd Ni-affinity column following His-tag removal.

3.5.3: SEC-MALLS: YabA

To determine the oligomeric state of recombinant YabA, analysis was carried out by size exclusion chromatography multi-angle laser light scattering (SEC-MALLS). The technique is used to determine accurate molecular masses in solution, allowing the oligomeric state of the protein to be inferred. Analysis of YabA at 2.5 mg ml^{-1} produced an elution profile with a single peak at ~26.5 min (Fig. 3.6). This corresponds to an empirical molecular mass of 58.7 kDa. The calculated mass of YabA is 14.3 kDa, thus YabA appears to be forming tetramers in solution (14.3 kDa x 4 = 57.3 kDa), consistent with the oligomeric state previously reported[163].

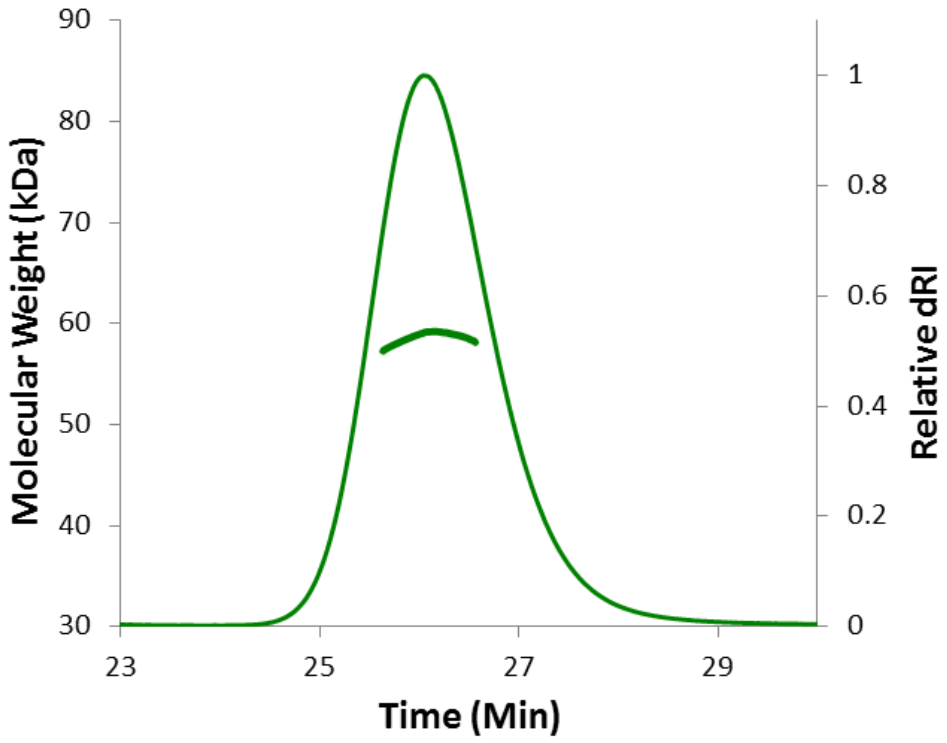


Figure 3.6: SEC-MALLS analysis of YabA : The thinner line displays the differential refractive index of the YabA containing elute from a Superdex 10/300 S200 column as a function of time. The thicker line show the weight average molecular mass of the eluting species, calculated from refractive index and light scattering measurements. Analysis was carried out at 2.5 mg ml^{-1}, revealing a species of approximately 58.7 kDa, indicating that YabA is a tetramer in solution.

3.5.4: AUC: YabA

The *in silico* dimer model of YabA has been proposed to form a tetramer of YabA in one of 3 ways: A) with all C-terminal domains together (beads on a string) B) with all N-terminal domains together (doggy bones) or C) with both N and C- terminal domains together (bunch of flowers) (Fig. 3.2). To help determine which model was more consistent with the conformation of YabA *in vitro*, analytical ultracentrifugation was used to provide information on the shape of YabA. Using a sedimentation velocity experiment, a frictional ratio and sedimentation coefficient was calculated for YabA, which was then compared to theoretical values for each of the three proposed YabA tetramer models (Table 3.4). YabA was found to have a sedimentation coefficient of approx. 3.0 and a frictional ratio of approx. 1.75. When compared to the theoretical values for the YabA tetramers, this was judged to be most consistent with the more elongated models of YabA; namely the 'beads on a string' and 'doggy bones' models and inconsistent with the theoretical values for the 'bunch of flowers' model. (Analysis performed by Andrew Leech).

Table 3.4: Analysis of YabA by AUC

	YabA (empirical)	**Beads on a Rail** (theoretical)	**Doggy Bones** (theoretical)	**Bunch of Flowers** (theoretical)
Frictional Ratio	1.75	1.71	1.64	1.44
Sedimentation Coefficient	3.0	3.08	3.21	3.68

These data allowed us to rule out the 'bunch of flowers model', but did not give a clear indication which of the other two models was more likely. However, we proposed that we could differentiate between the 'beads on a string' and doggy bones' models by analysing the respective oligomeric states of the N- and C-terminal domains of YabA. For the 'beads on a string' model, we would expect the C-terminal domains to form a tetramer, whist the N-terminal domains form dimers. However, for the 'doggy bones' model, we would expect the N-terminal domains to form a tetramer and the C-terminal domains to form dimers.

As the central residues of YabA (assumed to reside between the N- and C-terminal domains) are poorly conserved (Fig. 3.7) and contain many trypsin cleavage recognition sites, we initially attempted to separate the two domains by limited proteolysis. However

the results of limited proteolysis experiments on YabA were hard to reproduce, giving inconsistent proteolysis products. Thus, YabA N- and C-terminal domains were cloned to allow protein production, purification and ultimately, analysis of the domains separately.

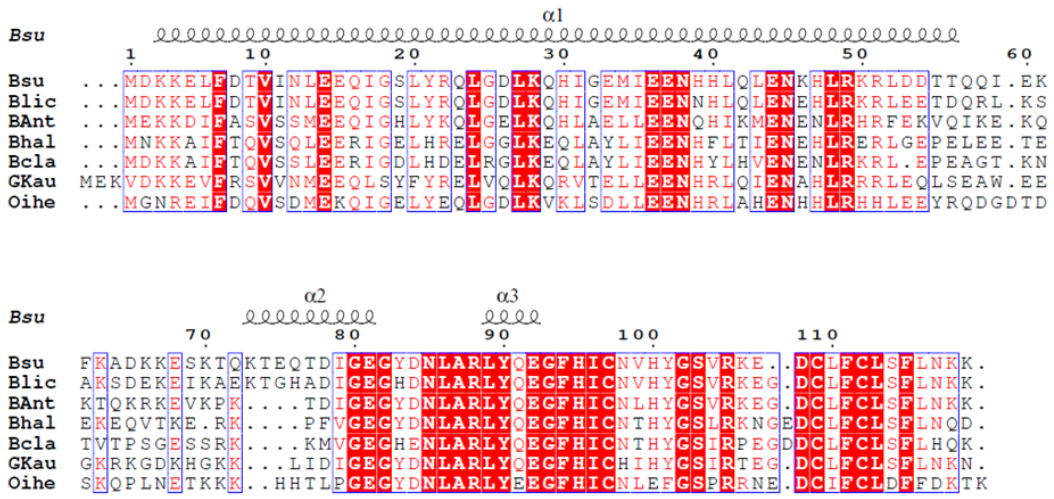


Figure 3.7: Alignment of the sequences of orthologues of YabA from selected *Bacillus* species: Bsu, *B. subtilis*; Blic, *B.licheniformis*; Bant, *B. anthracis*; Bhal, *B. halodurans*; Bcla, *B. clausii*; Gkau, *Geobacillus klaustophilus*; Oihe, *Oceanobacillus iheyensis*. Secondary structure elements taken from the *in silico* model of YabA and residue numberings are displayed above the alignment. The images were created using ESPript (Gouet, 2003).

3.5.5: Cloning of YabA N- and C-terminal domains

Sequence comparison of *Bacillus* YabA proteins shows that YabA is poorly conserved in a central region between the N-terminal leucine zipper domain and the C-terminal Zn-binding domain (Fig. 3.7). A comparison of this sequence conservation and the *in silico* YabA model (Fig. 3.1) informed the construct design for the cloning of the N- and C-terminal domains of YabA. The YabA N-terminal domain is well conserved up to residue 54, however the model predicts that the N-terminal helix extends to residue 58, thus a construct was created consisting of residues 1-58, hereafter known as YabA$^{1\text{-}58}$. YabA's C-terminus is conserved from residue 79 to the C-terminal residue, however the *in silico* model predicts helical structure between residues 72 and 82; thus a C-terminal construct was designed to also encompass these residues. A construct was created containing residues 70-119, hereafter known as YabA$^{70\text{-}119}$.

Vectors encoding YabA$^{1-58}$ and YabA$^{70-119}$ were created using a deletion mutagenesis method in which oligonucleotides were designed to create a deletion within the *yabA* gene on the pET-YSBLIC3C-YabA vector. PCR was used to amplify the pET-YSBLIC3C-YabA vector with oligonucleotides which would result in the deletion of a section of the y*abA* gene encoding residues 59-119 or encoding residues 1-69 for the YabA$^{1-58}$ and YabA$^{70-119}$ products, respectively. PCR products from this amplification can be seen in Fig. 3.8. The PCR products for YabA$^{1-58}$ and YabA$^{70-119}$ were expected to contain 5570 and 5546 base pairs, respectively. This is very close to the number of base pairs in the template (5570 bps) therefore it would be very difficult to differentiate between the two on an agarose gel. However, as a strong band can be seen around the expected number of base pairs for the product, and template concentration is very low, it was assumed that the bands seen in Fig. 3.8 were the expected products and that PCR had been successful.

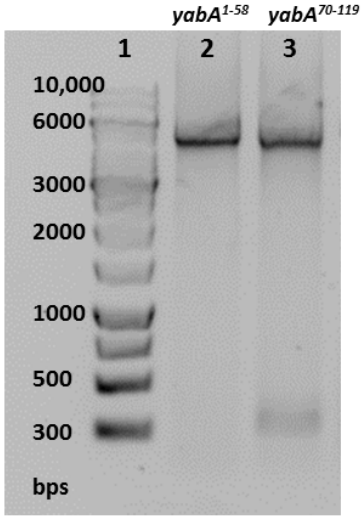


Figure 3.8: PCR of YabA$^{1-58}$ and YabA$^{70-119}$: Lane 1) 1kb DNA ladder; lane 2) 5 µl PCR product for *yabA$^{1-58}$*; 5 µl PCR product for *YabA$^{70-119}$*.

PCR products were treated with DpnI to digest the template, before transformation of Top10 cells. DNA was isolated from several of the resulting colonies. Double digests were performed as described in section 3.4.5 to identify recombinant plasmids containing products of the desired size (Fig. 3.9). A double digest of the PCR template DNA, pET-YSBLIC-YabA, with NcoI and NdeI is expected to produce two products; one of 5310 base pairs, and another of 443 base pairs, as can be seen in Fig. 3.9, lane 1. The same double digest of pET-YSBLIC-YabA$^{1-58}$ or pET-YSBLIC-YabA$^{70-119}$ is expected to yield products of 5310 base pairs and 260 bps or 236 bps, respectively. Bands in Fig. 3.9, lane 2, show the

expected number of base pairs for pET-YSBLIC-YabA$^{1-58}$ (5310 and 260 bps) and bands in Fig. 3.9, lane 5, show the expected number of base pairs for pET-YSBLIC-YabA$^{70-119}$ (5310 and 236 bps). These plasmids were subsequently confirmed as correct by DNA sequencing.

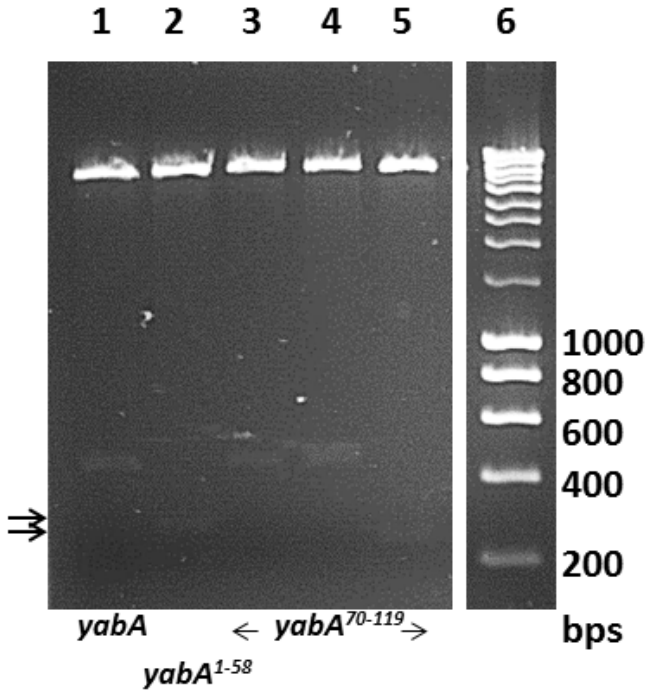


Figure 3.9: Double digests of YabA$^{1-58}$ and YabA$^{70-119}$: Lane 1) double digest products of 0.5 µg pET-YSBLIC3C-YabA; lane 2) double digest products of 0.5 µg pET-YSBLIC3C-yabA$^{1-58}$; lane3-5) double digest products of 0.5 µg pET-YSBLIC3C-yabA$^{70-119}$ I-III; lane 6) 12 µl Hyperladder I.

3.5.6: Protein production: YabA$^{1-58}$ and YabA$^{70-119}$

The recombinant plasmids pET-YSBLIC-YabA$^{1-58}$ and pET-YSBLIC-YabA$^{70-119}$ were introduced into *E. coli* BL21 (DE3) cells for protein production. Small scale expression tests carried out with post-induction growth at either 16 °C for ~20 hours or 37 °C for ~4 hours indicated that protein production was inducible at both temperatures and yielded a significant amount of over-expression of the gene(s) and that the protein was soluble when re-suspended in 'buffer A' (50mM Tris pH 8.0, 500 mM NaCl and 20 mM Imidazole) (Fig. 3.10). Post-induction growth at 37 °C for 4 hours was subsequently used for protein production in large scale cell cultures as these appeared to have a slightly increased yield over those grown at 16 °C post-induction.

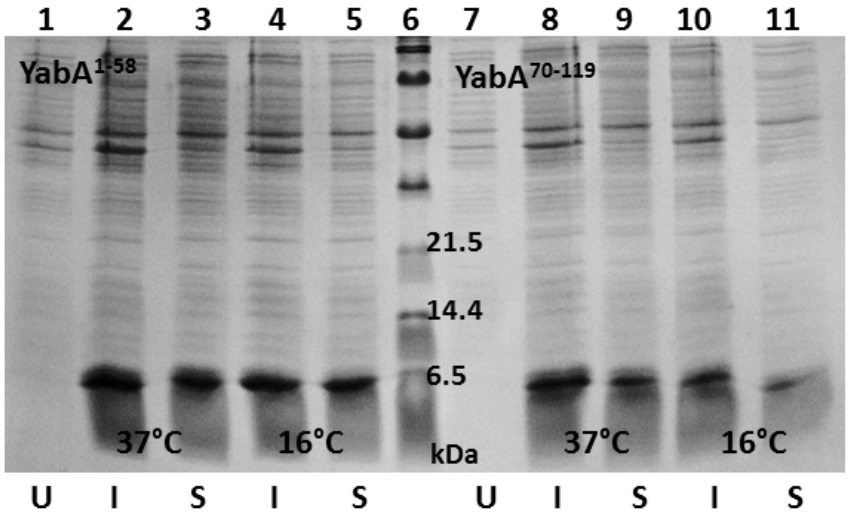


Figure 3.10: Expression of YabA$^{1-58}$ and YabA$^{70-119}$: lane 1) 10µl lysate from un-induced cell culture containing pET-YSBLIC3C-YabA$^{1-58}$; lane 2) 10µl lysate from cell culture induced to express *YabA$^{1-58}$* at 37 °C; lane 3) 10µl soluble lysate from cell culture induced to express *YabA$^{1-58}$* at 37 °C; ; lane 4) 10µl lysate from cell culture induced to express *YabA$^{1-58}$* at 16 °C; lane 5) 10µl soluble lysate from cell culture induced to express *YabA$^{1-58}$* at 16 °C; lane 6) broad-range markers; lane 7) 10µl lysate from un-induced cell culture containing pET-YSBLIC3C-YabA$^{70-119;}$ lane 8) 10µl lysate from cell culture induced to express *YabA$^{70-119}$* at 37 °C; lane 9) 10µl soluble lysate from cell culture induced to express *YabA$^{70-119}$* at 37 °C; ; lane 10) 10µl lysate from cell culture induced to express *YabA$^{70-119}$* at 16 °C; lane 11) 10µl soluble lysate from cell culture induced to express *YabA$^{70-119}$* at 16 °C.

3.5.7: YabA$^{1-58}$ and YabA$^{70-119}$ Purification

YabA$^{1-58}$ and YabA$^{70-119}$ can be purified in an analogous manner to YabA (see section 3.4.2). Proteins were assessed as pure according to staining on SDS-PAGE (Fig. 3.11).

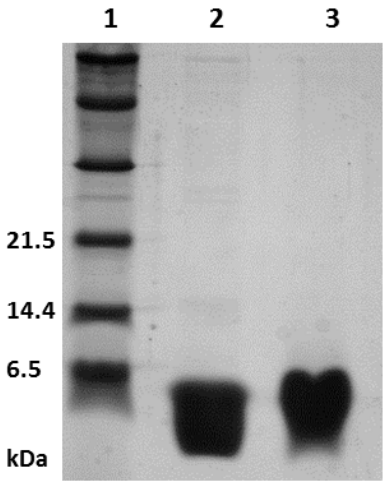


Figure 3.11: YabA$^{1-58}$ and YabA$^{70-119}$ following purification: lane 1) broad-range molecular weight markers; Lane 2) 10 µg of YabA$^{1-58}$; 10 µg of YabA$^{1-58}$.

3.5.8: Mass Spectrometry

YabA$^{1-58}$ and YabA$^{70-119}$ were analysed by electrospray ionisation – mass spectrometry (ESI-MS) for molecular weight determination. Results are shown in Table 3.5.

For YabA$^{1-58}$, the main peak in the processed spectrum is a mass of 7195 Da. This is 1 Da away from the theoretical mass of recombinant YabA$^{1-58}$, which is 7196 Da (within experimental error). The main peak in the processed spectrum for YabA$^{70-119}$ is at 6067 Da, exactly the theoretical mass of recombinant YabA$^{70-119}$. In both cases this indicates that the protein has been produced as designed and been isolated to a high purity.

Table 3.5: Mass spectrometry analysis of YabA$^{1-58}$ and YabA$^{70-119}$

Sample	Peak	Observed Mass (Da)	Theoretical Mass (Da)	Species
YabA$^{1-58}$	1	7195	7196	YabA$^{1-58}$
YabA$^{70-119}$	1	6067	6067	YabA$^{70-119}$

3.5.9: Circular Dichroism: YabA, YabA$^{1-58}$ and YabA$^{70-119}$

Circular dichroism was used to determine that YabA, YabA$^{1-58}$ and YabA$^{70-119}$ were folded *in vitro* and to assess the overall secondary structure composition of YabA. Circular dichroism spectra were recorded for YabA, YabA$^{1-58}$ and YabA$^{70-119}$ (Fig. 3.12). The spectra show strong secondary structure characteristics in all cases: all species show shallow minima in their molar ellipticity across the wavelength range of 205-225 nm and increase in molar ellipticity at wavelengths lower than 205 nm. This indicates that the proteins are folded in solution. In addition, spectra from all 3 proteins are consistent with a high percentage of α-helical character.

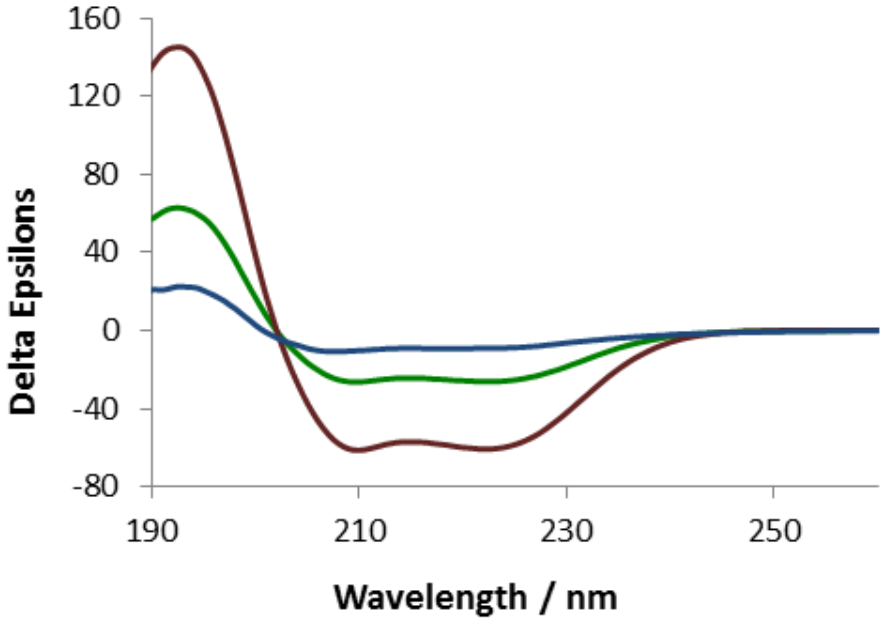


Figure 3.12: Overlay of CD spectra for YabA, YabA$^{1-58}$ and YabA$^{70-119:}$: In Blue: spectrum for YabA; in red: CD spectrum for YabA$^{1-58}$ and in green: CD spectrum for YabA$^{70-119}$.

3.5.10: SEC-MALLS: YabA$^{1-58}$ and YabA$^{70-119}$

To further our understanding of YabA's tetrameric conformation, YabA$^{1-58}$ and YabA$^{70-119}$ were analysed by SEC-MALLS to determine their oligomeric state *in vitro*.

The N-terminal fragment, YabA$^{1-58}$, elutes at ~21.5 min, with an associated empirical molecular weight of 27.7 kDa (Fig. 3.13). This is approximately 4 times the calculated mass of YabA$^{1-58}$ (7.2 KDa x 4 = 28.8 KDa) indicating that YabA's N-terminal domain (YabA$^{1-58}$) is a tetramer. Interestingly, the C-terminal fragment, YabA$^{70-119}$, elutes from the column in a single peak at a much later retention time of ~28 min. This peak has an associated empirical mass of 6.2 kDa. This is very close to its calculated mass of 6.0 KDa, indicating that the YabA C-terminal domain exists as a monomer. This experiment implies therefore that the principle quarternary structure determinants of YabA must reside in the N-terminal domain of the molecule.

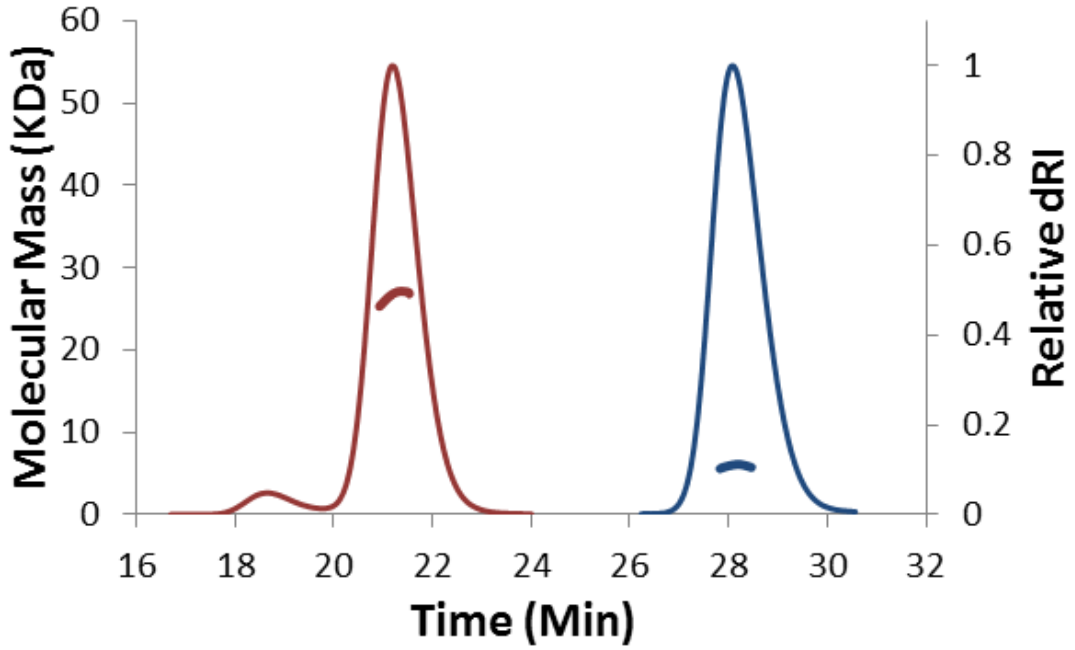


Figure 3.13: Overlay of SEC-MALLS analysis for YabA$^{1-58}$ and YabA$^{70-119}$: the thinner lines display the differential refractive index of the YabA$^{1-58}$ or YabA$^{70-119}$ containing elute from a Superdex 10/300 S75 column as a function of time. Thicker lines show the weight average molecular mass of the eluting species, calculated from refractive index and light scattering measurements. Red shows analysis of YabA$^{1-58}$ at 5 mg ml^{-1}, revealing a species of approximately 27.7 kDa, indicating that YabA$^{1-58}$ is a tetramer in solution. Blue shows analysis of YabA$^{70-119}$ at 5 mg ml^{-1}, revealing species of approximately 6.2 kDa, indicating that YabA$^{70-119}$ is a monomer in solution.

3.6: Discussion

YabA is a negative regulator of DNA replication in *B. subtilis*. Known to interact with both the DNA polymerase clamp, DnaN, and the DNA replication initiator protein DnaA[162,163], YabA controls the timing and synchrony of DNA replication in vegetatively growing *B. subtilis* cells[161]. YabA has been shown to affect the co-operativity of DnaA binding at *oriC* [34], and DnaN has been implicated in the YabA-dependent regulation of DNA replication events [34,216]. As there are no known homologous protein structures of YabA available, little structural information is available about YabA which can help to infer its mechanism of action. An *in silico* model of a YabA dimer has been proposed based on homology to YabA's predicted leucine zipper and Zn-binding domains (personal communication, L. Felicori and F. Molina). However, YabA is known to form tetramers, and this is believed to be its biologically active form. In order to reconcile this discrepancy with the dimer model, it was proposed that the model could form a tetramer from a dimer of dimers. This presented three obvious possibilities for the formation of the tetramer; 'beads on a rail', 'doggy bones' or 'bunch of flowers' (Fig. 3.2).

The results presented above are least consistent with the 'bunch of flowers' model; initial analysis of full length YabA by SEC-MALLS and AUC demonstrated that the YabA tetramer is more consistent with the elongated shape found in the 'beads on a rail' and 'doggy bones' models. Analysis of the individual N- and C-terminal domains of YabA indicates that the N-terminal leucine zipper domain forms a tetramer, consistent with both the 'doggy bones' and 'bunch of flowers' models. The C-terminal Zn-binding domain, by contrast, appears to be monomeric in solution. Although none of the tetramer models directly predict a ‘monomeric’ Zn-binding domain, the predictions are not exhaustive. An independent Zn-binding domain is most consistent with the 'doggy bones' model, where these domains could have more freedom to extend away from one another without disruption of the tetramer (beads on a rail) or steric hindrance (bunch of flowers). Thus, collectively the results are most consistent with a 'doggy bones' model which has both an elongated shape, and tetrameric determinants which reside in the N-terminal domain.

The DnaA and DnaN binding determinants of YabA have been shown to reside in YabA's C-terminal domain[163]. The binding determinants are predicted to be close together[163], and when mapped onto the *in silico* dimer model, these binding sites would overlap. A ternary complex between YabA, DnaA and DnaN has been observed by yeast 3 hybrid screening (*in vivo*)[163], suggesting that the production of this complex is possible, and potentially

physiologically relevant. (In one model of inhibition, YabA has been proposed to tether DnaA to the replisome via DnaN in a ternary complex[215]). Conceptually, the formation of a ternary complex is easiest to imagine in the experimentally-supported doggy bones model, where DnaA and DnaN could bind at either end of the YabA tetramer. In the other tetramer models, 'beads on a rail' and 'bunch of flowers', more steric hindrance would be anticipated in order to allow simultaneous DnaA and DnaN binding (this is especially true for the beads on a rail model where some of the DnaA and DnaN binding determinants would be obscured by the YabA-YabA interface). Although this model supports the possibility of a ternary complex, the complex itself has yet to be proved physiologically relevant. It is important to note therefore that the doggy bones model in no way excludes the possibility of a competitive model of binding for DnaA and DnaN, as has been proposed in recent literature[34]. It is necessary to improve our current understanding of YabA's mode and stoichiometry of binding to DnaA and DnaN in order to further elucidate YabA's mechanism of action.

Chapter 4 : Crystallisation of YabA Constructs

4.1: Abstract

YabA is a key regulator of DNA replication initiation in vegetatively growing *B. subtillis* cells. An inhibitor of DNA replication, it is known to affect both the synchrony and timing of replication initiation via a mechanism which involves both DnaA and DnaN. As no known homologues of YabA are available in the PDB, little structural information is available to aid the elucidation of YabA's mechanism. Here, following failed attempts at full-length YabA crystallisation, a structure of YabA's N-terminal domain was solved, revealing a 4-stranded anti-parallel coiled-coil domain with 2 C-terminal extensions; determining the location of 2 C-terminal domains at either end of the coiled-coil. The work presented here, and collaborative studies which provide additional structural and functional information delineate the overall architecture of YabA and provide insight into the ability of YabA to bind multiple binding partners during the regulation of DNA replication initiation.

4.2: Introduction

YabA regulates DNA replication events in the Gram-positive organism *B. subtilis*. A negative regulator of DNA replication, YabA affects both timing and synchrony in vegetatively growing cells[161], and interacts with DnaA and the DNA polymerase clamp, DnaN[162,163]. YabA is known to affect co-operative binding of DnaA at *oriC* [34], and has been shown to be capable of disrupting DnaA oligomerisation *in vitro* [109] however YabA's mechanism of action remains unclear. Models proposed include a mechanism by which YabA tethers DnaA to the replisome via DnaN (thus preventing re-initiaion by DnaA during replication)[215], and a mechanism where YabA antagonises DnaA's interaction with *oriC* only in the absence of DnaN[34]; in this model, competitive binding of 'free' DnaN to YabA removes YabA from *oriC*/DnaA, allowing DnaA to co-operatively bind *oriC* (and initiate DNA replication). As YabA has no known homologues, no structural evidence is available to support either mechanism. YabA is known to form tetramers[163], however the stoichiometry of its interaction with DnaN and DnaA is unknown and it remains unclear whether or not a ternary complex would be physically possible - especially considering YabA's binding determinants to DnaA and DnaN lie sequentially close together (suggesting overlapping binding sites)[163]. Thus, understanding the tertiary and quaternary structure of YabA may provide insight into its interactions with both DnaA and DnaN, and consequently its mechanism of action. Here we aim to solve the structure of YabA by X-ray crystallography.

Prior work carried out by M. Fogg at YSBL identified expression and purification conditions for YabA, and resulted in the formation of crystals which diffracted to ~2.7Å resolution. Data were collected at Diamond Light Source, however it was not possible to solve this data set due to a lack of phasing possibilities; no molecular replacement model is available due to lack of homology with other proteins and attempts to produce SeMet-derivative YabA in *E. coli* B834 cells was unsuccessful because the derivative protein separated into the insoluble fraction following cell lysis.

4.3: Aims

Elucidating the structure of YabA will provide a framework for exploring and understanding its interactions with DnaA and DnaN; shedding new light onto its mechanism of action. Here, we aim to solve the structure of YabA using X-ray crystallography. The initial aims of this work were to establish robust conditions for *B. subtilis* YabA crystallisation, and establish expression conditions for a soluble form of SeMet YabA, which could be crystallised to facilitate structure solution.

4.4: Methods

4.4.1: Cloning: *Geobacillus stearothermophilus* YabA, DnaA^{III-IV}, and DnaN

The coding sequences for *yabA*, *dnaN* and domains III-IV of *dnaA* (determined by sequence similarity to *B. subtilis dnaA*) were amplified by PCR from *G. stearothermophilus* genomic DNA using gene specific primers appended to sequences which facilitate entry into the pET-YSBLIC3C vector by ligation independent cloning (See Tables 4.1, 4.2 and 4.3)[217]. PCR products were run on a 1% agarose gel and gel purified using a QIAquick Gel Extraction kit (Qiagen). 0.2 pmol of gel purified PCR product was incubated with 0.4 µl T4 DNA polymerase (2.5 units μl^{-1}) in the presence of 2.5 mM dATP (total volume 20 µl) at 22°C for 30 mins, followed by incubation at 75°C for 20 mins. This generates 5' single stranded overhangs as a result of T4 DNA polymerases' combined 3' to 5' exonuclease and its 5' to 3' DNA polymerase activities. Complementary overhangs were generated on linearised pET-YSBLIC3C vector using T4 DNA polymerase and dTTP under the same conditions. 2µl of Insert and 1ul vector were annealed by incubation for 20 mins at room temperature. 50 µl *E. coli* XL-1 Blue cells were transformed with 2 µl annealed product via heatshock at 42°C and plated onto LB agar supplemented with 30 µg ml^{-1} kanamycin and incubated overnight at 37 °C. Colonies were grown to a diameter of ~1 mm. Overnight cultures were set up in LB supplemented with 30 µg ml^{-1} kanamycin and containing 1 transformant colony. Plasmid DNA was isolated from each overnight culture using a QIAprep Spin Miniprep kit (Qiagen). Plasmid sequences were confirmed to be correct by DNA sequencing. The resultant plasmids, pET-YSBLIC3C-*Gst*YabA, pET-YSBLIC3C-*Gst*DnaN and pET-YSBLIC3C-*Gst*DnaA^{III-IV} encode YabA, DnaN or domains III-IV of DnaA, respectively, each fused to an N-terminal 3C cleavable hexahistidine tag.

Table 4.1: PCR components for cloning from *G. stearothermophilus*

Reactant	Volume/ µl (50µl reaction)
G. stearothermophilus genomic DNA (50 ng ul^{-1})	0.5
Forward Primer (20 pmol μl^{-1})	1
Reverse Primer (20 pmol μl^{-1})	1
KOD Hot Start Polymerase Buffer (10x)	5
dNTPs (2mM)	5
KOD Hot Start DNA Polymerase (1,000 U/ ml)	1
$MgSO_4$ (25mM)	2
H_2O	34.5

Table 4.2: PCR cycling parameters for cloning from *G. stearothermophilus*

Step	Temperature /°C	Duration /seconds	Cycles
Initial Denature	94	120	1
Denature	94	30	35
Annealing	52	30	
Extension	72	60	
Final Extention	72	180	1

Table 4.3: Primers used for cloning from *G. stearothermophilus*

Construct	Encodes	Cloning Primers	
pET-YSBLIC3C-*Gst*YabA	YabA	Fwd	5'- **CCAGGGACCAGCA**GTAGATAAAAAAGA AGTGTTTCGATCAGTGGCC-3'
		Rev	5'-**GAGGAGAAGGCGCGTTA**TTTGTTCAAA AACGACAAGCAAAACAAACAATCGC-3'
pET-YSBLIC3C-*Gst*DnaN	DnaN	Fwd	5'- **CCAGGGACCAGCA**GTGAACATTTCCAT TGACCGCGAAGC -3'
		Rev	5'- **GAGGAGAAGGCGCGTTA**ATATGTTCTC ACCGGCAAAATGAGCTGA -3'
pET-YSBLIC3C-*Gst*DnaA^{III-IV}	DnaA residues 114-448	Fwd	5'- **CCAGGGACCAGCA**AATCCGAAATACAC GTTCGATACGTTTGTCATCG -3'
		Rev	5'- **GAGGAGAAGGCGCGTTA**CTTCAGCTTT TCTTGAATCTCTTGTATATGCCTTTGC-3'

4.4.2: Cloning: YabA$^{1-58}$, YabA$^{70-119}$ and YabA$^{\Delta 59-69}$

YabA$^{1-58}$ and YabA$^{70-119}$ were cloned as detailed in section 3.4.4. YabA$^{\Delta 59-69}$ was cloned in a analogous manner, by a deletion mutagenesis method. A whole vector amplification of pET-YSBLIC-YabA was performed by PCR using oligonucleotides which would produce a deletion of residues 59-69 (inclusive) from the YabA coding region of the vector (Table 4.4). PCR conditions were as detailed in Table 3.3. 45µl of PCR product was incubated with 1.5 µl DpnI (15 units) at 37 °C for 4-5 hours, before transformation into Top10 competent cells. 1µl of digested product was added to 50 µl Top10 competent cells before incubation for 30 minutes on ice. Cells were heat-shocked at 42 °C for 50 seconds before re-incubation on ice for 10 minutes. 150 µl LB media was added and cells were incubated at 37°C for 60 minutes prior to plating of 100 µl of the culture onto LB agar supplemented with 30 µg ml^{-1} kanamycin. Plates were incubated for 18-20 hours at 37°C until the production of colonies around 1.0-1.5 mm in diameter. Cultures containing 5 ml LB supplemented with 30 µg ml^{-1}

kanamycin and a single colony were incubated at 37 °C overnight, before isolation of plasmid DNA using a mini-prep kit (Qiagen). Resultant plasmids were analysed by double digest and sequences confirmed to be correct by sequencing (GATC). The recombinant plasmid pET- YSBLIC3C-YabA$^{\Delta 59\text{-}69}$ directs the expression of residues 1-58 fused directly to residues 70-119 of YabA with a 3C cleavable N-terminal His-tag.

Table 4.4: YabA construct primers and coding regions of YabA$^{\Delta 59\text{-}69}$

Construct	Residues of YabA Encoded	Cloning Primers		Template Used
pET-YSBLIC3C-YabA$^{\Delta 59\text{-}69}$	1-58, 70-119	Fwd	5'-GGACGATACGACGCAGCAGAAA ACACAGAAGACAGAGC-3'	YSBLIC3C-YabA
pET-YSBLIC3C-YabA$^{\Delta 59\text{-}69}$	1-58, 70-119	Rev	5'-GCTCTGTCTTCTGTGTTTTCTGCT GCGTCGTATCGTCC-3'	YSBLIC3C-YabA

4.4.3: Double Digest

50 µg DNA was digested with 10 units of NcoI and 10 units of NdeI in NEB buffer 2 at a total volume of 25 µl. The mixture was incubated for 1 hour at 37 °C before a 20µl sample was combined with 5ul DNA loading dye and analysed on a 1% agarose gel.

4.4.4: Expression Testing (Small-Scale Cultures)

The relevant recombinant pET-YSBLIC3C plasmid was introduced into a strain of *E. coli* for over-expression. Small scale expression tests were carried out to assess the protein production and solubility of protein from each construct. For all constructs, expression tests were carried out in *E. coli* BL21 (de3) cells. For pET-YSBLIC3C-*Gst*YabA, additional expression tests were carried out in BL21 (plysS) and C41 (de3) cells. 200 µl of overnight LB culture was used to inoculate 10 ml LB cultures supplemented with 30 µg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (shaking at 180 rpm) before cells from 1 ml of culture were havested by centrifugation, and protein production was induced in the remaining culture with 1 mM IPTG. Following induction, cultures were grown at 37°C for 4 hours, or 16°C for 18-20 hours with shaking at 180 rpm. Cells from 1 ml of culture were harvested by centrifugation for analysis of protein production. Protein production and solubility was assessed by SDS-PAGE.

4.4.5: Protein Production

YabA, YabA$^{1-59}$ and YabA$^{70-119}$

For the large-scale production of *B. subtilis* YabA, YabA$^{1-58}$ or YabA$^{70-119}$ see section 3.4.6.

YabA$^{\Delta 59-69}$

For the production of YabA$^{\Delta 59-69}$, pET-YSBLIC3C- YabA$^{\Delta 59-69}$ was introduced into *E. coli* BL21 (de3) cells. Overnight LB cultures were used to inoculate 500 ml LB cultures supplemented with 30 µg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (shaking at 180 rpm) before protein production was induced with 1 mM IPTG. Following induction, cultures were grown at 37°C for 4 hours, with shaking at 180 rpm. Cells were subsequently harvested by centrifugation (Sorvall RC 5B Plus, 15 min at 15 K (SS34 rotor)). The pelleted cells were re-suspended in 50 mM Tris pH 8.0, 500 mM NaCl, 20 mM Imidazole (purification lysis buffer) prior to re-pelleting by centrifugation (10 min, 5K), for storage at -20°C.

SeMet-YabA

Overnight cultures of *E. coli* BL21 (DE3) harbouring pET-YSBLIC3C-YabA were used to inoculate 500 ml minimal media supplemented with 30 µg/ml kanamycin. Cultures were grown to an OD_{600} of 0.6-0.8 at 37 °C (180 rpm shaking) prior to the addition of an amino acid mixture (50 mg lysine, 50 mg phenylalanine, 50 mg threonine, 25 mg isoleucine, 25 mg leucine, 25 mg valine) to suppress methionine production [199], and 30 mg selenomethionine. Cultures were grown at 37 °C (180 rpm shaking) for a further 15 minutes prior to induction of recombinant protein production with 1 mM IPTG. Cultures were subsequently grown at 30 °C (180 rpm shaking) overnight (16-20 hours) before cells were harvested by centrifugation.

4.4.6: Protein Purification

Purifications of YabA, YabA$^{1-58}$, YabA$^{70-119}$ were carried out as detaited in section 3.4.7. SeMet-YabA and YabA$^{\Delta 59-69}$ can also be purified by an analogous procedure.

4.4.7: Mass Spectrometry

Purified YabA, SeMet-YabA and YabA$^{\Delta 59\text{-}69}$ were buffer exchanged into 2mM Tris for analysis by Electrospray Ionisation-Mass Spectrometry (ESI-MS). Samples were provided at a concentration of 5 mg ml^{-1}. Spectra were obtained and processed by Simon Grist.

4.4.8: SEC-MALLS: YabA$^{\Delta 59\text{-}69}$

To determine the oligomeric state of YabA$^{\Delta 59\text{-}69}$ analysis was carried out by SEC-MALLS. A 100 µl sample of protein at 2.5 mg ml^{-1} was loaded onto a size-exclusion column equilibrated with 50 mM TRIS pH 8.0, 150 mM NaCl. The eluate was successively analysed by a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor. Data was analysed with Astra software (Wyatt).

4.4.9: Crystallisation Screening

A range of commercially available 96-well crystallisation screens were utilised to screen for conditions from which a purified YabA derivative would crystallise. Each screen was set up as a sitting-drop experiment with 300 nl drops containing a 1:1 ratio of protein solution: reservoir solution. Reservoirs contained 54 µl of reservoir solution. Table 4.5 shows which crystal screens were used for each purified protein, and the concentration of the protein added to the crystallisation drop.

Table 4.5: 96-well crystallisation screens set up for YabA and YabA-derivatives

Purified Protein	Concentration	Crystal Screen
YabA	8 mg ml^{-1}	JCSG+, SALT RX, CSS (Tris pH 7.5)
	30 mg ml^{-1}	PACT, INDEX
	45 mg ml^{-1}	PACT, INDEX
	100 mg ml^{-1}	PACT, INDEX
YabA (Stored at 4°C for 2 Weeks)	30 mg ml^{-1}	PACT, INDEX
SeMet YabA	30 mg ml^{-1}	PACT, INDEX, Morpheus, Ammonium Sulfate
$YabA^{1-58}$	12.9 mg ml^{-1}	JCSG+, PACT
$YabA^{70-119}$	9.7 mg ml^{-1}	JCSG+, PACT
	20 mg ml^{-1}	PEG Ion, INDEX,
$YabA^{\Delta 59-69}$	11.5 mg ml^{-1}	JCSG+, PACT

4.4.10: Crystallisation and Structure Solution: $YabA^{1-58}$

$YabA^{1-58}$ crystals were grown in hanging drops containing a 1:1 ratio of concentrated protein solution at 9.75 mg ml^{-1} and a reservoir solution of 0.1 M Bis-Tris Propane pH 6.0, 0.35 M ammonium acetate and 20% PEG 6K. Crystals were transferred to a cryo-protectant solution consisting of 85% reservoir solution supplemented with 15% ethylene glycol prior to being cryo-cooled in liquid nitrogen. X-ray diffraction data were collected to 2.0 Å resolution on beamline i04 at the Diamond Light Source (DLS), Harwell. The diffracting crystal belongs to the space group I222 with unit cell dimensions of a = 49.08 Å, b = 73.77 Å, c = 88.64 α = β = γ = 90.00 °. Data was processed using the automated processing pipeline software XIA2 [201], which utilised the software XDS[202]. Data were subsequently merged using AIMLESS. The structure was solved by molecular replacement using the program PHASER[218], and $YabA^{1-62}$ as a search model (a partial YabA structure obtained by L. Terradot following protein degradation during crystallisation). Two $YabA^{1-62}$ molecules were found in the asymmetric unit. The $YabA^{1-58}$ model was refined by iterative rounds of maximum-likelihood refinement methods in REFMAC[210] and manual model building in COOT[211] to an R-factor of 24.2 (28.2). Data and refinement statistics are shown in Table 4.8.

4.5: Results

4.5.1: YabA Crystallisation

Commercially available 96-well screens were used in sitting-drops to screen for conditions from which YabA would crystallise. Initial screening produced crystals after 3 days from conditions B7, B8, B9 and F5 of the PACT screen[219] and F5 of the INDEX screen[220] from YabA protein at 30 mg ml^{-1}. However, this protein had been stored at 4 °C for 2 weeks prior to screening, and the same screens with freshly prepared protein (at the same concentration), did not produce crystals. Furthermore, in the first instance, crystals were reproducible in 24-well hanging-drop format with protein taken from the same preparation (also stored at 4°C). However if protein from a new preparation was used, crystals did not appear under the same conditions. Analysis of the 'crystallising protein' by SDS-PAGE revealed that it was heavily degraded, suggesting that these crystals were not of full-length protein (Fig. 4.1). Thus, optimisation experiments were terminated. All other (i.e. subsequent) crystallisation screens were set up with freshly prepared protein (<1 day post-purification or from protein which had been flash frozen and stored at -80°C immediately after purification), however, no alternative crystallisation conditions were obtained.

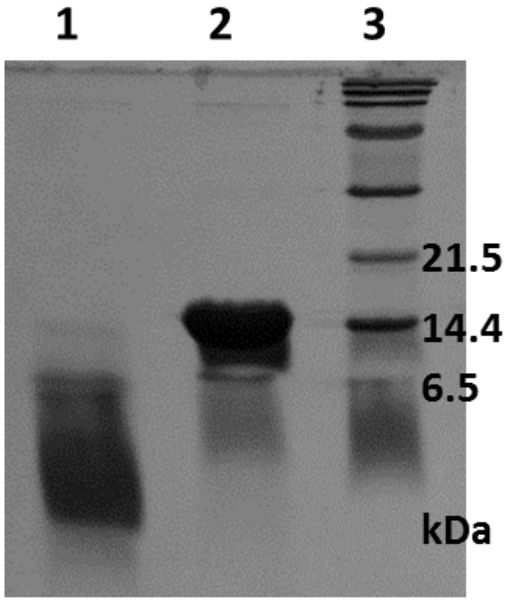


Figure 4.1: SDS-PAGE analysis of YabA after storage: lane 1) 1 µg YabA after storage at 4 °C for 27 days (crystallising 'YabA'); lane 2) 1 µg YabA after storage at 4 °C for 2 days; lane 3) Broad-range molecular weight markers.

4.5.2: SeMet YabA

To the best of our knowledge, there a no solved structures of proteins homologous to YabA available in the PDB; making structure solution by molecular replacement unlikely. Therefore phasing using heavy-atoms is a desirable course for structure solution. Unfortunately, previous efforts to produce SeMet-substituted YabA were unsuccessful as YabA partitioned into the insoluble fraction when produced in the methionine auxotroph *E. coli* strain, B834 (Mark Fogg, personal communication). Therefore, an alternative method for SeMet-substituted protein production was trialled, whereby proteins can be produced in any *E. coli* strain, but under conditions which inhibit normal methionine production. In this method, the amino acids leucine, isoleucine, lysine, phenylalanine and threonine are added to the culture to inhibit the natural methionine biosynthesis pathway[199]; SeMet is also added to cultures and becomes incorporated into the protein in the absence of normal methionine production. Using this method, SeMet-substituted YabA can be produced in *E. coli* BL21 (DE3) which partitions into the soluble fraction after cell lysis. (Fig. 4.2). (Selenium incorporation was later confirmed by mass spectrometry - see below.)

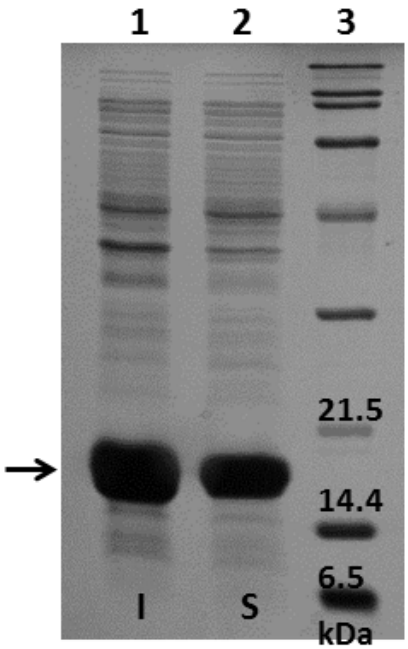


Figure 4.2: SDS-PAGE analysis of SeMet-substituted YabA production: lane 1) lysate from cells induced to produce SeMet-YabA; lane 2) Soluble lysate from cells induced to produce SeMet-YabA; lane 3) broad-range molecular weight markers.

4.5.3: Purification

SeMet-substituted YabA can be purified in an analogous manner to native YabA, (see sections 3.4.2 and 3.5.2). SeMet substituted protein is judged pure by staining after SDS-PAGE (Fig. 4.3).

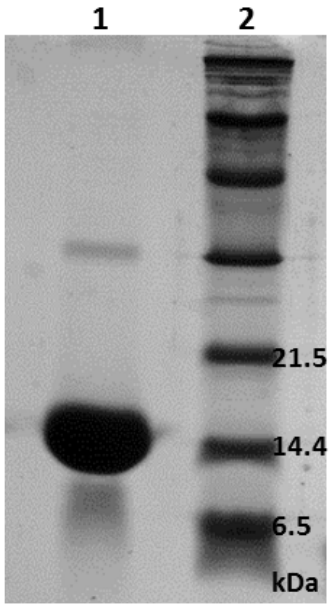


Figure 4.3: SeMet YabA following purification: lane 1) 10 µg of SeMet YabA following purification; lane 2) broad-range molecular weight markers.

4.5.4: Mass Spectrometry

Samples of YabA and SeMet-YabA were analysed by electrospray ionisation mass spectrometry for molecular weight determination. These results were used to infer correct production and purification of the recombinant protein, and in the case of the SeMet-derivative, that Se had been incorporated. Results are shown in Table 4.6.

For YabA, the main peak in the processed spectrum showed a mass of 14,324 Da. This is the same as the theoretical mass of recombinant YabA (14,324 Da) suggesting that the protein has been produced as anticipated and isolated to a high purity. For SeMet-YabA, the main peak in the spectrum shows a mass of 14,419 Da, 1 Da away from the theoretical mass of SeMet-YabA, assuming incorporation of 2 selenium atoms at methionine residues 1 and 34. The observed mass of SeMet-YabA is 95 Da higher than native YabA, indicating that incorporation of 2 Se atoms has occurred; selenium is 47 Da heavier than sulphur, thus an addition of 94 Da to YabA's mass indicates incorporation of 2 selenium atoms.

Table 4.6: Mass spectrometry analysis of YabA and SeMet-YabA

Sample	Peak	Observed Mass (Da)	Theoretical Mass (Da)	Species
YabA	1	14,324	14,324	YabA
SeMet-YabA	1	14,419	14,418	SeMet-YabA

4.5.5: Crystallisation: SeMet-YabA

Four 96-well crystallisation screens were set up with SeMet-YabA at 30 mg ml^{-1}. A crystal which diffracted to ~7 Å spacing in-house was obtained from the PACT screen, condition C8 (0.1 M Hepes pH 7.0, 0.2 M ammonium chloride and 20% PEG 6K), however the crystal was judged as too weak for data collection and crystals were not reproducible in either 24-well or 96-well optimisation screens. Furthermore, the original crystal had taken 13 days to form, so it is unclear whether or not it may have been a product of YabA degradation.

4.5.6: *Geobacillus stearothermophilus* YabA, DnaN and $DnaA^{III-IV}$

Thermophilic bacteria produce proteins with increased stability at higher temperatures[221,222]. For this reason, proteins from thermophilic species of bacteria are frequently utilised for crystallisation as they are more likely to carry an orthologue which has increased rigidity, improving the likelihood of crystallisation (or producing crystals with better packing (and thus obtaining diffraction to a higher resolution) over their non-thermophilic orthologues. YabA was cloned from the thermophilic bacterial species *Geobacillus stearothermophilus* in the hope that it might prove to be a better candidate for crystallisation. YabA's interaction partners DnaN, and domains III-IV of DnaA ($DnaA^{III-IV}$) were concurrently cloned from the same species to facilitate interaction assays or co-crystallisation in the future.

4.5.7: Cloning: *Geobacillus stearothermophilus* YabA, DnaN and $DnaA^{III-IV}$

yabA, $dnaA^{III-IV}$ and *dnaN* were amplified from the *Geobacillus stearothermophilus* genome by PCR (Fig. 4.4, A). Although the PCR products appeared somewhat dirty, attempts to amplify the genes using a higher annealing temperature during PCR cycles did not yield any

products. Thus, PCR products were gel purified using a QIAquick Gel Extraction kit (Qiagen). Purified PCR products can be seen in Figure 4.4, B. Purified products were inserted into the pET-YSBLIC3C vector via ligation independent cloning. Recombinant plasmids were amplified by growth in the *E. coli* strain XL1-blue and purified using a QIAprep Spin Miniprep kit (Qiagen). Correct sequences and vector insertions were confirmed by DNA sequencing (TF, University of York).

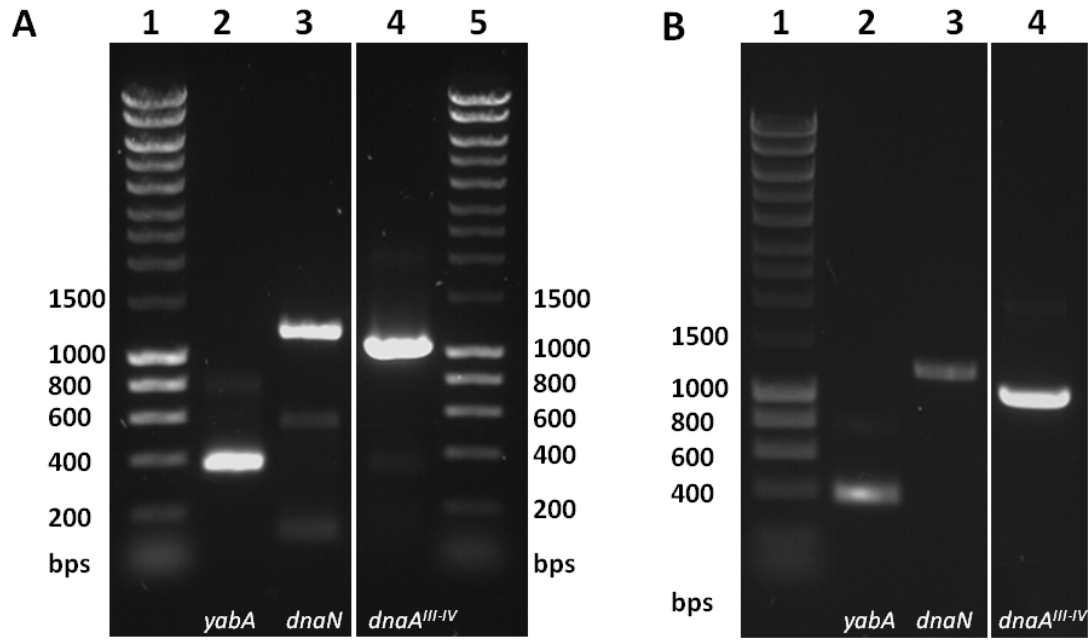


Figure 4.4: 1% Agarose gel showing PCR products for cloning of *G. stearothermophilus yabA, dnaN, dnaAIII-IV*: A) PCR products: Lane 1) 6 µl hyperladder I; lane 2) 5 µl PCR product from amplification of *Gst yabA* ; lane 3) 5 µl PCR product from amplification of *Gst DnaN;* lane 4) 5 µl PCR product from amplification of *Gst dnaA*$^{III-IV}$; lane 6) 6 µl hyperladder I. **B**) Gel extraction products: lane 1) 6 µl hyperladder I; lane 2) 5 µl gel-purified *Gst yabA*; lane 3) 5 µl gel-purified *Gst dnaN*; lane 4) 5 µl gel-purified *Gst dnaA*$^{III-IV}$.

4.5.8: Protein Production: *G. stearothermophilus* YabA, DnaN and DnaA^{III-IV}

Small scale trial cultures for the production of *G. stearothermophilus* YabA, DnaN and DnaA^{III-IV} were carried out as detailed in section 4.4.4 to test for conditions under which soluble proteins could be produced. As an initial test, recombinant plasmids harbouring the YabA, DnaN and DnaA^{III-IV} genes, respectively, were introduced into *E. coli* BL21 (DE3) cells for protein production. Unfortunately, these tests did not yield a positive result. Inducible YabA expression is observed when cells are grown at both 16 °C and 37 °C post-

induction (bands at approx. 14 kDa, Fig. 4.5, lanes 3 and 5), however, this protein is not soluble in the chosen re-suspension buffer of 50 mM Tris pH 8.0, 150 mM NaCl and 20 mM Imidazole. Production of DnaA^{III-IV} and DnaN is less clear (Fig. 4.5, lanes 7-14); bands indicating protein production would be expected at approx. 51 kDa for DnaA^{III-IV} and 42 kDa for DnaN, respectively. It is, however, unclear whether any of the bands around these weights correspond to DnaA^{III-IV} or DnaN, as no clear induction of expression is detectable (Fig. 4.5, lanes 8, 11 and 13).

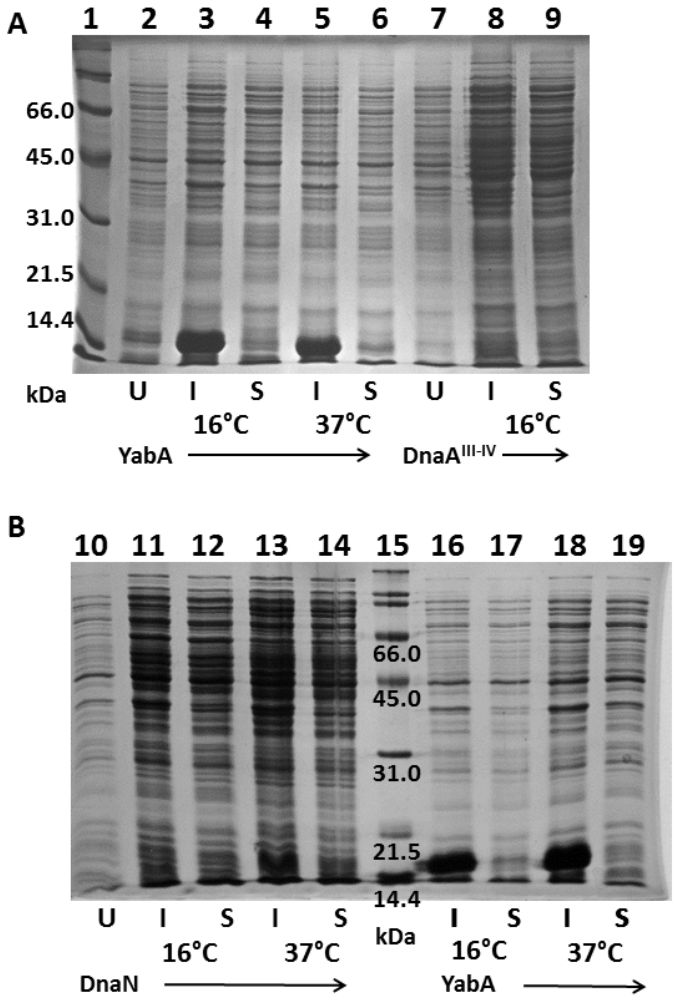


Figure 4.5: Expression of *G. stearothermophilus yabA, dnaN* and *dnaAIII-IV* in *E. coli* BL21 (DE3): A) YabA (low salt buffer) and DnaA^{III-IV} and **B)** DnaN and YabA (high salt buffer). Samples in lanes 1-14 were re-suspended in low-salt buffer, samples in lanes 16-19 in high salt buffer: Lanes 1 and 15) broad-range molecular weight markers; lanes 2-6 and 16-19) Expression of *Gst YabA*. Lanes 7-9) Expression of *Gst dnaA^{III-IV}*. Lanes 10-14) Expression of *Gst dnaN*. In each case, samples are shown in adjacent ascending lanes of 10µl lysate from un-induced cell culture containing relevant vector; 10µl lysate from cells induced to produce protein at 16°C; 10µl soluble lysate from cells induced to produce protein at 16°C; 10µl lysate from cells induced to produce protein at 37°C and 10µl soluble lysate from cells induced to produce protein at 37°C (with the exception of DnaA^{III-IV}, where production could not be tested at 37°C).

As YabA is the major target in this project, efforts were focused on trying to improve its solubility. Re-suspension in a buffer with increased salt concentration was trialled, making the re-suspension buffer the same as that used for *B. subtilis* YabA (50mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole). Unfortunately, YabA was still insoluble under these conditions (Fig. 4.5, lanes 16 to 19). Trials of gene expression in different *E. coli* cell strains BL21 (pLysS) and C41 were also carried out, however the YabA protein produced in these strains was also insoluble (re-suspension buffer: 50 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole) Fig. 4.6. At this stage, work with the *G. stearothermophilus* proteins was terminated. Although there are many other cell strains and solubilisation buffers which could be trialled, as conditions for all three proteins required optimisation, this would be time expensive. Conditions for significant levels of expression and solubility were already known for constructs expressing *B. subtilis* YabA, DnaA and DnaN, thus it was decided that optimisation of *G. stearothermophilus* protein expression should be deferred in favour of working with the *B. subtilis* counterparts.

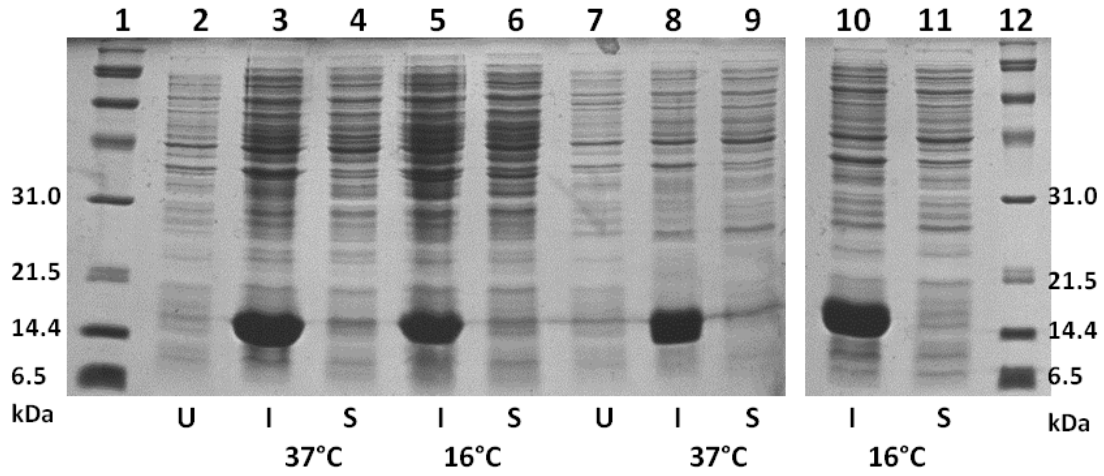


Figure 4.6: Expression of *G. stearothermophilus yabA* in *E.coli* BL21 (pLysS) and C41 : Lanes 2-6 show samples of expression *in E. coli* BL21 (pLysS), lanes 7-11 show samples of expression in *E. coli* C41. Lane 1) broad-range molecular weight markers; lane 2) 10µl lysate from un-induced BL21 (pLysS) cell culture containing pET-YSBLIC3C-*Gst*YabA; lane 3) 10µl lysate from BL21 (pLysS) cells induced to produce *Gst* YabA at 37°C; lane 4) 10µl soluble lysate from BL21 (pLysS) cells induced to produce *Gst* YabA at 37°C; lane 5) 10µl lysate from BL21 (pLysS) cells induced to produce *Gst* YabA at 16°C; lane 6) 10µl soluble lysate from BL21 (pLysS) cells induced to produce *Gst* YabA at 16°C; lane 7) 10µl lysate from un-induced C41 cell culture containing pET-YSBLIC3C-*Gst*YabA; lane 8) 10µl lysate from C41 cells induced to produce *Gst* YabA at 37°C; lane 9) 10µl soluble lysate from C41 cells induced to produce *Gst* YabA at 37°C; lane 10) 10µl lysate from C41 cells induced to produce *Gst* YabA at 16°C; lane 11) 10µl soluble lysate from C41 cells induced to produce *Gst* YabA at 16°C; lane 12) broad-range molecular weight markers.

4.5.9: Cloning of YabA$^{\Delta 59\text{-}69}$

As outlined in section 3.1, YabA contains a poorly conserved central region around amino acids 59-69. This region is predicted to be disordered and is likely to be vulnerable to proteolysis, thus producing a construct with a deletion in this central region may create a more ordered product which is easier to crystallise. A YabA construct with a deletion of residues 59-69 (inclusive) was produced, using the same deletion mutagenesis cloning method described in section 3.5.5), to create a vector encoding residues 1-58 directly fused to residues 70-119 of YabA with an N-terminal cleavable Histidine tag. The deletion product was produced by PCR (Fig. 4.7, A, lane 2) and introduced into *E. coli* Top10 cells for vector amplification. Double digests were performed on purified plasmid DNA to assess for the deletion (33 bases) (Fig. 4.7, B). A small reduction in size over the full-length *yabA* gene (lane 1) can be observed (lanes 2-6), indicating the production of a product with the desired deletion. Production of a product encoding residues 1-58 fused to residues 70-119 of YabA was subsequently confirmed by DNA sequencing.

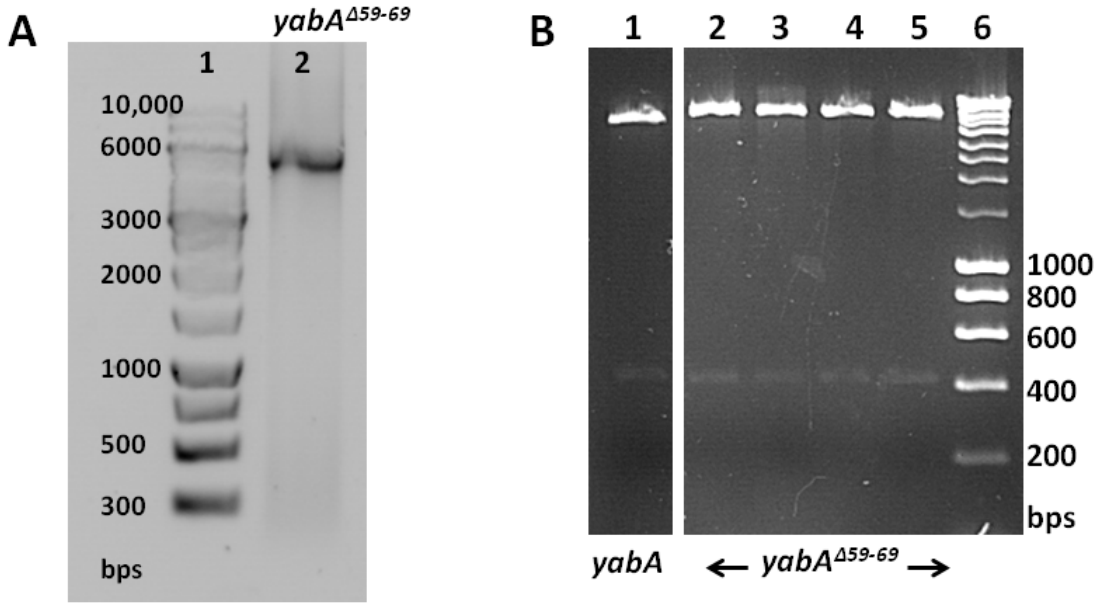


Figure 4.7: Cloning of *yabA*$^{\Delta 59\text{-}69}$: **A**) PCR product for *yabA*$^{\Delta 59\text{-}69}$*:* lane 1) 1kb DNA ladder; lane 2) 5 µl PCR product for *yabA*$^{\Delta 59\text{-}69}$*.* **B**) Test digests for recombinant plasmids**:** lane 1) double digest products of 0.5 µg pET-YSBLIC3C-YabA; lane 2-5) double digest products of 0.5 µg pET-YSBLIC3C-yabA$^{\Delta 59\text{-}69}$; lane 6) 12 µl Hyperladder I.

4.5.10: Protein Production: YabA$^{\Delta 59\text{-}69}$

Small scale expression tests were carried out as described in section 4.4.4 to determine suitable expression and solubilisation conditions for YabA$^{\Delta 59\text{-}69}$. The recombinant plasmid pET-YSBLIC-YabA$^{\Delta 59\text{-}69}$ was introduced into *E. coli* BL21 (DE3) cells for test expression in small scale cultures. Post-induction growth was carried out at 16 °C for ~20 hours and 37 °C for 4 hours. Growth at both temperatures showed a significant level of inducible gene expression (Fig. 4.8, lanes 3 and 5) and that a significant proportion of this protein was soluble in the chosen re-suspension buffer of 50 mM Tris pH8.0, 500 mM NaCl, 20 mM imidazole. Protein production at 37 °C gave a higher yield of protein than at 16 °C, therefore subsequent large-scale protein production was carried out at 37 °C.

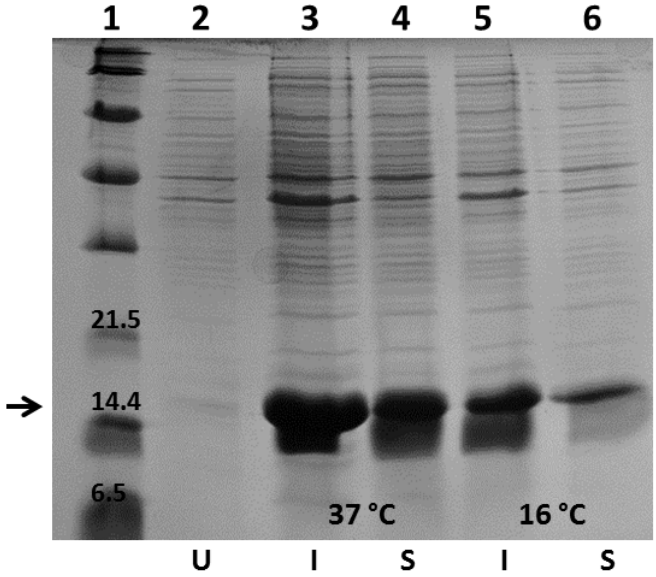


Figure 4.8: Expression of yabA$^{\Delta 59\text{-}69}$: lane 1) broad-range markers; lane 2) 10µl lysate from un-induced cell culture containing pET-YSBLIC3C-YabA$^{\Delta 59\text{-}69;}$ lane 3) 10µl lysate from cell culture induced to express *YabA*$^{\Delta 59\text{-}69}$ at 37 °C; lane 4) 10µl soluble lysate from cell culture induced to express *YabA*$^{\Delta 59\text{-}69}$ at 37 °C; ; lane 5) 10µl lysate from cell culture induced to express *YabA*$^{\Delta 59\text{-}69}$ at 16 °C; lane 6) 10µl soluble lysate from cell culture induced to express *YabA*$^{\Delta 59\text{-}69}$ at 16 °C.

4.5.11: Purification: YabA$^{\Delta 59\text{-}69}$

YabA$^{\Delta 59\text{-}69}$ was purified in an analogous manner to YabA (section 3.4.2). Purified protein is shown in Figure 4.9.

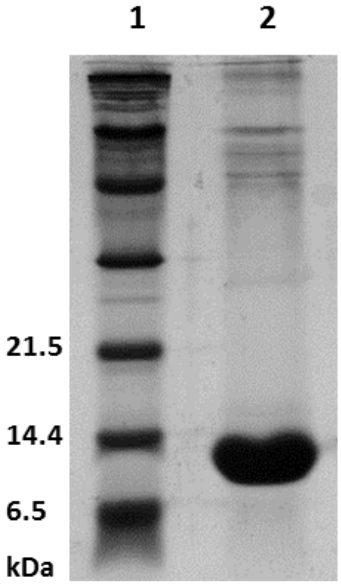


Figure 4.9: YabA$^{\Delta 59\text{-}69}$ following purification: lane 1) broad-range molecular weight markers; Lane 2) 10 µg of YabA$^{\Delta 59\text{-}69}$ following purification.

4.5.12: SEC-MALLS: YabA $^{\Delta 59\text{-}69}$

The oligomeric state of YabA$^{\Delta 59\text{-}69}$ was analysed by SEC-MALLS (Fig. 4.10). Native YabA forms a tetramer *in vitro,* and this is thought to be its functional state. Thus, to gain any functional insight from a structure (or otherwise) of YabA$^{\Delta 59\text{-}69}$, it must be able to retain YabA's tetrameric state.

YabA $^{\Delta 59\text{-}69}$ elutes at approx. 18.5 minutes from a Superdex S75 column, with an associated molecular weight of 50.6 kDa. This is approximately four times the calculated mass of YabA$^{\Delta 59\text{-}69}$ (13.0 kDa x 4 = 52.0 kDa) indicating that YabA$^{\Delta 59\text{-}69}$ is a tetramer in solution. We conclude therefore, that deletion of residues 59-69 of YabA, does not affect its oligomeric state (consistent with the oligomerisation determinants of YabA residing in the N-terminal domain).

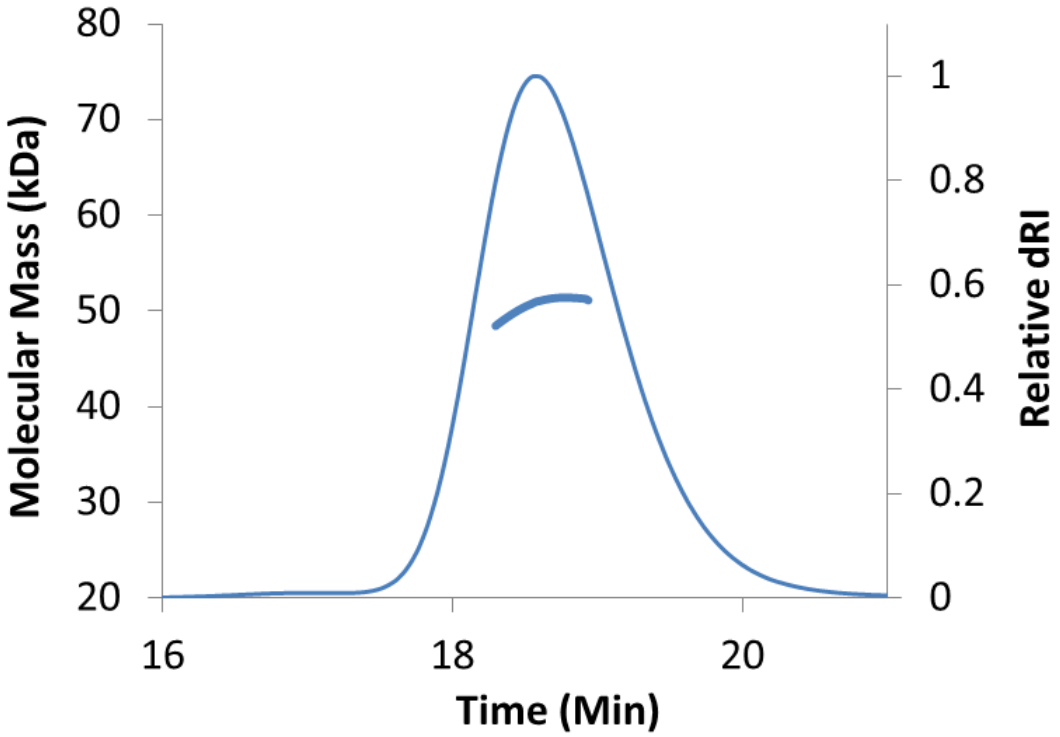


Figure 4.10: SEC-MALLS analysis of YabA$^{\Delta 59\text{-}69}$: the thinner line displays the differential refractive index of the YabA$^{\Delta 59\text{-}69}$ containing eluate from a Superdex 10/300 S75 column as a function of time. The thicker line shows the weight average molecular mass of the eluting species, calculated from refractive index and light scattering measurements. Analysis was carried out at a concentration of 2.5 mg ml^{-1}. The major species elutes at a mass of 50 kDa indicating that YabA$^{\Delta 59\text{-}69}$ is a tetramer in solution.

4.5.13: Crystallisation: YabA$^{1\text{-}58}$, YabA$^{70\text{-}119}$ and YabA$^{\Delta 59\text{-}69}$

96-well commercial crystallisation screens were set up for each of the proteins YabA$^{1\text{-}58}$, YabA$^{70\text{-}119}$ and YabA$^{\Delta 59\text{-}69}$. Although crystallisation conditions were not identified for YabA$^{70\text{-}119}$ and YabA$^{\Delta 59\text{-}69}$, crystals of YabA$^{1\text{-}58}$ formed from multiple conditions in PACT and JCSG+ screen(s). Optimisation was carried out in 24-well format, allowing crystals suitable for data collection to be obtained.

4.3.14: Diffraction Testing and Data Collection of YabA$^{1\text{-}58}$

In house testing of YabA$^{1\text{-}58}$ crystals showed diffraction to ~3.0 Å. Test images from the two 'best' crystals were indexed in MOSFLM, revealing two crystal forms: I222 and P622. Unit cell dimensions are shown in Table 4.8. Strikingly, the same crystal forms and unit cell dimensions had been observed for putative full-length YabA data collected by Mark Fogg *c.*

2011 (Table 4.7). In that case, crystals were grown in 24-well plates, in hanging drops containing 1µl YabA at 26 mg ml^{-1} and 1µl reservoir solution. Crystals with space group I222 grew from conditions of 0.1 M Bis-Tris propane pH 6.0, 23 % (w/w) PEG 5000 MME and 0.1 M NaSCN (somewhat similar to $YabA^{1-58}$ crystal growth conditions of 0.1 M Bis-Tris Propane pH 6.0, 0.35 M ammonium acetate and 20% PEG 6K) and crystals with space group P622 grew from 0.1 M sodium formate, 20% PEG 3350. Crystals formed after a few weeks. (M. Fogg, personal communication). The fact that the $YabA^{1-58}$ crystals and putative full-length YabA crystals grown in 2011 share the same space group and unit cell dimensions demonstrates that data collected previously must have been obtained from crystals of a degraded form of the protein. This is in agreement with the extended time required to grow those crystals. A full dataset of $YabA^{1-58}$ was collected on beamline i04 at Diamond Light Source, Harwell.

Table 4.7: Unit cell dimensions for $YabA^{1-58}$ and putative full-length YabA crystals

Crystal	Space Group	α β γ (°)	a (Å)	b (Å)	c (Å)
$YabA^{1-58}$	I222	90, 90, 90	49	74	89
YabA	I222	90, 90, 90	49	74	88
$YabA^{1-58}$	P622	90, 90, 120	83	83	107
YabA	P622	90, 90, 120	83	83	108

4.3.15: N-terminal Domain Structure

Concurrently and independently of the crystallisation of $YabA^{1-58}$, our collaborators at the University of Lyon, France, solved a fragment YabA structure using a Pt-derivative for phasing. Crystals had formed following protein degradation in drops containing YabA and DnaN. This model was used to solve the structure of $YabA^{1-58}$ by molecular replacement methods (Table 4.8). The $YabA^{1-58}$ crystal used for structure solution crystallised in space group I222, with unit cell parameters of a = 48.95 Å, b = 73.50 Å, c = 88.78 α = β = γ = 90.00 °. Two $YabA^{1-58}$ molecules are seen in the asymmetric unit. The refined model encompasses residues 1 to 54 of YabA in chain A and residues 3 to 52 of YabA in chain B.

The structure delineates the N-terminal domain of YabA, which comprises of a single extended α-helix (~13 turns), with a slight curvature. Two of these molecules make up the asymmetric unit, a parallel coiled-coil motif which crosses near the C-terminus in a configuration resembling a pair of tweezers (Fig. 4.11, A). The molecules in the asymmetric unit form a tetramer with two symmetry related molecules, A' and B' creating an overall structure which can be described as having a 4-stranded anti-parallel coiled-coil architecture (Fig. 4.11, B). This helical bundle is expected to be the functional assembly (tetramer confirmed *in vitro* by SEC-MALLS). The arrangement of the helices in the tetramer is such that the bundle comprises two pairs of parallel helices, with a closely associated region across ~7 turns at the N-terminus (Fig. 4.11, B and C).

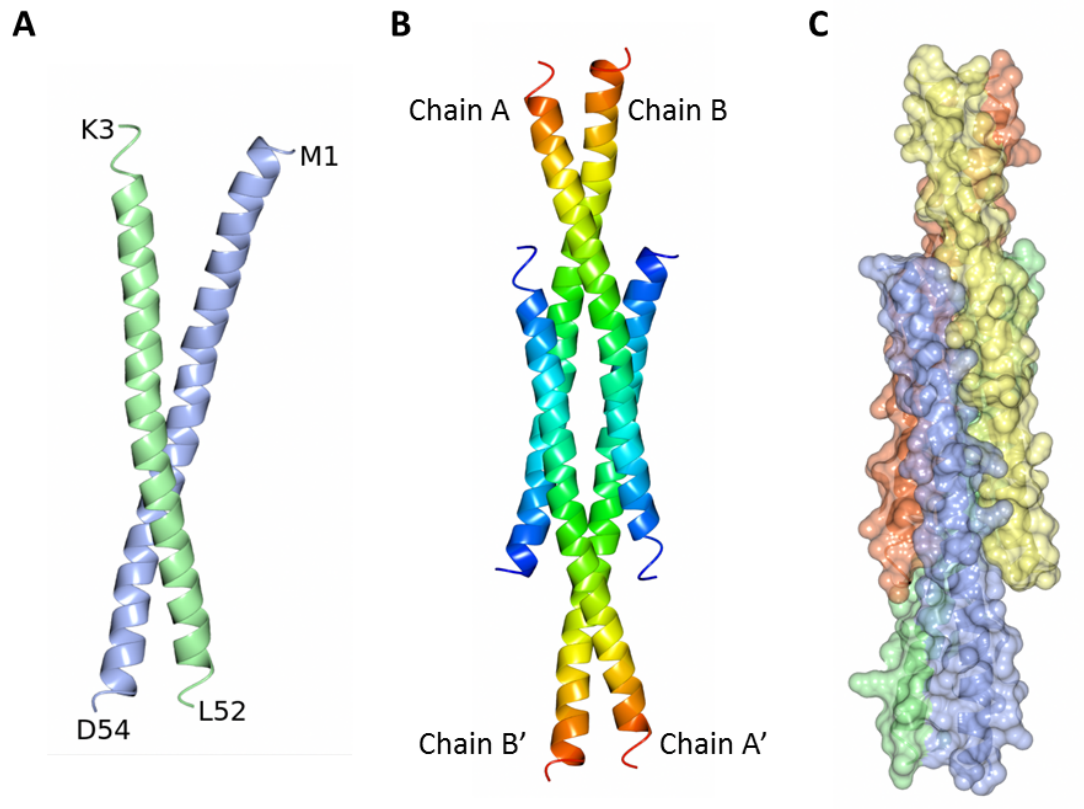


Figure 4.11: YabA$^{1-58}$ structure: A) Ribbon diagram of the asymmetric unit of YabA$^{1-58}$; chain A (blue) spans residues Met 1 to Asp 54 of YabA and chain B (green) spans residues Lys 3 to Leu 52 of YabA – the two chains form a dimer which resembles a pair of tweezers. **B**) YabA Tetramer; ribbon diagram of tetrameric YabA$^{1-58}$ colour-ramped from blue (N-terminus) to red (C-terminus). **C**) The YabA$^{1-58}$ tetramer buries a large percentage of the available surface area: YabA$^{1-58}$ tetramer rendered with a partially transparent surface: chain A is in blue, chain B in green, chain A' in yellow, and B' in coral.

Table 4.8: Data and refinement statistics for YabA[1-58]

	YabA[1-58]
Data Collection	
X-ray source	DLS, i04
Wavelength (Å)	0.9795
Resolution Range (Å)	56.70 - 2.0
Space Group	I222
Unit Cell Parameters	
a b c (Å)	49.08, 73.77, 88.64
α β γ (°)	90, 90, 90
No. of reflections [o]	72719/5336
Completeness (%) [a]	100/100
Redundancy [a]	6.5/6.6
I/σ(I) [a]	19.1/2.5
R_{merge} [p] (%) [a]	4.5/62.7
Refinement and Model Statistics	
Resolution Range (Å)	56.7-2.0
R-factor[q] (R_{free} [r])	24.2 (28.2)
Reflections (working/R_{free})	10629/593
Outer-shell/High Res range	2.00-2.05
Outer-shell[s]/High Res R-factor[c] (R_{free})[d]	30.1/32.9
Outer-shell/High Res Reflections (working/free)	774/37
Molecules per asymmetric unit	2
rmsd from ideal geometry[t]	
Bond Lengths (Å)	0.033
Bond Angles (°)	2.7
Average B-factor (Å2)	38.3
Ramachandran Plot[ug]	97.0/2.0/1.0

o . The first number refers to the overall data set, the second refers the outer resolution shell (2.05-2.00).

p . Rmerge=$\sum_{hkl}\sum_i|I_i - \langle I \rangle|/\sum_{hkl}\sum_i \langle I \rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes *hkl* and $\langle I \rangle$ is the statistically weighted average reflection intensity.

q . R-factor =$\sum||F_o| - |F_c||/\sum|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively.

r . R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.

s . Outer shell for refinement corresponds to 2.05-2.00 Å.

t . Root-mean-square deviation of bond lengths and bond angles from ideal geometry.

u . Percentage of residues in most-favoured/disallowed/allowed regions of the Ramachandran plot.

The helical pairs extend away from one another in the tetramer, with 2 C-terminal extensions located at either end. The 4-stranded antiparallel coiled-coil region is formed by the N-termini of each chain (residues 1-31) and the extended regions span from residue 32-54. The YabA helical bundle has a buried interface of 8672 A^2 of solvent accessible surface area - a large proportion of the overall available area (21606 A^2) (Fig. 4.11, C).

Coiled-coiled proteins are composed of α-helices which contain a seven amino acid 'heptad' repeat sequence $(a\text{-}g)_n$, each of which spans 2 helical turns. The first (a) and fourth (d) residues of each repeat are predominantly hydrophobic and help to define the architecture of the fold. Three types of anti-parallel tetrameric coiled-coil architecture have been previously reported, defined by the interacting residues between chains within the heptadic repeat; these are either a-d core, a-d-e core or a-d-g core (Fig. 4.12) [223,224].

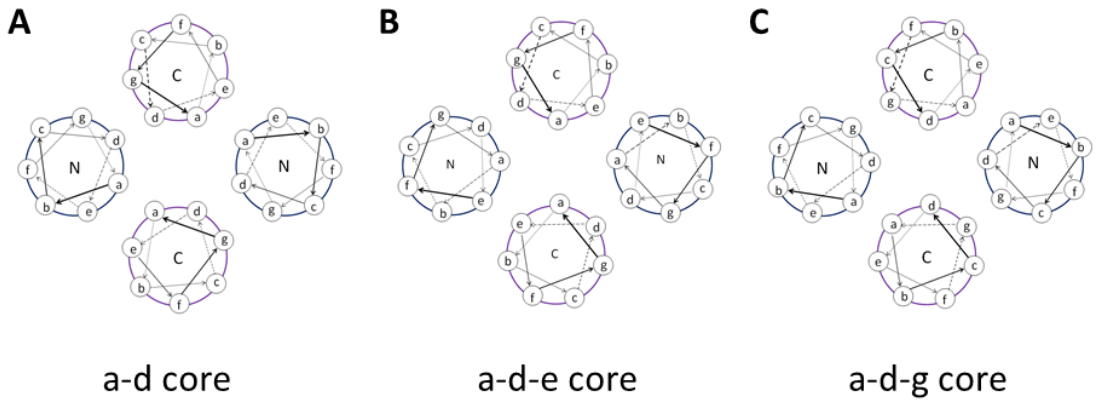


Figure 4.12: Possible core structures of 4-stranded antiparallel coiled-coil proteins : **A**) a-d core structure; side chains from residues a and d contribute to the core packing of the 4-stranded antiparallel coiled-coil structure. **B**) a-d-e core structure; side chains from residues a, d and e contribute to the core packing of the 4-stranded antiparallel coiled-coil structure. **C**) a-d-g core structure; side chains from residues a, d and g contribute to the core packing of the 4-stranded antiparallel coiled-coil structure.

YabA displays an a-d core structure, with a-a and d-d inter-chain interactions (Fig. 4.12, A). Leucine residues 6, 13, 20 and 27 in the d position form a classical leucine zipper, where the side chains between chains A and A' or B and B' display a typical "knobs into holes" style packing (Fig. 4.12, B and D). Val 10, Ile 17 and 31 and Leu 24 residues in the a positions display a similar side chain "knobs into holes" packing between chain A and chain B (Fig. 4.13, C and D).

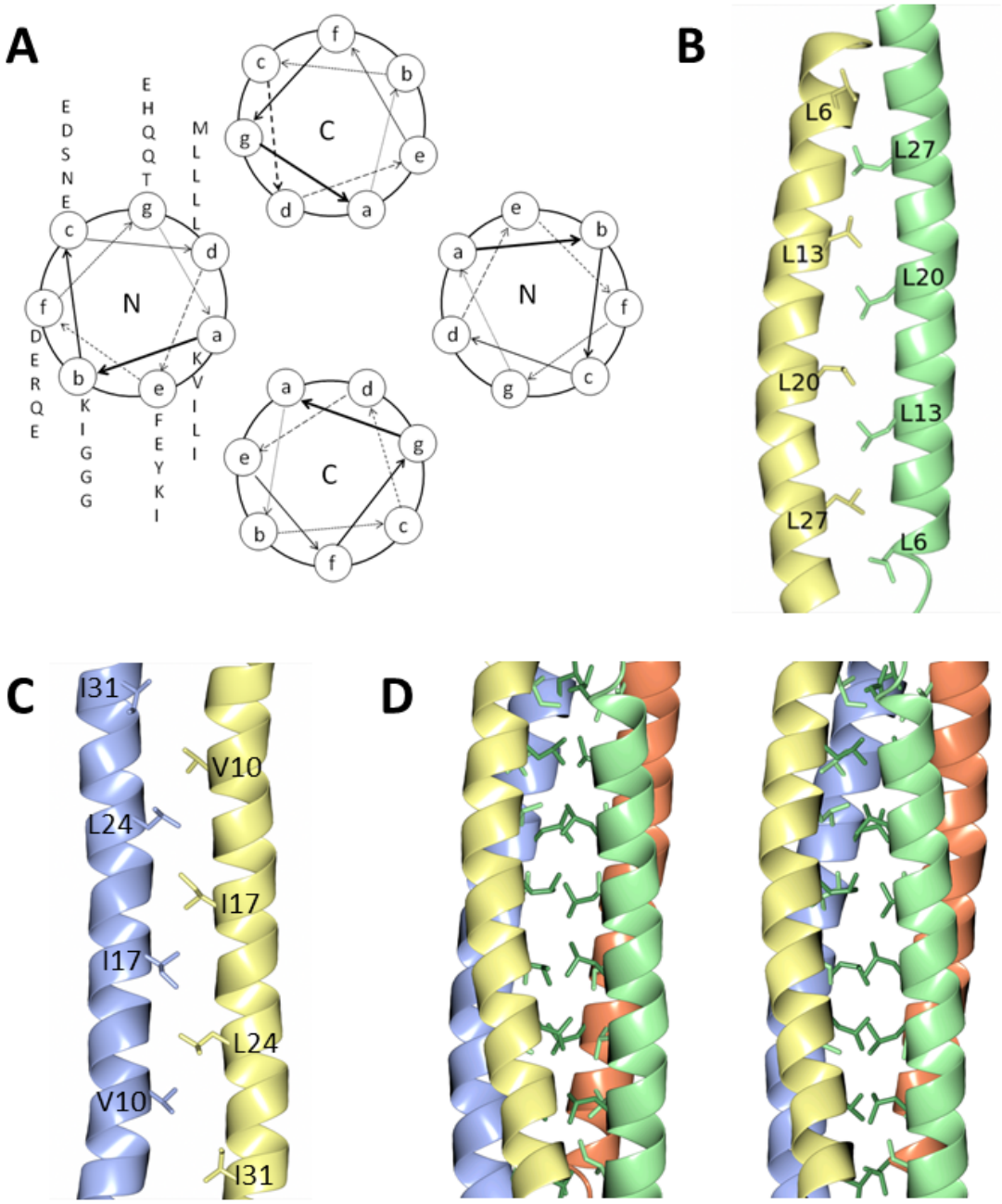


Figure 4.13: Hydrophobic packing in YabA$^{1-58}$ by side chains of residues at a and d positions: **A**) Helical wheel representation of the 4-stranded anti-parallel packing in YabA$^{1-58}$; **B**) Leucine residues 6, 13, 20 and 27 in the d positions of chain B (green) and A' (yellow) form a leucine zipper motif. **C**) Val 10, Ile 17 and 31 and Leu 24 residues in a positions display a "knobs into holes" packing between chain A and chain B. **D**) Stereo image of the YabA tetramer showing 'd' Leu residues (dark green), and 'a' Ile, Val and Leu residues (light green); chain A is in blue, chain B in green, chain A' in yellow, and B' in coral.

In this a-d core type structure, residues at the g position lie above those in the d position, partially buried and partially solvent exposed. Similarly, residues at e positions lie above those in position a (Fig. 4.14, A). In YabA$^{1-58}$ several stabilising hydrogen bonds can be observed between residues at g positions (Gln 16 and Gln 23 from bonds between chains A and B' and B and A') and e positions (Glu 14 forms a bond with Lys 28 between chains A and A' and B and B') (Fig. 4.14, A and B). Residues in positions b, c and f are predominantly solvent exposed.

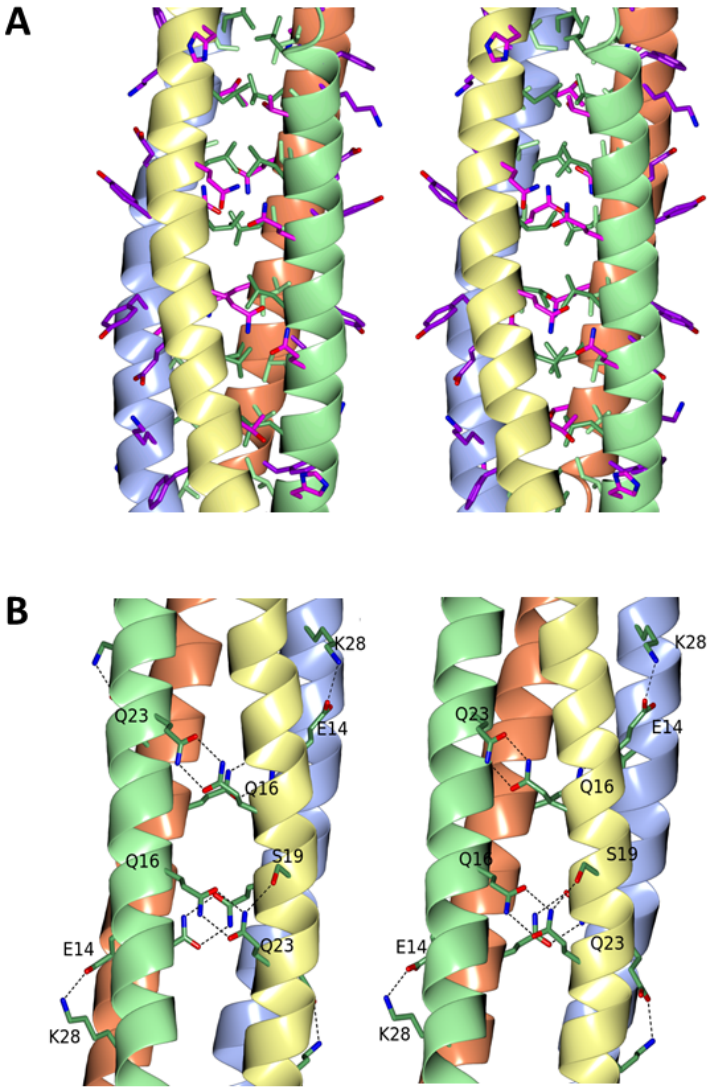


Figure 4.14: Stabilising inter-chain interactions form between e and g residues on YabA$^{1-58}$: Chain A is shown in blue, chain B in green, chain A' in yellow, and B' in coral. **A)** Stereo image showing side chain packing of core a residues (carbon atoms in light green), d residues (carbon atoms in dark green), e residues (carbon atoms in purple, oxygen atoms in red, nitrogen atoms in blue) and g residues (carbon atoms in pink, oxygen atoms in red and nitrogen atoms in blue). **B)** Stereo image of stabilising hydrogen bond interactions between YabA^{-58} chains: hydrogen bonds occur between residues Gln 16 and Gln 23 (chains A' to B and B' to A); and between residues Glu 14 and Lys 28, (A to A' and B to B').

4.5.16: Discussion

The structure of YabA$^{1-58}$ reveals a four-stranded anti-parallel coiled-coil domain with an 'a-d' core packing, in agreement with the YabA$^{1-62}$ structure solved by collaborators L. Terradot, M. Cherrier and A. Bazin. These structures confirm that the tetramerisation determinants of YabA reside in the N-terminal domain, in agreement with the SEC-MALLS data presented in Chapter 3 (see section 3.5.10) and that the tetramerisation determinants specifically lie in residues 1-31 of YabA.

YabA has been shown to interact with DnaA and DnaN via its C-terminal domain, thus structural information about the C-terminal domain and the full-length structure in addition to the N-terminal domain structure are of value for understanding how YabA regulates the initiation of replication. To address this question, our collaborators P. Roblin, M-F. Noirot-Grios, and L. Terradot used small angle X-ray scattering (SAXS) to obtain structural information about full-length YabA. SAXS can be used to determine low-resolution structures of proteins in solution. The analysis of YabA by SAXS revealed that it has a radius of gyration of 63.5 Å and maximum dimension (D_{max}) of 250 Å. These values are surprisingly high for a low molecular weight protein, even accounting for the fact that YabA forms tetramers. Analysis of the data in the form of a Kratky plot, which assesses the likelihood of unfolded regions within the protein, indicated flexibility between the domains of YabA. *Ab initio* modelling with an imposed P_{22} symmetry (as seen in the YabA$^{1-62}$ structure) generated a cylindrical model with pairs of elongated arms at either end (Fig. 4.15, A). Altogether, these results suggest that full-length YabA adopts an elongated form with flexibility between domains.

To generate a more accurate model of full-length YabA, a model was created using the N-terminal YabA$^{1-62}$ crystal structure and a truncated version of the *in silico* YabA model (presented in section 3.1) containing only the C-terminal domain of YabA from residue 76 to 119 (Fig. 4.15, B). The intermediate residues 62-76 were modelled as a flexible linker between the N- and C-terminal domains to account for the large maximum dimension described above. The resultant model (Fig. 4.15, C) fits the experimental scattering pattern and shows a similar domain organisation to the *ab initio* model. Several C-terminal domain positions within the model were found to fit the data with equal agreement (Fig. 4.15, D); this verifies an intrinsic flexibility in the linker between the N- and C-terminal domains of YabA. These results define a full-length structure of YabA, where the N-terminal domains form a tetrameric helical 'core' structure which is flanked on either end by two pseudo-

monomeric C-terminal domains, one attached to each N-terminal domain via a flexible linker region.

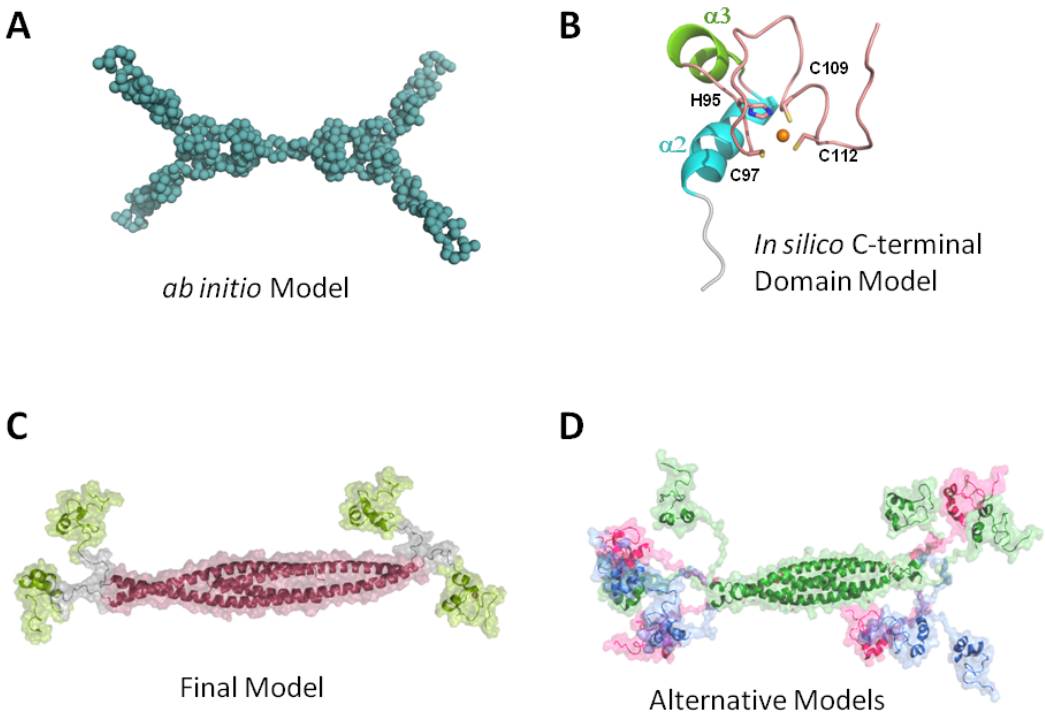


Figure 4.15: Low resolution full-length YabA structure defined by SAXS : A) *Ab initio* model of YabA generated from SAXS data with an imposed P22 symmetry. **B**) *In silico* model of the YabA C-terminal domain. **C**) Full-length YabA model in agreement with SAXS scattering. **D**) Alternative YabA models which agree equally with SAXS scattering data, confirming flexibility in the linker region between the N- and C-terminal domains of YabA.

The C-terminal domain of YabA has previously been shown to carry the determinants on YabA for interacting with DnaA and DnaN[163] and full length YabA can form a ternary complex with DnaA and DnaN as judged by their interaction in a yeast three-hybrid assay[162]. As the C-terminal domain has been characterised to be spatially independent of the N-terminal domain, this posed the question of whether or not it was also functionally independent. Experiments were carried out to assess whether or not an independent C-terminal domain could form interactions with DnaA and DnaN, or if the N-terminal domain was also required for binding. M-F Noirot Gros, T. Garcia and M. Ventroux carried out yeast two and three-hybrid experiments which assess the interaction of the C-terminal domain

(residues 62-119) with DnaA and DnaN. These reveal that the C-terminal domain alone can interact with DnaA or DnaN, but that it is unable to form a simultaneous interaction with both DnaA and DnaN *in vivo*. In contrast, full-length YabA can from a simultaneous interaction with DnaA and DnaN. These results therefore determine a functional role for the N-terminal domain of YabA which appears to provide a structural 'hub' from which the C-terminal domains can extend to form interactions with both DnaA and DnaN on separate domains.

M-F. Noirot Gros and colleagues have also mapped the DnaA and DnaN binding residues of YabA onto the C-terminal *in silico* model and identified additional residues involved in interactions with YabA using yeast two-hybrid screening. These results delineate two partially overlapping surfaces on the YabA C-terminal domain for binding DnaA and DnaN. Furthermore, introduction of these mutations into YabA affects its localisation *in vivo*. YabA has previously been shown to form foci at the mid-cell for most of the cell-cycle, indicating localisation with the replication machinery. Mutations of YabA which cause a loss of interaction with either DnaA or DnaN have been shown to affect localisation of YabA and also loss of initiation control. Remarkably, complementation of a YabA mutant which cannot interact with DnaN with a YabA mutant which cannot interact with DnaA restores the formation of YabA foci at the mid-cell; this provides evidence for the physiological role of a DnaA-YabA-DnaN complex, and highlights the requirement of the tetrameriation determinants of YabA to allow interaction with multiple partners *in vivo*.

YabA Function

YabA has been suggested to function by inhibiting co-ooperative binding of DnaA at the origin[34], and also by tethering DnaA to the replication machinery via a ternary complex with DnaN[164]. How YabA is able to coordinate these activities and regulate replication initiation is currently poorly understood and knowledge of the structural organisation of YabA and its interactions with DnaA and DnaN are vital to our understanding of its mechanism. The work described here and in Chapter 3, along with SAXS and *in silico* experiments carried out by our collaborators, defines the structural architecture of YabA. *In vivo* experiments carried out by M-F. Noirot-Gros and colleagues strongly suggest a mode of interaction where YabA binds DnaA and/or DnaN via separate 'independent' C-terminal domains at either end of a tetrameric core. The existence of several interaction

modules within YabA provides the protein with the opportunity to interact with several DnaA (or DnaN) molecules simultaneously, enabling its function. For example, YabA could inhibit co-operative binding of DnaA at *oriC* by binding several DnaA molecules (Fig. 4.16, A), or interact with DnaA and DnaN at opposite ends of the tetramer to form a ternary complex at the replication machinery (Fig. 4.16, B). Further study is required to understand the nature and stoichiometry of the interactions YabA forms with DnaA and DnaN. The structure of YabA presented here provides a framework for understanding how YabA interacts with multiple protein partners during the regulation of DNA replication initiation.

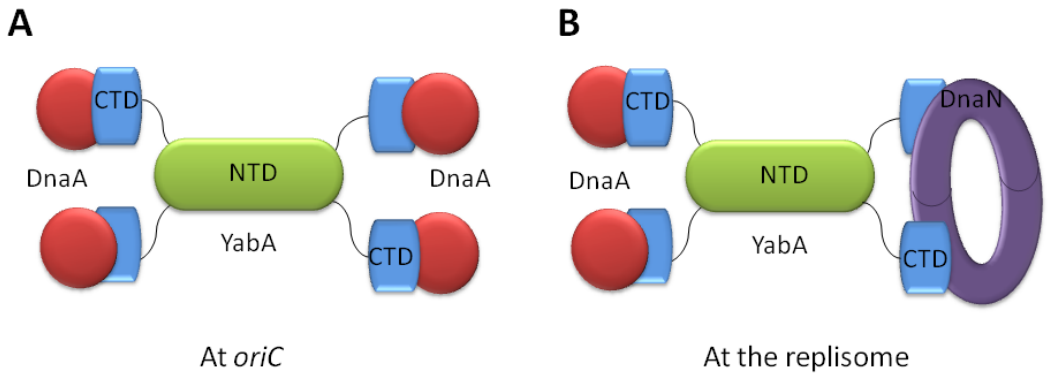


Figure 4.16: Theoretical modes of interaction of YabA with DnaA and DnaN:A) YabA binding multiple DnaA molecules at *oriC*. **B)** YabA forming a ternary complex with DnaA and DnaN at the replication machinery.

Structural Similarity to Geminin

YabA shares structural similarity to the protein Geminin[225], an inhibitor of origin licencing in eukaryotes. Origin licencing is the term given to the strict spatial and temporal control of DNA replication in eukaryotes and involves the assembly of a pre-replicative complex on chromatin before DNA replication can commence. A protein called Cdt1 is a crucial regulator of pre-replicative complex assembly as it recruits the replicative helicase, mini-chromosome maintenance (MCM) complex, the final step in pre-replicative complex assembly[226]. The activity of Cdt1 is in turn inhibited by Geminin during parts of the cell cycle as a method of control over origin licencing. However, there appear to be some significant differences in the oligomerisation states and mechanisms of Geminin and

YabA[226]. Geminin is a three domain protein whose central domain forms a dimeric parallel coiled coil (this is somewhat reminiscent of the asymmetric unit observed for YabA). However, unlike YabA, Geminin's other domains are believed to be unstructured and the functional (inhibitory) amino acids lie at the C-terminal end of its coiled-coil domain[227]. It forms a 'permissive' heterotrimeric complex with one molecule of Cdt1, and two of these heterotrimers are thought to form an inhibitory heterohexameric Geminin-Cdt1 4:2 complex[227]. The interaction between two Geminin dimers in this complex is mediated by two Cdt1 molecules and does not at all resemble the four-stranded anti-parallel helix observed in YabA$^{1-58}$. Furthermore, the functional determinants of YabA – i.e. residues which interact with DnaN and DnaA, have been mapped to its C-terminal domain (away from the coiled-coil), suggesting that YabA's function must be differently mediated to that of Geminin[227]. Thus, it seems unlikely that the structural similarity between Geminin and YabA has any bearing on their functional activities. It is more likely to be a result of the diversity in structure and function observed in coiled-coil domains. Nevertheless, it is interesting to observe the presence of a similar structural motif acting as a DNA replication inhibitor in both eukaryotes and prokaryotes, and it will be of interest to understand if there are any similarities in their mechanisms.

Antiparallel Coiled-Coil Domains

Coincidentally, the structure of YabA$^{1-58}$ may be of interest to protein scientists who study coiled-coil domains with the goal of understanding the relationship between primary sequence and tertiary structure, ultimately aiding rational protein design. Coiled-coil domains have been extensively studied for this purpose due to their ability to form a wide range of oligomers and topologies[223,228]. Strikingly, changes as small as a single amino acid substitution have been shown to induce a tetrameric coiled-coil domain topology to switch from parallel to anti-parallel[229]. Therefore the ability to accurately predict the quaternary state of coiled-coil domains is an attractive goal. Given the large number of elucidated coiled-coil structures, surprisingly few anti-parallel tetramers have been solved compared to parallel tetramers. Although modelling capabilities are improving, with programs being developed which can predict parallel and anti-parallel topology[230], YabA$^{1-58}$ may be of interest in the validation of these programs.

Chapter 5 : Conclusions and Future Perspectives

The regulators of DNA replication, YabA and SirA, form direct interactions with the replication initiator protein DnaA, to negatively regulate replication initiation. Although both have been predicted to act by preventing the binding of DnaA at the replication origin, the proteins interact with different domains of DnaA and appear to orchestrate replication control by different mechanisms. This is not surprising as YabA controls replication initiation during vegetative growth, where replication control must be dynamic to allow the cell to reinitiate replication when required, whilst SirA regulates DNA replication during sporulation; a situation in which the cell does not need to initiate any further replication events.

Perspectives from SirA

The structure of SirA in complex with domain I of DnaA presented in Chapter 2 delineates the interaction interface between the proteins. This work was supported by *in vivo* functional studies performed by N. Rostami and H. Murray which demonstrated that the SirA-DnaA^{DI} interface identified in the crystal structure was functionally relevant. Intriguingly, the SirA interaction determinants on DnaA^{DI} where found to be at a structurally equivalent site to those for two positive regulators of initiation, DiaA in *E. coli* and HobA in *H.pylori*, despite their differing effects on replication initiation[73,157,214]. The relevance of this binding site may extend further as two other known interaction partners of DnaA^{DI}, the *E. coli* DNA replication regulator Hda and the *E. coli* helicase, have also been shown to interact with specific residues within a structurally equivalent surface of DnaA^{DI} [26,82]. Furthermore, our inability to produce alanine mutations at residues Thr 26, Trp 27 and Phe 49 on DnaA at the SirA-DnaA^{DI} interface implies that *B. subtilis* DnaA may ultilise this site in an essential interaction with another component of the initiation machinery – perhaps an, as yet, unidentified interaction partner at this site. Significantly, no interaction partner of DnaA^{DI} has been found to interact with any other location on DnaA^{DI}. This poses the question of what makes this site on DnaA^{DI} a prime location for its interaction with protein partners, and how, if at all, it relates to the role of DnaA^{DI} during replication initiation?

Despite its many interaction partners, the role of DnaA^{DI} during the initiation process is not well established. DnaA^{DI} has a KH-domain fold commonly found in ssDNA binding proteins,

and one study has suggested that it has a weak but specific affinity for single stranded *oriC* DNA[26]. However, it is not known if this ssDNA binding plays a functional role *in vivo*. In *E. coli*, DnaA^{DI} has long been thought to play a role in the oligomerisation of DnaA molecules at *oriC* [231].This is based on the observation that it can functionally replace the dimerisation domain of the bacteriophage λ CI repressor in a "one-hybrid-system" [231] (i.e. a chimeric λ CI-DnaA^{DI} protein is able to inhibit λ replication as efficiently as λ cI) and because it affects DnaA oligomerisation in a gluteraldehyde cross-linking assay[76] (removal of domain I reduces the amount of observed cross-linking).

Based mainly on the observation that other KH-domains can form dimers, it has been suggested that DnaA^{DI} may also form dimers[26]. In *E. coli* DnaA^{DI}, it is proposed that dimerisation is mediated by an essential tryptophan at position 6 (and other residues of the same helix)[75,76] on one surface of the domain. This helix is located on the opposite face of DnaA^{DI} to the binding site of the interaction partners of DnaA^{DI} (Fig. 5.1). However, the assignment of this surface as a site for self-interaction must be treated cautiously because it is unclear if mutations at these residues specifically interfere with oligomerisation or if they could also affect the fold of the protein. In the *B. subtilis* DnaA^{DI} structure presented in Chapter 2, which has an equivalent tryptophan at position 8 (Fig. 5.1, A), the tryptophan side chain is completely buried in the protein core, forming significant interactions with side chains of residues in adjacent helices; including those involved in binding interaction partners of DnaA^{DI}. In the NMR structure of *E. coli* [26] DnaA^{DI}, this tryptophan is also partially buried (Fig. 5.1, B). Thus, a mutation of this residue is likely to cause movement of helices adjacent to the tryptophan side chain and destabilisation of the DnaA^{DI} structure. This would inevitably disrupt the interactions of DnaA^{DI} with its binding partners - such as the helicase, which interacts with the residue Glu 21[26] (Fig. 5.1, B). Helicase loading is an essential step in the initiation of replication and therefore blocking loading of the helicase is inhibitory to replication events. Thus, loss of the DnaA-helicase interaction could account for the *in vivo* effects seen with an *E. coli* Trp6Ala mutant; where Trp 6 was judged to be essential for DNA replication based on the fact that cells harbouring the Ala substitution at Trp 6 are unable to form colonies in a plating assay[75]. As demonstrated by the DnaA^{DI} Ala50Val mutation which prevents the binding of SirA to DnaA[83], even small perturbations of the protein core can have significant effects on the binding ability of DnaA^{DI}. Thus, interpretation of the results with regards to a self-interaction between two N-terminal domains of DnaA must be made cautiously.

Nevertheless, self-interaction may provide an explanation for a common binding site on DnaA^{DI}; this may be the only accessible surface of domain I during initiation if other surfaces are obscured due to a self-interaction. A related theory is that the orientation of domain I with respect to the rest of DnaA is somehow important for initiation. Given the flexible nature of DnaA Domain II, it is unclear how this could be orchestrated other than by self-interaction. Nevertheless, orientation could also play a role in dictating the surface of DnaA^{DI} available to bind interaction partners, and provides a mechanism by which positive regulators such as HobA/DiaA could promote replication initiation by promoting the correct orientation of domain I, and negative regulators such as SirA could inhibit the correct orientation of domain I.

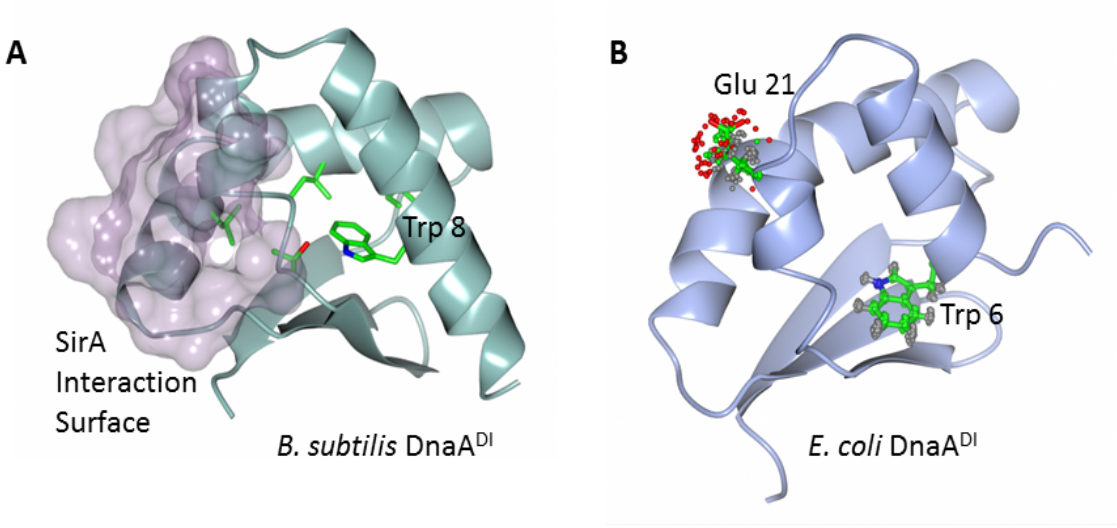


Figure 5.1: Structures of DnaA^{DI} from *B. subtilis* and *E. coli*: A) Crystal structure of *B. subtilis* DnaA^{DI} displaying the location of the buried tryptophan residue at position 6, and the close proximity to other buried side chains on adjacent helices. The SirA interaction site is displayed as a surface representation in lilac, as an example of the location of interaction for various partners of DnaA^{DI}. **B)** NMR structure of *E. coli* DnaA^{DI} displaying Trp 6 and Glu 21. The partially buried tryptophan 6 has been implicated in self-interaction of domain I, and Glu 21 has been shown to be important for DnaB helicase binding.

Regardless, the SirA-DnaA^{DI} structure highlights the importance of understanding the role of DanA^{DI} during replication initiation, and may signal additional essential interactions with DnaA^{DI} which have not yet been found in *B. subtilis*. Ultimately, to further our understanding of the initiation process and elucidate the mechanism of not only SirA, but of many other interaction partners of DnaA, more must be understood about the role of

the N-terminal domain of DnaA and how it orchestrates its function with the rest of the initiator.

Perspectives from YabA

The structural and biophysical characterisation of YabA described in Chapters 3 and 4, together with the results of our collaborators, delineates the overall architecture of YabA and provides insights into the mode of interaction between YabA, DnaA and DnaN. The architecture of the YabA tetramer comprises a four-helix anti-parallel coiled-coil core, flanked at either end by two 'independent' pseudo-monomeric C-terminal domains which are tethered to the N-terminal helices via a flexible linker. YabA is able to interact with DnaA and DnaN via the distal C-terminal domains in a manner which allows multiple interactions to take place simultaneously. However, to further elucidate how YabA regulates initiation, it is essential to know more about the stoichiometry and architecture of the complex(es) it forms with DnaA and DnaN. These questions could be addressed by a number of additional structural and biophysical experiments as outlined below.

Although we have defined the overall architecture of YabA, our understanding of the C-terminal domain interaction with DnaA and DnaN is still somewhat limited. The *in silico* model used for low-resolution structure determination by SAXS is not based on a direct homolog of YabA, but on proteins carrying similar Zn-binding motifs. This means that high-resolution structural information (e.g. locations of side chains) cannot be inferred. Elucidation of a high-resolution structure of the C-terminal domain of YabA would allow greater understanding of the interaction determinants on YabA for binding DnaA and DnaN, through mapping of identified loss of interaction residues.

The YabA C-terminal domain construct produced here, $YabA^{70-119}$, was not readily crystallisable. However, it is not immediately obvious where the secondary structure of the C-terminal domain begins, meaning this construct may contain residues of the flexible linker (between the N- and C-terminal domains of YabA) that are inhibiting crystallisation. As the C-terminal domain has a low molecular weight and appears as a monomer in solution, it is an ideal candidate for NMR spectroscopy[232]. Even obtaining a simple ^{1}H 1D spectrum would provide information about the level of structure present within the construct[233]. A full assignment could also be undertaken to elucidate the structure by NMR

and/or identify all unstructured regions within the C-terminal domain[232]. This would provide information for making a more 'crystallisable' construct; which could then be used in co-crystallisation trials with DnaA or DnaN. Formation of a complex of the YabA C-terminal domain with DnaA or DnaN may be facilitated by co-expression.

A crystal structure of full-length YabA is likely to be unobtainable. Crystallisation attempts with full-length protein described in section 4.5.1, by M. Fogg in York and by L. Terradot and colleagues in Lyon, all resulted in crystals containing only the N-terminal domain of YabA following *in situ* proteolysis. This highlights the sensitivity of the protein to proteolysis. Furthermore, the flexible linker between the N- and C-terminal domains is also likely to inhibit crystallisation of a full-length construct. Thus any future crystallisation studies carried out with YabA should include only the C-terminal domain or a construct of YabA which has the flexible linker removed.

Study of the interactions of YabA with DnaA and DnaN could also be undertaken with SAXS, or by AUC; if the proteins form stable complexes in solution. This could allow the stoichiometry of the complexes to be determined. Ultimately, if co-crystal structures of YabA C-terminal domain with DnaA or DnaN are possible, analysis of the ternary complex by SAXS could allow a low resolution structural envelope to be generated, into which the co-crystal structures could be fitted. This would provide significant insight into the nature of the DnaA-YabA-DnaN complex, and ultimately the molecular mechanism YabA uses to regulate replication initiation.

List of Abbreviations

Å	Angstrom
βME	β-Mercaptoethanol / 2-Mercaptoethanol
µg	Microgram
µl	Microlitre
Ala	Alanine
ADP	Adenosine diphosphate
Asn	Asparagine
Asp	Aspartic Acid
ATP	Adenosine triphosphate
B. anthracis	*Bacillus anthracis*
Bis-Tris Propane	2,2'-(Propane-1,3-diyldiimino)bis-[2(hydroxymethyl)propane-1,3-diol]
B. subtilis	*Bacillus subtilis*
C	Celsius
C. crescentus	*Caulobacter crescentus*
C. difficile	*Clostridium difficile*
CDP	Cytosine diphosphate
Da	Dalton
DARS	DnaA reactivating sequence
DDBH1	DnaD DnaB homology domain 1
DDBH2	DnaD DnaB homology domain 2
DLS	Diamond Light Source

D_{max}	Maximum dimension
DNA	Deoxyribonucleic acid
$DnaA^{DI}$	Domain I of DnaA
dsDNA	Double-stranded DNA
DUE	DNA unwinding element
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionisation
Glu	Glutamic acid
Gln	Glutamine
GFP	Green fluorescent protein
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
His	Histidine
H. pylori	*Helicobacter pylori*
Ile	Isoleucine
IPTG	Isopropyl-β-D-1-thiogalactoside
Kan	Kanamycin
KCl	Potassium chloride
kDa	Kilo-Dalton
KH	K-homology
l	Litre
LB	Lysogeny Broth
Leu	Leucine

MCM	Mini-chromosome maintenance
MES	2-ethanesulfonic acid
Met	Methionine
mg	Milligram
Min	Minutes
ml	Millilitre
mM	Micromolar
MS	Mass Spectrometry
ng	Nanogram
OD_{600}	Optical density at a wavelength of 600 nm
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein databank
PEG	Polyethylene glycol
Phe	Phenylalanine
Pro	Proline
RMSD	Root-mean-square deviation
rpm	Revolutions per minute
SAD	Single anomalous dispersion
SAXS	Small angle x-ray scattering
SDS	Sodium dodecyl sulphate

SEC-MALLS	Size exclusion chromatography – multi-angle laser light scattering
SeMet	Selenomethionine
Ser	Serine
ssDNA	Single-stranded DNA
Thr	Threonine
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
Zn	Zinc

References

1. Reyes-Lamothe, R., Nicolas, E. & Sherratt, D. J. Chromosome Replication and Segregation in Bacteria. *Annu. Rev. Genet.* **46,** 121–143 (2011).

2. Voet, D. & Voet, J. G. in *Biochemistry* 1171–1259 (2011).

3. Leonard, A. C. & Grimwade, J. E. The orisome: structure and function. *Front. Microbiol.* **6,** 1–13 (2015).

4. Zorman, S., Seitz, H., Sclavi, B. & Strick, T. R. Topological characterization of the DnaA-*oriC* complex using single-molecule nanomanipuation. *Nucleic Acids Res.* **40,** 7375–7383 (2012).

5. Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat. Struct. Mol. Biol.* **13,** 676–83 (2006).

6. Fuller, R. S., Funnell, B. E. & Kornberg, A. The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38,** 889–900 (1984).

7. Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J.* **8,** 4335–4344 (1989).

8. Bramhill, D. & Kornberg, A. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52,** 743–55 (1988).

9. Marszalek, J. & Kaguni, J. M. DnaA protein directs the binding of DnaB protein in initiation of DNA replication in *Escherichia coli*. *J. Biol. Chem.* **269,** 4883–4890 (1994).

10. Koboris, J. A. & Kornberg, A. *Escherichia coli* dnaC Gene Product. *Biol. Chem.* **257,** 13770–13775 (1982).

11. Fang, L., Davey, M. J. & O'Donnell, M. Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol. Cell* **4,** 541–553 (1999).

12. Fukuoka, T., Moriya, S., Yoshikawa, H. & Ogasawara, N. Purification and characterization of an initiation protein for chromosomal replication, DnaA, in *Bacillus subtilis*. *J. Biochem.* **107,** 732–9 (1990).

13. Nishida, S. A Nucleotide Switch in the *Escherichia coli* DnaA Protein Initiates Chromosomal Replication. *J. Biol. Chem.* **277,** 14986–14995 (2002).

14. Sekimizu, K., Bramhill, D. & Kornberg, A. ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the *E. coli* chromosome. *Cell* **50,** 259–65 (1987).

15. Margulies, C. & Kaguni, J. M. Ordered and sequential binding of DnaA protein to *oriC*, the chromosomal origin of *Escherichia coli*. *J. Biol. Chem.* **271,** 17035–17040 (1996).

16. Langer, U., Richter, S., Roth, A., Weigel, C. & Messer, W. A comprehensive set of DnaA-box mutations in the replication origin, *oriC*, of *Escherichia coli*. *Mol. Microbiol.* **21,** 301–311 (1996).

17. Miller, D. T., Grimwade, J. E., Betteridge, T., Rozgaja, T., Torgue, J. J.-C. & Leonard, A. C. Bacterial origin recognition complexes direct assembly of higher-order DnaA oligomeric structures. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 18479–84 (2009).

18. Rozgaja, T. A., Grimwade, J. E., Iqbal, M., Czerwonka, C., Vora, M. & Leonard, A. C. Two oppositely oriented arrays of low affinity recognition sites in *oriC* guide progressive binding of DnaA during *E. coli* pre-RC assembly. **82,** 475–488 (2012).

19. McGarry, K. C., Ryan, V. T., Grimwade, J. E. & Leonard, A. C. Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA-ATP. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2811–6 (2004).

20. Kaur, G., Vora, M. P., Czerwonka, C. A., Rozgaja, T. A., Grimwade, J. E. & Leonard, A. C. Building the bacterial orisome: high affinity DnaA recognition plays a role in setting the conformation of *oriC* DNA. *Mol. Microbiol.* **91,** 1148–1163 (2014).

21. Mott, M. L. & Berger, J. M. DNA replication initiation: mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* **5,** 343–54 (2007).

22. Speck, C. & Messer, W. Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* **20,** 1469–76 (2001).

23. Duderstadt, K. E., Chuang, K. & Berger, J. M. DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478,** 209–213 (2011).

24. Arias-Palomo, E., O'Shea, V. L., Hood, I. V. & Berger, J. M. The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* **153,** 438–448 (2013).

25. Zhang, W., Marszalek, J., Zhang, W., Hupp, T. R., Margulies, C., Carr, K. M., Cherry, S. & Kaguni, J. M. Domains of DnaA Protein Involved in Interaction with DnaB Protein, and in Unwinding the *Escherichia coli* Chromosomal Origin. *J. Biol. Chem.* **271,** 18535–18542 (1996).

26. Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T. & Ueda, T. Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on *oriC*. *J. Biol. Chem.* **282,** 17816–27 (2007).

27. Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135,** 623–34 (2008).

28. Makowska-Grzyska, M. & Kaguni, J. M. Primase Directs the Release of DnaC from DnaB. *Mol. Cell* **37,** 90–101 (2010).

29. Bell, S. P. & Kaguni, J. M. Helicase loading at chromosomal origins of replication. *Cold Spring Harb. Perspect. Biol.* **5,** 1–20 (2013).

30. Soultanas, P. Loading mechanisms of ring helicases at replication origins. *Mol. Microbiol.* **84,** 6–16 (2012).

31. Wolanski, M., Donczew, R., Zawilak-Pawlik, A. & Zakrzewska-Czerwinska, J. *oriC*-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* **5,** 1–14 (2015).

32. Briggs, G. S., Smits, W. K. & Soultanas, P. Chromosomal replication initiation machinery of low-G+C-content Firmicutes. *J. Bacteriol.* **194,** 5162–70 (2012).

33. Scholefield, G., Errington, J. & Murray, H. Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J.* **31,** 1–14 (2012).

34. Merrikh, H. & Grossman, A. D. Control of the replication initiator DnaA by an anti-cooperativity factor. *Mol. Microbiol.* **82,** 434–446 (2011).

35. Rotoli, S. M., Biswas-Fiss, E. & Biswas, S. B. Quantitative analysis of the mechanism of DNA binding by *Bacillus* DnaA protein. *Biochimie* **94,** 2764–2775 (2012).

36. Bruand, C., Ehrlich, S. D. & Jannière, L. Primosome assembly site in *Bacillus subtilis*. *EMBO J.* **14,** 2642–50 (1995).

37. Zhang, W., Carneiro, M. J. V. M., Turner, I. J., Allen, S., Roberts, C. J. & Soultanas, P. The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodelling activities. *J. Mol. Biol.* **351,** 66–75 (2005).

38. Smits, W. K., Goranov, A. I. & Grossman, A. D. Ordered association of helicase loader proteins with the *Bacillus subtilis* origin of replication *in vivo*. *Mol. Microbiol.* **75,** 452–61 (2010).

39. Velten, M., McGovern, S., Marsin, S., Ehrlich, S. D., Noirot, P. & Polard, P. A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol. Cell* **11,** 1009–20 (2003).

40. Ioannou, C., Schaeffer, P. M., Dixon, N. E. & Soultanas, P. Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in *Bacillus subtilis*. *Nucleic Acids Res.* **34,** 5247–5258 (2006).

41. Liu, B., Eliason, W. K. & Steitz, T. A. Structure of a helicase-helicase loader complex reveals insights into the mechanism of bacterial primosome assembly. *Nat. Commun.* **4,** 1–8 (2013).

42. Bailey, S., Eliason, W. K. & Steitz, T. A. Structure of Hexameric DnaB. **459,** 459–463 (2007).

43. Reyes-Lamothe, R., Sherratt, D. J. & Leake, M. C. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science* **328,** 498–501 (2010).

44. Robinson, A., J. Causer, R. & E. Dixon, N. Architecture and Conservation of the Bacterial DNA Replication Machinery, an Underexploited Drug Target. *Curr. Drug Targets* **13,** 352–372 (2012).

45. Beattie, T. R. & Reyes-Lamothe, R. A Replisome's journey through the bacterial chromosome. *Front. Microbiol.* **6,** 1–12 (2015).

46. Corn, J. E., Pelton, J. G. & Berger, J. M. Identification of a DNA primase template tracking site redefines the geometry of primer synthesis. *Nat. Struct. Mol. Biol.* **15,** 163–169 (2008).

47. Corn, J. E., Pease, P. J., Hura, G. L. & Berger, J. M. Crosstalk between primase subunits can act to regulate primer synthesis in *trans*. *Mol. Cell* **20,** 391–401 (2005).

48. Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., Jannière, L. & Ehrlich, S. D. Two essential DNA polymerases at the bacterial replication fork. *Science* **294,** 1716–1719 (2001).

49. Sanders, G. M., Dallmann, H. G. & McHenry, C. S. Reconstitution of the *B. subtilis* Replisome with 13 Proteins Including Two Distinct Replicases. *Mol. Cell* **37,** 273–281 (2010).

50. Rannou, O., Le Chatelier, E., Larson, M. a., Nouri, H., Dalmais, B., Laughton, C., Jannière, L. & Soultanas, P. Functional interplay of DnaE polymerase, DnaG primase and DnaC helicase within a ternary complex, and primase to polymerase hand-off during lagging strand DNA replication in Bacillus subtilis. *Nucleic Acids Res.* **41,** 5303–5320 (2013).

51. Hill, T. M., Henson, J. M. & Kuempel, P. L. The terminus region of the *Escherichia coli* chromosome contains two separate loci that exhibit polar inhibition of replication. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 1754–1758 (1987).

52. Hidaka, M., Kobayashi, T., Takenaka, S., Takeya, H. & Horiuchi, T. Purification of a DNA replication terminus (*ter*) site- binding protein in *Escherichia coli* and identification of the structural gene. *J. Biol. Chem.* **264,** 21031–21037 (1989).

53. Neylon, C., Kralicek, A. V, Hill, T. M. & Dixon, N. E. Replication Termination in *Escherichia coli*: Structure and Antihelicase Activity of the Tus-*Ter* Complex. *Microbiol. Mol. Biol. Rev.* **69,** 501–526 (2005).

54. Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N. & Morikawa, K. Structure of a replication-terminator protein complexed with DNA. *Nature* **383,** 598–603 (1996).

55. Mulcair, M. D., Schaeffer, P. M., Oakley, A. J., Cross, H. F., Neylon, C., Hill, T. M. & Dixon, N. E. A Molecular Mousetrap Determines Polarity of Termination of DNA Replication in *E. coli*. *Cell* **125,** 1309–1319 (2006).

56. Berghuis, B. A., Dulin, D., Xu, Z.-Q., van Laar, T., Cross, B., Janissen, R., Jergic, S., Dixon, N. E., Depken, M. & Dekker, N. H. Strand separation establishes a sustained lock at the Tus–Ter replication fork barrier. *Nat. Chem. Biol.* **11,** 579–585 (2015).

57. Hill, T. M. Arrest of bacterial DNA replication. *Annu. Rev. Microbiol.* **46,** 603–633 (1992).

58. Lewis, P. J., Ralston, G. B., Christopherson, R. I. & Wake, R. G. Identification of the replication terminator protein binding sites in the terminus region of the *Bacillus subtilis* chromosome and stoichiometry of the binding. *J. Mol. Biol.* **214,** 73–84 (1990).

59. Vivian, J. P., Porter, C. J., Wilce, J. A. & Wilce, M. C. J. An Asymmetric Structure of the *Bacillus subtilis* Replication Terminator Protein in Complex with DNA. *J. Mol. Biol.* **370,** 481–491 (2007).

60. Langley, D. B., Smith, M. T., Lewis, P. J. & Wake, R. G. Protein-nucleoside contacts in the interaction between the replication terminator protein of *Bacillus subtilis* and the DNA terminator. *Mol. Microbiol.* **10,** 771–779 (1993).

61. Kaguni, J. M. Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* **15,** 606–13 (2011).

62. Leonard, A. C. & Grimwade, J. E. Regulation of DnaA assembly and activity: taking directions from the genome. *Annu. Rev. Microbiol.* **65,** 19–35 (2011).

63. Zawilak-Pawlik, A., Kois, A., Majka, J., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Messer, W. & Zakrzewska-Czerwińska, J. Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria. *Biochem. J.* **389,** 471–481 (2005).

64. Ogasawara, N. & Yoshikawa, H. Genes and their organization in the replication origin region of the bacterial chromosome. *Mol. Microbiol.* **6,** 629–634 (1992).

65. Moriya, S., Atlung, T., Hansen, F. G., Yoshikawa, H. & Ogasawara, N. Cloning of an autonomously replicating sequence (*ars*) from the *Bacillus subtilis* chromosome. *Mol. Microbiol.* **6,** 309–315 (1992).

66. Donczew, R., Weigel, C., Lurz, R., Zakrzewska-Czerwińska, J. & Zawilak-Pawlik, A. *Helicobacter pylori oriC*-the first bipartite origin of chromosome replication in Gram-negative bacteria. *Nucleic Acids Res.* **40,** 9647–9660 (2012).

67. Krause, M., Rückert, B., Lurz, R. & Messer, W. Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. *J. Mol. Biol.* **274,** 365–80 (1997).

68. Woelker, B. & Messer, W. The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.* **21,** 5025–5033 (1993).

69. Zawilak, A., Durrant, M. C., Jakimowicz, P., Backert, S. & Zakrzewska-Czerwińska, J. DNA Binding Specificity of the Replication Initiator Protein, DnaA from *Helicobacter pylori*. *J. Mol. Biol.* **334,** 933–947 (2003).

70. Rajewska, M., Wegrzyn, K. & Konieczny, I. AT-rich region and repeated sequences - the essential elements of replication origins of bacterial replicons. *FEMS Microbiol. Rev.* **36,** 408–434 (2012).

71. Messer, W., Blaesing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tüngler, D., Wegrzyn, G., Weigel, C., Welzeck, M. & Zakrzewska-Czerwińska, J. Functional domains of DnaA proteins. *Biochimie* **81,** 819–25 (1999).

72. Sutton, M. D. & Kaguni, J. M. The *Escherichia coli dnaA* gene: four functional domains. *J. Mol. Biol.* **274,** 546–61 (1997).

73. Natrajan, G., Noirot-Gros, M.-F., Zawilak-Pawlik, A., Kapp, U. & Terradot, L. The structure of a DnaA/HobA complex from *Helicobacter pylori* provides insight into regulation of DNA replication in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21115–20 (2009).

74. Lowery, T. J., Pelton, J. G., Chandonia, J.-M., Kim, R., Yokota, H. & Wemmer, D. E. NMR structure of the N-terminal domain of the replication initiator protein DnaA. *J. Struct. Funct. Genomics* **8,** 11–7 (2007).

75. Felczak, M. M., Simmons, L. A. & Kaguni, J. M. An essential tryptophan of *Escherichia coli* DnaA protein functions in oligomerization at the *E. coli* replication origin. *J. Biol. Chem.* **280,** 24627–33 (2005).

76. Simmons, L. A., Felczak, M. & Kaguni, J. M. DnaA Protein of *Escherichia coli*: Oligomerization at the *E. coli* chromosomal origin is required for initiation and involves specific N-terminal amino acids. *Mol. Microbiol.* **49,** 849–858 (2003).

77. Seitz, H., Weigel, C. & Messer, W. The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Mol. Microbiol.* **37,** 1270–9 (2000).

78. Carr, K. M. & Kaguni, J. M. Stoichiometry of DnaA and DnaB protein in initiation at the *Escherichia coli* chromosomal origin. *J. Biol. Chem.* **276,** 44919–25 (2001).

79. Weigel, C. & Seitz, H. Strand-specific loading of DnaB helicase by DnaA to a substrate mimicking unwound *oriC*. *Mol. Microbiol.* **46,** 1149–1156 (2002).

80. Keyamura, K., Abe, Y., Higashi, M., Ueda, T. & Katayama, T. DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J. Biol. Chem.* **284,** 25038–50 (2009).

81. Zawilak-Pawlik, A., Kois, A., Stingl, K., Boneca, I. G., Skrobuk, P., Piotr, J., Lurz, R., Zakrzewska-Czerwińska, J. & Labigne, A. HobA - a novel protein involved in initiation of chromosomal replication in *Helicobacter pylori*. *Mol. Microbiol.* **65,** 979–94 (2007).

82. Su'etsugu, M., Harada, Y., Keyamura, K., Matsunaga, C., Kasho, K., Abe, Y., Ueda, T. & Katayama, T. The DnaA N-terminal domain interacts with Hda to facilitate replicase clamp-mediated inactivation of DnaA. *Environ. Microbiol.* **15,** 3183–95 (2013).

83. Rahn-Lee, L., Merrikh, H., Grossman, A. D. & Losick, R. The sporulation protein SirA inhibits the binding of DnaA to the origin of replication by contacting a patch of clustered amino acids. *J. Bacteriol.* **193,** 1302–7 (2011).

84. Nozaki, S. & Ogawa, T. Determination of the minimum domain II size of *Escherichia coli* DnaA protein essential for cell viability. *Microbiology* **154,** 3379–84 (2008).

85. Molt, K. L., Sutera, V. A., Moore, K. K. & Lovett, S. T. A role for nonessential domain II of initiator protein, DnaA, in replication control. *Genetics* **183,** 39–49 (2009).

86. Roth, A. & Messer, W. The DNA binding domain of the initiator protein DnaA. *EMBO J.* **14,** 2106–2111 (1995).

87. Schaper, S. & Messer, W. Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.29.17622

88. Fujikawa, N., Kurumizaka, H., Nureki, O. & Terada, T. Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res.* **31,** 2077–2086 (2003).

89. Blaesing, F., Weigel, C., Welzeck, M. & Messer, W. Analysis of the DNA-binding domain of *Escherichia coli* DnaA protein. *Mol. Microbiol.* **36,** 557–569 (2000).

90. Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure - diverse function. *Genes Cells* **6,** 575–97 (2001).

91. Katayama, T. Roles for the AAA+ motifs of DnaA in the initiation of DNA replication. *Biochem. Soc. Trans.* **36,** 78–82 (2008).

92. Erzberger, J. P., Pirruccello, M. M. & Berger, J. M. The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J.* **21,** 4763–73 (2002).

93. Ogura, T., Whiteheart, S. W. & Wilkinson, A. J. Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol.* **146,** 106–112 (2004).

94. Duderstadt, K. E., Mott, M. L., Crisona, N. J., Chuang, K., Yang, H. & Berger, J. M. Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* **285,** 28229–39 (2010).

95. Costa, A., Hood, I. V & Berger, J. M. Mechanisms for initiating cellular DNA replication. *Annu. Rev. Biochem.* **82,** 25–54 (2013).

96. Bruand, C., Farache, M., McGovern, S., Ehrlich, S. D. & Polard, P. DnaB, DnaD and DnaI proteins are components of the *Bacillus subtilis* replication restart primosome. *Mol. Microbiol.* **42,** 245–55 (2001).

97. Marston, F. Y., Grainger, W. H., Smits, W. K., Hopcroft, N. H., Green, M., Hounslow, A. M., Grossman, A. D., Craven, C. J. & Soultanas, P. When simple sequence comparison fails: the cryptic case of the shared domains of the bacterial replication initiation proteins DnaB and DnaD. *Nucleic Acids Res.* **38,** 6930–42 (2010).

98. Ishigo-oka, D., Ogasawara, N. & Moriya, S. DnaD Protein of *Bacillus subtilis* Interacts with DnaA, the Initiator Protein of Replication. *J. Bacteriol.* **183,** 1–4 (2001).

99. Zhang, W., Allen, S., Roberts, C. J. & Soultanas, P. The *Bacillus subtilis* primosomal protein DnaD untwists supercoiled DNA. *J. Bacteriol.* **188,** 5487–93 (2006).

100. Zhang, W., Machón, C., Orta, A., Phillips, N., Roberts, C. J., Allen, S. & Soultanas, P. Single-molecule atomic force spectroscopy reveals that DnaD forms scaffolds and enhances duplex melting. *J. Mol. Biol.* **377,** 706–14 (2008).

101. Schneider, S., Zhang, W., Soultanas, P. & Paoli, M. Structure of the N-Terminal Oligomerization Domain of DnaD Reveals a Unique Tetramerization Motif and Provides Insights into Scaffold Formation. *J. Mol. Biol.* **376,** 1237–1250 (2008).

102. Carneiro, M. J. V. M., Zhang, W., Ioannou, C., Scott, D. J., Allen, S., Roberts, C. J. & Soultanas, P. The DNA-remodelling activity of DnaD is the sum of oligomerization and DNA-binding activities on separate domains. *Mol. Microbiol.* **60,** 917–924 (2006).

103. Bruand, C., Velten, M., McGovern, S., Marsin, S., Sérèna, C., Dusko Ehrlich, S. & Polard, P. Functional interplay between the *Bacillus subtilis* DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Mol. Microbiol.* **55,** 1138–1150 (2005).

104. Collier, C., Machón, C., Briggs, G. S., Smits, W. K. & Soultanas, P. Untwisting of the DNA helix stimulates the endonuclease activity of *Bacillus subtilis* Nth at AP sites. *Nucleic Acids Res.* **40,** 739–750 (2012).

105. Rokop, M. E., Auchtung, J. M. & Grossman, A. D. Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol. Microbiol.* **52,** 1757–67 (2004).

106. Hoshino, T., McKenzie, T., Schmidt, S., Tanaka, T. & Sueoka, N. Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for DNA replication initiation and membrane attachment. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 653–657 (1987).

107. Grainger, W. H., Machón, C., Scott, D. J. & Soultanas, P. DnaB proteolysis *in vivo* regulates oligomerization and its localization at *oriC* in *Bacillus subtilis*. *Nucleic Acids Res.* **38,** 2851–2864 (2010).

108. Bonilla, C. Y. & Grossman, A. D. The primosomal protein DnaD inhibits cooperative DNA binding by the replication initiator DnaA in *Bacillus subtilis*. *J. Bacteriol.* **194,** 5110–5117 (2012).

109. Scholefield, G. & Murray, H. YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Mol. Microbiol.* **90,** 147–159 (2013).

110. Helmstetter, C. E. & Pierucci, O. DNA synthesis during the division cycle of three substrains of *Escherichia coli* B/r. *J. Mol. Biol.* **102,** 477–486 (1976).

111. Boye, E. & Nordström, K. Coupling the cell cycle to cell growth. *EMBO Rep.* **4,** 757–760 (2003).

112. Atlung, T., Løbner-Olesen, A. & Hansen, F. G. Overproduction of DnaA protein stimulates initiation of chromosome and minichromosome replication in *Escherichia coli*. *Mol. Gen. Genet.* **206,** 51–59 (1987).

113. Flåtten, I., Fossum-Raunehaug, S., Taipale, R., Martinsen, S. & Skarstad, K. The DnaA Protein Is Not the Limiting Factor for Initiation of Replication in Escherichia coli. *PLOS Genet.* **11,** e1005276 (2015).

114. Murray, H. & Koh, A. Multiple Regulatory Systems Coordinate DNA Replication with Cell Growth in Bacillus subtilis. *PLoS Genet.* **10,** 1–15 (2014).

115. Messer, W. & Weigel, C. DnaA initiator - a transcription factor. **24,** 1–6 (1997).

116. Goranov, A. I., Katz, L., Breier, A. M., Burge, C. B. & Grossman, A. D. A transcriptional response to replication status mediated by the conserved bacterial replication protein DnaA. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 12932–12937 (2005).

117. Collier, J., Murray, S. R. & Shapiro, L. DnaA couples DNA replication and the expression of two cell cycle master regulators. *EMBO J.* **25,** 346–56 (2006).

118. Atlung, T., Clausen, E. S. & Hansen, F. G. Autoregulation of the *dnaA* gene of *Escherichia coli* K12. *Mol. Gen. Genet. MGG* **200,** 442–450 (1985).

119. Breier, A. M. & Grossman, A. D. Dynamic association of the replication initiator and transcription factor DnaA with the *Bacillus subtilis* chromosome during replication stress. *J. Bacteriol.* **191,** 486–493 (2009).

120. Fernandez-Fernandez, C., Gonzalez, D. & Collier, J. Regulation of the activity of the dual-function DnaA protein in *Caulobacter crescentus*. *PLoS One* **6,** e26028 (2011).

121. Kaguni, J. M. DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol.* **60,** 351–75 (2006).

122. Katayama, T., Ozaki, S., Keyamura, K. & Fujimitsu, K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat. Rev. Microbiol.* **8,** 163–70 (2010).

123. Camara, J. E., Breier, A. M., Brendler, T., Austin, S., Cozzarelli, N. R. & Crooke, E. Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of *Escherichia coli* DNA replication. *EMBO Rep.* **6,** 736–741 (2005).

124. Katayama, T., Kubota, T., Kurokawa, K., Crooke, E. & Sekimizu, K. The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. *Cell* **94,** 61–71 (1998).

125. Kato, J. & Katayama, T. Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *EMBO J.* **20,** 4253–62 (2001).

126. Takata, M., Kubota, T., Matsuda, Y. & Katayama, T. Molecular mechanism of DNA replication-coupled inactivation of the initiator protein in *Escherichia coli* : interaction of DnaA with the sliding clamp-loaded DNA and the sliding clamp-Hda complex. *Genes to Cells* **9,** 509– 522 (2004).

127. Kurz, M., Dalrymple, B., Wijffels, G. & Kongsuwan, K. Interaction of the Sliding Clamp Beta-Subunit and Hda, a DnaA-Related Protein. *J. Bacteriol.* **186,** 3508–3515 (2004).

128. Su'etsugu, M., Shimuta, T.-R. R., Ishida, T., Kawakami, H. & Katayama, T. Protein associations in DnaA-ATP hydrolysis mediated by the Hda-replicase clamp complex. *J. Biol. Chem.* **280,** 6528–36 (2005).

129. Fujimitsu, K., Su'etsugu, M., Yamaguchi, Y., Mazda, K., Fu, N., Kawakami, H. & Katayama, T. Modes of overinitiation, *dnaA* gene expression, and inhibition of cell division in a novel cold-sensitive *hda* mutant of *Escherichia coli*. *J. Bacteriol.* **190,** 5368–5381 (2008).

130. Su'etsugu, M., Nakamura, K., Keyamura, K., Kudo, Y. & Katayama, T. Hda Monomerization by ADP Binding Promotes Replicase Clamp-mediated DnaA-ATP Hydrolysis. *J. Biol. Chem.* **283,** 36118–36131 (2008).

131. Xu, Q., McMullan, D., Abdubek, P., Astakhova, T., Carlton, D., Chen, C., Chiu, H.-J., Clayton, T., Das, D., Deller, M. C., Duan, L., Elsliger, M.-A., Feuerhelm, J., Hale, J., Han, G. W., Jaroszewski, L., Jin, K. K., Johnson, H. A., Klock, H. E., Knuth, M. W., Kozbial, P., Sri Krishna, S., Kumar, A., Marciano, D., Miller, M. D., Morse, A. T., Nigoghossian, E., Nopakun, A., Okach, L., Oommachen, S., Paulsen, J., Puckett, C., Reyes, R., Rife, C. L., Sefcovic, N., Trame, C., van den Bedem, H., Weekes, D., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A. & Wilson, I. A. A structural basis for the regulatory inactivation of DnaA. *J. Mol. Biol.* **385,** 368–80 (2009).

132. Nakamura, K. & Katayama, T. Novel essential residues of Hda for interaction with DnaA in the regulatory inactivation of DnaA: unique roles for Hda AAA + Box VI and VII motifs. *Mol. Microbiol.* **76,** 302–317 (2010).

133. Keyamura, K. & Katayama, T. DnaA Protein DNA-binding Domain Binds to Hda Protein to Promote Inter-AAA+ Domain Interaction Involved in Regulatory Inactivation of DnaA. *J. Biol. Chem.* **286,** 29336–29346 (2011).

134. Ryan, V. T., Grimwade, J. E., Camara, J. E., Crooke, E. & Leonard, A. C. *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Mol. Microbiol.* **51,** 1347–59 (2004).

135. Kasho, K., Fujimitsu, K., Matoba, T., Oshima, T. & Katayama, T. Timely binding of IHF and Fis to DARS2 regulates ATP-DnaA production and replication initiation. *Nucleic Acids Res.* **42,** 13134–13149 (2014).

136. Kasho, K. & Katayama, T. DnaA binding locus datA promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 936–941 (2012).

137. Filutowicz, M., Ross, W., Wild, J. & Gourse, R. L. Involvement of Fis protein in replication of the *Escherichia coli* chromosome. *J. Bacteriol.* **174,** 398–407 (1992).

138. Pratt, T. S., Steiner, T., Feldman, L. S., Walker, K. A. & Osuna, R. Deletion analysis of the *fis* promoter region in *Escherichia coli*: antagonistic effects of integration host factor and Fis. *J. Bacteriol.* **179,** 6367–6377 (1997).

139. Flåtten, I. & Skarstad, K. The Fis protein has a stimulating role in initiation of replication in *Escherichia coli in vivo*. *PLoS One* **8,** 1–9 (2013).

140. Cassler, M. R., Grimwade, J. E. & Leonard, A. C. Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J.* **14,** 5833–5841 (1995).

141. Grimwade, J. E., Ryan, V. T. & Leonard, A. C. IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. *Mol. Microbiol.* **35,** 835–844 (2000).

142. Kitagawa, R., Mitsuki, H., Okazaki, T. & Ogawa, T. A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol. Microbiol.* **19,** 1137–47 (1996).

143. Kitagawa, R., Ozaki, T., Moriya, S. & Ogawa, T. Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev.* **12,** 3032–3043 (1998).

144. Ogawa, T., Yamada, Y., Kuroda, T., Kishi, T. & Moriya, S. The *datA* locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in *Escherichia coli*. *Mol. Microbiol.* **44,** 1367–1375 (2002).

145. Nozaki, S., Yamada, Y. & Ogawa, T. Initiator titration complex formed at *datA* with the aid of IHF regulates replication timing in *Escherichia coli*. *Genes to Cells* **14,** 329–341 (2009).

146. Fujimitsu, K., Senriuchi, T. & Katayama, T. Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA. *Genes Dev.* **23,** 1221–33 (2009).

147. Lu, M., Campbell, J. L., Boye, E. & Kleckner, N. SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* **77,** 413–26 (1994).

148. Von Freiesleben, U., Rasmussen, K. V. & Schaechter, M. SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol. Microbiol.* **14,** 763–772 (1994).

149. Campbell, J. L. & Kleckner, N. *E. coli oriC* and the *dnaA* gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* **62,** 967–79 (1990).

150. Nievera, C., Torgue, J. J.-C., Grimwade, J. E. & Leonard, A. C. SeqA Blocking of DnaA-*oriC* Interactions Ensures Staged Assembly of the *E. coli* Pre-RC. *Mol. Cell* **24,** 581–592 (2006).

151. Slater, S., Wold, S., Lu, M., Boye, E., Skarstad, K. & Kleckner, N. *E. coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell* **82,** 927–936 (1995).

152. Han, J. S., Kang, S., Lee, H., Kim, H. K. & Hwang, D. S. Sequential binding of SeqA to paired hemi-methylated GATC sequences mediates formation of higher order complexes. *J. Biol. Chem.* **278,** 34983–34989 (2003).

153. Waldminghaus, T. & Skarstad, K. The *Escherichia coli* SeqA protein. *Plasmid* **61,** 141–150 (2009).

154. Kang, S., Lee, H., Han, J. S. & Hwang, D. S. Interaction of SeqA and Dam methylase on the hemimethylated origin of *Escherichia coli* chromosomal DNA replication. *J. Biol. Chem.* **274,** 11463–11468 (1999).

155. Helgesen, E., Fossum-Raunehaug, S., Saetre, F., Schink, K. O. & Skarstad, K. Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. *Nucleic Acids Res.* **43,** 2730–2743 (2015).

156. Ishida, T., Akimitsu, N., Kashioka, T., Hatano, M., Kubota, T., Ogata, Y., Sekimizu, K. & Katayama, T. DiaA, a novel DnaA-binding protein, ensures the timely initiation of *Escherichia coli* chromosome replication. *J. Biol. Chem.* **279,** 45546–55 (2004).

157. Zawilak-Pawlik, A., Donczew, R., Szafrański, S., Mackiewicz, P., Terradot, L. & Zakrzewska-Czerwińska, J. DiaA/HobA and DnaA: a pair of proteins co-evolved to cooperate during bacterial orisome assembly. *J. Mol. Biol.* **408,** 238–51 (2011).

158. Keyamura, K., Fujikawa, N., Ishida, T., Ozaki, S., Su, M., Fujimitsu, K., Kagawa, W., Yokoyama, S. & Kurumizaka, H. The interaction of DiaA and DnaA regulates the replication cycle in *E . coli* by directly promoting ATP – DnaA-specific initiation complexes. *Genes Dev.* 2083–2099 (2007).

159. Natrajan, G., Hall, D. R., Thompson, A. C., Gutsche, I. & Terradot, L. Structural similarity between the DnaA-binding proteins HobA (HP1230) from *Helicobacter pylori* and DiaA from *Escherichia coli*. *Mol. Microbiol.* **65,** 995–1005 (2007).

160. Terradot, L. & Zawilak-Pawlik, A. Structural insight into *Helicobacter pylori* DNA replication initiation. *Gut Microbes* **1,** 330–334 (2010).

161. Hayashi, M., Ogura, Y., Harry, E. J., Ogasawara, N. & Moriya, S. *Bacillus subtilis* YabA is involved in determining the timing and synchrony of replication initiation. *FEMS Microbiol. Lett.* **247,** 73–9 (2005).

162. Noirot-Gros, M.-F., Dervyn, E., Wu, L. J., Mervelet, P., Errington, J., Ehrlich, S. D. & Noirot, P. An expanded view of bacterial DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8342–7 (2002).

163. Noirot-Gros, M.-F., Velten, M., Yoshimura, M., McGovern, S., Morimoto, T., Ehrlich, S. D., Ogasawara, N., Polard, P. & Noirot, P. Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 2368–73 (2006).

164. Cho, E., Ogasawara, N. & Ishikawa, S. The functional analysis of YabA, which interacts with DnaA and regulates initiation of chromosome replication in *Bacillus subtils*. *Genes Genet. Syst.* **83,** 111–25 (2008).

165. Murray, H. & Errington, J. Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. *Cell* **135,** 74–84 (2008).

166. Scholefield, G., Whiting, R., Errington, J. & Murray, H. Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation. *Mol. Microbiol.* **79,** 1089–100 (2011).

167. Leonard, T. A., Butler, P. J. & Löwe, J. Bacterial chromosome segregation: structure and DNA binding of the Soj dimer--a conserved biological switch. *EMBO J.* **24,** 270–82 (2005).

168. Okumura, H., Yoshimura, M., Ueki, M., Oshima, T., Ogasawara, N. & Ishikawa, S. Regulation of chromosomal replication initiation by *oriC*-proximal DnaA-box clusters in *Bacillus subtilis*. *Nucleic Acids Res.* **40,** 220–234 (2012).

169. Ringgaard, S., van Zon, J., Howard, M. & Gerdes, K. Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19369–19374 (2009).

170. Gerdes, K., Howard, M. & Szardenings, F. Pushing and pulling in prokaryotic DNA segregation. *Cell* **141,** 927–42 (2010).

171. Lee, P. S. & Grossman, A. D. The chromosome partitioning proteins Soj (ParA) and Spo0J (ParB) contribute to accurate chromosome partitioning, separation of replicated sister origins, and regulation of replication initiation in *Bacillus subtilis*. *Mol. Microbiol.* **60,** 853–869 (2006).

172. Sullivan, N. L., Marquis, K. A. & Rudner, D. Z. Recruitment of SMC by ParB-parS Organizes the Origin Region and Promotes Efficient Chromosome Segregation. *Cell* **137,** 697–707 (2009).

173. Gruber, S. & Errington, J. Recruitment of Condensin to Replication Origin Regions by ParB/SpoOJ Promotes Chromosome Segregation in *B. subtilis*. *Cell* **137,** 685–696 (2009).

174. Errington, J. Regulation of endospore formation in *Bacillus subtilis*. *Nat. Rev. Microbiol.* **1,** 117–26 (2003).

175. Higgins, D. & Dworkin, J. Recent progress in *Bacillus subtilis* sporulation. *FEMS Microbiol. Rev.* **36,** 131–48 (2012).

176. Kay, D. & Warren, S. C. Sporulation in *Bacillus subtilis*. Morphological changes. *Biochem. J.* **109,** 819–24 (1968).

177. Jiang, M., Shao, W., Perego, M. & Hoch, J. A. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol. Microbiol.* **38,** 535–42 (2000).

178. Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S. & Losick, R. The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* **50,** 1683–1701 (2003).

179. Veening, J.-W., Murray, H. & Errington, J. A mechanism for cell cycle regulation of sporulation initiation in *Bacillus subtilis*. *Genes Dev.* **23,** 1959–1970 (2009).

180. Mandelstam, J. & Higgs, S. A. Induction of sporulation during synchronized chromosome replication in *Bacillus subtilis*. *J. Bacteriol.* **120,** 38–42 (1974).

181. Dunn, B. G., Jeffs, P. & Ma, N. H. The Relationship Between DNA Replication and the Induction of Sporulation in *Bacillus subtilis*. 189–195 (1978).

182. Narula, J., Kuchina, A., Lee, D. D., Fujita, M., Süel, G. M. & Igoshin, O. A. Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. *Cell* **162,** 328–337 (2015).

183. Fujita, M., González-pastor, J. E., Gonza, E. & Losick, R. High- and Low-Threshold Genes in the Spo0A Regulon of *Bacillus subtilis*. **187,** (2005).

184. Fujita, M. & Losick, R. Evidence that entry into sporulation in *Bacillus subtilis* is governed by a gradual increase in the level and activity of the master regulator Spo0A. *Genes Dev.* **19,** 2236–2244 (2005).

185. Burkholder, W. F., Kurtser, I. & Grossman, A. D. Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. *Cell* **104,** 269–79 (2001).

186. Bick, M. J., Lamour, V., Rajashankar, K. R., Gordiyenko, Y., Robinson, C. V. & Darst, S. A. How to Switch Off a Histidine Kinase: Crystal Structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *J. Mol. Biol.* **386,** 163–177 (2009).

187. Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trewhella, J. & Langley, D. B. The Structure of the KinA-Sda Complex Suggests an Allosteric Mechanism of Histidine Kinase Inhibition. *J. Mol. Biol.* **368,** 407–420 (2007).

188. Ruvolo, M. V., Mach, K. E. & Burkholder, W. F. Proteolysis of the replication checkpoint protein Sda is necessary for the efficient initiation of sporulation after transient replication stress in *Bacillus subtilis*. *Mol. Microbiol.* **60,** 1490–1508 (2006).

189. Levine, J. H., Fontes, M. E., Dworkin, J. & Elowitz, M. B. Pulsed feedback defers cellular differentiation. *PLoS Biol.* **10,** (2012).

190. Chapman, J. W. & Piggot, P. J. Analysis of the inhibition of sporulation of *Bacillus subtilis* caused by increasing the number of copies of the spo0F gene. *J. Gen. Microbiol.* **133,** 2079–2088 (1987).

191. Grimshaw, C. E., Huang, S., Hanstein, C. G., Strauch, M. A., Burbulys, D., Wang, L., Hoch, J. a. & Whiteley, J. M. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in *Bacillus subtilis*. *Biochemistry* **37,** 1365–1375 (1998).

192. Rahn-Lee, L., Gorbatyuk, B., Skovgaard, O. & Losick, R. The conserved sporulation protein YneE inhibits DNA replication in *Bacillus subtilis*. *J. Bacteriol.* **191,** 3736–9 (2009).

193. Wagner, J. K., Marquis, K. A. & Rudner, D. Z. SirA enforces diploidy by inhibiting the replication initiator DnaA during spore formation in *Bacillus subtilis*. *Mol. Microbiol.* **73,** 963–74 (2009).

194. Boonstra, M., de Jong, I. G., Scholefield, G., Murray, H., Kuipers, O. P. & Veening, J.-W. W. Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*. *Mol. Microbiol.* **87,** 925–938 (2013).

195. Castilla-Llorente, V., Muñoz-Espín, D., Villar, L., Salas, M. & Meijer, W. J. J. Spo0A, the key transcriptional regulator for entrance into sporulation, is an inhibitor of DNA replication. *EMBO J.* **25,** 3890–9 (2006).

196. Leonard, A. C. & Grimwade, J. E. Regulating DnaA complex assembly: it is time to fill the gaps. *Curr. Opin. Microbiol.* **13,** 766–72 (2010).

197. Kurokawa, K., Mizumura, H., Takaki, T., Ishii, Y., Ichihashi, N., Lee, B. L. & Sekimizu, K. Rapid exchange of bound ADP on the *Staphylococcus aureus* replication initiation protein DnaA. *J. Biol. Chem.* **284,** 34201–34210 (2009).

198. Grahl, A. Expression, Purification and Crystallisation of the DnaA-Interacting Protein and Sporulation Factor SirA from *Bacillus subtilis*. (University of Jena, 2011).

199. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530 (1997).

200. Colledge, V. L., Fogg, M. J., Levdikov, V. M., Leech, A., Dodson, E. J. & Wilkinson, A. J. Structure and organisation of SinR, the master regulator of biofilm formation in *Bacillus subtilis*. *J. Mol. Biol.* **411,** 597–613 (2011).

201. Winter, G. Xia2 : an Expert System for Macromolecular Crystallography Data Reduction. *J. Appl. Crystallogr.* **43,** 186–190 (2009).

202. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–32 (2010).

203. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D. Biol. Crystallogr.* **62,** 72–82 (2006).

204. Pannu, N. S., Waterreus, W. J., Skubák, P., Sikharulidze, I., Abrahams, J. P. & de Graaff, R. A. G. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. D. Biol. Crystallogr.* **67,** 331–7 (2011).

205. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

206. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A.* **64,** 112–22 (2008).

207. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D. Biol. Crystallogr.* **52,** 30–42 (1996).

208. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 470–8 (2010).

209. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D. Biol. Crystallogr.* **62,** 1002–11 (2006).

210. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D. Biol. Crystallogr.* **53,** 240–55 (1997).

211. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60,** 2126–32 (2004).

212. Vagin, A. & Teplyakov, A. MOLREP : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).

213. Leslie, A. G. W. & Powell, H. R. *Processing diffraction data with MOSFLM*. *Evolving methods for macromolecular Crystallography* (2007).

214. Jameson, K. H., Rostami, N., Fogg, M. J., Turkenburg, J. P., Grahl, A., Murray, H. & Wilkinson, A. J. Structure and interactions of the *Bacillus subtilis* sporulation inhibitor of DNA replication, SirA, with domain I of DnaA. *Mol. Microbiol.* **93,** 975–91 (2014).

215. Soufo, C. D., Soufo, H. J. D., Noirot-Gros, M.-F. F., Steindorf, A., Noirot, P. & Graumann, P. L. Cell-Cycle-Dependent Spatial Sequestration of the DnaA Replication Initiator Protein in *Bacillus subtilis*. *Dev. Cell* **15,** 935–41 (2008).

216. Goranov, A. I., Breier, A. M., Merrikh, H. & Grossman, A. D. YabA of *Bacillus subtilis* controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Mol. Microbiol.* **74,** 454–66 (2009).

217. Fogg, M. J. & Wilkinson, A. J. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.* **36,** 771–775 (2008).

218. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

219. PACT Crystal Screen.

220. Hampton Research INDEX Crystal Screen. at <https://hamptonresearch.com/product_detail.aspx?cid=1&sid=24&pid=5>

221. Jaenicke, R. & Böhm, G. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8,** 738–748 (1998).

222. Radestock, S. & Gohlke, H. Protein rigidity and thermophilic adaptation. *Proteins Struct. Funct. Bioinforma.* **79,** 1089–1108 (2011).

223. Deng, Y., Liu, J., Zheng, Q., Eliezer, D., Kallenbach, N. R. & Lu, M. Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. *Structure* **14,** 247–255 (2006).

224. Fujiwara, Y. & Minor, D. L. X-ray Crystal Structure of a TRPM Assembly Domain Reveals an Antiparallel Four-stranded Coiled-coil. *J. Mol. Biol.* **383,** 854–870 (2008).

225. Thépaut, M., Maiorano, D., Guichou, J. F., Augé, M. T., Dumas, C., Méchali, M. & Padilla, A. Crystal structure of the coiled-coil dimerization motif of geminin: Structural and functional insights on DNA replication regulation. *J. Mol. Biol.* **342,** 275–287 (2004).

226. Caillat, C. & Perrakis, A. Cdt1 and Geminin in DNA Replication Initiation. *Subcell. Biochem.* **62,** 71–87 (2012).

227. De Marco, V., Gillespie, P. J., Li, A., Karantzelis, N., Christodoulou, E., Klompmaker, R., van Gerwen, S., Fish, A., Petoukhov, M. V, Iliou, M. S., Lygerou, Z., Medema, R. H., Blow, J., Svergun, D. I., Taraviras, S. & Perrakis, A. Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19807–19812 (2009).

228. Moutevelis, E. & Woolfson, D. N. A Periodic Table of Coiled-Coil Protein Structures. *J. Mol. Biol.* **385,** 726–732 (2009).

229. Yadav, M. K., Leman, L. J., Price, D. J., Brooks III, C. L., Stout, D. & Ghadiri, M. R. Coiled-Coils at the Edge of Configurational Heterogeneity. Structural Analyses of Parallel and Antiparallel Homotetrameric Coiled-Coils Reveal Configurational Sensitivity to a Single Solvent-Exposed Amino Acid Substitution. **27,** 417–428 (2009).

230. Rämisch, S., Lizatović, R. & André, I. Exploring alternate states and oligomerization preferences of coiled-coils by de novo structure modeling. *Proteins Struct. Funct. Bioinforma.* **83,** 235–247 (2015).

231. Weigel, C., Schmidt, A., Seitz, H., Tüngler, D., Welzeck, M. & Messer, W. The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Mol. Microbiol.* **34,** 53–66 (1999).

232. Kwan, A. H., Mobli, M., Gooley, P. R., King, G. F. & Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. **278,** 687–703 (2011).

233. Bieri, M., Kwan, A. H., Mobli, M., King, G. F., MacKay, J. P. & Gooley, P. R. Macromolecular NMR spectroscopy for the non-spectroscopist: Beyond macromolecular solution structure determination. *FEBS J.* **278,** 704–715 (2011).

Bibliography

The peer-reviewed publication listed below were published (or submitted for review during the course of this thesis.

Jameson, K. H., Rostami, N., Fogg, M. J., Turkenburg, J. P., Grahl, A., Murray, H. & Wilkinson, A. J. Structure and interactions of the *Bacillus subtilis* sporulation inhibitor of DNA replication, SirA, with domain I of DnaA. *Mol. Microbiol.* **93,** 975–91 (2014).

Felicori, L., Jameson, K. H., Roblin, P., Fogg, M. J. Garcia, T., Ventroux, M., Cherrier, M.V., Bazin, A., Noirot, P., Wilkinson, A. J., Molina, F., Terradot, L., Noirot-Gros, M-F. Tetramerization and interdomain flexibility of the replication initiation controller YabA enables simultaneous binding to multiple partners. *Nucleic Acids Research*, **44**, 449–463 (2016).

Molecular Microbiology (2014) 93(5), 975–991 ■
doi:10.1111/mmi.12713
First published online 5 August 2014

Structure and interactions of the *Bacillus subtilis* sporulation inhibitor of DNA replication, SirA, with domain I of DnaA

Katie H. Jameson,[1†] Nadia Rostami,[2†] Mark J. Fogg,[1] Johan P. Turkenburg,[1] Anne Grahl,[1] Heath Murray[2*] and Anthony J. Wilkinson[1]**
[1]*Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK.*
[2]*Centre for Bacterial Cell Biology, Institute for Cell & Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4AX, UK.*

Summary

Chromosome copy number in cells is controlled so that the frequency of initiation of DNA replication matches that of cell division. In bacteria, this is achieved through regulation of the interaction between the initiator protein DnaA and specific DNA elements arrayed at the origin of replication. DnaA assembles at the origin and promotes DNA unwinding and the assembly of a replication initiation complex. SirA is a DnaA-interacting protein that inhibits initiation of replication in diploid *Bacillus subtilis* cells committed to the developmental pathway leading to formation of a dormant spore. Here we present the crystal structure of SirA in complex with the N-terminal domain of DnaA revealing a heterodimeric complex. The interacting surfaces of both proteins are α-helical with predominantly apolar side-chains packing in a hydrophobic interface. Site-directed mutagenesis experiments confirm the importance of this interface for the interaction of the two proteins *in vitro* and *in vivo*. Localization of GFP–SirA indicates that the protein accumulates at the replisome in sporulating cells, likely through a direct interaction with DnaA. The SirA interacting surface of DnaA corresponds closely to the HobA-interacting surface of DnaA from *Helicobacter pylori* even though HobA is an activator of DnaA and SirA is an inhibitor.

Accepted 8 July, 2014. For correspondence. *E-mail heath.murray@newcastle.ac.uk; Tel. (+44) 191 222 5234, Fax (+44) 191 222 7424. **E-mail tony.wilkinson@york.ac.uk; Tel. (+44) 1904 326261; Fax (+44) 1904 328266. †Both authors contributed equally to this work.

Introduction

Across the kingdoms of life, DNA replication is tightly regulated to ensure co-ordination with cell growth and development. Failure to maintain and control chromosome copy number is frequently associated with disease or cell death. Regulation of DNA replication is mainly exerted at the initiation step when an initiator protein binds to the origin of replication and promotes the assembly of a nucleoprotein complex from which replication forks diverge.

In prokaryotes, the DNA replication initiator protein is DnaA. In its ATP-bound state, DnaA assembles at the origin of replication, *oriC*, by binding to a number of 9 bp recognition sequences termed DnaA-boxes (Yoshikawa and Ogasawara, 1991). Recruitment of DnaA to *oriC* is believed to generate a helical oligomer of DNA-bound DnaA that promotes duplex unwinding at an AT-rich region within the origin termed the DNA unwinding element (Duderstadt *et al.*, 2011). DnaA is a member of the AAA+ (ATPases associated with diverse cellular activities) protein superfamily and is made up of four distinct domains (Kaguni, 2006). Domain I is known to have a number of interaction partners, including replication regulators and DNA helicase (Seitz *et al.*, 2000; Abe *et al.*, 2007). Domain II is thought to be a flexible linker, that may allow nuances in regulatory control (Molt *et al.*, 2009). Domain III binds and hydrolyses ATP, mediates DnaA oligomerization (Erzberger *et al.*, 2006) and binds single stranded DNA thus aiding duplex unwinding (Duderstadt *et al.*, 2011). Domain IV binds double stranded DNA, interacting with the DnaA-box motifs (Fujikawa *et al.*, 2003). The binding of a threshold level of DnaA–ATP at *oriC* leads to duplex unwinding and the recruitment of other initiation proteins (Leonard and Grimwade, 2011).

Although DnaA is conserved in bacteria, many of the other initiation components are not. This is exemplified by differences in replication initiation observed between *E. coli* and *B. subtilis*, organisms which provide our most thorough understanding of the control of replication initiation in Gram-negative and Gram-positive bacteria, respectively. An important early step following duplex unwinding in both organisms, is the recruitment of a DNA helicase (*Ec* DnaB/*Bsu* DnaC) to the origin where it is loaded onto the DNA by a helicase loader (*Ec* DnaC/*Bsu*

DnaI). This is followed by the binding of the primase, DnaG. Curiously, replication initiation in *B. subtilis* requires two additional proteins, DnaD and DnaB, neither of which is present in *E. coli* (Ishigo-oka *et al.*, 2001; Rokop *et al.*, 2004) – it should be noted that *B. subtilis* DnaB is unrelated to *E. coli* DnaB. DnaD is recruited to the origin through an interaction with DnaA, moreover DnaD binding has been shown to be accompanied by pronounced bending of origin DNA. Unwinding of the duplex appears to be assisted by the formation of DnaD scaffolds, which may provide further anchorage points for DnaA (Zhang *et al.*, 2008). DnaB is subsequently recruited, appearing to play a role in helicase loading along with the helicase loader DnaI (Velten *et al.*, 2003).

During rapid growth, bacteria reinitiate replication before the previous cycle of replication is complete, giving rise to multiple replication forks. Daughter cells thus inherit chromosomes that are already undergoing replication. This emphasizes the need for precise mechanisms to control the frequency of initiation of DNA replication so that it matches the frequency of cell division and nutrient availability. Regulatory mechanisms take the form of proteins and *cis*-acting DNA elements which typically act on DnaA or *oriC*. Protein regulators vary between genera. Notably, *B. subtilis* and *E. coli* employ a range of replication regulators that lack known homologues in the other species (*E. coli* Hda, DiaA, SeqA; *B. subtilis* YabA, Soj, SirA, Spo0A) (Katayama *et al.*, 2010; Briggs *et al.*, 2012), reflecting differences in their mechanisms of regulatory control. For example, in *E. coli* Hda inactivates DnaA by promoting ATP hydrolysis in DnaA–ATP, whereas its functional homologue in *B. subtilis*, YabA, acts by both sequestering DnaA at the replication fork, and by inhibiting DnaA oligomerization (Scholefield and Murray, 2013) (Soufo *et al.*, 2008).

An additional specialized DNA replication checkpoint exists in Gram-positive bacteria of the genera *Bacilli* and *Clostridia* during sporulation under conditions of nutrient depletion. Sporulation begins with an asymmetric cell division producing genetically identical daughter cells of unequal size. The larger mother cell and the smaller forespore must each inherit a complete copy of the genome in order to drive the developmental program. Therefore, DNA replication is regulated and monitored at the onset of sporulation to ensure that two intact copies of the chromosome are present in the pre-divisional cell. DnaA contributes to this DNA replication check-point through its role as a transcription factor. In *B. subtilis*, DnaA activates the expression of *sda*, which encodes an inhibitor of the sporulation sensor kinases, KinA and KinB. Sda thus serves to delay sporulation by limiting the phosphorylation of the master sporulation response regulator, Spo0A. Sda is an intrinsically unstable protein whose levels fluctuate with the cell cycle, reaching a minimum immediately prior to the initiation of a new round of DNA replication (Burkholder *et al.*, 2001; Veening *et al.*, 2009). This creates a small 'window of opportunity', for a threshold concentration of Spo0A~P to be achieved and for sporulation to commence.

SirA, a protein produced under Spo0A~P regulation, has been identified as an inhibitor of DNA replication that plays a specific role in preventing replication re-initiation in cells committed to sporulation (Rahn-Lee *et al.*, 2009). Although single deletion mutants of *sirA*, like those of *sda*, display only mild phenotypes, under conditions of nutrient depletion leading to sporulation *sda*/*sirA* double mutants are severely impaired in chromosome copy number control, indicating a shared role in controlling DNA replication (Veening *et al.*, 2009). Artificial induction of expression of *sirA* in vegetatively growing *B. subtilis* blocks replication and causes cell death in a DnaA-dependent manner (Wagner *et al.*, 2009). Cells artificially induced to sporulate under conditions of rapid growth undergo a marked decrease in chromosome copy number which is partially relieved by deletion of *sirA* (Rahn-Lee *et al.*, 2009). These experiments imply that SirA is an inhibitor of DNA replication during sporulation that acts by binding to DnaA. Furthermore, the SirA binding determinants of DnaA have been mapped to its N-terminal domain, $DnaA^{DI}$ (Rahn-Lee *et al.*, 2011).

SirA has no significant sequence similarity to other proteins besides orthologues in *Bacilli*. Here, we have solved the structure of SirA from *B. subtilis* in complex with $DnaA^{DI}$ providing the first structure of a DnaA domain in an inhibitory complex. The structure reveals a heterodimer with an α-helical interface. The importance of this interface for SirA–DnaA interaction *in vitro* and *in vivo* has been demonstrated by analysis of a panel of site-directed mutants. Furthermore, localization of GFP–SirA within sporulating cells indicates that the protein accumulates at the replisome, likely through a direct interaction with DnaA. Interestingly, the structure reveals a conserved binding site on $DnaA^{DI}$ that is used by DnaA in *H. pylori* and *E. coli* to bind the replication activators HobA and DiaA, respectively, implying this surface is functionally important in DNA replication initiation.

Results

Coexpression with $DnaA^{DI}$ confers solubility on recombinant SirA

Attempts to produce recombinant SirA in *E. coli* yielded disappointing results; although SirA could be produced at high levels in a number of *E. coli* expression strains, the protein always partitioned into the insoluble fraction upon cell lysis. Variations in growth conditions or lysis procedures failed to overcome the insolubility of SirA. Following

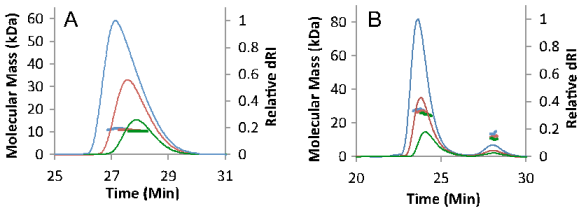


Fig. 1. Molecular mass measured from SEC-MALLS analysis. In A and B, the thinner lines trace the differential refractive index of the eluate from a Superdex 10/30 S75 column as a function of time. The thicker lines represent the weight average molecular weight of the species in the eluate, calculated from refractive index and light-scattering measurements.
A. Overlay of chromatograms for DnaA^{DI} at 3 concentrations: 1 mg ml^{-1} (green), 2.5 mg ml^{-1} (red) and 5 mg ml^{-1} (blue), revealing species of mass 11 kDa indicating that DnaA^{DI} is a monomer.
B. Overlay of chromatograms for SirA–DnaA^{DI} at 3 concentrations: 0.5 mg ml^{-1} (green), 1.0 mg ml^{-1} (red) and 2.5 mg ml^{-1} (blue). The derived M_w values for the principal species are 25–28 kDa indicating that the SirA–DnaA^{DI} complex is a 1:1 heterodimer. There is evidently, excess/dissociated DnaA^{DI} giving rise to the minor peak eluting at ~28 min.

a report that the determinants of SirA binding to DnaA reside in its N-terminal domain (DnaA^{DI}) (Rahn-Lee *et al.*, 2011), we generated a coexpression construct in which sequences encoding DnaA^{DI} (fused to an N-terminal cleavable polyhistidine-tag) and SirA were expressed from separate promoters on the same vector. Strikingly, this coexpression strategy led to the appearance of SirA in the soluble fraction following cell lysis, presumably the result of its interaction with DnaA^{DI}. Consistent with the notion that recombinant SirA and DnaA^{DI} were forming a complex, the two proteins co-purified. SirA was retained on an immobilized nickel affinity column with His-tagged DnaA^{DI} and co-eluted with the latter. Moreover, the two proteins eluted together following gel filtration chromatography.

SirA and DnaA^{DI} form a heterodimer

The stoichiometry of the SirA–DnaA^{DI} complex was determined using size-exclusion chromatography with multi-angle laser light scattering (SEC-MALLS). In these experiments, samples are fractionated on a gel-filtration column and the absorbance at 280 nm and the refractive index of the eluate are monitored together with the multi-angle laser light scattering of the sample. This enables the weight average molecular weight (M_w) of species in the eluate to be calculated continuously. Samples of DnaA^{DI} and SirA–DnaA^{DI} were analysed at a series of protein concentrations. As shown in Fig. 1A, DnaA^{DI} elutes from the size exclusion column as a single A_{280} peak at ~ 27.5 min and has an experimentally determined molecular mass of ~ 11 kDa. This suggests that DnaA^{DI} from *B. subtilis* is a monomer (calculated molecular mass = 9.7 kDa) in contrast to *E. coli* DnaA^{DI} which is reported to form dimers under similar conditions (Abe *et al.*, 2007).

The SirA–DnaA^{DI} elution profile has two peaks: a major peak at ~ 24 min and a minor peak at ~ 28 min (Fig. 1B). The minor peak comprises 8–12% of the total protein content and the analysis above suggests this is monomeric DnaA^{DI}. The major peak (88–92% of the total protein content), which corresponds to a molecular mass of 25–28 kDa, is consistent with a 1:1 heterodimer of SirA : DnaA^{DI} (calculated molecular mass = 28.7 kDa). A discernible shift in the relative sizes of the major and minor peaks occurs as the protein concentration changes in these experiments. There is an increasing area under the minor peak as the protein concentration is lowered, accompanied by a small shift in the elution time associated with the major peak and by a decrease in its associated molecular mass at lower concentrations. This is consistent with increasing complex dissociation at lower protein concentrations. It is clear, however, that under these experimental conditions the SirA–DnaA^{DI} heterodimer is the predominant species.

The crystal structure of the SirA–DnaA^{DI} complex

The crystal structure of SirA–DnaA^{DI} was solved to 1.7 Å resolution using single anomalous dispersion (SAD) phasing techniques (Table 1). Native and selenomethionine-substituted SirA–DnaA^{DI} crystals grew under different conditions from PEG 3350 containing solutions (see *Experimental procedures*). Although both crystals belong to space group $P2_1$, the crystals are different (Table 1). SeMet-derivative crystals contain one complex per asymmetric unit (one molecule of SirA, one molecule of DnaA^{DI}) while the asymmetric unit of the native crystals contains two complexes. The SeMet structure was solved and partially refined to allow solution of the native structure by molecular replacement. The refined model

Table 1. X-ray data collection and refinement statistics.

	SirA–DnaADI SeMet	SirA–DnaADI Native I
Data collection		
X-ray source	DLS, i24	DLS, i03
Wavelength (Å)	0.9789	0.9763
Resolution range (Å)	40.8–2.09	62.76–1.65
Space group	$P2_1$	$P2_1$
Unit cell parameters		
a, b, c (Å)	51.35, 35.63, 63.27	77.29, 34.69, 84.74
$\alpha = \gamma$, β (°)	90, 92.77	90, 102.09
No. of unique reflections[a]	13 549/989	52 893/2598
Completeness (%)[a]	98.7/99.1	98.9/99.6
Redundancy[a]	3.2/3.3	2.8/2.8
$I/\sigma(I)$[a]	11.9/1.9	12.5/1.9
R_{merge}[b] (%)[a]	7.4/79.9	3.9/45.8
Refinement and model statistics		
Resolution range (Å)		62.84–1.65
R-factor[c] (R_{free}[d])		13.1 (19.7)
Reflections (working/R_{free})		50172/2706
Outer-shell/high resolution range		1.69–1.65
Outer-shell[e]/high resolution R-factor[c] (R_{free})[d]		19.0 (27.9)
Outer-shell/high resolution reflections (working/free)		3677/214
Molecules per asymmetric unit		4
rmsd from ideal geometry[f]		
Bond lengths (Å)		0.017
Bond angles (°)		1.8
Average B-factor (Å2)		27.8
Ramachandran plot[g]		98.16/0.92/0.92

a. The first number refers to the overall data set, the second refers the outer resolution shells; Native: 1.68-1.65 Å; SeMet: 2.15-2.09 Å.
b. $R_{merge} = \sum_{hkl}\sum_i |I_i - \langle I \rangle| / \sum_{hkl}\sum_i \langle I \rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I \rangle$ is the statistically weighted average reflection intensity.
c. R-factor $= \sum||F_o| - |F_c||/\sum|F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes respectively.
d. R-free is the R-factor calculated with 5% of the reflections chosen at random and omitted from refinement.
e. Outer shell for refinement corresponds to 1.69–1.65 Å.
f. Root-mean-square deviation of bond lengths and bond angles from ideal geometry.
g. Percentage of residues in most-favoured/allowed/disallowed regions of the Ramachandran plot.

encompasses residues 2–141 of SirA (Met^1 and residues 142–148 being disordered) in both molecules A and C. Residues 1–81 of $DnaA^{DI}$ are defined in molecules B and D (the C-terminal Gln^{82} being disordered). A vestigial N-terminal Gly-Pro-Ala sequence inherited from the $DnaA^{DI}$ polyhistidine-tag can be seen in molecule D, and an additional N-terminal Ala is visible in molecule B. SirA–$DnaA^{DI}$ is seen as a heterodimer, with molecules A (SirA) and B ($DnaA^{DI}$) forming one heterodimer and molecules C and D forming the other. The electron density maps reveal a 2-mercaptoethanol (BME) molecule linked through a disulphide bond to Cys^{125} of both SirA chains in the asymmetric unit (Fig. S1). The presence of this adduct explains species observed in the electrospray ionization mass spectrum of SirA of 18 776 Da, 79 Da larger than that of SirA; 18 697 Da. BME was present during the purification steps as it was found to improve the solubility of the $DnaA^{DI}$–SirA complex.

SirA consists of a single globular domain comprising seven β-strands and five α-helices in the order β1-α1-α2-α3-α4-β2-β3-β4-β5-α5-β6-β7 (Fig. 2A and C). The SirA fold consists of a central seven-stranded twisted β-sheet with strand order β2-β3-β4-β5-β1-β6-β7, flanked on either side by two α-helical regions, one comprising helices α1, α2 and α3 and the other of helices α4 and α5. Comparative analysis of the SirA chain topology using PDBeFold identified the kinase associated domain 1 from the protein KCCP4 (PDB entry 3osm) as the highest Q-scoring hit with 79 Cα atoms overlaying with a positional root mean squared deviation of 2.4 Å. This domain has been identified as a membrane association domain that binds acidic phospholipids (Moravcevic *et al.*, 2010). The region of structural similarity spans residues 2–9 and 53–124 covering the β1-α4-β2-β3-β4-β5-α5 segment. The other highest scoring matches, the core domain of the human ribosomal protein L10 (2pa2) and the yeast mitochondrial protein frataxin (4ec2), exhibit structural similarity to the same region of SirA.

$DnaA^{DI}$ from *B. subtilis* (Fig. 2B) consists of four alpha helices and three beta strands in the order α1-α2-β1-β2-α3-α4-β3 with a β-sheet topology of β1-β2-β3. It has a closely similar topology to the previously determined structures of the corresponding domains of DnaA from *E. coli* (Abe *et al.*, 2007), *M. genitalium* (Lowery *et al.*, 2007)

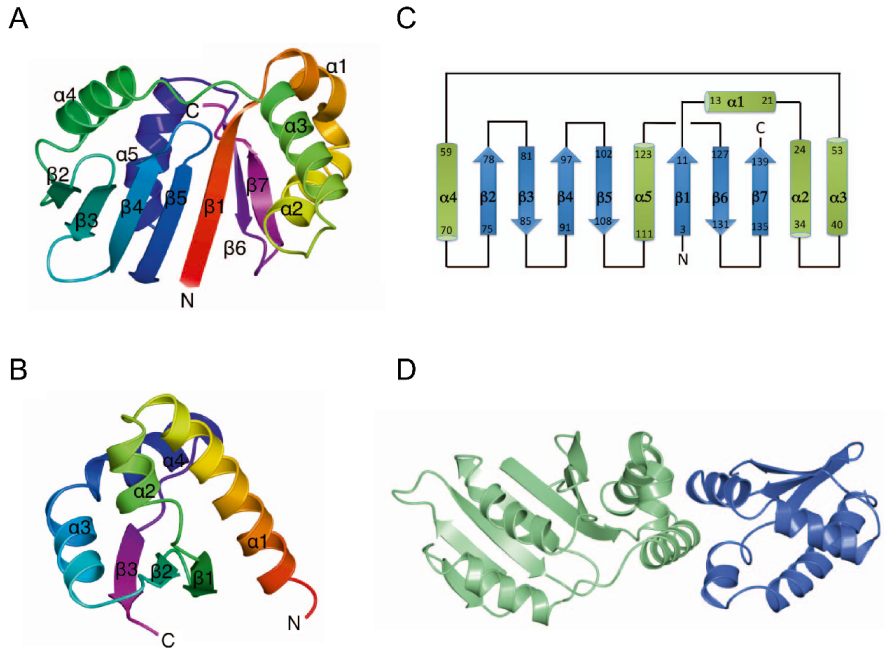


Fig. 2. The structure of the SirA–DnaA^{DI} complex.
A and B. Ribbon diagram of the SirA (A) and DnaA^{DI} (B) chains from the complex. In each case, the chain is colour-ramped from its N-terminus (red) to the C-terminus (magenta) and the secondary structure elements are labelled. These and subsequent structure figures were produced using the program CCP4mg (McNicholas *et al.*, 2011).
C. The polypeptide chain topology in SirA.
D. Ribbon diagram of the SirA–DnaA^{DI} complex with SirA shown in light green and DnaA^{DI} shown in blue.

and *H. pylori* (Natrajan *et al.*, 2009). Thus, it shares the K homology domain motif that is widespread in single-stranded nucleic acid binding proteins.

The SirA–DnaA^{DI} interface

The binding of SirA and DnaA^{DI} is mediated by helices α1, α2 and α3 of SirA and α2 and α3 of DnaA^{DI} (Figs 2D and 3A). As an extensively α-helical interface the interactions of the two proteins are dominated by side-chain–side-chain contacts (Fig. 3A). Seventeen residues on each chain contribute to the interface which constitutes 12% and 8% of the surface areas of DnaA^{DI} and SirA, respectively. In the complex, 1240 Å^2 of otherwise accessible surface area is buried in the interface. This buried surface area is at the lower end of the range observed in non-obligate dimeric protein-protein complexes (Janin *et al.*, 2008). In the SirA binding surface of DnaA^{DI}, Thr26, Trp27 and Phe49 contribute to the core of the interface with residues Pro22, Ser23, Glu25, Ser30, Pro46, Asn47, Glu48, Asp52, Ser56 and Trp53 prominent in the rim. As shown in Fig. 3C, these residues are very strongly conserved in a set of DnaA orthologues from endospore-forming bacteria. On the corresponding DnaA^{DI} binding surface of SirA, Phe14, Tyr18, Gln48 and Ile52 contribute to the core and Glu13, His17, Val24, Leu28, Gln41, Met44, Lys47 and Tyr51 are prominent in the rim. Again core residues are well conserved with some variation observed in the residues constituting the rim (Fig. 3D).

The SirA binding determinants of DnaA have previously been explored using genetic approaches. Induction of *sirA* expression under nutrient rich conditions inhibits growth of *B. subtilis*. Four strains were identified which

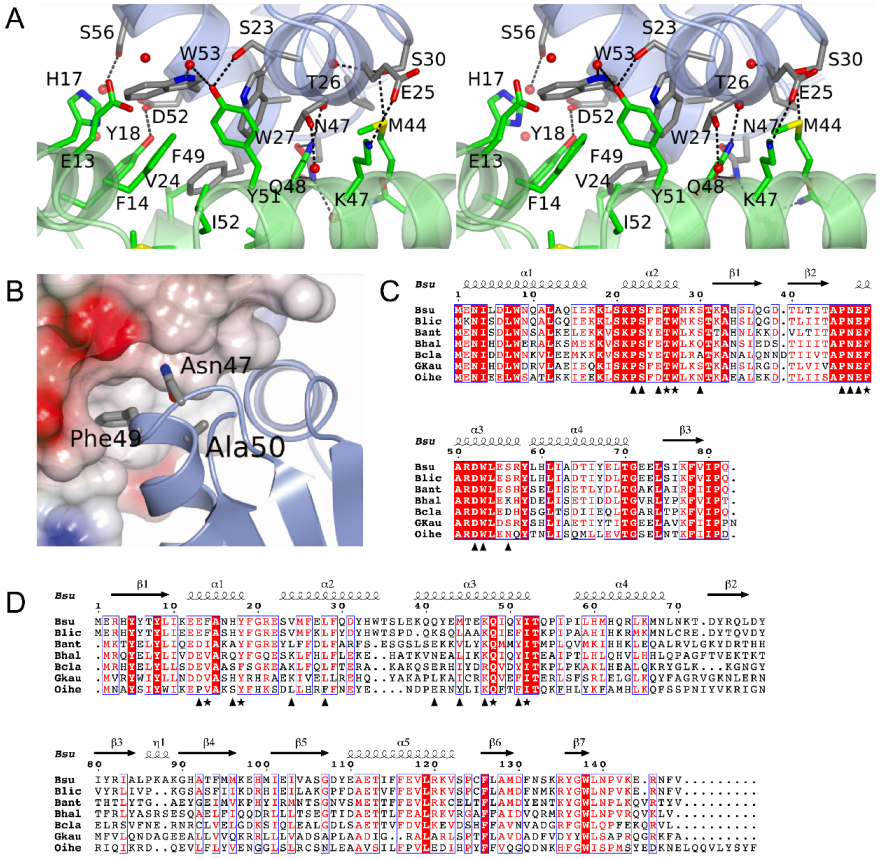


Fig. 3. The SirA–DnaA^{DI} interface.

A. Stereoview of the complex between DnaA^{DI} (chain D) and SirA (chain C) represented as light blue and light green ribbons, respectively. Side-chains of labelled residues are displayed in cylinder format and coloured by atom type with nitrogen (blue), oxygen (red), sulphur (yellow) and carbons coloured in grey for DnaA^{DI} and green for SirA. Water molecules are represented as red spheres, and polar interactions are denoted by dashed lines.

B. Mapping onto the structure of DnaA^{DI} the sites corresponding to mutations in *dnaA* that allow growth of *B. subtilis* even when *sirA* is being overexpressed (Rahn-Lee *et al.*, 2011). SirA is rendered as a partially transparent electrostatic surface and DnaA^{DI} as a ribbon with the side-chains of residues Asn47, Phe49 and Ala50 in cylinder format.

C and D. Alignment of the sequences of orthologues of DnaA^{DI} (C) and SirA (D) from selected *Bacillus* species; Bsu, *B. subtilis*; Blic, *B. licheniformis*; Bant, *B. anthracis*; Bhal, *B. halodurans*; Bcla, *B. clausii*; Gkau, *Geobacillus klaustophilus*; Oihe, *Oceanobacillus iheyensis*. Symbols below the alignments indicate interfacial residues in the respective molecules contributing to the core (asterisks) and the rim (triangles). Secondary structure elements and residue numberings are displayed above the alignment. The images were created using ESPript (Gouet, 2003).

harbour mutations in *dnaA* that were able to suppress this slow growth phenotype accompanying *sirA* induction (Rahn-Lee *et al.*, 2011). Analysis of the sequence of these *dnaA* alleles revealed point mutations giving rise to Asn47Asp, Phe49Tyr, Ala50Val and Ala50Thr substitutions (Rahn-Lee *et al.*, 2011). Yeast two hybrid analysis confirmed that mutations at these residues prevent DnaA from interacting with SirA, suggesting they affect SirA–$DnaA^{DI}$ complex formation (Rahn-Lee *et al.*, 2011). The SirA–$DnaA^{DI}$ structure reveals that Asn47 and Phe49 make direct interactions with SirA (Fig. 3A and B). Asn47 of DnaA forms a pair of hydrogen bonds with Gln48 on helix α3 of SirA, while the side-chain of Phe49 of DnaA projects into a hydrophobic pocket created by helices α1, α2 and α3 of SirA. In contrast, Ala50 does not contact SirA in the complex; instead it is buried within $DnaA^{DI}$ in such a way that it determines the structure of the interface (Fig. 3B). It is expected that mutations at this position which introduce bulkier side-chains, such as valine or threonine, will alter the structure of the interaction surface leading to lower affinity binding of SirA. In summary, the structure of SirA–$DnaA^{DI}$ confirms previous interpretations of the genetic data.

Substitutions at the SirA–$DnaA^{DI}$ interface affect the SirA–DnaA interaction in vitro

To confirm the importance of the protein-protein interface observed in the SirA–$DnaA^{DI}$ crystals, we assayed the interactions of site-directed mutants of *sirA* and *$dnaA^{DI}$ in vitro*. Mixing experiments using the purified proteins are not possible because we are unable to produce soluble SirA in the absence of coexpression with $DnaA^{DI}$. Instead we took advantage of the dependence of SirA solubility on its co-production with, and binding to, $DnaA^{DI}$ in developing a qualitative binding assay. We hypothesized that disruption of the interaction between SirA and $DnaA^{DI}$ following coexpression would reduce or abolish SirA solubility.

Site-directed mutagenesis was used to introduce alanine substitutions into the pET-YSBLIC3C-$DnaA^{DI}$ SirA coexpression vector at codons that specify the SirA–$DnaA^{DI}$ interface in the crystal structure. Three alanine substitutions were introduced into SirA: Phe14Ala, Tyr18Ala and Gln48Ala. These residues contribute 45 Å^2, 45 Å^2 and 80 Å^2 of buried surface area respectively to the interface with the phenolic hydroxyl of Tyr18 forming a charge-dipole interaction with Asp52 of DnaA and the side-chain amide of Gln48 forming a pair of hydrogen bonds with the amide of Asn47 of DnaA (Fig. 3A). A further three substitutions were introduced into $DnaA^{DI}$, these being Thr26Ala, Trp27Ala and Phe49Ala. Thr26, Trp27 and Phe49 contribute 75 Å^2, 50 Å^2 and 135 Å^2 of surface area respectively to the interface. After sequencing to confirm the presence of the mutations, the mutated plasmids were introduced into *E. coli* BL21 and expression experiments were performed. The solubility of the recombinant proteins was compared to that of the wild-type proteins by SDS-PAGE of cell fractions following lysis (Fig. S2).

These experiments show that there are reduced levels of SirA in the soluble lysate fractions (labelled S in Fig. S2) of cells producing $SirA^{F14A}$ and $SirA^{Y18A}$ and negligible levels of SirA in these fractions from cells producing $DnaA^{DI,W27A}$ and $DnaA^{DI,F49A}$. This suggests weaker binding of SirA by $DnaA^{DI}$. However, interpretation of these experiments is complicated by variability in the levels of $DnaA^{DI}$ present in these fractions. Thus, the effect of each mutation on the interaction of SirA with $DnaA^{DI}$ was further probed using a pull down assay where the soluble fraction of the cell lysate was loaded onto a Ni-affinity column. The latter was washed extensively with loading buffer and bound proteins were eluted in a buffer containing a high concentration of imidazole. pET-YSBLIC3C-$DnaA^{DI}$SirA directs expression of $DnaA^{DI}$ with a hexahistidine tag together with untagged SirA. Thus, any retention of SirA is expected to result from its interaction with the histidine-tagged $DnaA^{DI}$. The eluate (E) fractions shown in Fig. S2 were diluted to normalize to an approximately equivalent amount of $DnaA^{DI}$ and the samples again resolved by SDS-PAGE with Coomassie staining. As can be seen in Fig. 4A, the $DnaA^{DI,W27A}$ and $DnaA^{DI,F49A}$ mutations have the most striking effect, with little discernable SirA eluted from the Ni-NTA column in the high imidazole fraction. In marked contrast, $DnaA^{DI,T26A}$ supports wild-type levels of SirA recovery after the nickel pull down. For the three SirA mutant proteins, the effects are more modest. Quantification of the $DnaA^{DI}$ and SirA band intensities in Fig. 4A using the software ImageJ revealed, relative to the wild type SirA, 1.5-fold lower recovery of $SirA^{F14A}$ and $SirA^{Q48A}$ and a 2.5-fold lower recovery of $SirA^{Y18A}$.

Collectively, these results correlate with the SirA–DnaA interface in the crystal structure. Residues Phe14, Trp18 and Gln48 of SirA form contacts with $DnaA^{DI}$ which would be weakened upon truncation of these side-chains to alanine (Fig. 4B). The side-chains of residues Trp27 and Phe49 in $DnaA^{DI}$ project away from the surface of the protein into a hydrophobic groove on the SirA surface, forming extensive van der Waals contacts across the SirA–DnaA interface (Fig. 4C). The results indicate that truncation of either of these large hydrophobic residues strongly affects the SirA–DnaA interaction due to the loss of these contacts. By contrast, Thr26 of $DnaA^{DI}$ binds at the edge of a hydrophobic groove and although it is largely buried, its hydroxyl is able to form a hydrogen bond to a recessed water molecule on the protein surface. Moreover, substitution of threonine with alanine is a less

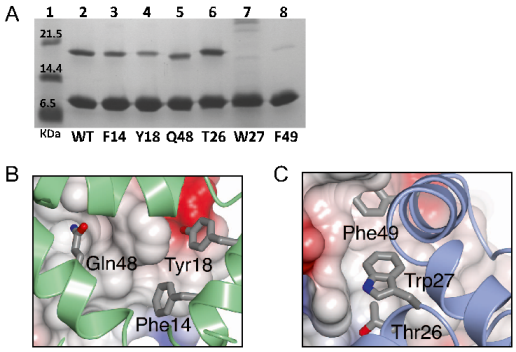


Fig. 4. The SirA–DnaA^{DI} interface analysed by site-directed mutagenesis.
A. SDS-PAGE. Cultures of cells harbouring plasmids encoding wild type and alanine-substituted variants of His-tagged DnaA^{DI} and SirA were grown. Soluble cell lysates were prepared and loaded onto a Ni-NTA column. High imidazole eluate fractions were collected for analysis. Samples of the eluate fractions containing approximately normalized levels of DnaA^{DI} were loaded so that the efficiency of SirA pull-down could be compared. Lane 1 contains molecular weight markers. Lane 2: wild type DnaA^{DI}–SirA. Lanes 3–5: Native DnaA^{DI} and the SirA variants, loaded as follows; Lane 3: SirA(F14A), Lane 4: SirA(Y18A), Lane 5: SirA(Q48A). Lanes 6–8: Samples of native SirA and the DnaA^{DI} variants, loaded as follows; Lane 6: DnaA^{DI}(T26A), Lane 7: DnaA^{DI}(W27A), Lanes 8: DnaA^{DI}(F49A).
B and C. Core residues from the DnaA ^{DI}-interacting surface of SirA (B) and the SirA interacting surface of DnaA^{DI} (C) which were sites of alanine substitution. (B) DnaA^{DI} is shown as an electrostatic surface with SirA represented as a green ribbon with the side-chains of F14, Y18 and Q48 displayed as cylinders. (C) SirA is shown as an electrostatic surface with DnaA^{DI} represented as a blue ribbon with the side-chains of T26, W27 and F49 displayed as cylinders.

drastic change, evidently allowing DnaA$^{DI,\ T26A}$ to maintain its interaction with SirA.

GFP–SirA colocalizes with the replisome in sporulating cells

The biochemical analysis of SirA–DnaA^{DI} complex formation provides strong evidence that the interaction observed in the crystal structure is also formed between the two proteins in solution. To study the physiological relevance of the proposed SirA–DnaA interface, SirA activity was examined *in vivo*. Visualization of SirA was achieved by replacing the endogenous gene with *gfp–sirA* (expressed from its native transcriptional and translational regulatory sequences; Fig. 5A), inducing cells to sporulate by nutrient deprivation, and detecting the localization of GFP–SirA using epifluorescence microscopy (Fig. 5B). A time-course experiment showed that GFP–SirA foci began to appear approximately 90 min after cells were resuspended in starvation medium. By 150 min, the number of cells containing a GFP–SirA focus reached its maximum (~ 20%; Fig. 5C). In the majority of cases (> 80%) GFP–SirA foci were located near mid-cell. Previous studies suggested that SirA inhibits new rounds of DNA replication by inhibiting the binding of DnaA to the origin of replication (Wagner *et al.*, 2009). However, the mid-cell localization of GFP–SirA foci is contrary to origin positioning in sporulating cells where the two origins are positioned towards the cell poles, suggesting that SirA was not accumulating at *oriC*.

In order to further investigate GFP–SirA localization we constructed a strain that allowed visualization of both origin regions and SirA. An array of *tet* operators was inserted near *oriC* and the Tet repressor was fused to a red fluorescent protein (TetR-mCherry); interaction of TetR-mCherry with the *tetO* array produces a fluorescent focus near each origin of replication. Cells were induced to sporulate by nutrient deprivation and the localization of GFP–SirA was determined in respect to the origin regions. In the majority of cells with a GFP–SirA focus there was no colocalization of SirA with *oriC* (78% with non-overlapping signals; Fig. 5D). Only 8% of the GFP–SirA foci appeared to colocalize with the origin regions, with the remaining 14% of cells containing a GFP–SirA focus that partially overlapped with the origin marker (Fig. 5D). These results show that during sporulation GFP–SirA mainly accumulates away from the replication origin.

A previous study in *B. subtilis* reported that DnaA colocalizes with the replisome at mid-cell via the YabA–DnaN complex during DNA replication (Soufo *et al.*, 2008). We hypothesized that SirA might be interacting with DnaA when it is bound to the replication machinery. To begin testing this model we examined GFP–SirA localization in cells that contained a fusion of a red fluorescent protein to

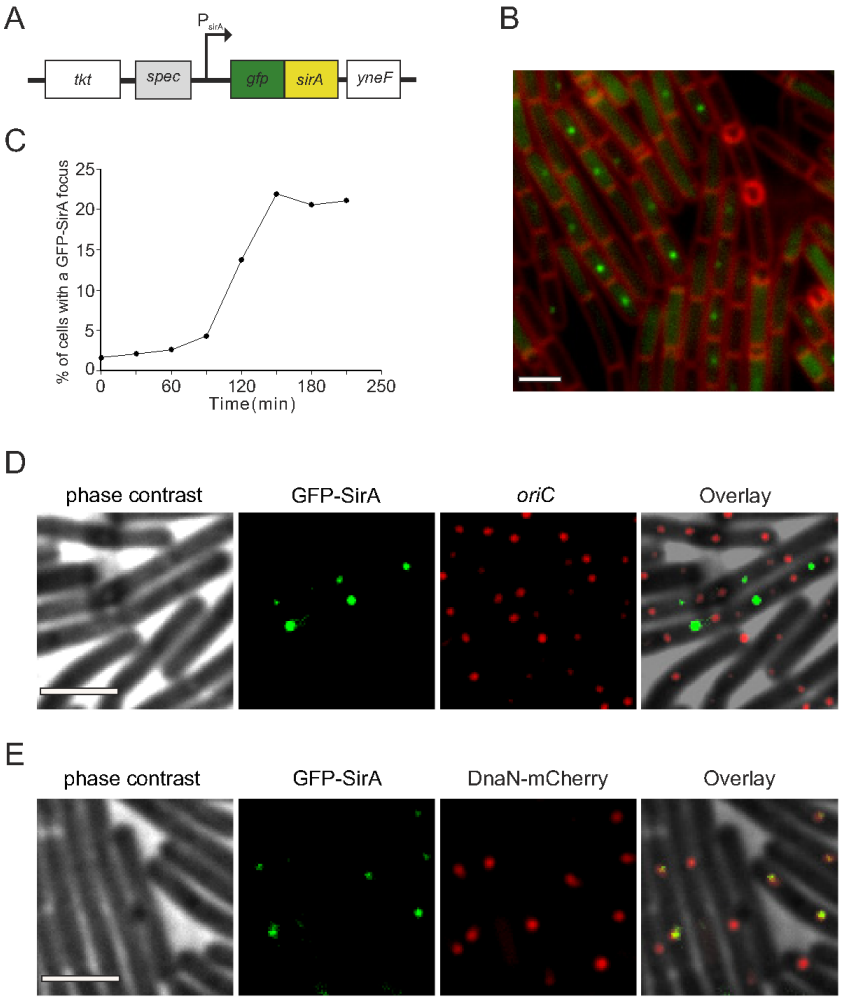


Fig. 5. Localization of GFP–SirA *in vivo*.
A. Schematic diagram showing the modified *sirA* locus used for localization studies. Chromosomal *sirA* was replaced with *gfp–sirA* under the control of its native expression system.
B. GFP–SirA localization during sporulation of *B. subtilis* 150 min post resuspension in starvation media. Membrane dye FM5-95 was used to highlight the outlines of the cells. Scale bar = 3 μm. *gfp–sirA* (NR3).
C. Temporal analysis of GFP–SirA foci formation during sporulation. At least 500 cells were analysed at each time-point.
D. Colocalization of GFP–SirA with *oriC* during sporulation (150 min post resuspension in starvation media). More than 200 cells were analysed and a representative image is shown. Scale bar = 3 μm. *gfp–sirA oriC*$^{tetO/TetR\ mCherry}$ (NR164).
E. Colocalization of GFP–SirA with the replisome during sporulation (150 min post resuspension in starvation media). More than 100 cells were analysed and a representative image is shown. Scale bar = 3 μm. *gfp–sirA dnaN–mCherry* (NR168).

the sliding clamp of the replisome (DnaN–mCherry). Cells were induced to sporulate by nutrient deprivation and the localization of GFP–SirA was determined in respect to the replisome. Strikingly, in cells containing a GFP–SirA focus the vast majority colocalized with DnaN–mCherry (88%; Fig. 5E). This result indicates that GFP–SirA accumulates at the replisome.

Substitutions at the SirA–DnaA^{DI} interface affect GFP–SirA localization in vivo

To investigate whether the localization of GFP–SirA at the replisome was dependent upon an interaction with DnaA, first the wild-type *sirA* (from the *gfp–sirA* chimera) was replaced with *sirA* mutants altering the residues identified in the SirA–DnaA^{DI} structure implicated in complex formation (*gfp–sirA^{F14A}*, *gfp–sirA^{Y18A}* or *gfp–sirA^{Q48A}*) and the localization GFP–SirA proteins was determined during sporulation. All of the *sirA* mutants caused a significant decrease in the number of cells containing a fluorescent focus (Fig. 6A). Both *gfp–sirA^{F14A}* and *gfp–sirA^{Y18A}* mutants reduced the number of GFP foci to background levels (i.e. – in the absence of a GFP fusion), while the *gfp–sirA^{Q48A}* mutant decreased foci formation 2.5-fold. These results indicate that the amino acid residues in SirA identified in the structure at the interface with DnaA^{DI} are required for GFP–SirA localization at the replisome.

Next we attempted to replace *dnaA* with *dnaA^{T26A}*, *dnaA^{W27A}* and *dnaA^{F49A}*; however, we were unable to isolate any of these mutants (see *Discussion*). In contrast, mutations in *dnaA* at locations that were previously shown to inhibit SirA activity *in vivo* (*dnaA^{A50V}* and *dnaA^{N47H}*) could be readily generated; therefore, GFP–SirA localization was determined using these *dnaA* alleles. Figure 6B shows that both DnaA variants inhibited GFP–SirA foci formation, with DnaA^{A50V} reducing foci formation to background levels and DnaA^{N47H} decreasing foci formation 2.4-fold. Immunoblot analysis of GFP-SirA proteins showed similar levels in all mutants tested, indicating that the absence of foci formation was not due to altered protein expression (Fig. 6C). Taken together with the analysis of the *sirA* mutants, these results show that the SirA–DnaA^{DI} interface identified in the crystal structure is critical for GFP–SirA localization *in vivo* and they suggest that GFP–SirA localization is mediated through a direct interaction with replisome-bound DnaA.

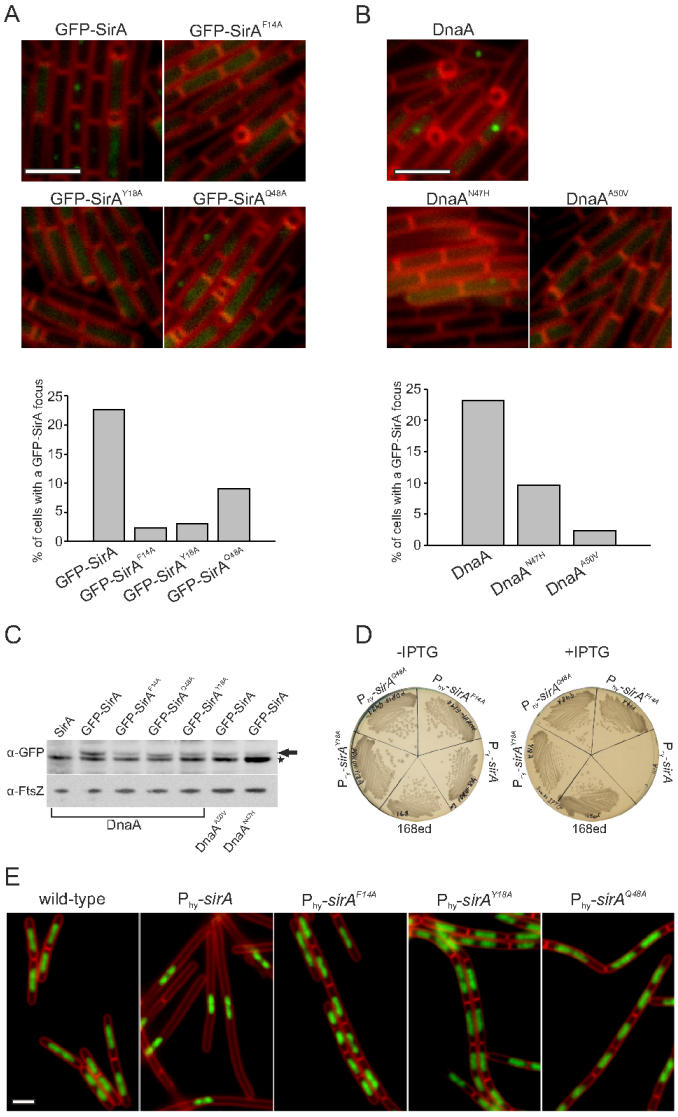
Substitutions at SirA–DnaA^{DI} interface render cells resistant to lethal effects of SirA overexpression in vegetatively growing cells

We attempted to determine whether GFP–DnaA colocalized with the replisome during sporulation, but unfortunately the previously published *gfp–dnaA* reporter strain displayed a severe sporulation defect (Soufo *et al.*, 2008). Therefore, to test whether SirA variants that displayed decreased foci formation were also defective in DnaA regulation, SirA proteins were overexpressed. Wild-type and mutant *sirA* genes were placed under the control of an IPTG-inducible promoter (P$_{hyperspank}$) integrated at an ectopic locus. Induction of wild-type SirA inhibited cell growth on solid media, in contrast to the SirA variants (SirA^{F14A}, SirA^{Y18A}, SirA^{Q48A}; Fig. 6D). Induction of wild-type SirA during vegetative growth in liquid media inhibited DNA replication, producing elongated cells that contained a single nucleoid (Fig. 6E). Induction of SirA variants did not affect DNA distribution or cell morphology, and these cells were indistinguishable from a control strain lacking the ectopic *sirA* construct (Fig. 6E). These results show that amino acid residue substitutions in SirA that impair protein localization also affect the ability of SirA to inhibit DnaA activity.

A conserved binding site on DnaA^{DI}

The structure of the SirA–DnaA^{DI} complex and that formed between DnaA^{DI} from *H. pylori* and HobA

Fig. 6. Examination of the SirA–DnaA interface *in vivo*.
A. Amino acid substitutions in SirA inhibit GFP–SirA foci formation. For each strain over 700 cells were analysed and the experiment was repeated at least three times. Quantification of a representative dataset is shown below. Scale bar = 3 μm. *gfp–sirA* (NR3); *gfp–sirA^{F14A}* (NR130); *gfp–sirA^{Y18A}* (NR156); *gfp–sirA^{Q48A}* (NR131).
B. Amino acid substitutions in DnaA inhibit GFP–SirA foci formation. For each strain over 700 cells were analysed and the experiment was repeated at least three times. Quantification of a representative dataset is shown below. Scale bar = 3 μm. *gfp–sirA* (NR3); *gfp–sirA dnaA^{A50V}* (NR5); *gfp–sirA dnaA^{N47H}* (NR154).
C. Immunoblot analysis showing levels of GFP-tagged SirA proteins. Cell samples were taken 150 min post resuspension in starvation media. The arrow points to GFP–SirA, the star highlights a contaminating band. Immunoblot of FtsZ was utilized to standardize the amount of protein from different samples. Wild-type (168ed); *gfp–sirA* (NR3); *gfp–sirA^{F14A}* (NR130); *gfp–sirA^{Y18A}* (NR156); *gfp–sirA^{Q48A}* (NR131); *gfp–sirA dnaA^{A50V}* (NR5); *gfp–sirA dnaA^{N47H}* (NR154).
D. Wild-type and mutant *sirA* were placed under control of an IPTG-inducible promoter and streaked on nutrient agar plates in the presence and absence of IPTG (3 mM). Wild-type (168ed); P$_{hyperspank}$-*sirA* (NR171); P$_{hyperspank}$-*sirA^{F14A}* (NR172); P$_{hyperspank}$-*sirA^{Q48A}* (NR173); P$_{hyperspank}$-*sirA^{Y18A}* (NR174).
E. The effects of overexpressing various SirA proteins at the single cell level were studied by growing cells in liquid CH medium. The images were taken 180 min after induction of gene expression with IPTG (3 mM). Membrane dye FM5-95 was used to highlight the outlines of the cells, DAPI was used to stain the DNA. Scale bar = 3 μm. Wild-type (168ed); P$_{hyperspank}$-*sirA* (NR171); P$_{hyperspank}$-*sirA^{F14A}* (NR172); P$_{hyperspank}$-*sirA^{Q48A}* (NR173); P$_{hyperspank}$-*sirA^{Y18A}* (NR174).



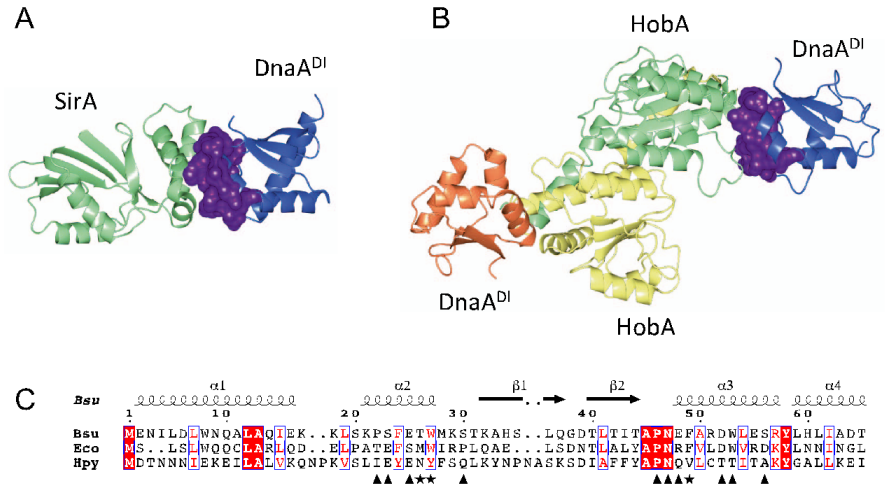


Fig. 7. Comparison of *B. subtilis* SirA–DnaA^{DI} with *H. pylori* HobA-DnaA^{DI}.
A. Ribbon diagram of the *B. subtilis* SirA–DnaA^{DI} complex with SirA shown in light green and DnaA^{DI} shown in blue.
B. Ribbon diagram of the *H. pylori* HobA-DnaA^{DI} complex (PDB id code: 2wp0) with two molecules of HobA shown in yellow and light green, and two molecules of DnaA^{DI} shown in blue and coral. The SirA and HobA binding surfaces on DnaA^{DI} in (A) and (B) respectively are shown in purple.
C. Alignment of DnaA^{DI} from *B. subtilis*, *E. coli* and *H. pylori*. Symbols below the alignments indicate interfacial residues on DnaA^{DI} in the SirA–DnaA^{DI} structure contributing to the core (asterisks) and the rim (triangles). Secondary structure elements and residue numberings are displayed above the alignment. The images were created using ESPript (Gouet, 2003).

(Natrajan *et al.*, 2009), a regulator of DNA replication in this pathogen, were compared with each other (Fig. 7). It is apparent that HobA and SirA bind to the same structural site on DnaA^{DI}, burying equivalent surface residues (Fig. 7A and B). This is surprising given the divergent effects on DnaA exerted by SirA and HobA. HobA is an essential stimulator of replication initiation in *H. pylori*, in contrast to SirA which is a replication inhibitor. Thus, HobA and SirA achieve opposing regulatory functions by binding to the same structural site on DnaA^{DI}. Despite their close tertiary structural correspondence, residues on the regulatory protein binding site on DnaA^{DI} are poorly conserved between *H. pylori* and *B. subtilis*, perhaps reflecting a divergence in their respective regulatory mechanisms (Fig. 7C). Nevertheless, this indicates an important structural site on DnaA^{DI} for the regulation of replication initiation.

Discussion

The initiator of bacterial DNA replication, DnaA, is stringently regulated so that DNA replication is co-ordinated with cell growth and differentiation. Five negative regulators of DNA replication have been identified in *B. subtilis:* SirA (Rahn-Lee *et al.*, 2009), YabA (Noirot-Gros *et al.*, 2006), Spo0A (Castilla-Llorente *et al.*, 2006) Soj (Murray and Errington, 2008) and DnaD (Bonilla and Grossman, 2012). YabA, Soj and DnaD bind to domain III of DnaA and are thought to block the assembly of helical DnaA filaments at *oriC* (Cho *et al.*, 2008; Scholefield *et al.*, 2012; Scholefield and Murray, 2013). Phosphorylated Spo0A binds to a set of Spo0A-boxes at *oriC* which overlap with DnaA-boxes, suggesting that Spo0A~P occludes DnaA from the replication origin (Boonstra *et al.*, 2013). SirA is distinct and represents the first, and so far only, *B. subtilis* regulator that interacts with domain I of DnaA (Rahn-Lee *et al.*, 2011). In other organisms however, regulators have been identified which interact with DnaA^{DI} namely, *E. coli* DiaA (Keyamura *et al.*, 2009) and Hda (Su'etsugu *et al.*, 2013), and *H. pylori* HobA (Natrajan *et al.*, 2009). For *E. coli*, DnaA^{DI} has also been shown to interact with the DNA helicase, DnaB (Sutton *et al.*, 1998; Seitz *et al.*, 2000), and to play a role in the oligomeriation of DnaA (Weigel *et al.*, 1999; Felczak *et al.*, 2005), forming dimers *in vitro* (Abe *et al.*, 2007). Here we have elucidated the structure of the SirA–DnaA^{DI} from *B.*

subtilis revealing DnaA^{DI} bound in an inhibitory complex. This structure complements that of HobA-DnaA^{DI} from *H. pylori* in which DnaA^{DI} is bound in a complex that leads to activation (Natrajan *et al.*, 2009).

As previously inferred (Rahn-Lee *et al.*, 2011), SirA binds to a site on DnaA^{DI} that corresponds closely to that bound by the regulators HobA from *H. pylori* (Natrajan *et al.*, 2009) and DiaA from *E. coli* (Keyamura *et al.*, 2009). HobA and DiaA are structural homologues which form tetramers that promote DnaA oligomerization and activate the initiation of DNA replication (Zawilak-Pawlik *et al.*, 2011). Each HobA/DiaA tetramer binds to four DnaA^{DI} molecules in a way that is thought to facilitate DnaA-binding to the array of DnaA-boxes distributed at *oriC* (Natrajan *et al.*, 2009). In marked contrast, SirA binds a single molecule of DnaA^{DI} and inhibits DNA replication initiation. Although SirA and HobA/DiaA have quite different three dimensional structures, each buries a structurally equivalent site on DnaA^{DI}. It is intriguing therefore that this elicits different regulatory outcomes.

It has been previously proposed that SirA inhibits DnaA binding to *oriC*, based on the observations that SirA disrupts DnaA–GFP localization at *oriC* during vegetative growth, and that there is a SirA-dependent decrease in the amount of DnaA at *oriC* following artificial induction of sporulation (Wagner *et al.*, 2009; Rahn-Lee *et al.*, 2011). At first glance, SirA may achieve this by inhibiting DnaA-oligomerization at *oriC*, since domain I fragments of *E. coli* DnaA form dimers *in vitro*, and the dimerization surface has been identified (Felczak *et al.*, 2005). However, the corresponding surface in *B. subtilis* DnaA^{DI} is located on the opposite side of DnaA^{DI} to the SirA binding surface so that SirA binding would not be expected to prevent dimer formation. Furthermore, we did not observe dimers or oligomers of *B. subtilis* DnaA^{DI} *in vitro*. Thus, it seems unlikely that SirA influences DnaA assembly simply by inhibiting DnaA^{DI}-DnaA^{DI} interactions.

Our localization studies indicate that SirA accumulates away from *oriC* and with the replisome near mid-cell during sporulation. We hypothesize that SirA could interact with replisome-bound DnaA to stabilize replisome–DnaA complexes, thereby inhibiting DnaA rebinding at the origin. This is reminiscent of a DnaA-tethering model proposed for YabA (Soufo *et al.*, 2008). Alternatively, our finding that mutations directing alanine substitutions of three residues on the SirA binding surface of DnaA could not be introduced into *dnaA* suggests that SirA could act by inhibiting a critical interaction of DnaA with other components of the initiation complex. In *E. coli*, DnaA^{DI} is implicated directly in the recruitment of the helicase to the nascent replicative complex (Abe *et al.*, 2007). There is no evidence for a DnaA-helicase interaction in *B. subtilis* however, in which helicase recruitment involves additional DNA remodelling proteins (Zhang *et al.*, 2005). DnaD forms multimeric scaffolds on the DNA and recruits DnaB, which in turn is thought to bridge an interaction with the helicase–helicase loader (Zhang *et al.*, 2008; Smits *et al.*, 2010). Thus, we speculate that SirA may inhibit a DnaA–DnaD interaction arresting assembly of the initiation complex.

In summary, this work has defined the interaction surfaces of SirA and DnaA^{DI} and the stoichiometry of their complex. Moreover, we have shown that their interaction is required for GFP-SirA foci formation at the replisome in sporulating *B. subtilis*. These observations will help elucidate the mechanism of action of SirA, the understanding of which is currently limited by imprecise knowledge of the function of domain I of DnaA in this organism.

Experimental procedures

Cloning

DNA fragments encoding SirA and domain I of DnaA (DnaA^{DI}) were amplified from *B. subtilis* genomic DNA by the polymerase chain reaction (PCR) and inserted into the expression vector pET-YSBLIC3C (Fogg and Wilkinson, 2008) using a ligation independent cloning method. Two constructs were created: one encoding SirA with an N-terminal, 3C protease cleavable His-tag (pET-YSBL3C-SirA) and the other a duet construct containing fragments encoding DnaA^{DI} and SirA cloned upstream and downstream respectively of a LIC Duet Minimal Adaptor (Novagen). The recombinant plasmid (pET-YSBLIC3C-DnaA^{DI}SirA) directs the expression of DnaA^{DI} fused to a 3C protease cleavable N-terminal His-tag (His–DnaA^{DI}) and SirA from separate bacteriophage T7 promoters. Alanine substitution mutations were introduced into pET-YSBLIC3C-DnaA^{DI}SirA by site-directed mutagenesis. The sequences of oligonucleotides used for the cloning and subsequent site-directed mutagenesis of pET-YSBLIC3C-DnaA^{DI}SirA are listed in Tables S1 and S2 respectively.

Expression

The plasmid pET-YSBLIC3C-DnaA^{DI}SirA was introduced into *E. coli* BL21 (DE3) cells for the co-overproduction of SirA and His-tagged DnaA^{DI}. Overnight cell cultures were used to inoculate 500 ml of Luria–Bertani (LB) media supplemented with 30 μg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6–0.9 at 37°C with shaking at 180 rpm before protein production was induced with 1 mM isopropyl-β-D-1-thiogalactoside (IPTG). Following induction, cultures were grown at 37°C (180 rpm shaking) for a further 4 h before cells were harvested by centrifugation. SirA/DnaA^{DI} proteins harbouring site-directed mutations were produced in an analogous manner.

For SeMet substituted protein production, overnight cultures of *E. coli* BL21 (DE3) harbouring pET-YSBLIC3C-DnaA^{DI}SirA were used to inoculate 500 ml minimal media supplemented with 30 μg ml^{-1} kanamycin. Cultures were grown to an OD_{600} of 0.6–0.8 at 37°C (180 rpm shaking) prior to the addition of an amino acid mixture (50 mg lysine, 50 mg

phenylalanine, 50 mg threonine, 25 mg isoleucine, 25 mg leucine, 25 mg valine) to suppress methionine production (Doublié, 1997), and 30 mg selenomethionine. Cultures were grown at 37°C (180 rpm shaking) for a further 15 min prior to induction of recombinant protein production with 1 mM IPTG. Cultures were subsequently grown at 30°C (180 rpm shaking) overnight (16–20 h) before cells were harvested by centrifugation.

DnaA^{DI}–SirA purification

The protein purification procedure was identical for native and SeMet substituted proteins. Harvested cells were resuspended in 50 mM Tris pH 8.5, 200 mM KCl, 20 mM imidazole and 10 mM BME, and an EDTA-free protease inhibitor cocktail tablet (Roche) was added. Resuspended cells were lysed by sonication and the lysate clarified by centrifugation. The cell lysate was loaded on to a His Trap FF crude Ni-affinity column (GE Healthcare) and bound protein eluted with an increasing imidazole concentration gradient (20–500 mM). This step was followed by size-exclusion chromatography on a HiLoad 16/60 Superdex S75 column (GE Healthcare) equilibrated with 50 mM Tris pH 8.5, 200 mM KCl, 10 mM BME. The chromatogram exhibited two protein peaks. SDS-PAGE analysis of the peak fractions showed the earlier eluting peak corresponded to the SirA : His–DnaA^{DI} complex with the later eluting peak containing His–DnaA^{DI} which is produced in excess in the duet expression system. The protein complex and DnaA domain I fractions were combined separately and the N-terminal histidine tag was removed from DnaA^{DI} in both cases by incubation with 3C protease overnight (protease : protein ratio of 1:50). Passage through a second Ni-affinity column to remove the histidine tag and protease yielded pure protein in a buffer of 50 mM Tris pH 8.5, 200 mM KCl, 10 mM BME. The proteins were judged to be pure according to Coomassie staining of SDS-polyacrylamide gels.

Crystallization and structure solution

Native crystals of SirA–DnaA^{DI} were grown in hanging-drops containing a 1:1 ratio of concentrated protein solution and reservoir solution. Native crystals suitable for data collection were obtained following mixing of a protein solution of 6.4 mg ml^{-1} and a reservoir solution of 100 mM HEPES pH 7.5, 200 mM NH_4 acetate, 25% (w/v) PEG 3350, 1% (v/v) DMF. Crystals were transferred to a cryoprotectant solution consisting of the reservoir solution containing 20% (v/v) glycerol before being cryocooled in liquid nitrogen. X-ray diffraction data were collected to 1.7 Å resolution on beamline I03 at the Diamond Light Source (DLS), Harwell. The crystal belongs to space group $P2_1$ with unit cell dimensions $a = 77.3$ Å, $b = 34.7$ Å, $c = 84.7$ Å and $\alpha = \gamma = 90°$, $\beta = 102.1°$. SeMet crystals were grown in hanging-drops containing a 2:1 ratio of concentrated protein solution : reservoir solution. SeMet-substituted crystals suitable for data collection were obtained using a protein concentration of 1.9 mg ml^{-1} and a reservoir solution of 100 mM MMT pH 6.0 (DL-malic acid, MES and Tris buffers in a molar ratio of 1:2:2), 20% (w/v) PEG 3350, 2% (v/v) DMF. Crystals were soaked in a cryoprotectant solution consisting of the reservoir solution containing 20% (v/v) glycerol before being cryocooled in liquid nitrogen. X-ray diffraction data were collected to 2.1 Å resolution on beamline I24 at DLS. The crystal belongs to space group $P2_1$ with cell dimensions $a = 51.4$ Å, $b = 35.6$ Å, $c = 63.3$ Å and $\alpha = \gamma = 90°$, $\beta = 92.8°$.

Diffraction datasets obtained from the SeMet derivative and native crystals were processed using the automated data processing pipeline Xia2 (Winter, 2009) with options that run XDS (Kabsch, 2010). Data were merged using AIMLESS (Evans, 2006). The structure of SirA–DnaA^{DI} was solved by single-wavelength anomalous dispersion (SAD) phasing. Heavy atom substructure determination, density modification and model building were carried out using the CRANK (Pannu *et al.*, 2011) pipeline available within the Collaborative Computational Project No. 4 (CCP4) program suite (Winn *et al.*, 2011). Nine selenium atom sites were identified using SHELX C/D (Sheldrick, 2008) and their positions refined using BP3. The correct hand for the phases was identified using SOLOMON (Abrahams and Leslie, 1996) and density modification carried out in PARROT (Cowtan, 2010) before atomic model building in BUCCANEER (Cowtan, 2006). The SeMet–SirA–DnaA^{DI} model was partially refined using maximum-likelihood methods in REFMAC (Murshudov *et al.*, 1997) and manual model building in COOT (Emsley and Cowtan, 2004). The partially refined SeMet–SirA–DnaA^{DI} model was used as a model for the solution of native SirA–DnaA^{DI} by molecular replacement with MOLREP (Vagin and Teplyakov, 1997), the search for SirA molecules preceding that for DnaA^{DI} domains. The native SirA–DnaA^{DI} model was refined through iterative cycles of refinement in REFMAC and manual model building in COOT to an *R*-factor of 13.1 (R_{free} = 19.7). Refinement statistics are shown in Table 1.

The atomic co-ordinates and crystallographic structure factors have been deposited in the Protein Data Bank with the Accession Code 4TPS.

SEC-MALLS

SEC-MALLS analysis of DnaA^{DI} and SirA–DnaA^{DI} was carried out at a range of protein concentrations: DnaA^{DI} was analysed at 1.0 mg ml^{-1}, 2.5 mg ml^{-1} and 5.0 mg ml^{-1} and SirA–DnaA^{DI} was analysed at 0.5 mg ml^{-1}, 1.0 mg ml^{-1} and 2.5 mg ml^{-1}. For each run, 100 µl of sample was loaded onto a Superdex 75 HR 10/30 size-exclusion column equilibrated with 50 mM Tris pH 8.5, 200 mM KCl at a flow rate of 0.5 ml min^{-1}. The eluate was analysed successively by a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor as described previously (Colledge *et al.*, 2011). Data were analysed with Astra software (Wyatt).

Solubility assay

E. coli BL21 (DE3), harbouring wild type and mutated pET-YSBLIC3C-DnaA^{DI}SirA plasmids were grown in 200 ml LB-kanamycin until the A_{600} reached ~ 0.6. A portion of cells (uninduced) was set aside and grown separately while IPTG was added to the remaining cells. After a further 4 h growth, aliquots of the uninduced (U) and induced (I) cells were taken

and used to prepare total cell samples. The remaining cells from the induced culture were harvested by centrifugation and the cell pellets were re-suspended in 20 ml of 50 mM Tris pH 8.5, 200 mM KCl, 20 mM imidazole, 10 mM BME. Cells were lysed by sonication and the lysate clarified by centrifugation. An aliquot of this soluble fraction (S) was retained. The remaining lysate was loaded onto a 1 ml HisTrap FF crude Ni-affinity column (GE Healthcare), washed with 6 ml re-suspension buffer, and bound protein eluted with 4 ml of 50 mM Tris pH 8.5, 200 ml KCl, 500 mM imidazole, 10 mM BME and the eluate (E) was collected. For each of the wild type and alanine variants, samples of the total fractions from uninduced (U) and induced (I) cells together with the soluble lysis (S) and high imidazole column eluate (E) fractions were analysed by SDS-PAGE followed by Coomassie blue staining.

B. subtilis *strains, media and growth conditions*

Strains used in this study are listed in Table S3. Nutrient agar (Oxoid) was used as the solid media for growth of *B. subtilis*. LB medium was used for growing cells to extract genomic DNA. Antibiotics were added to the growth media as required: chloramphenicol (5 μg ml^{-1}), spectinomycin (50 μg ml^{-1}). To induce sporulation *Bacillus subtilis* cells grown in hydrolysed casein media at 37°C were induced to sporulate according to the resuspension method of Sterlini and Mandelstam (1969) as modified by Partridge and Errington (1993).

Microscopy

After induction of sporulation, samples were taken every 30 min and visualized using fluorescence microscopy. Microscopy was performed using glass slides covered with a ~ 1.5% agarose pad containing sporulation media. A glass coverslip (0.17 mm VWR) covered cells immobilized on the agarose pad. The dye FM5-95 was added to the agar pads to visualize the membrane (2.9 μg ml^{-1} final). To visualize the nucleoid the DNA was stained with 4′-6-diamidino-2-phenylindole (DAPI 2.5 μg ml^{-1} final). Microscopy was performed on an inverted epifluorescence microscope (Zeiss Axiovert 200M) fitted with a Plan-Neofluar objective (Zeiss 100×/1.30 Oil Ph3). Light was transmitted from a 300 Watt xenon arc-lamp through a liquid light guide (Sutter Instruments) and images were collected using a Sony CoolSnap HQ cooled CCD camera (Roper Scientific). All filters were Modified Magnetron ET Sets from Chroma and details are available upon request. Digital images were acquired and analysed using METAMORPH software (version V.6.2r6).

Western blot analysis

Proteins were separated by electrophoresis using a NuPAGE 4–12% Bis-Tris gradient gel run in MES buffer (Life Technologies) and transferred to a Hybond-P PVDF membrane (GE Healthcare) using a semi-dry apparatus (Hoefer Scientific Instruments). Proteins of interest were probed with polyclonal primary antibodies and then detected with an anti-rabbit horseradish peroxidase-linked secondary antibody using an ImageQuant LAS 4000 mini digital imaging system (GE Healthcare).

Acknowledgements

This work was supported by the EU-funded BaSysBio project LSHG-CT-2006-037469, by BBSRC studentships awarded to K.J. at York and N.R. at Newcastle and by a Royal Society University Research Fellowship to H.M. A.G. was funded by an ERASMUS programme. We would like to thank the DIAMOND Light Source (Harwell) and the Biology Technology Facility at York for access to equipment and experimental advice. We also wish to thank Sheila Taylor and Sam Hart for technical assistance, Jean Whittingham for advice and Alan Koh for providing a plasmid.

References

Abe, Y., Jo, T., Matsuda, Y., Matsunaga, C., Katayama, T., and Ueda, T. (2007) Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on *oriC*. *J Biol Chem* **282:** 17816–17827.

Abrahams, J.P., and Leslie, A.G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr* **52:** 30–42.

Bonilla, C.Y., and Grossman, A.D. (2012) The primosomal protein DnaD inhibits cooperative DNA binding by the replication initiator DnaA in *Bacillus subtilis*. *J Bacteriol* **194:** 5110–5117.

Boonstra, M., de Jong, I.G., Scholefield, G., Murray, H., Kuipers, O.P., and Veening, J.-W. (2013) Spo0A regulates chromosome copy number during sporulation by directly binding to the origin of replication in *Bacillus subtilis*. *Mol Microbiol* **87:** 925–938.

Briggs, G.S., Smits, W.K., and Soultanas, P. (2012) Chromosomal replication initiation machinery of low-G + C-content Firmicutes. *J Bacteriol* **194:** 5162–5170.

Burkholder, W.F., Kurtser, I., and Grossman, A.D. (2001) Replication initiation proteins regulate a developmental checkpoint in *Bacillus subtilis*. *Cell* **104:** 269–279.

Castilla-Llorente, V., Muñoz-Espín, D., Villar, L., Salas, M., and Meijer, W.J.J. (2006) Spo0A, the key transcriptional regulator for entrance into sporulation, is an inhibitor of DNA replication. *EMBO J* **25:** 3890–3899.

Cho, E., Ogasawara, N., and Ishikawa, S. (2008) The functional analysis of YabA, which interacts with DnaA and regulates initiation of chromosome replication in *Bacillus subtilis*. *Genes Genet Syst* **83:** 111–125.

Colledge, V.L., Fogg, M.J., Levdikov, V.M., Leech, A., Dodson, E.J., and Wilkinson, A.J. (2011) Structure and organisation of SinR, the master regulator of biofilm formation in *Bacillus subtilis*. *J Mol Biol* **411:** 597–613.

Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62:** 1002–1011.

Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66:** 470–478.

Doublié, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **276:** 523–530.

Duderstadt, K.E., Chuang, K., and Berger, J.M. (2011) DNA stretching by bacterial initiators promotes replication origin opening. *Nature* **478:** 209–213.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.
Erzberger, J.P., Mott, M.L., and Berger, J.M. (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* **13:** 676–683.
Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62:** 72–82.
Felczak, M.M., Simmons, L.A., and Kaguni, J.M. (2005) An essential tryptophan of *Escherichia coli* DnaA protein functions in oligomerization at the *E. coli* replication origin. *J Biol Chem* **280:** 24627–24633.
Fogg, M.J., and Wilkinson, A.J. (2008) Higher-throughput approaches to crystallization and crystal structure determination. *Biochem Soc Trans* **36:** 771–775.
Fujikawa, N., Kurumizaka, H., Nureki, O., and Terada, T. (2003) Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res* **31:** 2077–2086.
Gouet, P. (2003) ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31:** 3320–3323.
Ishigo-oka, D., Ogasawara, N., and Moriya, S. (2001) DnaD protein of *Bacillus subtilis* interacts with DnaA, the initiator protein of replication. *J Bacteriol* **183:** 1–4.
Janin, J., Bahadur, R.P., and Chakrabarti, P. (2008) Protein-protein interaction and quaternary structure. *Q Rev Biophys* **41:** 133–180.
Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125–132.
Kaguni, J.M. (2006) DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol* **60:** 351–375.
Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010) Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nat Rev Microbiol* **8:** 163–170.
Keyamura, K., Abe, Y., Higashi, M., Ueda, T., and Katayama, T. (2009) DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J Biol Chem* **284:** 25038–25050.
Leonard, A.C., and Grimwade, J.E. (2011) Regulation of DnaA assembly and activity: taking directions from the genome. *Annu Rev Microbiol* **65:** 19–35.
Lowery, T.J., Pelton, J.G., Chandonia, J.-M., Kim, R., Yokota, H., and Wemmer, D.E. (2007) NMR structure of the N-terminal domain of the replication initiator protein DnaA. *J Struct Funct Genomics* **8:** 11–17.
McNicholas, S., Potterton, E., Wilson, K.S., and Noble, M.E.M. (2011) Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67:** 386–394.
Molt, K.L., Sutera, V.A., Moore, K.K., and Lovett, S.T. (2009) A role for nonessential domain II of initiator protein, DnaA, in replication control. *Genetics* **183:** 39–49.
Moravcevic, K., Mendrola, J., and Schmitz, K. (2010) Kinase associated-1 (KA1) domains drive MARK/PAR1 kinases to membrane targets by binding acidic phospholipids. *Cell* **143:** 966–977.
Murray, H., and Errington, J. (2008) Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. *Cell* **135:** 74–84.
Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.
Natrajan, G., Noirot-Gros, M.-F., Zawilak-Pawlik, A., Kapp, U., and Terradot, L. (2009) The structure of a DnaA/HobA complex from *Helicobacter pylori* provides insight into regulation of DNA replication in bacteria. *Proc Natl Acad Sci USA* **106:** 21115–21120.
Noirot-Gros, M.-F., Velten, M., Yoshimura, M., McGovern, S., Morimoto, T., Ehrlich, S.D., *et al.* (2006) Functional dissection of YabA, a negative regulator of DNA replication initiation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* **103:** 2368–2373.
Pannu, N.S., Waterreus, W.J., Skubák, P., Sikharulidze, I., Abrahams, J.P., and de Graaff, R.A.G. (2011) Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr D Biol Crystallogr* **67:** 331–337.
Partridge, S.R., and Errington, J. (1993) The importance of morphological events and intercellular interactions in the regulation of prespore-specific gene expression during sporulation in *Bacillus subtilis*. *Mol Microbiol* **8:** 945–955.
Rahn-Lee, L., Gorbatyuk, B., Skovgaard, O., and Losick, R. (2009) The conserved sporulation protein YneE inhibits DNA replication in *Bacillus subtilis*. *J Bacteriol* **191:** 3736–3739.
Rahn-Lee, L., Merrikh, H., Grossman, A.D., and Losick, R. (2011) The sporulation protein SirA inhibits the binding of DnaA to the origin of replication by contacting a patch of clustered amino acids. *J Bacteriol* **193:** 1302–1307.
Rokop, M.E., Auchtung, J.M., and Grossman, A.D. (2004) Control of DNA replication initiation by recruitment of an essential initiation protein to the membrane of *Bacillus subtilis*. *Mol Microbiol* **52:** 1757–1767.
Scholefield, G., and Murray, H. (2013) YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Mol Microbiol* **90:** 147–159.
Scholefield, G., Errington, J., and Murray, H. (2012) Soj/ParA stalls DNA replication by inhibiting helix formation of the initiator protein DnaA. *EMBO J* **31:** 1542–1555.
Seitz, H., Weigel, C., and Messer, W. (2000) The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*. *Mol Microbiol* **37:** 1270–1279.
Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr A* **64:** 112–122.
Smits, W.K., Goranov, A.I., and Grossman, A.D. (2010) Ordered association of helicase loader proteins with the *Bacillus subtilis* origin of replication *in vivo*. *Mol Microbiol* **75:** 452–461.
Soufo, C.D., Soufo, H.J.D., Noirot-Gros, M.-F., Steindorf, A., Noirot, P., and Graumann, P.L. (2008) Cell-cycle-dependent spatial sequestration of the DnaA replication initiator protein in *Bacillus subtilis*. *Dev Cell* **15:** 935–941.
Sterlini, J.M., and Mandelstam, J. (1969) Commitment to sporulation in *Bacillus subtilis* and its relationship to development of actinomycin resistance. *Biochem J* **113:** 29–37.
Su'etsugu, M., Harada, Y., Keyamura, K., Matsunaga, C., Kasho, K., Abe, Y., *et al.* (2013) The DnaA N-terminal domain interacts with Hda to facilitate replicase clamp-mediated inactivation of DnaA. *Environ Microbiol* **15:** 3183–3195.

Sutton, M.D., Carr, K.M., Vicente, M., and Kaguni, J.M. (1998) *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J Biol Chem* **273:** 34255–34262.

Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30:** 1022–1025.

Veening, J.-W., Murray, H., and Errington, J. (2009) A mechanism for cell cycle regulation of sporulation initiation in *Bacillus subtilis*. *Genes Dev* **23:** 1959–1970.

Velten, M., McGovern, S., Marsin, S., Ehrlich, S.D., Noirot, P., and Polard, P. (2003) A two-protein strategy for the functional loading of a cellular replicative DNA helicase. *Mol Cell* **11:** 1009–1020.

Wagner, J.K., Marquis, K.A., and Rudner, D.Z. (2009) SirA enforces diploidy by inhibiting the replication initiator DnaA during spore formation in *Bacillus subtilis*. *Mol Microbiol* **73:** 963–974.

Weigel, C., Schmidt, A., Seitz, H., Tüngler, D., Welzeck, M., and Messer, W. (1999) The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. *Mol Microbiol* **34:** 53–66.

Winn, M., Ballard, C., Cowtan, K., Dodson, E., Emsley, P., Evans, P., *et al.* (2001) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67:** 235–242.

Winter, G. (2009) Xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43:** 186–190.

Yoshikawa, H., and Ogasawara, N. (1991) Structure and function of DnaA and the DnaA-box in eubacteria: evolutionary relationships of bacterial replication origins. *Mol Microbiol* **5:** 2589–2597.

Zawilak-Pawlik, A., Donczew, R., Szafrański, S., Mackiewicz, P., Terradot, L., and Zakrzewska-Czerwińska, J. (2011) DiaA/HobA and DnaA: a pair of proteins co-evolved to cooperate during bacterial orisome assembly. *J Mol Biol* **408:** 238–251.

Zhang, W., Carneiro, M.J.V.M., Turner, I.J., Allen, S., Roberts, C.J., and Soultanas, P. (2005) The *Bacillus subtilis* DnaD and DnaB proteins exhibit different DNA remodelling activities. *J Mol Biol* **351:** 66–75.

Zhang, W., Machón, C., Orta, A., Phillips, N., Roberts, C.J., Allen, S., and Soultanas, P. (2008) Single-molecule atomic force spectroscopy reveals that DnaD forms scaffolds and enhances duplex melting. *J Mol Biol* **377:** 706–714.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Published online 28 November 2015 *Nucleic Acids Research, 2016, Vol. 44, No. 1* **449–463**
doi: 10.1093/nar/gkv1318

Tetramerization and interdomain flexibility of the replication initiation controller YabA enables simultaneous binding to multiple partners

Liza Felicori[1,2,†], Katie H. Jameson[3,†], Pierre Roblin[4], Mark J. Fogg[3], Transito Garcia-Garcia[5,6], Magali Ventroux[5,6], Mickaël V. Cherrier[7,8,9], Alexandre Bazin[7,8,9], Philippe Noirot[5,6], Anthony J. Wilkinson[3], Franck Molina[2,*], Laurent Terradot[7,8,9,*] and Marie-Françoise Noirot-Gros[5,6,*]

[1]Departamento de Bioquimica e Imunologia, Universidade Federal de Minas Gerais, UFMG, 31270-901, Belo Horizonte, MG, Brazil, [2]Sys2Diag FRE3690—CNRS/ALCEDIAG, Montpellier, France, [3]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5DD, UK, [4]Synchrotron SOLEIL—L'Orme des Merisiers Saint-Aubin— BP 48 91192 GIF-sur-YVETTE CEDEX, France, [5]INRA, UMR1319 Micalis, F-78350 Jouy-en-Josas, France, [6]AgroParisTech, UMR1319 Micalis, F-78350 Jouy-en-Josas, France, [7]CNRS, UMR 5086 Bases Moléculaires et Structurales de Systèmes Infectieux, Institut de Biologie et Chimie des Protéines, 7 Passage du Vercors, F-69367 Lyon, France, [8]Université de Lyon, F-69622 Lyon, France and [9]Université Claude Bernard Lyon 1, F-69622 Villeurbanne, France

Received August 24, 2015; Revised November 6, 2015; Accepted November 10, 2015

ABSTRACT

YabA negatively regulates initiation of DNA replication in low-GC Gram-positive bacteria. The protein exerts its control through interactions with the initiator protein DnaA and the sliding clamp DnaN. Here, we combined X-ray crystallography, X-ray scattering (SAXS), modeling and biophysical approaches, with *in vivo* experimental data to gain insight into YabA function. The crystal structure of the N-terminal domain (NTD) of YabA solved at 2.7 Å resolution reveals an extended α-helix that contributes to an intermolecular four-helix bundle. Homology modeling and biochemical analysis indicates that the C-terminal domain (CTD) of YabA is a small Zn-binding domain. Multi-angle light scattering and SAXS demonstrate that YabA is a tetramer in which the CTDs are independent and connected to the N-terminal four-helix bundle via flexible linkers. While YabA can simultaneously interact with both DnaA and DnaN, we found that an isolated CTD can bind to either DnaA or DnaN, individually. Site-directed mutagenesis and yeast-two hybrid assays identified DnaA and DnaN binding sites on the YabA CTD that partially overlap and point to a mutually exclusive mode of interaction. Our study defines YabA as a novel structural hub and explains how the protein tetramer uses independent CTDs to bind multiple partners to orchestrate replication initiation in the bacterial cell.

INTRODUCTION

In all living organisms, chromosome replication is highly regulated to ensure only one initiation event per chromosome per cell cycle (1). Bacteria have developed various strategies to prevent inappropriate re-initiation, principally by regulating the activity and/or the availability of the master initiator protein DnaA. DnaA assembles at specific DNA sequences within *oriC* to promote the opening of the DNA duplex, and directs the assembly of the replisome machinery by first recruiting DNA helicase (1,2). DnaA is a member of the AAA+ (ATPases associated with diverse cellular activities) superfamily that binds to and hydrolyses adenosine triphosphate (ATP). Although both ATP and adenosine diphosphate (ADP) bound forms of DnaA are proficient in *oriC* binding, only the ATP-bound form is active in replication initiation (3).

In bacteria, multiple homeostatic mechanisms contribute to coordinate DNA replication with the cellular cycle (1,4). In Gram negative bacteria, proteins have been identified

*To whom correspondence should be addressed. Tel: +1 630 252 1759; Fax: +1 630 252 3387; Email: marie-francoise.gros@jouy.inra.fr
Correspondence may also be addressed to Laurent Terradot. Tel: +33 472722652; Fax: +33 472722604; Email: laurent.terradot@ibcp.fr
Correspondence may be also addressed to Frank Molina. Tel: +33 467047463; Fax: +33 467047401; Email: franck.molina@sysdiag.cnrs.fr
†These authors contributed equally to the paper as first authors.
Present address: Marie-Françoise Noirot-Gros, Biosciences Division, Argonne National Laboratory, Lemont, IL 60439, USA.

which regulate positively or negatively the initiation of replication by forming a complex with DnaA. Positive regulators such as DiaA in *Escherichia coli*, or the structural homolog HobA in *Helicobacter pylori* promote initiation by stimulating the assembly of ATP–DnaA at *oriC* (5–7). In *E. coli*, the major mechanism of initiation control is the regulatory inactivation of DnaA (RIDA), mediated by the ADP-activated protein Hda (homolog of DnaA) and the β-clamp subunit of the replicative DNA polymerase DnaN. The interaction between Hda, DnaN and ATP-bound DnaA, promotes the hydrolysis of ATP and the accumulation of inactive ADP–DnaA (8). RIDA has also been reported in *Caulobacter crescentus*, a Gram-negative proteobacteria phylogenetically distant from *E. coli* (9). In this bacterium, the Hda homolog HdaA prevents over-initiation and co-localizes with the replication machinery in the stalked cell. However, the presence of Hda seems to be restricted to proteobacteria suggesting that other bacterial species have developed different mechanisms for the negative regulation of DnaA.

In the Gram positive and spore forming bacterium *Bacillus subtilis*, replication initiation control is fulfilled by various proteins, which bind to DnaA and modulate its activity during the different life-styles of the bacilli. In cells committed to sporulation, the protein SirA prevents replication re-initiation by antagonizing DnaA binding to *oriC* (10–12). In vegetative cells, the primosomal protein DnaD and the ParA-like protein Soj were recently found to down regulate replication initiation by preventing the formation of a DnaA nucleofilament at *oriC* (13,14). The main regulatory protein YabA was also found to prevent over-initiation by inhibiting the cooperative binding of DnaA to *oriC* (13,14). Furthermore, YabA downregulates replication initiation as part of a multimeric complex with DnaA and DnaN that is associated with the replication factory (15–18). Thus, YabA is likely to act by trapping DnaA in a manner that can be both dependent and independent of DnaN during the cell replication cycle. However, although YabA, like Hda, regulates initiation through coupling to the elongation of replication, several pieces of evidence point to a distinct mode of action. In contrast to Hda, YabA has no sequence similarity to DnaA, and YabA does not belong to the AAA+ superfamily. YabA is a small protein of 119 residues, with an unusual predicted organization composed of a leucine zipper at its N-terminus and a putative zinc-binding domain at its C-terminus (15). Curiously, it lacks the classical bacterial DnaN-binding pentapeptide consensus motif (19). Given that YabA is conserved in low-GC Gram-positive bacteria, the protein represents the prototype of a distinct family of replication controller proteins.

To gain insight into the mechanism by which YabA interacts with multiple partners to control DNA replication, we have combined the X-ray structure of the N-terminal domain (NTD) of YabA, *in silico* modeling of the C-terminal domain (CTD) and small angle X-ray scattering (SAXS) of the full-length protein to derive a model of the YabA structure. The crystal structure of the NTD of YabA revealed an extended α-helix, four of which assemble into a helical bundle. We found that the CTD of YabA binds to a single Zn^{2+} ion with a zinc binding motif resembling known short Zn binding domains. Solution studies demonstrated that YabA is a tetramer in which the monomeric CTDs are separated from the tetrameric NTDs by a flexible linker resulting in an extended conformation. We then showed that the CTD is sufficient for interaction with either DnaA or DnaN. The interacting surfaces with DnaA and DnaN have been mapped within the CTD and unveil an atypical motif for binding to the β-clamp. Our study provides a structural explanation for the capacity of YabA to bind multiple partners, hereby defining a novel protein hub central to DNA replication in low GC Gram-positive bacteria.

MATERIALS AND METHODS

Proteins expression and purification

The *yabA* gene sequence was amplified from *B. subtilis* (strain 168) genomic DNA by polymerase chain reaction (PCR) and inserted into pET151/D-TOPO (Invitrogen) to generate the plasmid pET151-*yabA*. *E. coli* BL21 (DE3) cells (Invitrogen) carrying pET151-*yabA* were grown in LB medium (with ampicillin at 100 μg/l) at 37°C until OD_{600} = 0.6. Protein expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 16 hours at 20°C. Cells were centrifuged and resuspended in buffer L (30 mM Tris pH 8, 150 mM NaCl, 1% Triton, 1 mM $ZnCl_2$ and 5% Glycerol (V/V)) with protease inhibitor tablet (complete EDTA-free, Roche), lysozyme (Roche) and Dnase (Sigma-Aldrich). The cells were lysed by sonication and centrifuged at 16 000 *g* for 20 min. The soluble fraction was loaded on a HisTrap™ 5 ml column equilibrated with buffer A (30 mM Tris pH 8, 150 mM NaCl, 5% glycerol) and the protein eluted using a 0–100% gradient of buffer B (buffer A + 500 mM imidazole). Fractions containing YabA were incubated with TEV protease in the presence of 0.5 mM dithiothreitol (DTT) and 0.5 mM ethylenediaminetetraacetic acid (EDTA) and dialyzed for 16 h against buffer A at 4°C. The cleaved histidine tag was subsequently removed by passage of the dialysate through a HisTrap™ column. The flow-through fraction was concentrated and injected onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare) equilibrated with buffer A. Fractions containing YabA were pooled and concentrated to 5 $mg.ml^{-1}$.

To produce the YabA/DnaN complex, a vector (pET Lic_duet_yabA-dnaN) was used which directs the co-expression of N-terminally histidine tagged YabA and DnaN. *E. coli* BL21 (DE3) cells (Invitrogen) harboring (pET Lic_duet_yabA-dnaN) were grown in LB medium (with kanamycin at 40 μg/l) at 37°C until OD_{600} = 0.6. Protein expression was induced by addition of 1 mM IPTG and cells were incubated 3 h at 37°C. Cells were harvested by centrifugation and resuspended in buffer L with a protease inhibitor tablet (complete EDTA-free, Roche), lysozyme (Roche) and Dnase (Sigma-Aldrich). The cells were lysed by sonication and centrifuged at 16 000 *g* for 20 min. The soluble fraction was diluted four times with buffer A and applied on a HisTrap™ 5 ml column equilibrated with buffer A. The protein complex was eluted using a 0 to 100% gradient of buffer B. Fractions containing the YabA/DnaN complex were pooled and applied to a Superdex 200 10/300 GL gel filtration column (GE Healthcare) equilibrated with buffer A. The pure YabA/DnaN complex was concentrated to 5 $mg\ ml^{-1}$.

For the preparation of YabA$^{1\ 58}$ and YabA$^{70\ 119}$, the coding region of *yabA* was amplified from *B. subtilis* genomic DNA (strain 168) by PCR using gene-specific primers, with appended sequences to facilitate ligation independent cloning. The PCR product was inserted into the vector pET-YSBLIC3C by ligation-independent cloning methods (20). Plasmids encoding N- and C-terminal fragments of YabA, residues 1–58 (YabA$^{1\ 58}$) and residues 70–119 (YabA$^{70\ 119}$) respectively, were created using a deletion mutagenesis method. A whole vector amplification of pET-YSBLIC3C-YabA was performed by PCR using the primers listed in Supplementary Table S5. The recombinant plasmids pET-YSBLIC3C-YabA$^{1\ 58}$ and pET-YSBLIC3C-YabA$^{70\ 119}$ direct the expression of YabA$^{1\ 58}$ or YabA$^{70\ 119}$, respectively, each fused to a 3C cleavable N-terminal His-tag. The recombinant YabA fragments were overproduced in *E. coli* BL21 (DE3) and following cell lysis, purified by steps of immobilized nickel affinity and gel filtration chromatography similar to those described above with the affinity tag subsequently being removed by HRV 3C protease.

Crystallization, structure determination and refinement

Two crystals forms of YabA were obtained using vapor diffusion methods in hanging drops consisting of 1.0 μl of protein or protein complex and 1.0 μl of reservoir solution containing 20% PEG 3350 (W/V), 100 mM Bis-Tris propane pH 6.5 and 200 mM potassium thiocyanate. Crystal form I was obtained using the purified YabA/DnaN complex (5 mg.ml^{-1}). A second crystal form (II) was obtained using YabA (5 mg.ml^{-1}) and by including 0.025% (V/V) dichloromethane in the drop. For structure determination, form II crystals were soaked in reservoir solution supplemented with 10 mM platinum potassium thiocyanate for 16 h. Crystals were then flash-cooled in liquid nitrogen (100K) using mother liquor containing step-wise additions of glycerol (15% final concentration) as cryoprotectant. X-ray diffraction data were collected at the ID14EH4 (form I) and ID23EH1 (form II) beamlines of the European Synchrotron Radiation Facility (ESRF, Grenoble). Form I crystals belonged to the space group $P3_221$ with unit cell dimensions of a = b = 83.4 Å and c = 64.7 Å and diffracted to a resolution of 2.7 Å. Crystal form II belonged to the space group $P6_122$ with unit cell dimensions of a = b = 83 Å and c = 110 Å and diffracted to a resolution of 4 Å according to $CC_{1/2}$ (21). The diffraction data were indexed and integrated using XDS (22) and scaled with SCALA from the CCP4 program suite (23). Data collection statistics are given in Table 1.

A single high redundancy dataset was collected at the Pt edge on a form II crystal soaked into platinum thiocyanate. The structure was solved using Autosol from the PHENIX package (24). The experimental map obtained at 4 Å and the two Pt sites identified were used in Autobuild which resulted in a non crystallographic symmetry (NCS) averaged, solvent flattened and improved map (24). The initial model consisted of two polyalanine helices of 30 residues and was subsequently extended manually in the improved low-resolution experimental map. The model was then used for molecular replacement using the data obtained from crystal form I. The model was built manually using COOT (25) and refined using PHENIX (24). Phasing and refinement statistics are indicated in Table 1. Figures depicting the structure were generated with PyMOL (26). The coordinates of YabA$^{1\ 62}$ have been deposited in the protein databank (pdb code 5DOL).

SEC-MALS

Samples (100 μl) of YabA, YabA$^{1\ 58}$ and YabA$^{70\ 119}$ were loaded onto gel-filtration columns equilibrated with 50 mM Tris pH 8.0, 150 mM NaCl. YabA was loaded at a concentration of 2.5 mg.ml^{-1} onto a Superdex 200 HR 10/30 column and YabA$^{1\ 58}$ and YabA$^{70\ 119}$ were loaded at concentrations of 5 mg.ml^{-1} onto a Superdex 200/75 HR 10/30 column. In each case, the column eluate was successively analyzed by a SPD20A UV/Vis detector, a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor. Data were analyzed with Astra software (Wyatt).

Circular dichroism

Circular dichroism spectra were recorded at 20°C on a Jasco J-810 CD spectrophotometer using a quartz cell with a 0.1 cm path length. Experiments were carried out at protein concentrations of 0.2 mg.ml^{-1} in a buffer of 20 mM potassium phosphate, pH 8.0. Spectra were recorded across the wavelength range of 260–185 nm. A buffer scan was also recorded and subtracted from the protein spectra to remove any contribution from the buffer. Analysis of YabA CD spectra was performed using Dicroweb (http://dichroweb.cryst.bbk.ac.uk) (27).

Atomic absorption spectroscopy

For AAS we used a Phillips PU9200 double beam spectrometer equipped with a flame volatilization system. Samples of YabA dissolved in deionized water at 0.5 mg.ml^{-1} were analyzed in triplicate using a zinc-specific lamp of wavelength 213.9 nm and the absorbance values were compared to those for a standard curve derived from samples of zinc powder dissolved in HCl.

In silico molecular modeling of CTD

Template Search and Secondary Structure Prediction: the amino acid sequence analysis was performed using the Position Specific Iterated BLAST (PSI-BLAST) and Pattern Hit Initiated BLAST (PHI-BLAST) methods at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and using the All Non-Redundant (NR) amino acid sequence database from April 2000, which includes SwissProt, CDS translation of GenBank (gb), EMBL (emb), the DNA database of Japan (dbj) and the Protein Structure Database (pdb) (28). Default amino acid replacement matrices and gap penalties were used in all database searches. Secondary structure predictions were made using JPRED (http://www.compbio.dundee.ac.uk/jpred) through the Jalview version 2 software (29).

Full length YabA and two fragments encompassing residues 1–70 and 71–119 were scanned against the protein databank (PDB) by PSI-BLAST over five iterations.

Table 1. Data collection, phasing and refinement statistics

	Native	Pt derivative
ESRF beamline	ID14EH4	ID23EH1
Wavelength (Å)	0.93222	1.0717 Å
Space group	$P\,3_2\,2\,1$	$P\,6_1\,2\,2$
Unit cell parameters (Å)		
	a = 83.48	a = 83
	b = 83.48	b = 83
	c = 64.71	c = 110
Number of protein molecules per asymmetric unit	2	2
Resolution range (Å)	36.1–2.7	43.69–4.0
	(2.8–2.7)	(4.2–4.0)
Completeness (%)	98.5 (99.9)	99.9 (100.0)
I/σ(I)	18.02 (1.24)	11.5 (2.6)
No. of measured reflections	51889 (5561)	45757 (6834)
Redundancy	7.1 (7.4)	21.2 (23.0)
Anomalous multiplicity		12.7 (13.0)
R_{SYM} (%)	7.3 (262.6)	14.1 (145.0)
$CC_{1/2}$ (%)	99.9 (52.7)	100 (74.6)
Heavy atom sites		2
FOM (after Phaser)		0.24
FOM (after Resolve)		0.58
R_{work}/R_{free} factor (%)	22.23/27.01	
N°. of non hydrogen atoms		
Macromolecules	1063	
Water	4	
Wilson B-factor (Å^2)	90.1	
Average B-factor (Å^2)	126.6	
RMSD		
Bonds (Å)	0.005	
Angles (°)	0.616	
Ramachandran (%)		
Favored region	97.5	
Allowed region	2.5	
Outliers region	0.0	

PHI-BLAST was also used to search templates for the C-terminal part of YabA. The templates selected by the BLAST approach were aligned with YabA sequence using ISPALIGN (Intermediate Sequence Profile Alignment) (29). A manual alignment adjustment was performed with the help of Geneious Pro 4.8.4 software (Biomatters Ltd). The identity and similarity between YabA-CTD and its templates (aminoacids 76–119) were computed using SIAS tool (http://imed.med.ucm.es/Tools/sias.html) with a BLOSUM62 matrix. Based on alignment results, a 3D model of YabA was obtained by homology modeling using the software package Modeller 9v8 (30). Image rendering and H95 position refinement were performed using PyMOL (DeLano Scientific LLC) sculpting properties. The Qmean package was used to evaluate model quality (http://swissmodel.expasy.org/qmean) (31,32).

SAXS data collection

Protein samples were centrifuged for 10 min at 10 000 rpm prior to X-ray analysis in order to eliminate all aggregates. Sample concentration was measured by UV absorption at λ = 280 nm on a Thermo Scientific NanoDrop 1000 Spectrophotometer. For each sample, a stock solution was prepared at a final concentration of 8 mg/ml and stored at 4°C and then directly used for the experiments. SAXS experiments were conducted on the SWING beamline at the SOLEIL synchrotron (λ = 1.033 Å). The Aviex charge-coupled device detector was positioned to collect data in the q-range 0.006–0.5 Å^{-1} ($Q = 4\pi\sin\theta.\lambda^{-1}$, where 2θ is the scattering angle). The solution was injected in a fixed-temperature (15°C) quartz capillary with a diameter of 1.5 mm and a wall thickness of 10 μm, positioned within a vacuum chamber. A total of 80 μl of monodisperse protein sample were loaded onto a size-exclusion column (SEC-3 300 Agilent), using an Agilent HPLC system and eluted directly into the SAXS flow-through capillary cell at a flow rate of 0.2 ml min^{-1} (33). The elution buffer consisted of 50 mM Tris–HCl pH 8, 150 mM NaCl, 5% glycerol filtered and degassed. SAXS data were collected continuously, with a frame duration of 1.5 s and a dead time between frames of 0.5 s. Data reduction to absolute units, frame averaging and subtraction were done using FOXTROT, a dedicated homemade application. All subsequent data processing, analysis and modeling steps were carried out with PRIMUS and other programs of the ATSAS suite (34).

SAXS data processing, analysis and molecular modeling

The experimental SAXS data for all samples were linear in a Guinier plot of the low q region, indicating that the proteins had not undergone aggregation. The radius of gyration R_g was derived by the Guinier approximation $I(q) = I(0)\exp(-q^2R_g^2/3)$ for $qR_g < 1.0$ using PRIMUS (35). The program GNOM (36) was used to compute the pair-distance distribution functions, $p(r)$. This approach also features the maximum dimension of the macromolecule, D_{max}. The overall shapes of the entire assemblies were de-

rived from the experimental data using the program GASBOR (37). These models were averaged to determine common structural features and to select the most typical shapes using the programs DAMAVER (38) and SUPCOMB(39). An initial model of the YabA tetramer was built by assembling a tetramer of the NTD (residues 1–61) as seen in the crystal structure with four homology models of the CTD (residues 76–119). A linker of 14 residues between each NTD and CTD generated a complete full-length tetramer. To generate a model of full-length YabA using the crystal structure, and the homology model, which was compatible with the SAXS data, we performed atomic modeling using Dadimodo (40), a genetic algorithm based rigid-body refinement analysis program. The NTD and CTD structures were not modeled further but the linker region of YabA and the positions of the four CTDs were allowed to move. Continuity of the structure was assured by subsequent energy minimization using Dadimodo (40). A SAXS χ^2 value was then computed for each eligible structure, using CRYSOL. The averaged scattering curve of YabA and the fitting curve calculated from the model produced by Dadimodo were superimposed with CRYSOL (41).

Yeast two-hybrid and three-hybrid assays

Full length YabA and the C-terminal region (62–119), full length DnaA and an N-terminally truncated DnaA fragment (71–440), and DnaN proteins were expressed as fusions to the GAL4 binding domain BD or activating domain AD from the vectors pGBDU-C1 and PGAD-C1, respectively. pGBDU and PGAD derivative constructs were introduced by transformation into PJ694- (α) and (a) haploid strains, respectively. Binary interactions were tested by combinational mating of the strains expressing the BD and AD fusions as previously described (15). Interacting phenotypes are tested by the ability of the diploid cells to grow on selective media SD-LUH and SD-LUA. For 3HB experiments, DNA fragments encoding full length YabA or the C-terminal domain were inserted into the p3HB vector (18). The p3HB-YabA derivatives were co-expressed with either DnaA or DnaN in haploid (α) strains and the strains expressing the different combination pairs are then mated again with haploid (a) strains expressing the appropriate AD-fusion.The ability of a given BD-protein fusion to interact with a given AD-protein fusion conditional upon the presence of a third protein is attested by the appearance of growth on SD-LUWH and SD-LUWA selective media as previously described (18).

Site targeted mutagenesis and analysis of Loss-of-Interaction (LOI) phenotypes by yeast-two hybrid

Site-directed saturation mutagenesis of *yabA* was performed by PCR amplification and fragment joining using degenerate oligonucleotides containing a randomized codon at the targeted position. The mutated PCR amplified *yabA* coding sequences were cloned into the pGBDU vector, in frame with the BD domain of GAL4 as previously described (15). Mutations were identified by sequencing and pGBDU-*yabA* derivatives carrying mutations at the targeted codon leading to different amino-acid substitutions were introduced into the pJ69–4 (a) yeast haploid strain by transformation. Yeast haploid cells expressing the BD-*yabA* mutant derivatives were mated with haploid strains of the opposite mating type (pJ69–4 (α)), harboring the pGAD-partner constructs encoding interacting proteins DnaA, DnaN or YabA fused to the AD domain of GAL4, as previously described (15). Interacting phenotypes were assessed according to the ability of the diploid forms to grow on (-LUH and -LUA) selective media. Mutants were tested for their ability to express interacting phenotypes with each of the three partners. The screening for loss-of-interaction (LOI) mutations against a single partner while maintaining the ability to bind to the other partners precludes the isolation of mutations that destabilize the tertiary structure of YabA.

Construction of GFP fusions

Wild-type and mutant *yabA* genes were PCR amplified from the corresponding pGBDU-*yabA* constructs for subsequent cloning into the vector pGS1190 to generate *cfp-yabA* fusions under the control of the xylose-inducible promoter P_{xyl} (42). The *cfp-yabA* constructs were then integrated into the *amyE* locus of the *B. subtilis* Δ*yabA* strain, JJS142 (Supplementary Table S4). For the fluorescence complementation assay, genomic DNA from a strain expressing the *yfp-yabA-N85D* fusion under the control of its native promoter at the locus position, (17) was used to transform the recipient strains carrying the xylose inducible *cfp-yabA* constructs at *amyE* locus (see Supplementary Table S4), using chloramphenicol (5μg/ml) for selection.

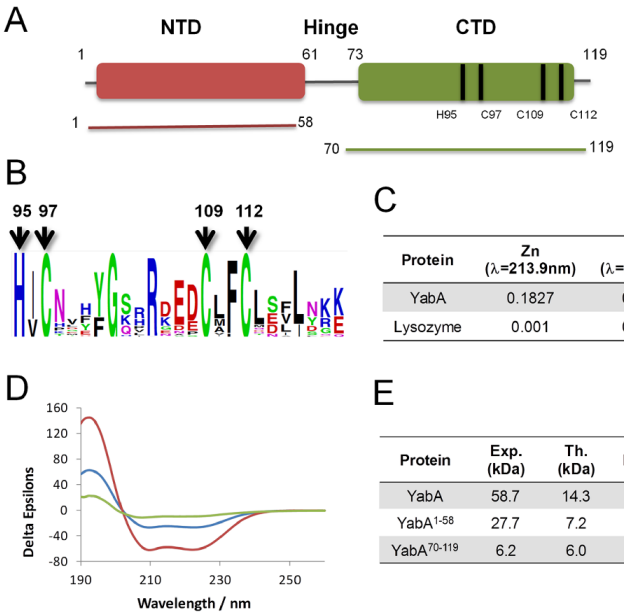
Fluorescence microscopy

Cells expressing the CFP-tagged YabA mutant derivatives and cells co-expressing CFP-YabA and YFP-YabA were grown and treated as previously described (15). For microscopic observations, the cells were mounted on agarose slides (43). Images were acquired by using an HQ snap digital charge-coupled device camera and analyzed using Metamorph v3 software. Appropriate filter sets to visualize the CFP and YFP fluorescence signals were obtained from Leica/WWR. DNA was stained with DAPI, and membranes were stained with FM5–95 (Molecular Probes).

RESULTS

YabA is a tetrameric, multidomain Zn-binding protein

Homology-based sequence analysis suggested that YabA consists of N-terminal (residues 1–56) and C-terminal (residues 73–119) structural domains separated by a region of low complexity and poorly conserved sequence, which we will refer to as a 'linker' or 'hinge' (Figure 1A, S1). The N-terminal region includes a conserved leucine heptad repeat (residues 6–48) and a regular distribution of charged residues, pointing to a potential leucine zipper-like domain. The C-terminal region exhibits higher sequence conservation, with three cysteine residues (C97, C109, C112) and one histidine (H95) being invariant in YabA sequences, suggesting they participate in the coordination of a zinc ion



Protein	Zn (λ=213.9nm)	Ca (λ=422.7nm)
YabA	0.1827	0.0046
Lysozyme	0.001	0.0022

Protein	Exp. (kDa)	Th. (kDa)	Multimers
YabA	58.7	14.3	4
YabA$^{1-58}$	27.7	7.2	4
YabA$^{70-119}$	6.2	6.0	1

Figure 1. (**A**) Schematic representation of YabA domain organization with the NTD (Indian red) and CTD (green) protein constructs used in the study. Bars indicate the position of Zn binding residues. (**B**) Web logo showing conservation of amino acid residues in the C-terminal region of YabA. Conserved histidines and cysteines are indicated by arrows. (**C**) Atomic Absorption Spectroscopy of YabA in solution. Zn measurements by AAAF. (**D**) Circular dichroism spectroscopy. Spectra were recorded for YabA (blue), YabA$^{1-58}$ (red) and YabA$^{70-119}$ (green). All three spectra exhibit features characteristic of a folded protein, with increasing molar ellipticity with decreasing wavelength below 205 nm and minima over the wavelength range 205–225 nm indicating a preponderance of α-helices. (**E**) Molecular mass measured from SEC-MALS analysis.

(Figure 1A and B and Supplementary Figure S1). A PSI-BLAST search of non-redundant sequence databases using the C-terminal region of YabA as seed allowed us to assign a YabA-like function to 46 additional proteins harboring the HCCC motif in firmicutes (Supplementary Table S1).

The presence of zinc atoms in purified *B. subtilis* YabA was thus studied by spectroscopic methods. Atomic absorption spectra of the purified protein dissolved in deionized water revealed strong absorption at 231.9 nm diagnostic of the presence of zinc (Figure 1C). Quantitative measurements calibrated against a zinc standard curve gave a zinc:YabA molar ratio of 0.94 indicating that one zinc ion is bound per YabA chain. These data suggest that YabA defines a novel class of bacterial Zn binding proteins in which the HCCC motif could be used to coordinate one Zn per protein chain.

To assess the overall secondary structure composition, circular dichroism spectra were recorded for purified YabA, as well as NTD and CTD recombinant fragments, YabA$^{1\ 58}$ and YabA$^{70\ 119}$ (Figure 1D). The spectra for all three species exhibit shallow minima in their molar ellipticity over the wavelength range 205–225 nm and increasing molar ellipticity with decreasing wavelength below 205 nm. These strong secondary structure characteristics indicate that the proteins are folded. Analysis of these spectra using Dicroweb confirmed this conclusion and indicated that the secondary structures of full length YabA and its domain fragments feature a high percentage of α-helices. To determine the quaternary structures of YabA, YabA$^{1\ 58}$ and YabA$^{70\ 119}$, the proteins were next analyzed by size exclusion chromatography and multi-angle laser light scattering (SEC-MALS). In these experiments, the protein samples are analyzed on a gel-filtration column and the absorbance at 280 nm and the refractive index of the eluting species are monitored together with the multi-angle laser light scattering of the sample. YabA eluted as a single peak from an S200 column with a retention time of ∼26.5 min (Figure 1E and Supplementary Figure S2) and a molecular mass of 58.7 kDa. As the theoretical molecular mass of YabA is 14.3 kDa, we concluded that recombinant YabA is a tetramer (57.2 kDa) in solution as previously reported (15). The N-terminal fragment YabA$^{1\ 58}$ had a retention time of ∼21.5 min on an S75 column and this peak was associated with an Mw of 27.7 kDa (Figure 1E and Supplementary Figure S2). This value

is in agreement with the mass calculated for a YabA$^{1\ 58}$ tetramer (28.8 kDa). Interestingly, YabA$^{70\ 119}$ eluted from the same column as a single peak with a retention time of ∼28 min and an associated Mw of 6.2 kDa (Figure 1E, Supplementary Figure S2), indicating a monomeric state for this C-terminal domain fragment (theoretical molecular mass of 6.0 kDa). We conclude that YabA tetramer formation relies predominantly on the N-terminal portion of the molecule.

Crystal structure of YabA NTD

We attempted to crystallize YabA and obtained two crystal forms (form I and II, see 'Materials and Methods' section for details). In both cases, analysis of the unit cell contents and lattice interactions indicated that proteolysis occurred during crystallization. The final model allowed us to delineate the NTD boundary of the crystallized fragment as residing around residue F62 in the form I crystals, fragment that will be thus described as YabA$^{1\ 62}$. The Form I and II crystals each contain two molecules of YabA$^{1\ 62}$ per asymmetric unit arranged almost identically. The structure obtained with form I crystals was refined to a final resolution of 2.7 Å with good geometry (Table 1). The YabA$^{1\ 62}$ protomer consists of a single 90 Å long α-helix comprising 15 turns with a pronounced curvature at its center (Figure 2A). YabA$^{1\ 62}$ assembles as a dimer of dimers in the crystal consisting of chains A and B and their symmetry equivalents A' and B' (Figure 2B). The A'B dimer resembles a pair of tweezers (Figure 2C) assembled via tight interactions of the C-terminal elements of the two helices whose arrangement can be described as a parallel coiled-coil. Two pairs of these tweezers (AB' and A'B) embrace each other to form the head-to-head tetramer (Figure 2D). The tetramer thus comprises a central four-helix bundle flanked by pairs of helical extensions arranged as coiled coils (Figure 2B and D). The packing of the molecules appears to be tight with ∼40% of the surface area of each chain buried by intermolecular interactions, with each subunit forming extensive contacts with its three neighbors. As a result, there is a total buried surface area of 9880 Å^2 in the tetramer. The four-helix bundle comprises residues 1–35 of the four chains with chains A and B' running antiparallel to chains A' and B. The interior of the bundle is formed by the close packing of the aliphatic side chains of L6, V10, L13, I17, L20, L24, L27, I31 and M34 from all four chains creating a substantial hydrophobic core (Figure 2E). The heptadic arrangement of L6, L13, L20 and L27 gives rise to a leucine zipper-type interdigitation of side chains projecting from pairs of adjacent chains (A-A' and B-B'). Sets of reciprocal charge-charge interactions (four in total) between the E14 and K28 side chains on the A chains and hydrogen bonds between Q16 and Q23 side chains on the B chains presumably lend further stability to the tetramer (Figure 2E).

The flanking coiled-coil regions are formed by the C-terminal halves of chains A and B' on one side of the four helical region and chains A' and B on the other (residues 36–62). These pairs of helices, which are splayed in the four helix bundle region, converge at their C-termini (Figure 2C) where hydrophobic interactions of the side chains of M34, L41, L48, L52 and I59 with the equivalent residues in the partner chain are augmented by hydrogen bonding interactions of the side chains of N38, N45, T55 and most strikingly by reciprocal salt-bridging interactions between the side chains of E44 and R49 (Figure 2F). The crystal structure of YabA$^{1\ 62}$ is consistent with the circular dichroism data which suggested a high proportion of α-helix in the YabA$^{1\ 58}$ fragment and the SEC-MALS data which clearly point to a tetramer. The structure also shows that residues 59–61 previously assigned to the linker region can adopt a α-helical conformation.

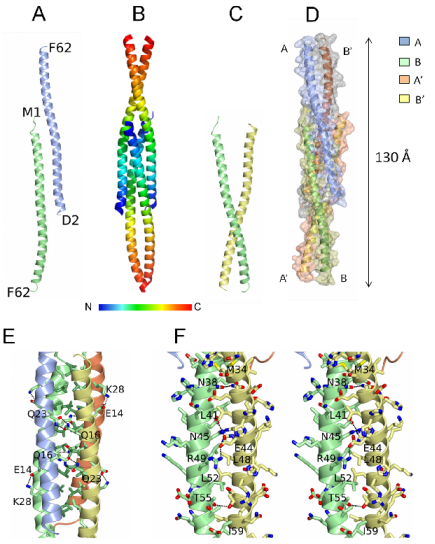


Figure 2. Structure of the YabA-NTD tetramer. (**A**) Ribbon representation of the two chains of YabA$^{1–62}$ in the asymmetric unit A (ice blue) and B (light green) with the N- and C-terminal residues labelled. (**B** and **D**) The YabA$^{1–62}$ tetramer. Each chain is shown as a ribbon either colored ramped as shown in the key (**B**) or colored by chain (**D**) with an accompanying surface rendering. (**C**) Ribbon representation of molecules A' and B illustrating the tweezer arrangement of these chains. (**E**) The four helix bundle region. The chains are displayed as ribbons colored by chain A (ice blue) B (light green) A' (lemon) and B' (coral). The Cα and side chains of apolar residues in the interior are colored by residue type, light green for Val and Ile and lawn green for Leu to emphasize the leucine zipper. Polar interactions between the chains are shown as dashed lines and residues forming the prominent interchain interactions referred to in the text are labeled. (**F**) Stereo image of the coiled coil region of chains B and A'. The Cα and the side chains of residues from the two chains are displayed with the carbon atoms colored according to the chain and with nitrogens in blue, oxygens in red and sulphurs in yellow. Polar interactions between the chains are shown as dashed lines.

YabA CTD *in silico* structure model

In the absence of a crystal structure for the CTD, we searched for relevant structural templates to build a three-dimensional model of YabA CTD. Using a PSI-BLAST

search we first identified *Saccharomyces cerevisiae* profilin (pdb code: 1YPR) which exhibited partial similarity to YabA CTD although it does not contain a zinc finger motif (44). In order to refine the template identification for the zinc finger motif, a hypothesis-driven pattern search was performed. Protein of known structures containing variations of the zinc-binding motif of YabA H-x-C-x(11)-C-x(2)-C were identified using PHI-BLAST (45). These include a series of zinc finger proteins that belong to the PHD, RING, CHY and FYVE domain families (Supplementary Table S2). The structure of the endosome targeting protein EEA1 (pdb code: 1JOC) (46) was of particular interest since the sequence similarities extend beyond the FYVE zinc-binding domain to the linker region and the NTD of YabA (Figure 3A and B). Based on a multiple alignment of the combination of homologous regions of interest from the structural templates, a 3D-model of YabA CTD was built using constraint-based homology modeling procedures implemented in Modeller 9v8 (24) (Figure 3B). To evaluate the quality of its fit to experimental structures the model was next validated by QMEAN (32). The determined Qmean Z-scores were -1.59 for CTD spanning amino acid residues 76–119 and −3 for the region corresponding to model structure from residues 63 to 119. This indicated that the structural features of the YabA CTD model were of a quality comparable to high resolution structures (32). In addition, the YabA CTD (residues 76–119) sequence has 44% similarity and 35% identity with profilin, and accordingly the YabA model exhibited two short alpha-helices (Figure 3A, α2 and α3). Structural alignment with EEA1 revealed 12% identity and 19% similarity with YabA CTD. Similar to EEA1, the YabA zinc finger residues H95 and C110 (Figure 3A, upper case residues) and C97 and C112 (Figure 3A, lower case residues) were, respectively, inaccessible and accessible to solvent (Figure 3A, orange dots). These results stress the structural relevance of the two templates used here to build the YabA CTD model. The 3D model consists of two short helices (α2 and α3) followed by zinc binding loops (Figure 3B). Interestingly, the model positions the residues of the HCCC in a fold similar to the short zinc-binding loops found in other proteins (47).

Next, the model was assessed using knowledge of surface exposed residues. A previous functional dissection of YabA identified single residue changes that selectively disrupted interactions with one partner protein, while maintaining interactions with its other protein partners (15). In view of the selective effects of these LOI mutations, the substituted residues are expected to be exposed on the surface where their mutation is unlikely to affect the overall structure and stability of the protein. The sites of the LOI residue substitutions were mapped onto the YabA model (Figure 3C). The residues important for interaction with DnaA and DnaN were found to be surface exposed, defining non-overlapping interacting surfaces on the C-terminal zinc-binding domain. Residues Y83, N85 and L86, important for interaction with DnaA, are clustered on the helical side of the CTD surface when viewed as in Figure 3C while residues V99, H100 and L110 involved in interaction with DnaN are grouped predominantly on a surface formed by protein loops. This mapping of residues important for interaction with DnaA and DnaN to the surface of the YabA CTD provides an important validation of our *in silico* structural model.

Low-resolution structure of the YabA tetramer

To gain insight into the overall architecture of the full-length tetrameric YabA, we used SAXS. YabA eluted from the online HPLC as a single peak and SAXS data indicated a radius of gyration (R_g) of 63.2 Å and a maximum dimension (D_{max}) of 250 Å (Figure 4A). These values are very high for a low molecular weight protein, even for a tetramer. The presence of a flexible region is suggested by the shape of the Kratky plot which displays a flattened bell shape at low values of s^2 corresponding to the folded part of the protein, followed by a continuous rise at the higher angle sections (48) (Supplementary Figure S3A). Such dual behavior is indicative of flexibility in multidomain proteins (49). An extended YabA structure is also strongly suggested by the autocorrelation function with a shape characteristic of an elongated object (Figure 4A). In the crystal structure of the NTD, the C-termini of chains A and B' (residues 62) are separated by ∼130 Å from the C-termini of chains A' and B. The CTD model encompassing residues 76–119 is of around 25 Å at its maximum radius. Coupled with the observation from the SEC-MALS experiments that $YabA^{70\ 119}$ is monomeric, these data suggest that residues 63–75 may act as an extended flexible linker between the NTD and the CTD. This is also supported by the fact that residues 60–80 are poorly conserved in YabA sequences (Supplementary Figure S1).

Using the scattering curve, a low-resolution structure of YabA was calculated by *ab initio* modeling with the program GASBOR without imposing symmetry constraints. In this case, we obtained a set of very different shapes (data not shown). To limit the range of possibilities, internal P_{22} symmetry was imposed, giving a shape composed of a cylindrical core with pairs of elongated arms attached to each end (Figure 4B). To gain better insights into the structure of the tetramer of YabA, we combined the crystal structure of the YabA NTD tetramer, with the CTD structure obtained by molecular modeling. Each N-terminal helix was connected to its cognate CTD domain by a linker composed of 14 residues (62–76). Modeling was performed with DADIMODO which allows a flexible linker region to be defined in the model while maintaining structural constraints imposed by molecular modeling processes and the SAXS data (see Materials and Methods, (40)). The model produced by DADIMODO fits the experimental scattering pattern (Figure 4C) and exhibits the same global organization as that obtained by *ab initio* modeling (Figure 4B and D). Individual runs revealed different solutions with variable positions of the CTDs, suggesting an intrinsic flexibility of the linker (Supplementary Figure S3B). Nevertheless, the *ab initio* and structure-guided modeling approaches converge to a common architecture for YabA: a core of four NTD α-helices each of which is connected to a globular CTD by a linker region in an extended conformation. The SAXS data and modeling experiments suggest that the linker confers a high degree of flexibility, leading to a multiplicity of conformations of the individual CTDs relative to the tetrameric core bundle Supplementary Figure S3B).

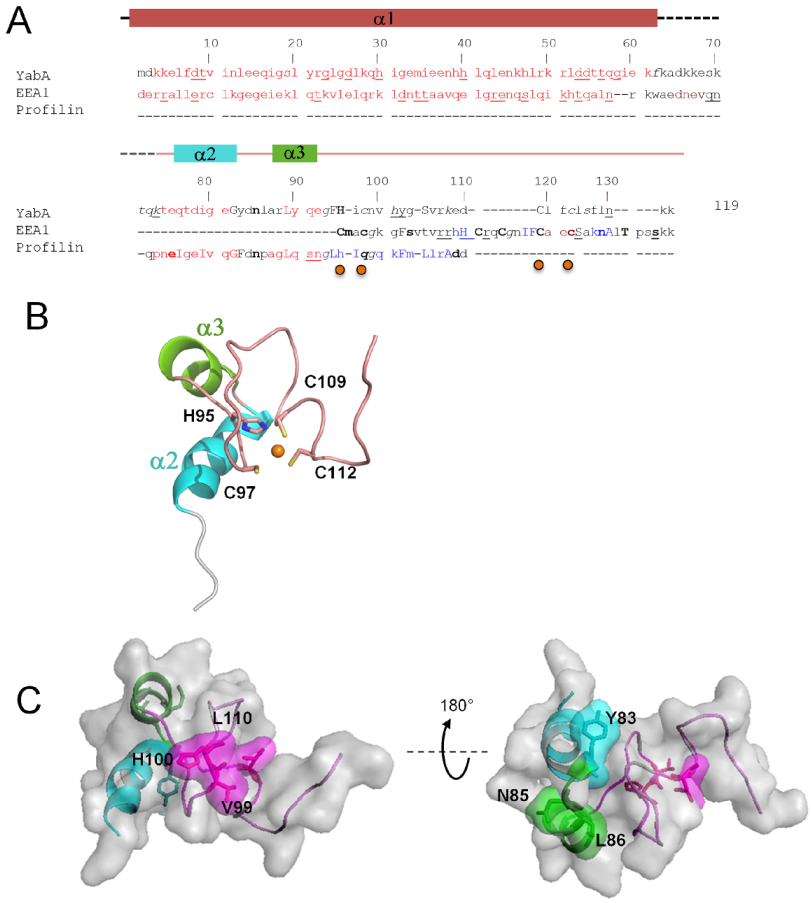


Figure 3. Homology model of the YabA-CTD monomer. (**A**) Structural sequence alignment of YabA model with the identified homologs EEA1 (pdb code 1JOC) (46) and profilin (1YPR) using the Joy algorithm (63), highlighting the structural similarities between YabA model and it's templates. Alpha helix (red); 3^{10} helix (maroon); beta strand (blue); solvent accessible residues (lower case); solvent inaccessible residues (upper case); hydrogen to bond main-chain amide(bold), hydrogen bond to main-chain carbonyl (underline). Secondary structures of YabA are indicated on top and correspond to the crystal structure (residues 1–62) and the homology model (76–119). The residues involved in Zn coordination are indicated by orange dots. (**B**) Cartoon representation of YabA CTD model colored as in (A). Side chains of the residues participating in Zn binding are indicated as ball and sticks. (**C**) Surface representation of YabA-CTD revealing surface-exposed residues important for interaction with DnaN (left) and DnaA (right) identified in a previous study (15).

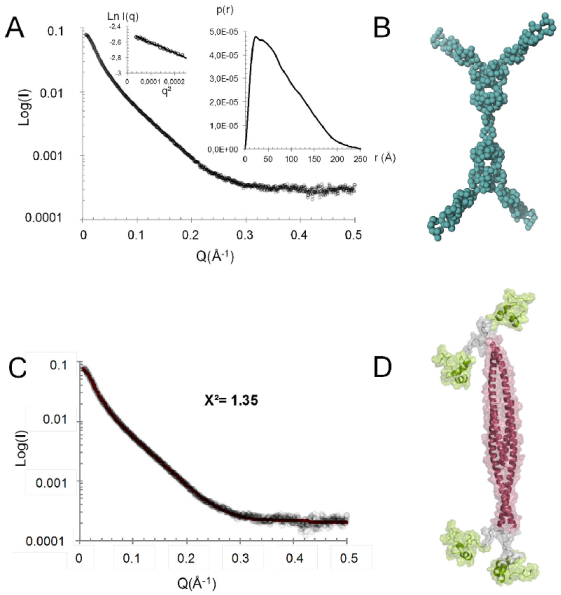


Figure 4. (**A**) Averaged scattering curve of the YabA protein coming from frames recorded on the main peak from SEC. The values of the radius of gyration Rg and the maximum dimension, Dmax, of the scattering particle are calculated from Guinier extrapolation (first box image) and autocorrelation function p(r) determination (second box image). (**B**) Bead model of YabA compatible with the SAXS data. (**C**) Averaged scattering curve of the YabA protein and the fitting curve calculated and superimposed with CRYSOL, coming from the model produced by Dadimodo under SAXS constraint. The logarithm of scattering intensity I is displayed as a function of the momentum transfer $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and $\lambda = 1.033$ Å is the wavelength. (**D**) YabA structure model.

Mutually exclusive interactions of DnaA and DnaN to the YabA C-terminal moiety

Previous studies revealed that YabA forms heterocomplexes with DnaA and DnaN *in vivo* (15,16). YabA was also shown to bridge DnaA and DnaN in a yeast three-hybrid assay (18). It was further shown that YabA LOI-mutations disrupting interactions with DnaA or with DnaN mapped to the CTD (15). Our structural and biophysical data established that the isolated CTD is monomeric, suggesting that these domains may function as independent binding sites in the YabA tetramer. We tested this assumption by determining whether YabA$^{62\ 119}$, comprising the CTD globular domain together with the linker, was sufficient for interaction with DnaA or with DnaN in a yeast two-hybrid assay (Supplementary Figure S4A). We found that indeed, YabA$^{62\ 119}$ was able to support interaction with either partner, implying that the NTD's quaternary structure is dispensable. This scheme raised the important question of whether a single CTD monomer could form simultaneous interactions with both DnaA and DnaN or whether the interactions with the two proteins are mutually exclusive. To answer this question, we performed a yeast 3HB assay in which a constitutively expressed YabA$^{62\ 119}$ was tested for its ability to trigger interacting phenotypes by bridging DnaA and DnaN. We found that in contrast to the full length YabA, YabA$^{62\ 119}$ was not able to bind simultaneously to DnaA and to DnaN whether they are fused with to the BD or the AD domains of Gal4. This observation sheds new light on the mode of binding of YabA to its partners. We conclude that DnaA and DnaN are able to bind concurrently to a YabA tetramer but not to a single CTD.

Structural and functional determinants underlying the binding of YabA to DnaA and DnaN

Next, the model of the YabA CTD was used to predict additional residues involved in interactions with the partner proteins DnaA and DnaN. The predictions were based on residue surface accessibility calculations, sequence conservation analysis, comparison with templates and previous work (15). Candidate residues were targeted for scanning mutagenesis and the interaction profiles of the corresponding YabA mutants were determined experimentally in a yeast two-hybrid assay, similar to that previously used to identify LOI mutations (15).

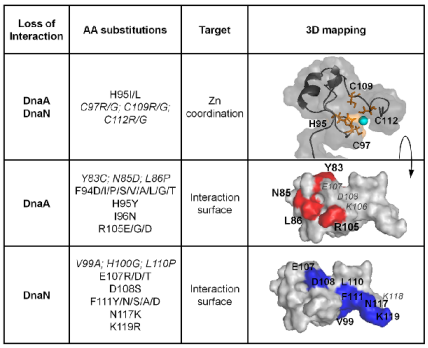


Figure 5. Surface representation of YabA showing key residues involved in interactions with partner proteins. Mapping of mutational data onto the structural model with interacting surfaces specified by color; DnaA (Blue) DnaN (red). Previous data is indicated in italic. Zn is illustrated by a cyan sphere. Residues with no interacting phenotypes are labeled in gray.

Previous work showed the importance of residues C97, C109 and C112 of the HCCC motif for interactions with both DnaA and DnaN (15). To experimentally investigate the role of the H95 of this motif, we tested the effects of targeted substitutions on the capacity of YabA to bind to its different partners. Replacement of H95 by I or L resulted in loss of interaction with both DnaA and DnaN while YabA–YabA self-interactions were preserved (Figure 5, Supplementary Table S3 and Figure S5). Similar effects were observed previously upon mutation of the conserved cysteine residues (15). This suggests that substitution of the polar H95 by hydrophobic residues results in structural defects that are confined to the CTD. Curiously, a H95G substitution did not exhibit an interaction defect in our yeast two hybrid assay (Supplementary Table S3). However this mutant, like the previously examined C97A mutant, exhibited a propensity to aggregate during expression in *E. coli* cells (data not shown). This is consistent with reduced stability caused by loss of Zn binding, although the direct participation of H95 in Zn coordination is not yet proved. Substitutions of the neighboring residues, F94 and I96 also impeded interaction with DnaA but had only moderate effects on interaction with DnaN (Supplementary Table S3). The hydrophobic residue F94, is highly conserved as a F or Y among YabA homologs and is buried in our model (Supplementary Figure S1). The F94Y mutation in YabA did not affect protein interactions. In contrast, substitution of F94 by a variety of polar and aliphatic residues led to specific defects in DnaA interactions (Figure 5, Supplementary Table S3 and Figure S5). These results suggest that the H95 flanking residues, F94 and I96, are also important for the folding of the CTD. Together these results support the importance of a structured domain in which Zn-coordination might play a role.

To further delineate the DnaA binding surface, we targeted residues R105 and K106, as these neighbor the surface-exposed residues N85 and L86 (previously shown to be important for DnaA binding). YabA mutants R105E, R105D and R105G lost the ability to bind to DnaA. However substitution of K106 by a range of amino acid residues had no effect on DnaA binding (Figure 5, Supplementary Table S3 and Figure S5). This suggests that R105 but not K106 is part of the DnaA binding site (Figure 5 and Supplementary Table S3). YabA substitutions of the neighboring residues E107and D108 did not exhibit LOI phenotypes for DnaA, indicating these residues are not part of the DnaA binding surface. Instead these substitutions produced a DnaN LOI phenotype. This suggests that the two interacting surfaces may be overlapping.

To further delineate the DnaN interaction surface, mutagenesis of F111, N117 and the terminal K118 and K119 residues was performed. Whilst, the substitution of K118 by various amino-acid residues had no effect on protein interactions, mutations targeting residues F111, N117 and K119 all resulted in DnaN LOI mutants (Figure 5, Supplementary Table S3 and Figure S5). The location of residues F111, N117 and K119 in the CTD model are consistent with their role in DnaN binding as they are predominantly surface exposed. Moreover, these residues are adjacent to V99 and L110 previously found to be important for DnaN binding. Taken together, these data clearly delineate the interaction determinants of DnaA and DnaN on two partially overlapping surfaces on the YabA CTD.

Effect of loss-of-interaction mutations on sub-cellular localization in *B. subtilis* cells

YabA forms a heterocomplex with DnaA and DnaN which co-localizes with the replication machinery for most of the cell cycle (15,17). To validate our structure predictions, we investigated the subcellular localization of several $YabA_{DnaA}$ and $YabA_{DnaN}$-LOI mutants in living cells by transferring the corresponding point mutations into a *yabA* gene fused in-frame to *cfp or yfp.* Examination of the cellular protein levels revealed that the YabA mutant derivatives were present at similar or higher levels compare to the wild-type (Supplementary Figure S6). Wild-type YabA localizes as one or two foci per cell, with single foci displaying a midcell localization (Figure 6, (15,17)). LOI with either DnaA or DnaN is associated with both loss of YabA localization and loss of initiation control (15). Here, we observed that F94S and R105E substitutions in a CFP-YabA fusion give a fluorescence signal dispersed throughout the cell, similar to that of the YFP-YabA-N85D mutant identified in a previous functional study (15) (Figure 6A and Supplementary Figure S7). Since F94S and R105E substitutions specifically impaired the binding of YabA to DnaA (but not DnaN) in the yeast two-hybrid assays, we conclude that these two residues contribute to the integrity of the YabA/DnaA interface together with residues N85 and L86 identified formerly. The R105E and F94S substitutions reduce but do not completely abolish foci formation since patches of increased fluorescence intensity can still be discerned in parts of the cells, in agreement with the observation of Goranov *et al.* that YabA foci can be detected in cells deficient in DnaA-mediated initiation of chromosomal replication (50). Thus, while these two substitutions cause

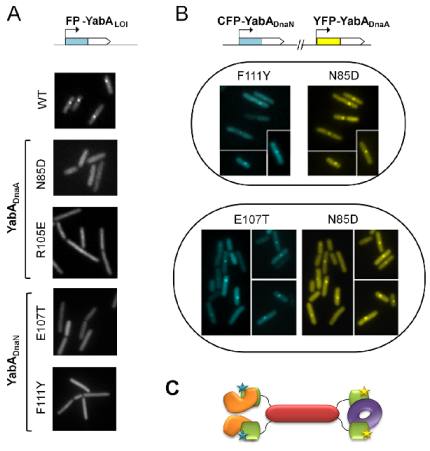


Figure 6. Cellular localization of YabA-WT and -LOI mutants by fluorescence monitoring of signals from CFP- (cyan) or YFP- (yellow) YabA fusions in living cells. (**A**) Impaired localization of YabA LOI mutants in DnaA (N85D and R105E) and in DnaN (E107T and F111Y) compared to YabA-WT. (**B**) Restoration of localization of $YabA_{DnaN}$-LOI mutants by cross-complementation with a YabA-$_{DnaA}$-LOI mutant. The co-expression in the same cell of the YabA-$_{DnaN}$ mutants F111Y or E107T fused to CFP (blue) with a YabA-$_{DnaA}$ N85D fused to YFP (yellow) restored the ability of each YabA mutant derivative to localize as a focus at the cell centre. The co-localization of foci indicates that they are part of a DnaA/YabA/DnaN complex, where the YabA tetramer is potentially composed of a mixture of YabA-$_{DnaN}$ (proficient for binding to DnaA) and YabA-$_{DnaA}$ (proficient for binding to DnaN) mutants. (**C**) Model illustrating cross complementation between YabA-$_{DnaA}$ (CTDs in green with yellow stars) and $YabA_{DnaN}$ (CTDs in green with blue stars) for the binding and the cellular localization of DnaN (purple doughnut) and DnaA (orange). The YabA NTD's tetrameric core is represented in Indian red.

significant dispersal of YabA throughout the cell, residual YabA co-localization with the replication machinery could still be observed. Examination of CFP-YabA fusions carrying the DnaN-LOI substitutions E107T, E107R, D108S or F111Y, revealed that foci formation was completely abolished as observed previously for DnaN-LOI substitutions of V99 and L110 (Figure 6A, Supplementary Figure S7). These observations support the assignment of these residues to the DnaN interacting surface. Overall, the data are consistent with the predictions of the structural model of YabA.

When co-expressed in the cell, DnaA- and DnaN-interaction deficient mutants of YabA can cross-complement and restore YabA cellular localization and function (15). Here we tested the capacity of the DnaA-LOI YabA mutant N85D of YabA, to restore the localization of the DnaN-LOI mutants (F111Y and E107T). The CFP-YabA-F111Y (or E107T) and the YFP-YabA-N85D fusion proteins are primarily dispersed in the cell when expressed alone (Figure 6A), but re-localized as sharp mid-cell foci containing both CFP and YFP fluorescence when co-expressed in the same cell (Figure 6B). This functional complementation of DnaA and DnaN LOI mutations indicates that the mutant YabA proteins are properly folded in the cell, and that the interaction deficiency of each mutant is complemented by the other in a functional heterocomplex that localizes correctly at mid-cell (Figure 6C). These results fully support the assignment of residues E107 and F111 to the DnaN-binding surface thus lending further experimental support to the structural model. These observations suggest that the CTD harbors the determinants for association with the replication machinery while the N-terminal domains account for the quaternary structure of YabA perhaps allowing additional levels of regulation of its activity.

DISCUSSION

In *B. subtilis*, YabA is proposed to exert its negative control of replication initiation by concurrent mechanisms at work at the replication origin and at the replication factory. YabA inhibits the formation of a DnaA-nucleoprotein structure at *oriC* by binding to free DnaA thus reducing its availability in the cell (13,14). YabA also tethers DnaA with the β-clamp DnaN in a heterocomplex, which co-localizes with the replication factory during most of the replication cycle (15–16,18). Thus, YabA acts to prevent premature re-initiation by trapping DnaA in a DnaN-independent or a DnaN-dependent manner, respectively. The question of how YabA coordinates its actions to allow the transition from initiation to elongation of replication remains obscure and a better understanding of the structural organization and dynamic character of the complexes it forms with DnaA and DnaN is needed.

Our study establishes that YabA assembles as a head-to-head dimer of dimers in solution, through coiled-coil interactions of its helical NTDs. The C-terminal portions of YabA fold as independent domains linked to their cognate NTDs by a short flexible region. YabA CTD is a zinc-associated domain that mediates binding to DnaA and to DnaN. Analysis of the surfaces involved in binding to DnaA and DnaN revealed they are partly overlapping, thus ruling out the possibility that the two proteins bind concomitantly to the same CTD. As a result, ternary interactions of YabA, DnaA and DnaN are observed only if the full length YabA, able to self-assemble through its NTD, is expressed in the yeast 3HB assay. Importantly, the capacity of $YabA_{DnaA}$ and $YabA_{DnaN}$ LOI mutants to complement each other's localization defects and restore foci at the replication factory when co-expressed in *B. subtilis* cells provides strong evidence to substantiate the hypothesis that DnaA, YabA and DnaN can be part of the same hetero complex. Altogether our results suggest strongly that DnaA and DnaN associate with separate CTDs of the YabA tetramer.

Examination of residues important for interaction with DnaA and DnaN showed that they clustered in partially overlapping patches. Protein–protein interfaces are generally composed of different types of residues that contribute unevenly to the binding free energy. Scanning mutagenesis experiments allow three classes of residues to be defined according to their contribution to protein–protein complex formation (51). Interfaces are composed of (i) core residues, which become extensively buried upon complex formation

and whose mutation leads to drastic lowering of complex stability. These residues have been defined as 'hot spots' and they are typically highly evolutionary conserved (52); (ii) rim residues which surround the core, sheltering the latter from the solvent, and whose mutation generally tends to have modest effects on complex stability (52,53) and (iii) residues which do not contribute directly to the interface but which play an important role in defining its structural integrity. Our refinement of the DnaA interacting surface revealed novel conserved residues that fall in two of these categories. Firstly, a three amino-acid residue cluster FHI at positions 94–96, encompassing the conserved H95 of the HCCC motif, was found to be important for interaction with DnaA. We propose that this cluster is most likely involved in the stabilization of the zinc-containing fold and contributes to the structural integrity of the binding surface. Secondly, we found that the surface-exposed and positively charged residue R105, is critical for DnaA binding to YabA. R105 is strictly conserved among the YabA-family of protein. We propose that R105 constitutes a binding hotspot, in agreement with the observation that arginines are frequently represented in hot spots (52). Together with N85 and L86, R105 is part of an interacting patch highly conserved in the YabA family of proteins.

YabA exerts part of its regulatory mechanism by preventing DnaA oligomerization at *oriC* (13,14). DnaA mutant derivatives involving residues H162 and A163 were found to be insensitive to YabA control (14). These two residues are located in domain III of DnaA, which is important variously for binding to ssDNA, for ATP hydrolysis and for DnaA oligomerization. In the three dimensional structure of DnaA, H162 and A163 are in close proximity to residue F120 which is known to be important for interactions with YabA and DnaD (1,14,54). Whether these residues play a role in binding to YabA directly or indirectly remains to be established. The available structural data on YabA and DnaA from *B. subtilis* is insufficiently detailed to allow precise inferences on the YabA:DnaA interface to be drawn. However, analysis of the of a crosslinked YabA:DnaA complexes prepared by Schoefield *et al.*, revealed the dominance of a 1:1 complex stochiometry with a low proportion of 2:1 stochiometry (14). We can anticipate that YabA could bind to one DnaA unit per YabA-CTD domain.

Proteins interacting with the bacterial sliding clamp usually possess a short linear binding consensus signature QLxLF (19). Remarkably, no sequence related to this consensus is present in YabA (15). Although hydrophobic residues were identified, the composition and spatial arrangement of the surface-exposed residues involved in binding to DnaN suggest a different mode of interaction for YabA. Among the $YabA_{DnaN}$ surface residues only one, F111, is highly conserved among the YabA family of proteins. Residues at this position are strictly hydrophobic, represented by F (82%) or L (18%) (Supplementary Table S1) suggesting this might contribute substantially to the free energy of binding. In YabA, further amino-acid residues including two hydrophobic residues (V99 and L110), two acidic residues (E107 and D108) and a positively charged residue at the C-terminus (K119) have been found to contribute to binding to DnaN. These less conserved residues might be important for tuning the affinity and/or the specificity of the interaction. In summary, our results point to a YabA β-binding mode that differs from other β-binding proteins such as the replicative DNA polymerase and presumably requires a different interacting surface. This result is in agreement with the DnaN-dependent localization of YabA at the replication machinery for most of the bacterial cell cycle (15,17).

The structural characterization of YabA reveals a hub architecture. The structure of YabA is reminiscent of that of the FtsZ associated protein ZapA which is recruited to the divisome machinery to stabilize the Z-ring during bacterial cytokinesis (55–57). ZapA assembles into an elongated anti-parallel tetramer, in which two subunits associate via their coiled coil CTDs and two dimers then associate into a tetramer. At both ends of the tetrameric helical bundle, globular NTDs comprise the determinants for binding to FtsZ (56). An important aspect of such homo-oligomeric proteins is their ability to contact and connect different protein partners. In *E. coli*, ZapA recruits ZapB to the Z-ring (58,59). In *B. subtilis*, ZapA interacts with different components of the division machinery including SepF and EzrA (60). The architecture of YabA is also similar to the eukaryotic protein Geminin, which downregulates DNA replication through the timely inhibition of the licensing factor Cdt1 (61,62). Geminin forms a parallel coiled-coil homodimer that binds to Cdt1. It has been proposed that regulation of initiation is exerted though a switch in quaternary structure from a permissive $Geminin_2/Cdt1$ heterotrimer to an inhibitory $Geminin_4/Cdt1_2$ heterohexamer (61). Auxiliary basic and acidic residues in Cdt1 are involved in binding to PCNA, a parallel which may be extended to the recruitment of DnaN by YabA. Although these structural resemblances between proteins regulating replication in bacterial and eukaryotic cells do not imply similar regulatory mechanisms, they suggest that the evolutionary constraints on preventing re-replication have led to targeted interactions with key factors involved in both initiation and elongation.

Thus YabA, Geminin and ZapA represent examples of the exploitation of a coiled-coil helical core as a protein hub allowing the assembly and disassembly of protein complexes in a dynamic way. The structural organization of YabA offers a simple explanation for how it forms heterocomplexes with DnaA and DnaN. The flexible C-terminal moieties provide YabA with the dynamic capacity to interact (i) with more than one DnaA molecule, preventing the helical assembly of the latter at *oriC*, or (ii) after initiation with both DnaA and DnaN at the replication machinery. In the latter case, we anticipate that the two CTDs (say the AB' pair) at one end of the YabA tetramer could contact symmetrical binding surfaces on the ring-shaped DnaN dimer, with up to two DnaA monomers, interacting with the CTD pair (A'B) pair at the opposite end of the complex as depicted in Figure 6C. Future studies are required to gain insights into the architecture and stoichiometry of the different complexes YabA forms with its partners. Nevertheless, our experimentally informed structural model of the YabA tetramer sheds light on how YabA can mediate multiple interactions, and lays the foundation for further understanding of how YabA coordinates its actions to down-regulate replication initiation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank the following for advice and/or technical assistance: Eleanor Dodson, Andrew Leech and Johan Turkenberg at York, and Stephen McGovern and Sara Santos at INRA. We acknowledge the use of the UMS3444 Protein Science platform and the access to beamlines SWING of the synchrotron SOLEIL and ID23EH1 at the ESRF.

FUNDING

European integrated project, BaSysBio; Biotechnology and Biological Sciences Research Council, UK Studentship (to K.H.J.); EU-Marie Curie Project AMBER FP7-People 317338 (to T.G.G.); CIBLE program 2011 from the région Rhônes-alpes (to L.T., A.B.). Funding for open access charge: MICALIS INRA.
Conflict of interest statement. None declared.

REFERENCES

1. Katayama,T., Ozaki,S., Keyamura,K. and Fujimitsu,K. (2010) Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. *Nat. Rev. Microbiol.*, **8**, 163–170.
2. Konieczny,I. (2003) Strategies for helicase recruitment and loading in bacteria. *EMBO Rep.*, **4**, 37–41.
3. Nishida,S., Fujimitsu,K., Sekimizu,K., Ohmura,T., Ueda,T. and Katayama,T. (2002) A nucleotide switch in the Escherichia coli DnaA protein initiates chromosomal replication: evidence from a mutant DnaA protein defective in regulatory ATP hydrolysis in vitro and in vivo. *J. Biol. Chem.*, **277**, 14986–14995.
4. Skarstad,K. and Katayama,T. (2013) Regulating DNA replication in bacteria. *Cold Spring Harb. Perspect. Biol.*, **5**, a012922.
5. Keyamura,K., Abe,Y., Higashi,M., Ueda,T. and Katayama,T. (2009) DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J. Biol. Chem.*, **284**, 25038–25050.
6. Keyamura,K., Fujikawa,N., Ishida,T., Ozaki,S., Su'etsugu,M., Fujimitsu,K., Kagawa,W., Yokoyama,S., Kurumizaka,H. and Katayama,T. (2007) The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.*, **21**, 2083–2099.
7. Natrajan,G., Noirot-Gros,M.F., Zawilak-Pawlik,A., Kapp,U. and Terradot,L. (2009) The structure of a DnaA/HobA complex from Helicobacter pylori provides insight into regulation of DNA replication in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 21115–21120.
8. Su'etsugu,M., Takata,M., Kubota,T., Matsuda,Y. and Katayama,T. (2004) Molecular mechanism of DNA replication-coupled inactivation of the initiator protein in Escherichia coli: interaction of DnaA with the sliding clamp-loaded DNA and the sliding clamp-Hda complex. *Genes Cells*, **9**, 509–522.
9. Collier,J. and Shapiro,L. (2009) Feedback control of DnaA-mediated replication initiation by replisome-associated HdaA protein in Caulobacter. *J. Bacteriol.* **191**, 5706–5716.
10. Jameson,K.H., Rostami,N., Fogg,M.J., Turkenburg,J.P., Grahl,A., Murray,H. and Wilkinson,A.J. (2014) Structure and interactions of the Bacillus subtilis sporulation inhibitor of DNA replication, SirA, with domain I of DnaA. *Mol. Microbiol.*, **93**, 975–991.
11. Rahn-Lee,L., Merrikh,H., Grossman,A.D. and Losick,R. (2011) The sporulation protein SirA inhibits the binding of DnaA to the origin of replication by contacting a patch of clustered amino acids. *J. Bacteriol.*, **193**, 1302–1307.
12. Wagner,J.K., Marquis,K.A. and Rudner,D.Z. (2009) SirA enforces diploidy by inhibiting the replication initiator DnaA during spore formation in Bacillus subtilis. *Mol. Microbiol.*, **73**, 963–974.
13. Merrikh,H. and Grossman,A.D. (2011) Control of the replication initiator DnaA by an anti-cooperativity factor. *Mol. Microbiol.*, **82**, 434–446.
14. Scholefield,G. and Murray,H. (2013) YabA and DnaD inhibit helix assembly of the DNA replication initiation protein DnaA. *Mol. Microbiol.*, **90**, 147–159.
15. Noirot-Gros,M.F., Velten,M., Yoshimura,M., McGovern,S., Morimoto,T., Ehrlich,S.D., Ogasawara,N., Polard,P. and Noirot,P. (2006) Functional dissection of YabA, a negative regulator of DNA replication initiation in Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 2368–2373.
16. Soufo,C.D., Soufo,H.J., Noirot-Gros,M.F., Steindorf,A., Noirot,P. and Graumann,P.L. (2008) Cell-cycle-dependent spatial sequestration of the DnaA replication initiator protein in Bacillus subtilis. *Dev. Cell*, **15**, 935–941.
17. Hayashi,M., Ogura,Y., Harry,E.J., Ogasawara,N. and Moriya,S. (2005) Bacillus subtilis YabA is involved in determining the timing and synchrony of replication initiation. *FEMS Microbiol. Lett.*, **247**, 73–79.
18. Noirot-Gros,M.F., Dervyn,E., Wu,L.J., Mervelet,P., Errington,J., Ehrlich,S.D. and Noirot,P. (2002) An expanded view of bacterial DNA replication. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 8342–8347.
19. Dalrymple,B.P., Kongsuwan,K. and Wijffels,G. (2007) Identification of putative DnaN-binding motifs in plasmid replication initiation proteins. *Plasmid*, **57**, 82–88.
20. Fogg,M.J. and Wilkinson,A.J. (2008) Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.*, **36**, 771–775.
21. Karplus,P.A. and Diederichs,K. (2012) Linking crystallographic model and data quality. *Science*, **336**, 1030–1033.
22. Kabsch,W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.*, **26**, 795–800.
23. Collaborative Computational Project-4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
24. Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1948–1954.
25. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
26. DeLano,W.L. (2002) http://pymol.sourceforge.net10.1093/nar/gkv1318.html.
27. Whitmore,L. and Wallace,B.A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**, 392–400.
28. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
29. Lu,Y. and Sze,S.H. (2008) Multiple sequence alignment based on profile alignment of intermediate sequences. *J. Comput. Biol.*, **15**, 767–777.
30. Eswar,N., Webb,B., Marti-Renom,M.A., Madhusudhan,M.S., Eramian,D., Shen,M.Y., Pieper,U. and Sali,A. (2006) Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics*, Chapter 5, Unit 5.6.
31. Benkert,P., Kunzli,M. and Schwede,T. (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Res.*, **37**, W510–W514.
32. Benkert,P., Biasini,M. and Schwede,T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27**, 343–350.
33. David,G. and Pérez,J. (2009) Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *J. Appl. Crystallogr.*, **42**, 892–900.
34. Konarev,P.V., Vladimir,M.V.P., Volkov,V. and Svergun,D.I. (2006) TSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **39**, 277–286.

35. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
36. Svergun,D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503.
37. Svergun,D.I., Petoukhov,M.V. and Koch,M.H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953.
38. Volkov,VV and Svergun,D.I. (2003) Uniqueness of ab initio shape determination in small angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.
39. Kozin,M.B. and Svergun,D.I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.
40. Evrard,G., Mareuil,F., Bontems,F., Sizun,C. and Perez,J. (2011) DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J. Appl. Crystallogr.* **44**, 1264–1271.
41. Svergun,D.I, Barberato,C. and Koch,M.H.J. (1995) CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.*, **28**, 768–773.
42. Feucht,A. and Lewis,P.J. (2001) Improved plasmid vectors for the production of multiple fluorescent protein fusions in Bacillus subtilis. *Gene*, **264**, 289–297.
43. Glaser,P., Sharpe,M.E., Raether,B., Perego,M., Ohlsen,K. and Errington,J. (1997) Dynamic, mitotic-like behavior of a bacterial protein required for accurate chromosome partitioning. *Genes Dev.*, **11**, 1160–1168.
44. Eads,J.C., Mahoney,N.M., Vorobiev,S., Bresnick,A.R., Wen,K.K., Rubenstein,P.A., Haarer,B.K. and Almo,S.C. (1998) Structure determination and characterization of Saccharomyces cerevisiae profilin. *Biochemistry*, **37**, 11171–11181.
45. Zhang,Z., Schaffer,A.A., Miller,W., Madden,T.L., Lipman,D.J., Koonin,E.V. and Altschul,S.F. (1998) Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.*, **26**, 3986–3990.
46. Dumas,J.J., Merithew,E., Sudharshan,E., Rajamani,D., Hayes,S., Lawe,D., Corvera,S. and Lambright,D.G. (2001) Multivalent endosome targeting by homodimeric EEA1. *Mol. Cell*, **8**, 947–958.
47. Krishna,S.S., Majumdar,I. and Grishin,N.V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.*, **31**, 532–550.
48. Bernado,P. and Svergun,D.I. (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.*, **8**, 151–167.
49. Bernado,P. (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.*, **39**, 769–780.
50. Goranov,A.I., Breier,A.M., Merrikh,H. and Grossman,A.D. (2009) YabA of Bacillus subtilis controls DnaA-mediated replication initiation but not the transcriptional response to replication stress. *Mol. Microbiol.*, **74**, 454–466.
51. Janin,J., Bahadur,R.P. and Chakrabarti,P. (2008) Protein-protein interaction and quaternary structure. *Q Rev. Biophys.*, **41**, 133–180.
52. Bogan,A.A. and Thorn,K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1–9.
53. Moreira,I.S., Fernandes,P.A. and Ramos,M.J. (2007) Hot spots–a review of the protein-protein interface determinant amino-acid residues. *Proteins*, **68**, 803–812.
54. Cho,E., Ogasawara,N. and Ishikawa,S. (2008) The functional analysis of YabA, which interacts with DnaA and regulates initiation of chromosome replication in Bacillus subtils. *Genes Genet. Syst.*, **83**, 111–125.
55. Mohammadi,T., Ploeger,G.E., Verheul,J., Comvalius,A.D., Martos,A., Alfonso,C., van Marle,J., Rivas,G. and den Blaauwen,T. (2009) The GTPase activity of Escherichia coli FtsZ determines the magnitude of the FtsZ polymer bundling by ZapA in vitro. *Biochemistry*, **48**, 11056–11066.
56. Roach,E.J., Kimber,M.S. and Khursigara,C.M. (2014) Crystal structure and site-directed mutational analysis reveals key residues involved in Escherichia coli ZapA function. *J. Biol. Chem.*, **289**, 23276–23286.
57. Small,E., Marrington,R., Rodger,A., Scott,D.J., Sloan,K., Roper,D., Dafforn,T.R. and Addinall,S.G. (2007) FtsZ polymer-bundling by the Escherichia coli ZapA orthologue, YgfE, involves a conformational change in bound GTP. *J. Mol. Biol.*, **369**, 210–221.
58. Galli,E. and Gerdes,K. (2012) FtsZ-ZapA-ZapB interactome of Escherichia coli. *J. Bacteriol.*, **194**, 292–302.
59. Galli,E. and Gerdes,K. (2010) Spatial resolution of two bacterial cell division proteins: ZapA recruits ZapB to the inner face of the Z-ring. *Mol. Microbiol.*, **76**, 1514–1526.
60. Ishikawa,S., Kawai,Y., Hiramatsu,K., Kuwano,M. and Ogasawara,N. (2006) A new FtsZ-interacting protein, YlmF, complements the activity of FtsA during progression of cell division in Bacillus subtilis. *Mol. Microbiol.*, **60**, 1364–1380.
61. De Marco,V., Gillespie,P.J., Li,A., Karantzelis,N., Christodoulou,E., Klompmaker,R., van Gerwen,S., Fish,A., Petoukhov,M.V., Iliou,M.S. *et al.* (2009) Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 19807–19812.
62. Lee,C., Hong,B., Choi,J.M., Kim,Y., Watanabe,S., Ishimi,Y., Enomoto,T., Tada,S., Kim,Y. and Cho,Y. (2004) Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. *Nature*, **430**, 913–917.
63. Mizuguchi,K., Deane,C.M., Blundell,T.L., Johnson,M.S. and Overington,J.P. (1998) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**, 617–623.