STRUCTURAL ANALYSIS OF THE SEQA–DNA COMPLEX

STRUCTURAL INSIGHTS INTO THE ROLES OF SEQA ON ORIGIN SEQUESTRATION AND CHROMOSOME ORGANIZATION

By

YU SEON CHUNG, B.SC.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

Doctor of Philosophy (2011) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Structural insights into the roles of SeqA on origin sequestration and chromosome organization

AUTHOR: Yu Seon Chung, B.Sc. (McMaster University)

SUPERVISOR: Dr. Alba Guarné

Number of Pages: xiii, 118

Dedicated to the memory of my grandmother Kim Dan Pil, 1936-2010.

For her never ending love and encouragement.

Abstract

DNA replication is a fundamental process that must be precisely regulated to ensure timely and faithful transmission of genetic material for proliferation of all organisms. Replication initiation is regulated through a series of precisely timed protein–DNA and protein–protein interactions. In *Escherichia coli*, one regulatory mechanism of replication initiation occurs through SeqA binding to specific sequences within the *oriC*, resulting in origin sequestration. SeqA also plays a role in chromosome organization at the replication forks. Despite the functional importance of SeqA in *E. coli*, its DNA binding mechanism has remained elusive. The work described in this thesis has shown for the first time the minimal functional unit of SeqA that forms a high-affinity complex with DNA through the loss of symmetry. This is a novel observation that explains how SeqA can distinguish template versus newly replicated strand of DNA. We have also identified a protein–protein interaction surface that separates the roles of SeqA at the origin in sequestration and at the replication forks in chromosome organization. The final contribution of the thesis is in the exploration of SeqA functions in other bacterial species and demonstrating the structural and functional similarities between *Vibrio cholerae* SeqA and *E. coli* SeqA. Together our work has made a crucial connection between the structural organization of the protein and its functional ability to bind DNA.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Alba Guarné for her support, guidance and encouragement throughout the entire duration of my research, without which this thesis would not have been possible. Her passion for science is truly inspiring, which I respect and admire. I am truly thankful for your dedication and confidence in me as a developing graduate student.

I would like to thank my committee members, Dr. Justin Nodwell and Dr. Yingfu Li who have always challenged me and provided guidance and insight.

All past and present members of the Guarné lab and the Junop lab who knew how to do science and have fun, I thank you. My roommate Sara, whom I have experienced every hurdle of being a graduate student together, I will miss our rants and thank goodness for Monday night specials.

I would like to thank my parents for always believing in me and having full confidence that I could achieve anything that I wished to. I couldn't have asked for a better environment to grow up in, and I am truly grateful for all the experiences that you exposed me to. I am also privileged to have a wonderful brother who I can depend on for anything and everything. Without your support and sacrifice, it would not have been possible for me to pursue my dream.

Lastly, I would like to thank my fiancé, Andrew Gadsden, for your love and patience throughout the long journey as a graduate student. Your support during one of the most difficult times in my life is invaluable and I cannot express the gratitude I feel

for having such a supportive partner. I am glad that we got to be graduate students together, adopt and love three cats, and always remembering to enjoy life a little throughout the constant hurdles and frustrations. I look forward to a life time of happiness and love with you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

β sliding clamp	β sliding clamp subunit of DNA polymerase III
AAA	ATPases Associated with various cellular Activities
ARS	Autonomously Replicating Sequence
bp	Base pairs
Cdc6	Cell division cycle 6
Cdt1	Cdc-10 dependent transcription factor 1
ChIP-Chip	Chromatin immunoprecipitation coupled with microarray analysis
Dam	DNA adenine methylase
DARS	DnaA-reactivating sequence
datA	DnaA titration site
DiaA	DnaA initiator-associating factor
DNA	Deoxyribonucleic Acid
DnaA	DNA replication initiator protein
DnaB	Replicative DNA helicase
DnaC	Replicative DNA helicase loader
DnaG	RNA primase
DUE	DNA Unwinding Element
EMSA	Electrophoretic mobility shift assay
EcSeqA	*E. coli* SeqA
Fis	Factor for inversion stimulation
FtsK	Filamentation temperature sensitive mutant K
Hda	Homologous to DnaA
HU	Histone-like protein
IHF	Integration Host Factor
Kb	Kilo bases
MCM	Mini Chromosome Maintenance
MreB	Murein cluster e
MukBEF	Mukaku Chromosome partitioning protein complex
NCBI	National Centre for Biotechnology Information
ORC	Origin Recognition Complex

oriC	Origin of replication
PAGE	Polyacrylamide Gel Electrophoresis
ParC	Plasmid partitioning protein C
PCNA	Proliferating Cell Nuclear Antigen
pre-RC	pre-Replication Complex
r.m.s	Root mean square
RFC	Replication Factor C
RIDA	Regulatory Inactivation of DnaA
RNA	Ribonucleic Acid
SeqA	Sequestration factor A
ssDNA	Single-stranded DNA
Topo IV	Topoisomerase IV
VcSeqA	*V. cholerae* SeqA

Chapter 1

Introduction

1.1 DNA is the Hereditary Material

DNA was first discovered in 1869 as 'nuclein', a substance that precipitated out of solution upon the addition of acid, but dissolved with the addition of alkali [1]. However, it took another 59 years to link DNA as the hereditary material by Frederick Griffiths [2]. This idea was confirmed by Avery, McLeod and McCarty, ruling out RNA and protein as the hereditary material [3]. The first insight into the structure of DNA came in 1953 when Francis Crick and James Watson proposed the three-dimensional, double helical structure of DNA [4]. Watson and Crick's model of DNA also suggested a mechanism for DNA replication where the two strands could function as templates to generate complementary strands, a prediction that was confirmed shortly after by Matthew Meselson and Franklin Stahl [5]. Semi-conservative manner of DNA replication was visualized in replicating *Escherichia coli* chromosome by autoradiography as a circular structure that resembled the Greek letter theta [6]. Further autoradiography studies of *E. coli* chromosome also revealed the bi-directional nature of 'theta replication'

establishing a single origin of replication, referred to as the *oriC* (Figure 1.1) [7]. These discoveries have been the driving force behind understanding the physical mechanism of DNA replication even till today.

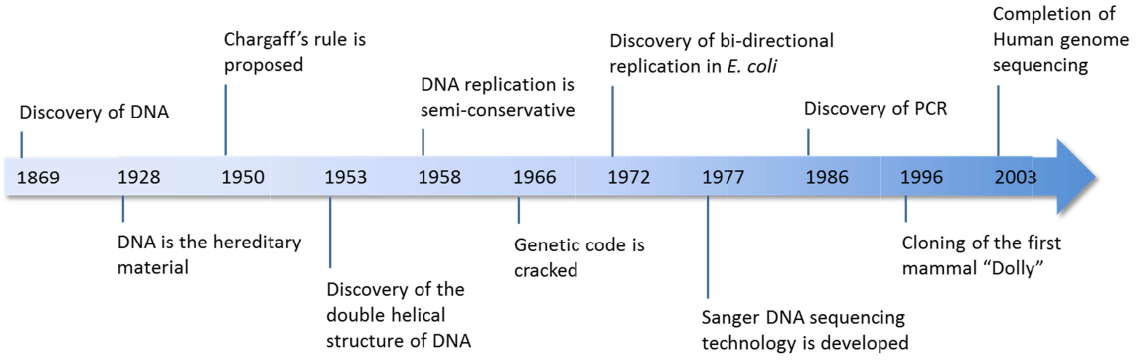

Figure 1.1. The DNA timeline. *This timeline illustrates significant discoveries of DNA made between 1869 and 2003.*

1.2 DNA Replication Overview

In every single organism, the cell cycle is organized into distinct processes involving cell growth, DNA replication, segregation and division to accomplish its most fundamental task of copying and passing on its genetic information. In bacteria, these processes can occur simultaneously in rapidly growing cells (Figure 1.2A). However, in eukaryotic cells, these processes are temporally separated into four distinct phases G_1, S, G_2 and M, where DNA replication only occurs during the S phase (Figure 1.2B). Nevertheless, chromosome replication in both bacteria and eukaryotes share a set of common steps: initiation, elongation and termination. Once replication initiates, elongation progresses at a steady rate determined by the processivity of the DNA polymerase, which acts as

the limiting factor till termination of replication. Hence, replication initiation provides the primary point of control over DNA replication to ensure that the genome is replicated once and only once during the cell cycle. Although bacteria and eukaryotes share a parallel in the basic process of DNA replication, the regulation of DNA replication is significantly more complex in eukaryotes than in bacteria due to the difference in the organization of the cell cycle.

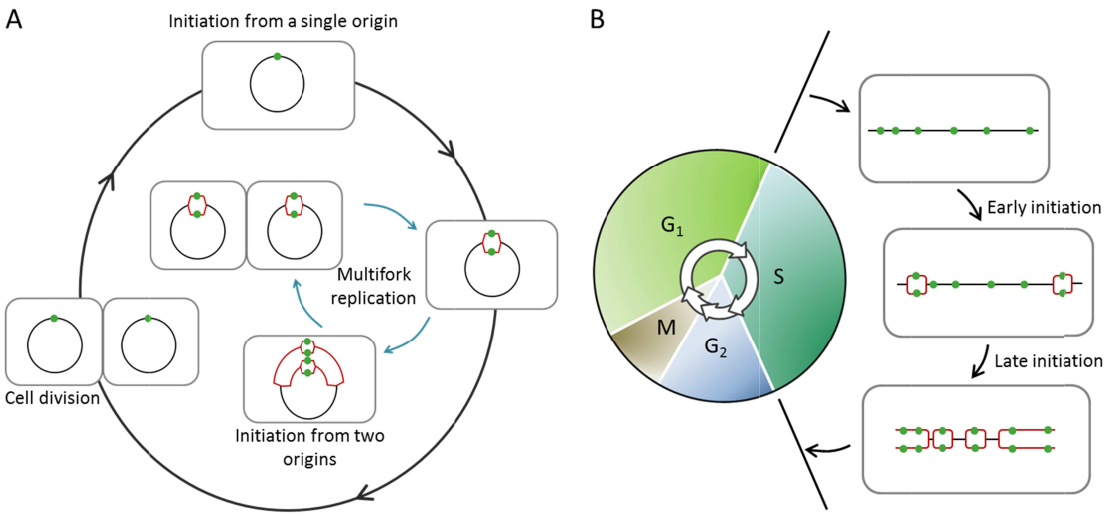

Figure 1.2. Cell cycle in bacteria and eukaryotes. *Chromosomes are shown in black with the origins shown as green circles. Bi-directional movement of the replication fork is shown in red.* ***A****. E. coli replication pattern with different growth rates. The outer cell cycle (black circle) represents a slow growth rate while the inner cell cycle (blue circle) represents a fast growth rate. Unlike slow growing E. coli, replication initiation and cell division can occur simultaneously in rapidly growing E. coli (inner cell cycle).*

Continued on pg. 4

Continued from pg. 3

***B**. Eukaryotic cell cycle is separated into four distinct phases, G_1, S, G_2 and M. DNA synthesis only occurs during S phase. Only a single chromosome is shown for simplicity and the origins shown as green circles. In Saccharomyces cerevisiae, replication initiates from multiple autonomously replicating sequences that initiates at different times during S phase. The diagrams are not to scale.*

1.2.1 Eukaryotic replication initiation

In eukaryotic cells, the genomic DNA is organized into multiple chromosomes due to its considerable size, which can range from 10^7 to 10^{11} bases. Most eukaryotic cells contain more DNA than can be duplicated from a single origin during S phase due to the limiting rate of DNA polymerase processivity [8]. Hence, replication in eukaryotic cells initiate from multiple origins scattered throughout the genome, terminating once the multiple replication forks converge together or reach the telomere [9]. The eukaryotic origin of replication has been best characterized in the budding yeast *Saccharomyces cerevisiae*, which has a distinct chromosomal element known as autonomously replicating sequence (ARS) element [10,11]. *S. cerevisiae* genome is ~12.5 Mbp and contains several hundred ARS elements throughout its genome. Origins of replication in higher eukaryotes have been significantly more difficult to define due to the lack of a clear consensus sequence. Despite the variations in the sequence and size of origins of replication, all eukaryotes utilize a conserved mechanism for initiating and regulating replication initiation [12].

DNA replication in eukaryotes begins with the sequential assembly of proteins into a pre-replication complex (pre-RC) at origins of replication throughout the genome (Figure 1.3). This process is regulated by an 'origin licensing' mechanism that ensures that each origin fires only once during S phase through tight regulation of multiple licensing factors. Pre-RC assembly begins in late M and early G_1 phase with the binding of an initiator protein, origin recognition complex (ORC) to the specific recognition sites within the origin of replication in an ATP dependent manner [13]. ORC is a heterohexameric complex consisting of six proteins Orc1–6, which recruits additional replication factors, a protein called cell division cycle 6 (Cdc6) and Cdc10-dependent transcription factor 1 (Cdt1) [14,15]. The binding of Cdc6 and Cdt1 to the origin in turn facilitates the loading of the mini chromosome maintenance (MCM) helicases onto the origin [16,17]. MCM is a heterohexamer consisting of Mcm2–7 and is an ATP dependent helicase that unwinds the DNA. With the binding of MCM to the origin, the pre-RC assembly is complete and fully 'licensed' for replication initiation.

Activation of the pre-RC for replication initiation requires phosphorylation of the MCM complex by additional kinases during the G_1/S phase transition, which promotes the helicase activity of MCM [18]. Once the origin is unwound, the replisome including DNA polymerases are recruited, resulting in the bi-directional movement of the replication forks. All replication origins are fired, either early or late in S phase staggered either stochastically or in a temporal order [19,20]. How the cell regulates their timing of origin activation remains largely unknown and further adds to the complexity of eukaryotic replication regulation, since every single origin within the chromosome must be prevented from re-initiations [21].

1.3 Prokaryotic Replication Initiation

Replication initiation in prokaryotes occurs in an analogous manner to that in eukaryotes, but with a fundamental difference. Bacterial replication initiates from a single origin rather than multiple origins resulting in a much less complex regulatory mechanism (Figure 1.2A). *Escherichia coli* and *Bacillus subtilis* have been model organisms for studying replication initiation in prokaryotes highlighting the conserved nature of the overall architecture of the replication machinery [22].

Replication initiation in *E. coli* initiates from the origin of replication *oriC*, when the initiator protein DnaA binds to specific recognition sites within the *oriC* (Figure 1.3B). Within the *oriC*, there are multiple DnaA binding sites, one class of binding sites that are bound by DnaA–ATP/DnaA–ADP throughout the cell cycle that is analogous to the eukaryotic ORC [23]. DnaA–ATP binding to another class of DnaA binding sites results in the assembly of the pre-RC [23]. Once pre-RC is assembled through the formation of large DnaA–*oriC* nucleoprotein complex, the origin is unwound at a specific DNA unwinding element (DUE) [24]. Following DUE unwinding, DnaA directs the recruitment of the hexameric replication helicase DnaB associated with DnaC helicase loader as a DnaB–DnaC–ATP complex to the unwound region [25,26]. Two hexameric DnaB molecules are loaded onto two strands of single-stranded DNA (ssDNA) following ATP hydrolysis induced dissociation of DnaC [27,28,29]. The hexameric DnaB on opposite strands promotes further duplex melting and elongates the ssDNA bi-directionally, which activates the RNA primase DnaG, synthesizing the RNA primer on the lagging strand [30]. DNA polymerase III (which consists of Pol III core subunit, clamp loader and the β sliding clamp subunit) is loaded onto the DNA, initiating synthesis of the complementary strand [31].

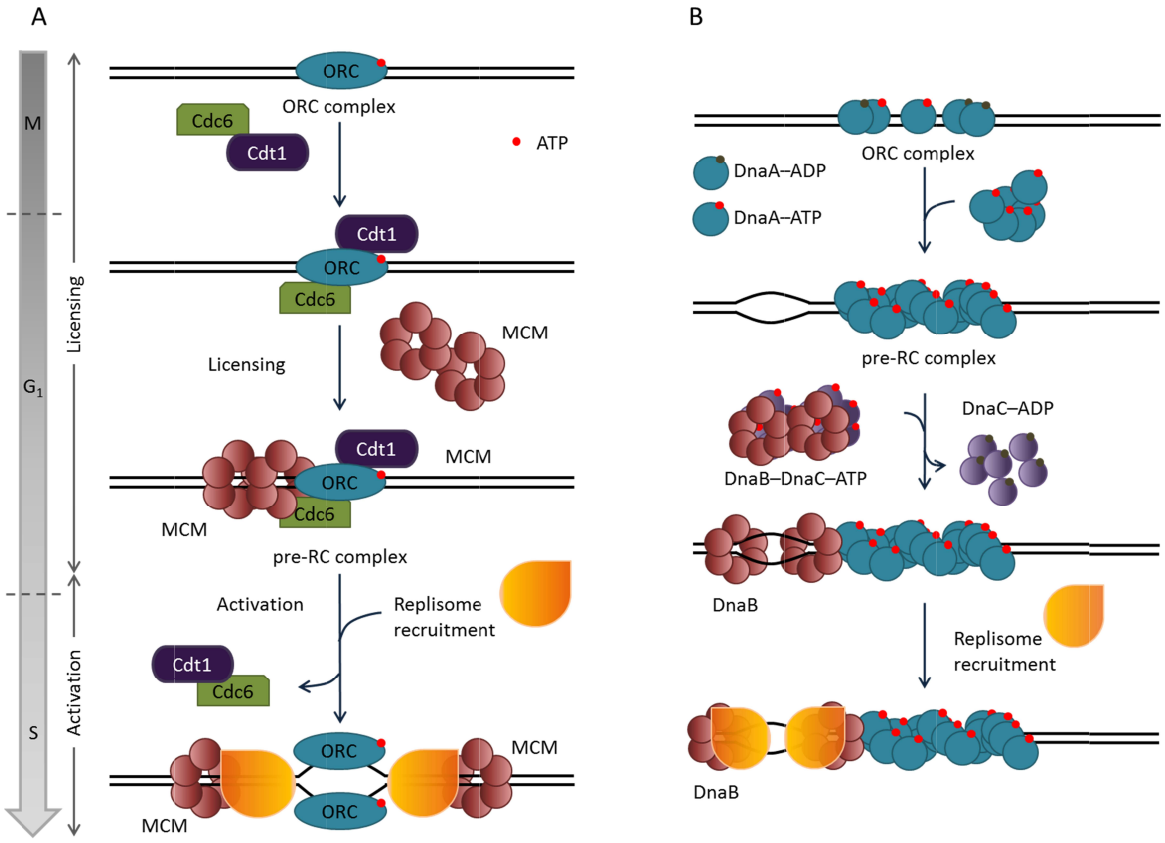

Figure 1.3. Mechanism of replication initiation in *S. cerevisiae* and *E. coli*. *The symbols for different proteins in the replication initiation pathway are indicated in the figure. Proteins that play analogous roles in S. cerevisiae* and *E. coli are shown in the same colouring. The diagrams are not to scale.* ***A****. Replication initiation in S. cerevisiae with only one origin shown for simplicity. In late M and early G_1 phase, Cdc6 and Cdt1 are recruited to the ORC bound to the ARS element. The MCM proteins are loaded onto the ORC during G_1 phase completing the pre-RC assembly. The fully licensed pre-RC becomes activated by phosphorylation of MCM by two additional kinases during the G_1/S phase transition. MCM unwinds the origin and replisomes are recruited for replication. Replications of all origins are completed during S phase.*

Continued on pg. 8

Continued from pg. 7

***B**. Replication initiation in E. coli. Replication initiator protein DnaA forms a complex with ATP or ADP and binds to specific binding sites in the oriC throughout the cell cycle similar to the eukaryotic ORC. Binding of DnaA–ATP molecules to additional binding sites result in the formation of a pre-RC, which unwinds the adjacent DUE. Once the origin is unwound, DnaB helicase is loaded onto the DUE in association with DnaC–ATP. Following loading of the helicase, replisomes are recruited for replication.*

1.3.1 Replication initiation protein DnaA

DnaA is a member of the ATPases Associated with various cellular Activities (AAA+) and is highly conserved in all bacteria for initiation of chromosome replication [32]. DnaA initiates replication through three consecutive steps: 1) binding to the DnaA boxes within the *oriC*, 2) unwinding of the DUE as a result of DnaA–*oriC* nucleoprotein complex formation and 3) recruitment of the helicase DnaB–DnaC complex which further unwinds the *oriC* for bi-directional replication. The activity of DnaA is regulated by the binding of adenine nucleotides, where DnaA–ATP is defined as the active form for its requirement for *oriC* unwinding [33]. The ATP is hydrolyzed to ADP through its intrinsic ATPase activity, rendering an inactive DnaA–ADP complex.

E. coli DnaA is a 52 KDa protein with four distinct functional domains (Figure 1.4A) [34]. Domain I (residues 1–90) mediates interactions with DNA helicase DnaB and DnaA initiator-associating factor (DiaA) that stimulates pre-RC assembly [35]. Domain I also facilitates interactions between DnaA molecules and shows weak affinity for *oriC* ssDNA

[36]. Domain II (residues 90–130) is a poorly conserved region that likely functions as a flexible linker [37].

Domains III and IV are the most highly conserved and the functional roles of these two domains are well characterized. Domain III (residues 131–346) contains the conserved AAA+ motifs: Walker A and B motifs, sensor I and II and box VII for ATP binding and its hydrolysis [38]. ATP binding by Domain III imposes a conformational change of DnaA from a monomeric state into a large oligomeric complex that forms a right-handed helical filament, which is essential for replication initiation [39]. Domain IV (347–467) contains the helix-turn-helix motif that binds to its specific recognition site with the consensus sequence $^{5'}$TGTGNT/AATT$^{'3}$ [40]. The crystal structure of ATP bound DnaA domain III–IV from *Aquifex aeolicus* suggests that DnaA interacts in a head-to-tail manner forming a right-handed helical filament that wraps the *oriC* DNA around itself (Figure 1.4B) [39]. The formation of the toroidal DNA wrap by the DnaA–ATP filament likely destabilizes the origin by introducing compensatory negative supercoils into the DUE, facilitating DNA melting.

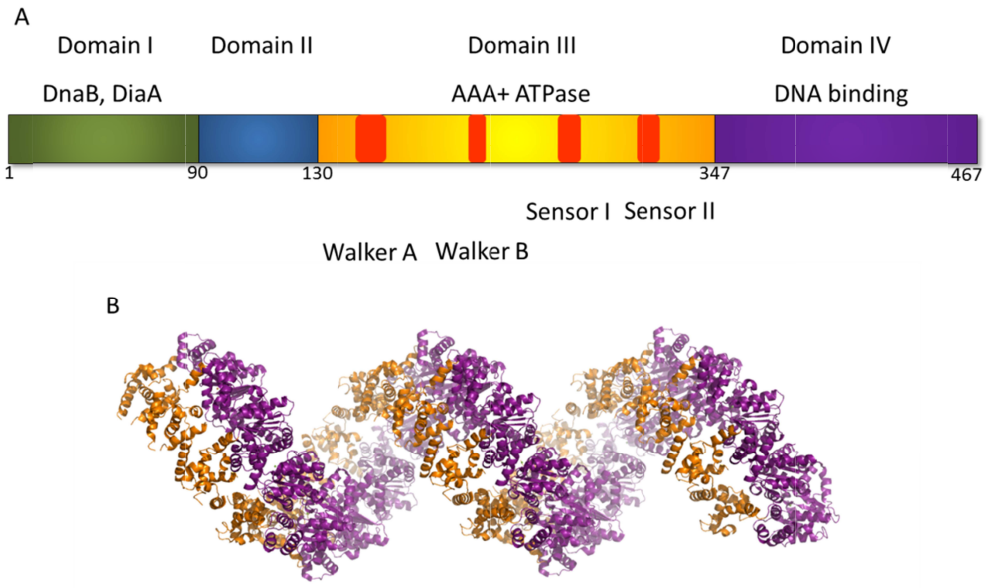

Figure 1.4. The domain architecture of E. coli DnaA. **A.** *This illustrates the domain structures of DnaA and their functional roles. The residue numbers corresponding to the domains are shown below the bar. This diagram is not to scale.* ***B****. Ribbon diagram of the AMPPCP bound DnaA filament structure of domains III and IV which are shown in the corresponding colours to* ***A*** *(PDB ID: 2HCB). The ATP dependent filament formation of DnaA is essential for replication initiation.*

1.3.2 *Escherichia coli* replication origin architecture

The *E. coli oriC* is 245 bp in length and contains a number of key DNA elements that are essential for replication initiation and its regulation (Figure 1.5A–B). The *oriC* contains multiple 9-mer sequences, termed DnaA boxes that are specific sites of recognition by DnaA [41,42]. DnaA boxes are the most prevalent protein binding sites within the *oriC* and are classified into two distinct classes based on affinity for DnaA: high- and low-

affinity binding sites. There are three high-affinity (K_d < 200 nM) DnaA boxes termed R boxes (R1, R2 and R4), with the consensus sequence of $^{5'}$TGTGGATAA$^{3'}$, that are bound by both the active DnaA–ATP or the inactive DnaA–ADP [23,43,44]. The low-affinity (K_d > 200 nM) DnaA boxes (I1–I3, C1–C3, τ1–2 and R5M) are preferentially bound by DnaA–ATP prior to replication initiation with the exception of the R5M site [23,45,46]. This is likely due to a single mismatch in the consensus sequence, as I, C and τ low-affinity DnaA boxes contain two or more mismatches and are bound by DnaA–ATP with up to four-fold higher affinity. [23,47,48]. The low-affinity DnaA boxes are located between the high-affinity DnaA boxes, where τ1, τ2, I1 and I2 are located between R1 and R2 while C3, C2, I3 and C1 are located between R2 and R4 (Figure 1.5A–B) [23,48,49].

In addition to DnaA boxes, the *oriC* also contains an AT rich DUE, which provides a distinct entry site for the replication initiation machinery [50]. DUE consists of three repeats of 13 bp sequences with the consensus sequence $^{5'}$GATCTNTTNTTTT$^{3'}$ termed L, M and R sites [50]. The AT rich nature of the DUE reduces the helical stability of the DNA resulting in the localized unwinding. In fact, the DUE can be unwound even in the absence of origin binding proteins, demonstrating that DNA supercoiling alone can modulate replication initiation via the DUE [50]. The 13 bp DUE consensus sequence is exclusively found in the origin and exhibits a special DNA topology that is critical for its function.

The *oriC* also contains binding sites for proteins that induce topological changes such as histone-like protein HU, integration host factor (IHF) and factor for inversion stimulation (Fis) protein. HU belongs to a family of DNA architectural proteins that stabilizes higher-order nucleoprotein complexes through inducing a bend in the DNA (~105–140°) [51]. IHF binds to the left half of the *oriC*, between R1 and τ1 DnaA binding

boxes, and severely bends the *oriC* DNA by > 160° [52,53]. Fis binds to the right half of the *oriC* between R2 and R4 high-affinity DnaA boxes, and bends the *oriC* DNA by > 50° [54,55]. The interplay of these DNA bending proteins that induce topological changes to the origin is essential for replication initiation.

An additional characteristic of the *oriC* is the presence of an exceptionally high number of GATC sequences that are interspersed throughout the 245 bp *oriC* (Figure 1.5A–B) [56]. There are approximately 20,000 GATC sequences in the *E. coli* genome, with a frequency of 4.1 GATC sequences per every 1,000 bp [57]. The number of GATC sequences within the *oriC* equals to a frequency of 45 GATC sequences per every 1,000 bp, which suggests the importance in its abundance at the origin. The GATC sequences are recognition sites for DNA adenine methylase (Dam) which adds a methyl group to the adenine in the newly synthesized DNA. At the time of replication initiation, the *oriC* is fully methylated, which becomes hemimethylated with the passage of the replication forks through the synthesis of an unmethylated daughter strand. This state of hemimethylation coordinates several regulatory processes, such as DNA mismatch repair, protein expression as well as DNA replication. Distribution of the eleven GATC sequences in the DUE and the lower-affinity DnaA boxes (τ1, R5M, τ2, I2 and I3) within the *oriC* functions in the regulation of replication initiation through a process known as sequestration. A protein termed sequestration factor A (SeqA) binds to these hemimethylated GATC sequences with high affinity, which in combination with the high density of GATC sequences renders the *oriC* inaccessible to DnaA for approximately one-third of the cell cycle, sequestering the origin from new rounds of replication initiations [58,59,60].

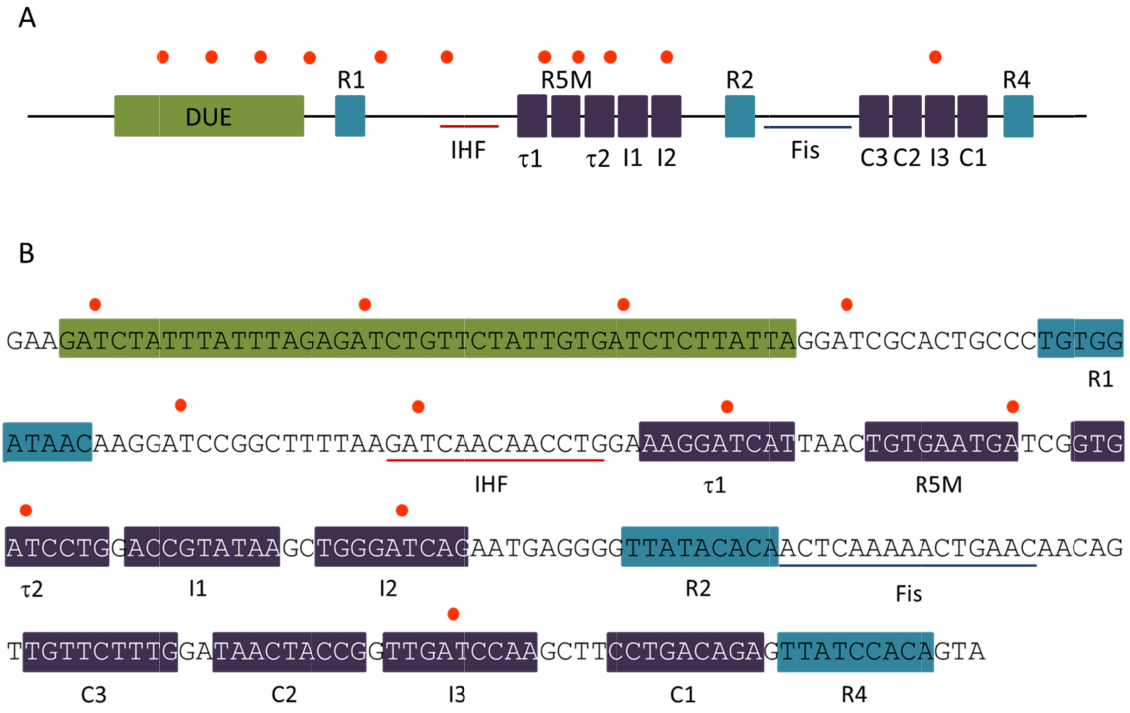

Figure 1.5. Map of the *E. coli oriC*. *A*. *Schematic diagram of the E. coli oriC illustrating the locations of DnaA boxes, DUE, GATC sequences, IHF and Fis binding site. The DUE is in green, the orange dots represent the GATC sequences and the IHF and Fis binding sites are underlined with red and blue, respectively. High-affinity DnaA boxes (R1, R2 and R4) are shown in blue and low-affinity DnaA boxes (τ1, R5M, τ2, I1, I2, C3, C2, I3 and C1) are shown in purple. The locations of GATC sequences are marked with orange circles. The illustration is not to scale* ***B****. Sequence of the E. coli oriC with the same colour scheme as in* ***A****. The sequences of C3, C2 and C1 are shown based on the consensus sequence and preliminary results reported by Leonard and Grimwade [46].*

1.3.3 Assembly of the pre-RC for replication initiation

During the cell cycle, the high-affinity DnaA boxes R1, R2 and R4 within the *oriC* are bound by DnaA–ATP and DnaA–ADP forming a nucleoprotein complex that is analogous to the eukaryotic ORC and functions as a platform for pre-RC assembly. The transition from an ORC like complex into a pre-RC for replication initiation requires the binding of DnaA–ATP to the low-affinity DnaA boxes (R5M, τ, I and C sites) along with the dynamic interplay between other *oriC* binding proteins IHF, Fis and SeqA. Both Fis and SeqA remain bound to their cognate binding sites for majority of the cell cycle preventing IHF binding and low-affinity DnaA–ATP interactions [60,61]. Fis alters *oriC* topology by bending the DNA while SeqA sequesters GATC sequences associated with low-affinity DnaA boxes and the IHF binding site, directly blocking their interactions [60,61].

Near the time of replication initiation, Fis and SeqA are displaced, promoting IHF binding to the *oriC* [61]. IHF bends the *oriC* into a configuration that stimulates additional interactions of DnaA–ATP with the low-affinity sites [46]. In addition to IHF, DiaA stabilizes the DnaA–ATP interactions by stimulating oligomerization [62]. The stabilization of DnaA–ATP binding to low-affinity DnaA boxes triggers the formation of a right-handed DnaA–ATP filament on the *oriC* through ATPase domain interactions [39]. This topological change likely causes helical instability to the AT-rich DUE region, resulting in the localized unwinding essential for replication initiation.

1.4 Regulation of Replication Initiation

Once replication has initiated, origin firing is regulated to ensure that chromosomal replication takes place only once per cell cycle. *E. coli* cells have established many mechanisms to regulate replication initiation, one of which is the regulation of the pre-RC assembly at the *oriC*, as described above. During pre-RC assembly at the *oriC*, DnaA interacts with many proteins in the replication initiation cascade, making it an exceptional target for regulation. There are several regulatory mechanisms that function at the *oriC* and outside the *oriC* that forms a regulatory network around the replication initiation protein, DnaA.

1.4.1 Regulatory inactivation of DnaA

Regulatory inactivation of DnaA (RIDA) stimulates ATP hydrolysis immediately after the onset of replication initiation, yielding inactive DnaA–ADP complexes [63]. RIDA reduces the level of active DnaA–ATP within the cell, which impedes replication initiation and is the predominant mechanism of negatively regulating replication initiation [64]. RIDA requires two components, the β sliding clamp subunit of DNA polymerase III (β sliding clamp) and a protein named Hda (Homologous to DnaA) for its homology to DnaA [65]. Hda belongs to the AAA+ ATPase family like DnaA and contains a conserved clamp binding motif as well as a conserved arginine finger, through which Hda can interact with β sliding clamp and DnaA–ATP, respectively [66,67,68]. However, DnaA–ATP hydrolysis only occurs when Hda associates with ADP and interacts with the DNA-loaded form of β sliding clamp [69,70]. This interaction is mediated by the DNA binding domain of DnaA, which recognizes the DNA flanking the β sliding clamp. In rapidly growing cells, new rounds of replication initiation generates new replication forks with subsequently more

DNA-loaded β sliding clamp, which would result in a corresponding increase in the inactivation of DnaA–ATP. Hence, the requirement for the DNA-loaded form of β sliding clamp ensures timely inactivation of DnaA–ATP coupled with active chromosome replication.

1.4.2 Regulating cellular levels of DnaA

DnaA levels within the cell are also regulated through the titration of DnaA within the cell. Titration of DnaA occurs through approximately 300 DnaA boxes that are evenly distributed throughout the *E. coli* genome [40]. While majority of the DnaA boxes are dispersed throughout the genome, there are several regions that contain clusters of DnaA boxes, one of them named the DnaA titration site (*datA)* for its particularly high capacity for DnaA molecules [71,72]. The *datA* locus contains five high-affinity DnaA boxes that attracts eight times more DnaA than the *oriC in vivo* [71]. Further analysis of the *datA* locus revealed a single IHF binding site as well as potential arrays of low-affinity DnaA boxes, indicating that the roles of IHF at the *oriC* in stimulating oligomerization of DnaA–ATP may be conserved elsewhere in the chromosome [49,73].

1.4.3 Regulation of *dnaA* transcription

In addition to the *datA* locus, there is another DnaA binding site situated within the transcriptional unit of the *dnaA* gene. There is at least one perfect consensus high-affinity DnaA box between the two promoters of the *dnaA* gene, along with several other low-affinity DnaA boxes near the two promoters that act as an autorepressor of gene transcription of both its promoters when bound by DnaA [74,75,76]. With increasing concentrations of DnaA within the cell, DnaA–ATP binding to these binding sites triggers the oligomerization of DnaA along the two promoters, directly inhibiting

RNA polymerase from binding to both promoters thereby reducing the level of transcription [77]. Similar to titration of free DnaA molecules within the cell, autoregulation of the *dnaA* gene directly results in lower levels of available DnaA within the cell, thus providing an added level of replication initiation regulation.

1.4.4 Reactivation of DnaA

While there are several regulatory mechanisms to ensure lower levels of active DnaA, there is another aspect in the regulation of DnaA levels which balances the active and inactive form of DnaA. During rapid cell growth in *E. coli*, multiple rounds of replication can begin prior to cell division, depleting the level of available DnaA–ATP within the cell. In a mechanism to provide active DnaA–ATP beyond the synthetic capabilities of the cell, 'recycling' of DnaA–ADP molecules occurs through a mechanism utilizing DnaA-reactivating sequences (DARS) [78]. The *E. coli* genome contains two DARS sites, DARS1 and DARS2, which directly promote the reactivation of DnaA through the nucleotide exchange of ADP to ATP [79]. DARS1 and DARS2 each contain a single DnaA box and two DnaA box-like sequences, the orientation of which is imperative for the reactivation event [78]. The DARS sites likely provide a platform for the interaction between DnaA–ADP proteins for the nucleotide exchange. The rate of reactivation of DnaA–ADP to DnaA–ATP is proportional to replication time, as more rounds of replication generate more copies of DARS within the cell. Hence, *E. coli* cells have evolved a very responsive mechanism to master the fine balance between inactivation and reactivation of its DNA replication initiator protein.

1.4.5 Regulation of replication initiation through sequestration

Another important mechanism of replication initiation regulation occurs through sequestration. SeqA protein binds preferentially to hemimethylated GATC sites that result from semi-conservative replication. At the *oriC*, SeqA binds to the eleven hemimethylated GATC sequence within the *oriC* for approximately one third of the cell cycle and prevents new rounds of replication initiation [59,80]. SeqA binding at the *oriC* functions to regulate replication initiation in two ways: 1) preventing remethylation of the GATC sequences by Dam, which regulates the timing of replication initiation, and 2) preventing the pre-RC assembly while allowing resetting of the ORC for the next round of replication initiation. However, the exact order of steps in which SeqA is released from the *oriC* is not known. Once *oriC* is released form sequestration, dam is able to methylate the GATC sequences on the unmethylated daughter strand, resulting in a fully methylated *oriC*.

SeqA also plays a role at the origin through regulating *dnaA* transcription levels by sequestrating the *dnaA* gene, located in close proximity to the *oriC* (~43 kb). Like the *oriC*, the *dnaA* gene contains a high frequency of GATC sequences with approximately 12 GATC sequences per 1000 bp, which is three times more than the average frequency of 4.1 GATC sequences per 1000 bp within the *E. coli* genome [57]. SeqA binds to the GATC sequences within *dnaA* gene for approximately one sixth of the cell cycle, thereby reducing the levels of DnaA within the cell. The *dnaA* gene and *oriC* sequestration is coordinated with replication as the movement of the *dnaA* gene to a location opposite from *oriC* results in excess replication initiations [81]. Hence, the coordinated sequestration of *dnaA* and *oriC* is required for regulating the precise timing of replication initiation.

The role of SeqA in regulating replication initiations at the *oriC* has been extensively studied by several independent research groups. However, the extent of SeqA function outside of the *oriC* has been long suspected given the presence of ~20,000 GATC sequences distributed within the *E. coli* genome and the phenotypes displayed in the absence of SeqA protein. Analysis of the genome-wide distribution of SeqA in *E. coli* using Chromatin immunoprecipitation coupled with microarray analysis (ChIP-Chip) has shown that SeqA binding correlates with the frequency and spacing of GATC sequences across the entire genome. There are 137 genes that show SeqA binding with a signal that is at least four times higher than background, and 24 of these genes are directly involved in DNA replication, repair, synthesis as well as in methyl group transfer [57]. This indicates that SeqA truly acts in a global fashion to regulate the transcription of specific genes through sequestration. The remainder of the introduction will focus on SeqA and outline its many functional roles outside of the origin and its influences on chromosome dynamics.

1.5 The Negative Regulatory Protein SeqA

The existence of a mechanism that regulates replication initiation was first proposed by Russell and Zinder in 1987 who found that *dam*$^-$ strains are efficiently transformed with *oriC* minichromosome plasmids only if the plasmid is unmethylated [82]. If the plasmid is fully methylated, the transformation efficiency is dramatically reduced as the plasmid undergoes one round of replication and accumulates in hemimethylated form [82]. Investigation of the GATC sequences within the *oriC* indicated that these sites are sequestered in a hemimethylation dependent manner by a cellular protein preventing remethylation by Dam [58]. The protein required for sequestration was identified as SeqA by Lu and others in 1994 from an *E. coli* mutant defective in sequestration, obtained as a derivative of a *dam*$^-$ strain that now efficiently accepts *oriC* minichromosomes [83,84].

1.5.1 Phenotypes of *seqA* mutants

Although *seqA* is not essential for viability, deletion of *seqA* abolishes origin sequestration leading to premature initiations and asynchronous replication [83,85]. Cells that replicate asynchronously contain irregular numbers of chromosomes due to random origin firing at multiple origins and display abnormal cell division and irregular cell size. Synchrony of replication initiation functions to fine tune the coordination between replication and cell division as a response to the cell growth. Hence, cells that lack the *seqA* gene face the consequence of asynchronous replication resulting in elevated levels of anucleate cells with segregation defects as well as aberrant nucleoid distribution [83,86]. Additionally, *seqA null* cells display smaller colony size as well as longer doubling times on any growth media, suggesting that loss of SeqA decreases the

overall fitness of the cells [83]. The *seqA* deletion mutants also exhibit increased negative superhelicity, revealing that SeqA also affects the superhelicity of chromosomes [87].

Similar to phenotypes observed in *seqA null* mutants, overexpression of SeqA results in delayed nucleoid segregation and cell division [86,88]. These phenotypes are likely a result of significantly longer hemimethylation periods at the *oriC* following initiation of replication, resulting in inhibition of new rounds of replication initiations. The remainder of the *E. coli* chromosomes are also hemimethylated for a prolonged period of time with excess SeqA, preventing segregation of sister nucleoids. Furthermore, prolonged binding by SeqA at replication forks may hinder the assembly of segregation machinery [88]. The phenotypes exhibited in *seqA null* mutants and SeqA overexpression further suggests that the function of SeqA in regulation of replication initiation is linked to chromosome segregation.

1.5.2 Role of SeqA beyond the origin

In addition to the global role of SeqA in the precise timing of replication synchrony and transcription repression, SeqA also plays an additional role in chromosome organization at the replication forks. The nearly 20,000 GATC sequences within the *E. coli* genome become transiently hemimethylated with the passage of the replication fork. These hemimethylated GATC sites behind the replication fork are bound by clusters of SeqA which can be visualized as discrete foci by fluorescence microscopy [89,90,91]. In new born cells, SeqA is visualized as a single focus localized at the midcell, followed by separation into two foci, which subsequently migrate in a bi-directional manner towards the cell quarter positions until cell division [89,91]. SeqA foci are not co-localized with the *oriC*, and can be visualized in the absence of the *oriC*, indicating that SeqA–DNA

complexes are formed at specific regions within the *E. coli* chromosome, but not at the *oriC* itself [89]. Furthermore, the formation of SeqA foci is also dependent on DNA methylation as well as ongoing DNA replication, which suggests that foci represent clusters of SeqA bound to newly hemimethylated GATC sequences trailing the replication forks [89,91]. Cells with SeqA mutants that can replicate synchronously but unable to form foci display abnormal nucleoid organization as well as segregation defects similar to phenotypes observed with *seqA null* mutants. This indicates that binding of SeqA at the replication forks is required for proper segregation of the cell and is distinct from its roles at the origin.

SeqA also plays a role in the organization of replication machinery during multi-fork replication through co-localization of the replication forks [92]. While the exact mechanism of how SeqA binding affects chromosome organization at the replication forks is not clear, one approach could be through direct interactions with a protein that regulates superhelicity and condensation of the nucleoid. Indeed, SeqA interacts with the C-terminal domain of ParC subunit of Topoisomerase IV (Topo IV) [93]. Topo IV is a type II topoisomerase that is essential for viability and is implicated in a variety of cellular processes including DNA replication, transcription and chromosome segregation through relaxing supercoils and decatenation of daughter strands [94]. SeqA stimulates activities of Topo IV, both in relaxing supercoiled plasmids and in converting DNA catenanes to monomers [93].

Topo IV is also known to interact with other proteins involved in chromosome organization and segregation. The MukB subunit of the MukB–MukE–MukF chromosome partitioning complex, a tubulin-like protein, FtsK and a bacterial actin ortholog, MreB all bind and interact with Topo IV stimulating its relaxation and

decatenation activities [95,96]. While these observations suggest a network of physical interactions linking Topo IV with other chromosome organizational proteins, it is not clear how all of these functional interactions are coordinated and orchestrated in chromosome segregation.

1.5.3 Biochemical properties of SeqA

SeqA from *E. coli* is a small protein consisting of 181 amino acids, organized into two functional domains; an N-terminal oligomerization domain (residues 1–35) and a C-terminal DNA binding domain (residues 64–181), interconnected by a flexible linker region (residues 36–63) [97,98,99]. SeqA can bind to both fully methylated as well as hemimethylated DNA, but shows highest affinity for hemimethylated DNA and binds cooperatively with a nanomolar dissociation constant and a Hill coefficient of 4.7 [59,80]. SeqA also requires a minimum of two GATC sequences that may be as far as 31 bp and shows optimal binding when the two GATC sequences are situated on the same face of the helix [100,101]. This ability of SeqA to bind distant GATC sequences is due to the flexibility of the linker region between the N- and C-terminal domains.

Structural determination of the SeqA C-terminal domain bound to a 12 bp DNA containing a single hemimethylated GATC sequence revealed that SeqA makes direct hydrogen bonds and van der Waals contacts with the central mA–T bp, but makes no significant interactions with the surrounding G–C bp (Figure 1.4) [97]. An additional crystal structure of SeqA C-terminal domain bound to a hemimethylated DNA with a mismatch in the central mA–T bp to a mA–G bp show that the two pairs are recognized in a similar manner. This indicates that SeqA recognition of the hemimethylated GATC sequence occurs directly through the methylated adenine. Compared to full-length SeqA that binds to DNA with a nanomolar dissociation constant, SeqA C-terminus binds to a

single hemimethylated GATC sequence with a K_d of ~7 μM determined by sedimentation equilibrium analysis [97]. This indicates that the C-terminal domain alone is incapable for carrying out SeqA function and requires its N-terminal domain counterpart.

The N-terminal domain mediates dimerization of SeqA which further associates to form a left-handed filament (Figure 1.5) [98,102]. The minimal structural unit is a SeqA dimer, which is sufficient for synchronous replication initiation when adequate levels of protein are present, suggesting that it is also the minimal functional unit [98]. This is coherent with the requirement of SeqA for a minimum of two hemimethylated GATC sites. Association of SeqA with DNA is cooperative and binding of SeqA to a pair of hemimethylated GATC sequences mediates formation of higher-order complexes [103,104]. This suggests that oligomerization is essential for its function.

The structural work prior to 2005 had given rise to the current model for SeqA function in *E. coli*. This model proposes that SeqA foci consist of SeqA filaments at the replication forks that bind irregularly spaced GATC sequences, thereby looping out intervening DNA [98]. Dynamic addition and removal of SeqA dimers from old to new hemimethylated GATC sequences allows the movement of SeqA foci as the replication fork progresses [98]. However, the exact mechanism through which the functional dimer of SeqA recognizes and forms a high-affinity complex with two GATC sequences remains elusive.

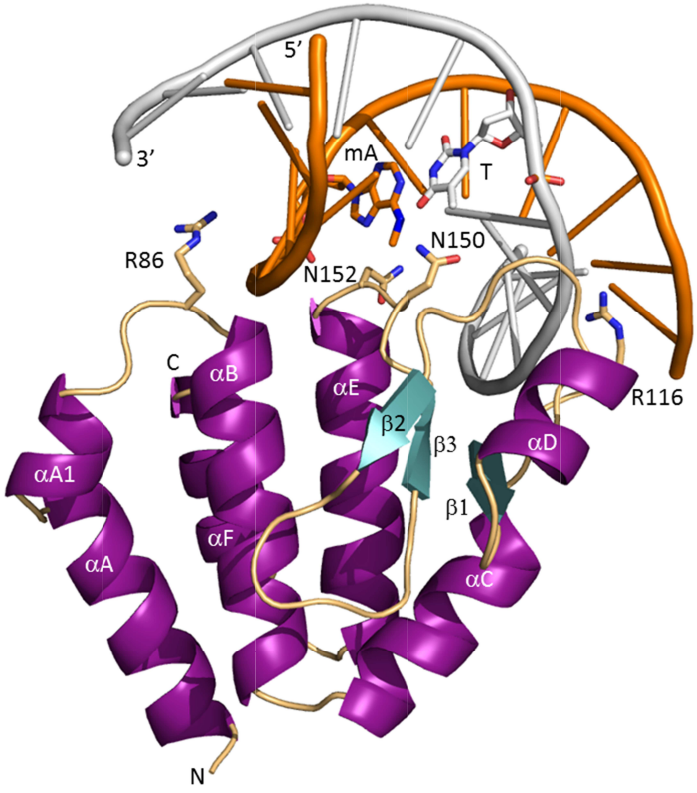

Figure 1.4. Structure of SeqA C-terminal domain in complex with DNA. *Ribbon diagrams of the SeqA C-terminal domain bound to a 12 bp DNA with a single hemimethylated GATC sequence with α-helices in purple and β-strands in teal. The hemimethylated DNA is shown as rods in orange (methylated) and in light grey (unmethylated). The central mA–T base pair is shown in sticks. Asn150 and Asn152 (light yellow sticks) interact with the central mA–T base pairs. Arg86 and Arg116 (light yellow sticks) interact with the minor groove surrounding the GATC sequence. The image was prepared with PDB ID: 1LRR using Pymol [105].*

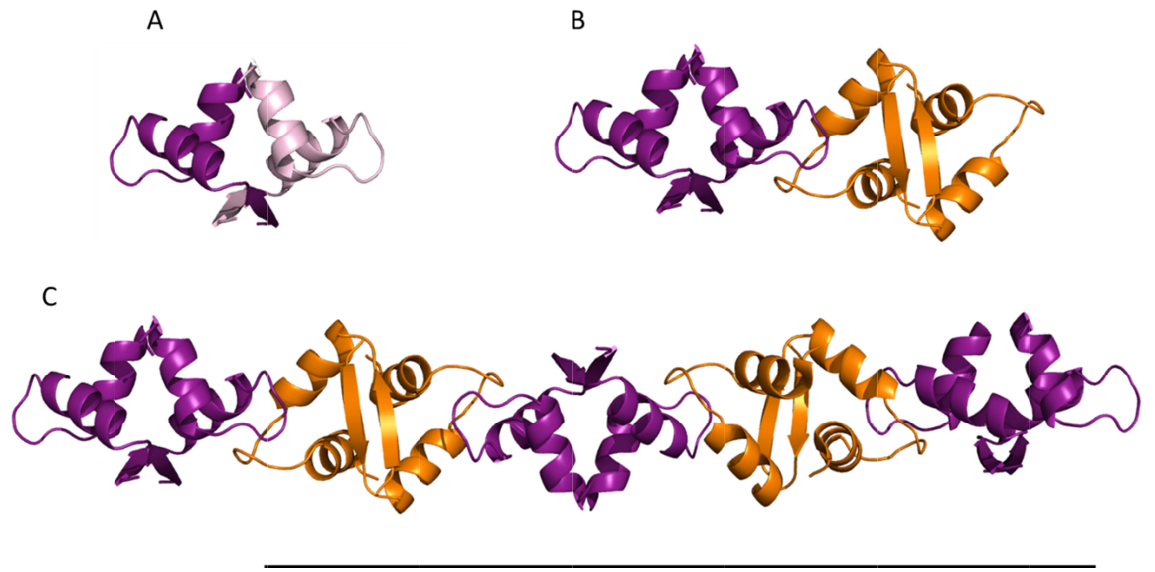

Figure 1.5. Structure of SeqA N-terminal domain. ***A****. Ribbon diagram of a SeqA N-terminal dimer. Each monomer is illustrated in purple and light pink. The SeqA N-terminus is stabilized upon dimerization, where the two N-terminal-strands cross over and hydrogen bonds to form an anti-parallel pair.* ***B****. SeqA N-terminal dimer (purple) forms a tetramer through reciprocal hydrophobic interactions with its adjacent dimer (orange).* ***C****. SeqA N-terminal dimers can form a left-handed helical filament through repetitions of the dimer-dimer interaction. The black line below the filament represents one complete helical turn consisting of four N-terminal dimers. The images were prepared with PDB ID: 1XRX using Pymol [105].*

1.5.4 Diversity of SeqA

SeqA is highly conserved within several families of the γ-proteobacteria class. However, investigations into understanding the roles of SeqA in other bacterial species have been limited. One interesting bacterial species with SeqA homologues is *Vibrio cholerae*,

which contains two chromosomes. The presence of two replication origins poses an intriguing question as to how replication of the two chromosomes is initiated and regulated. Several studies have suggested similar roles of SeqA in origin sequestration and in chromosome organization in *V. cholerae* despite the complexity of having two replication origins. There have also been reports implicating SeqA in the virulence of *Salmonella typhymurium* [106,107]. However, further investigation is required to examine if virulence is directly related to the function of SeqA.

1.6 Thesis Objectives and Organization

The primary focus of this thesis was to probe the existing model of SeqA function in *E. coli* by investigating the molecular interactions that trigger the formation of a high-affinity complex of SeqA with newly replicated DNA; and its roles in origin sequestration and chromosome organization. To this end, hemimethylated oligonucleotides were systematically optimized to obtain crystals of the SeqA–DNA complex and the optimization process is described in detail in chapter 2. Further biochemical characterizations of the SeqA–DNA complex and the structural analysis of the complex providing insight into the roles of SeqA at both the *oriC* and at the replication forks is described in chapter 4. Our work also extends into understanding the roles of SeqA in another gram-negative bacterium, *Vibrio cholerae* through structural and biochemical characterizations described in chapter 4.

Chapter 2

Crystallization and preliminary X-ray diffraction analysis of SeqA bound to a pair of hemimethylated GATC sites

The full citation is as follows:

Chung, YS and Guarné, A. (2008) Crystallization and preliminary X-ray diffraction analysis of SeqA bound to a pair of hemimethylated GATC sites. *Acta Cryst*, **F64**, 567-571.

This work has been reproduced with permission from the International Union of Crystallography. The references from the publication are combined with the references from the thesis.

2.1 Author's Preface

The work presented in chapter 2 has been published in the peer-reviewed *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* and appears in published format. This work describes the crystallization trials of SeqA bound to hemimethylated DNA to investigate how SeqA forms high-affinity complexes with DNA. I have shown that modifications in length, spacing between two hemimethylated GATC sites and the use of dinucleotide overhangs played an essential role in crystallization of the SeqA–DNA complex. I conducted all the work described in this chapter, prepared the figures and wrote the manuscript with Dr. Alba Guarné.

2.2 Abstract

Escherichia coli SeqA is a negative regulator of DNA replication. The SeqA protein forms a high-affinity complex with newly replicated DNA at the origin of replication and thus prevents premature re-initiation events. Beyond the origin, SeqA is found at the replication forks, where it organizes newly replicated DNA into higher ordered structures. These two functions depend on SeqA binding to multiple hemimethylated GATC sequences. In an effort to understand how SeqA forms a high-affinity complex with hemimethylated DNA, a dimeric variant of SeqA was overproduced, purified and crystallized bound to a DNA duplex containing two hemimethylated GATC sites. The preliminary X-ray analysis of crystals diffracting to 3 Å resolution is presented here.

2.3 Introduction

All the GATC sequences within the *Escherichia coli* chromosomes normally exist in their fully methylated state. Upon initiation of DNA replication, GATC sequences are transiently hemimethylated until the newly incorporated adenine is methylated by Dam methylase. Regulation of several important biological processes such as mismatch repair or timing of replication re-initiation depends on this period of hemimethylation [58,108,109]. The SeqA protein binds clusters of hemimethylated GATC sites and thus prevents premature re-initiation events in a process known as origin (*oriC*) sequestration [83]. Upon initiation of replication, SeqA prevents binding of the main initiator protein DnaA to the low-affinity DnaA binding boxes that include GATC sequences and hence SeqA sequestration resets the conformation of the newly replicated origin [60].

Interestingly, clusters of SeqA bound to DNA are detected at the replication forks rather than the origin. The formation of these foci depends on Dam methylation and ongoing replication, but does not require the presence of *oriC*, suggesting that SeqA has additional functional roles beyond sequestration [90,110,111]. Indeed, loss of *seqA* causes asynchrony of replication, hyper-initiation and abnormal localization of nucleoids [85,112]. Purified SeqA forms long and polydisperse linear polymers that are able to form a multivalent complex with DNA [98]. Formation of these SeqA filaments on newly replicated DNA has been postulated to restrain negative supercoils on the DNA [98,102]. Conversely, mutants of SeqA that have lost the ability to oligomerize introduce positive supercoils onto DNA, revealing a rich pattern of interactions between SeqA and newly replicated DNA [102].

The structure of the C-terminal domain of SeqA bound to a single hemimethylated GATC revealed the nature of the specific interactions with DNA and their dependence on methylation [97]. However, SeqA only interacts tightly with DNA when it is bound to multiple hemimethylated GATC sequences [100]. Moreover, the balance between positive/negative supercoils introduced by SeqA onto newly replicated DNA also relies on its ability to recognize multiple hemimethylated GATC sites [102]. In an effort to understand the topological constraints imposed by SeqA on newly replicated DNA, we have undertaken the structure determination of a dimeric SeqA mutant bound to a pair of hemimethylated GATC sequences. Here, we report the optimization of hemimethylated DNA duplexes, their crystallization in complex with the SeqAΔ(41–59)-A25R mutant and the preliminary characterization of SeqAΔ(41–59)-A25R–DNA cocrystals.

2.4 Experimental and Results

2.4.1 Cloning of SeqAΔ(41–59)-A25R

The *seqA* coding sequence was subcloned in pET-11a as previously described [100]. Wild-type SeqA forms long and polydisperse linear polymers [97,98]. However, we have previously shown that mutation of Thr18, Ile21 or Ala25 abrogates filament formation [98]. SeqA-A25R is a monodisperse dimer in solution, whereas SeqA-I21R and SeqA-T18E can form higher molecular-weight species at high concentrations [98]. Hence, we selected SeqA-A25R (pAG8015, Table A.1) to elucidate the crystal structure of the SeqA dimer.

Residues 35–50 were disordered in the structure of the oligomerization domain of SeqA [98], as were residues 51–63 in the structure of the DNA-binding domain [97]. Presumably, this region encompasses a flexible linker that provides plasticity to the SeqA–DNA interaction. In order to restrict the flexibility between the N- and C-terminal domains of SeqA-A25R, residues 41–59 were removed to generate the SeqAΔ(41–59)-A25R mutant (pAG8033, Table A.1). To this end, we designed self-complementary oligonucleotides that annealed with 15 nucleotides on each side of the deletion and used the QuikChange site-directed mutagenesis kit (Stratagene). The sequences of all mutants were verified by DNA sequencing (MOBIX Laboratory, McMaster University).

2.4.2 Overproduction and purification of SeqAΔ(41–59)-A25R

SeqAΔ(41–59)-A25R was overproduced in *E. coli* BL21(DE3) cells (Invitrogen) transformed with the pAG8033 plasmid (Table A.1). Cells were grown at 310 K to an OD_{600} of 0.7 and protein production was induced by the addition of 0.5 mM IPTG

(isopropyl β-D-1-thiogalactopyranoside) to the culture. Cells were harvested after 3 h and washed with phosphate-buffered saline. Cell lysis was performed in buffer A (20 mM Tris pH 8, 100 mM NaCl, 5 mM DTT, 0.5 mM EDTA and 5% glycerol) with lysozyme (0.5 mg ml^{-1}) and brief sonication. Lysates were clarified by centrifugation (39 000 g at 277 K for 40 min). The supernatant was loaded onto a heparin column (GE Healthcare) equilibrated with buffer A and SeqAΔ(41–59)-A25R was eluted from the column using a linear gradient to 1 M NaCl. The sample was further purified by ion exchange on a MonoS 10/100 GL column using the same buffers and gradient as for the heparin column. Purified SeqAΔ(41–59)-A25R was then concentrated to 3 mg ml^{-1} and stored in 20 mM Tris pH 8, 150 mM NaCl, 5 mM DTT, 0.5 mM EDTA and 5% glycerol. As expected from previous data [98], SeqAΔ(41–59)-A25R eluted at a volume consistent with a dimer from a size-exclusion column (Superdex 75 10/300 GL, GE Healthcare). All chromatographic steps were performed using an ÄKTA FPLC (GE Healthcare).

Table 2.1 Hemimethylated DNA duplexes crystallized with SeqAΔ(41–59)-A25R

	DNA Sequence[a, b]	Length/ spacing	Crystallization conditions	Res (Å)
A	5′GAGTCG**A**TCGGTGCG**A**TCCTTAG3′ 3′CTCAGCTAGCCACGCTAGGAATC5′	23/9	20% PEG 1000, 0.1M Tris pH 9	6
B	5′GAGTCG**A**TCGGTGCTG**A**TCCTTAG3′ 3′CTCAGCTAGCCACGACTAGGAATC5′	24/10	5–7% PEG 400, 0.15 M KCl, 0.01 M $MgCl_2$	12
C	5′GAGTCG**A**TCGGTGCTG**A**TCCTTA3′ 3′CTCAGCTAGCCACGACTAGGAAT5′	23/10	6.5 % PEG 10,000, 0.25 M ammonium acetate, 0.1 M Bis-TRIS pH 6.5	7
D	5′GAGTCG**A**TCGGTGCG**A**TCCTTA3′ 3′TCAGCTAGCCACGCTAGGAATC5′	21/9	10% PEG 400, 0.1 M sodium citrate, 0.1 M TRIS pH 7.5	5
E	5′GAGTCG**A**TCGGCGGG**A**TCCTTA3′ 3′TCAGCTAGCCGCCCTAGGAATC5′	21/9	10% PEG 400, 0.1 M sodium citrate, 0.1 M TRIS pH 7.5	5
F	5′GAGTCG**A**TCGGCGGG**A**TCCTTA3′ 3′CAGCTAGCCGCCCTAGGAATCT5′	20/9	14–20% MPD, 0.3–0.4 M ammonium acetate, 0.1 M sodium citrate pH 5.6	3

[a] Methylated adenines are indicated in bold. CG snap-fasteners are shaded dark grey and GATC sequences light grey.
[b] A comprehensive list of all oligonucleotides utilized in the thesis is listed in Table A.2.

2.4.3 Complex formation and crystallization

Both methylated and unmethylated oligonucleotides were purchased from W. M. Keck Foundation at Yale University. Single stranded oligonucleotides were purified over 10 % polyacrylamide gels and eluted from the gel in elution buffer (10 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA) overnight at 310 K. The eluted DNA was precipitated twice with 80 mM sodium acetate pH 7 and 70 % ethanol at 253 K. Purified oligonucleotides were resuspended in deionized water and annealed to yield hemimethylated DNA duplexes as indicated in Table 2.1. SeqAΔ(41–59)-A25–DNA complexes (1:1 ratio) were incubated at room temperature for 15 min and subsequently stored at 277 K. Complex formation was

monitored by size-exclusion chromatography (Superdex 200 10/300 GL, GE Healthcare) and electrophoretic mobility shift assays.

Protein–DNA cocrystals were grown in hanging drops using the vapour-diffusion method at 277 K. Screening for crystallization conditions was performed using the Index (Hampton Research), Wizard I and II (Emerald) and Classics (Qiagen Inc.) crystallization screens. Initial hits were optimized using the sparse-matrix approach (Figure 2.1). Diffraction-quality crystals were cryoprotected by addition of 20 % glycerol or increasing the amount of PEG 400 present in the crystallization drop where applicable.

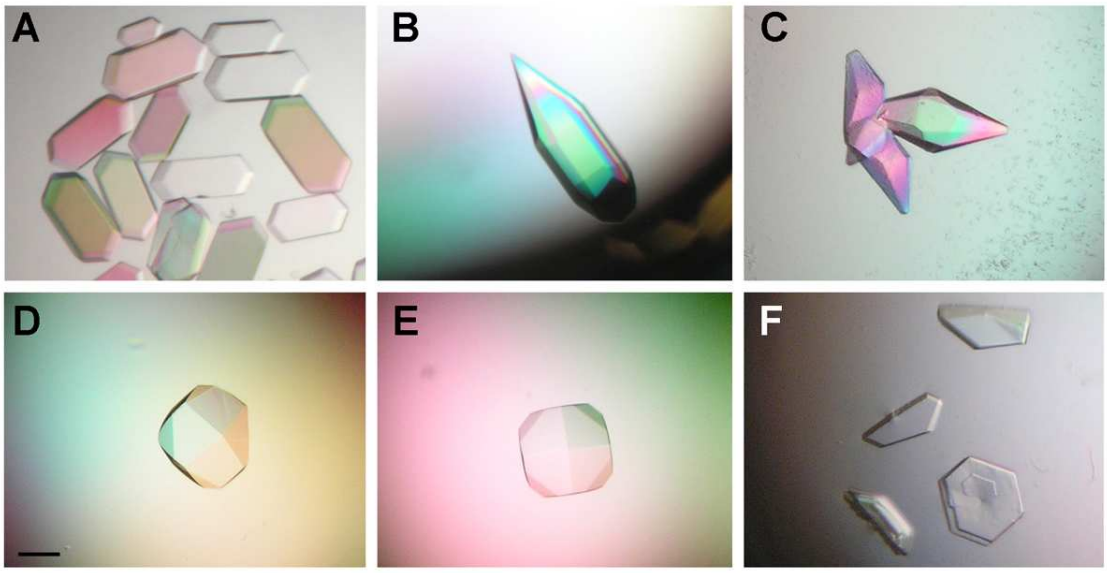

Figure 2.1. Crystals of SeqAΔ(41–59)-A25R bound to DNA. *The duplex name is indicated in the top left corner of each image with the same convention as in Table 1. All pictures were taken at the same magnification; the scale bar in the bottom left image indicates 100 μm.*

2.4.4 Optimization of target DNA sequences

SeqA requires at least two hemimethylated GATC sequences residing on the same face of the DNA helix to form a stable complex [100]. The maximum separation between adjacent hemimethylated GATC sites that allows SeqAΔ(41–59)-A25R binding is 12 base pairs. However, complexes with GATC sequences at this spacing did not yield crystals (28/12, duplex length/GATC spacing) or the crystals diffracted to a very poor resolution (24/12, ~20 Å). In subsequent crystallization trials, the spacing between adjacent hemimethylated sites was limited to either nine or ten base pairs (bp), which exhibited the strongest binding [98]. Additionally, the overall length was kept at about two helical turns (20–24 bp) to facilitate DNA packing within the crystal. Multiple combinations of these two parameters were screened using blunt-ended duplexes (Table 2.1; duplexes A–C and data not shown). All these hemimethylated duplexes formed stable complexes with SeqAΔ(41–59)-A25R and rendered nicely shaped crystals that diffracted X-rays to low resolution (Table 2.1 and Figure 2.1). To facilitate end-to-end stacking of DNA molecules, we designed a 21/9 duplex with one overhanging nucleotide on each end (Table 2.1; duplex D). Use of duplex D led to the formation of better quality crystals as judged by their ability to diffract X-rays and hence we maintained the overall parameters of this duplex in subsequent optimizations.

Since SeqA interacts with the hemimethylated GATC site through the major groove of the duplex [97], a complex between the SeqAΔ(41–59)-A25R dimer and any GATC pairs separated by 9–10 bp would leave one face of the DNA duplex completely exposed to the solvent. In such a situation, DNA–DNA contacts could contribute significantly to the crystal packing. Therefore, we modified duplex D to engineer a CG

snap-fastener on the opposite face of the GATC sites (duplex E in Table 2.1). CG dinucleotides have been previously shown to act as packing-driving boxes for oligonucleotides of different sizes [113]. However, this modification did not affect the crystallization conditions nor improve the diffraction quality of the crystals (Table 2.1), suggesting that the central part of the duplex did not mediate packing contacts.

Table 2.2. Data-collection statistics for SeqAΔ(41–59)-A25R–duplex F cocrystals

Experimental conditions[a]	
X-ray source	X29 (NSLS, BNL)
Wavelength (Å)	0.9795
Temperature (K)	100
Detector	ADSC Q315 CCD
No. of images	100
Exposure time (s)	5
Oscillation angle (˚)	1
Data processing	
No. of measured reflections	293,384
No. of unique reflections	30,853
Space group	R3
Unit cell (Å, ˚)	$a = b = 121.8$, $c = 279.1$, $\gamma = 120$
Resolution (Å)	30.00–3.0 (3.11–3.00)
Completeness (%)	99.8 (100.0)
Multiplicity	3.2 (3.2)
Mean $I/\sigma(I)$	13.5 (1.8)
R_{merge}	0.088 (0.527)
No. complexes per ASU	2
Matthews coeff. V_M ($Å^3$ Da^{-1})	4.01
Solvent content (%)	72.6

[a] Values in parentheses are for the highest resolution shell.

Lastly, we optimized the ends of the duplexes to favour end-to-end interactions between neighbouring DNA duplexes. To this end, identical duplexes with either one or two overhangs at each end were used (compare duplexes E and F in Table 2.1). Extending the single stranded part of the duplex from one nucleotide to two nucleotides had a dramatic influence on crystal formation. Indeed, crystals of SeqAΔ(41–59)-A25R bound to duplex F had a radically different morphology to all the other cocrystals (Figure 2.1).

2.4.5 Data collection, diffraction analysis and structure determination

X-ray diffraction data were collected from flash-frozen crystals on the X29 beamline of the National Synchrotron Light Source (Brookhaven National Laboratory). All data sets were indexed, integrated and scaled using the HKL-2000 package [114]. Crystals of SeqAΔ(41–59)-A25R bound to duplex F diffracted X-rays to 3.0 Å resolution and belonged to space group R3, with unit-cell parameters $a = b = 121.8$, $c = 279.1$ Å , $\gamma = 120^\circ$ (Figure 2.2 and Table 2.2).

The structure of this complex was determined by molecular replacement using the N- and C-terminal domains of SeqA (PDB ID: 1XRX and 1LRR). DNA was omitted from the search model in order to validate the quality of the molecular-replacement solution. Four DNA-binding domains and one dimerization domain were readily placed using Phaser [115]. The two DNA duplexes within the asymmetric unit were easily identified in the initial electron-density map, as was the second dimerization domain.

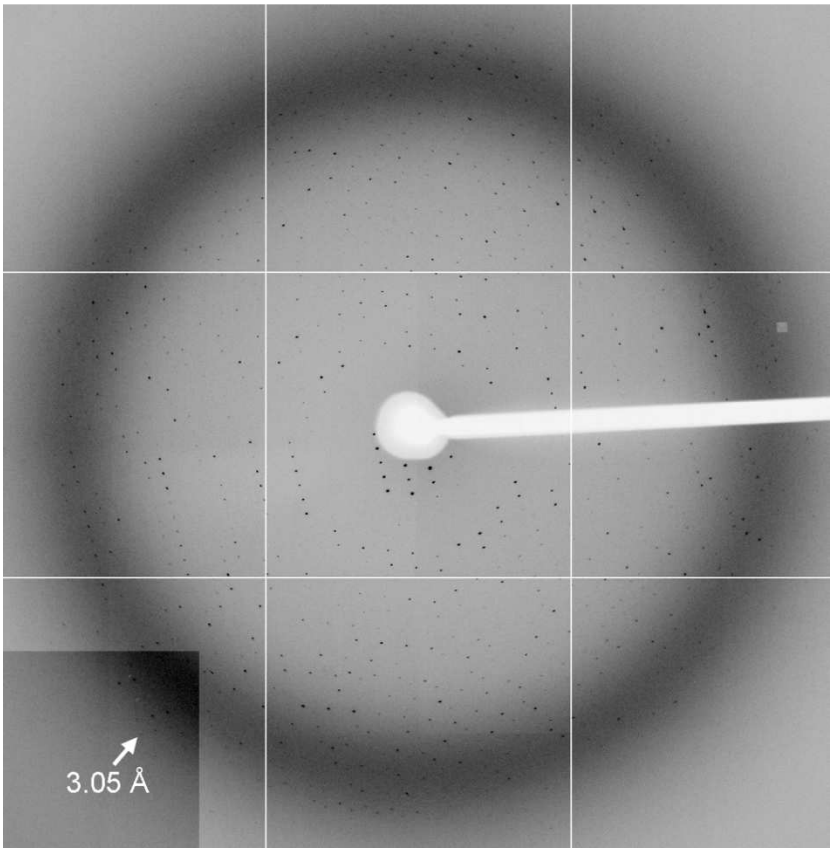

Figure 2.2. Representative X-ray diffraction pattern of SeqAΔ(41–59)-A25R bound to duplex F collected on beamline X29 (NSLS, BNL). *Data were collected at a crystal-to-detector distance of 430 mm and a wavelength of 0.9795 Å. The resolution at the detector edge is 2.9 Å. In the bottom left inset the contrast has been adjusted to make weak reflections more prominent. One reflection at 3.05 Å resolution is indicated with a white arrow for reference.*

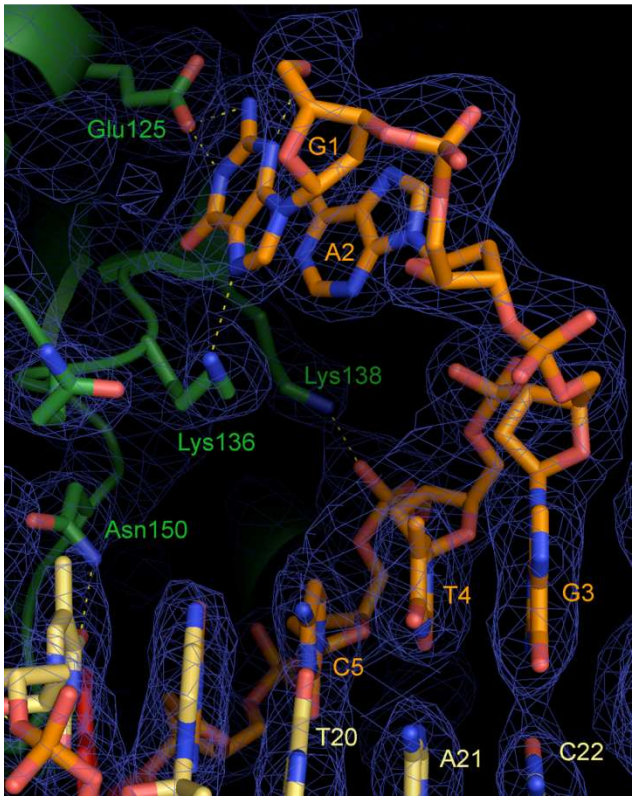

Figure 2.3. *2Fo–Fc* electron-density map superimposed on the SeqAΔ(41–59)-A25R-DNA model at an early stage of refinement. *The interactions between the two overhanging base pairs on the methylated strand and SeqAΔ(41–59)-A25R are illustrated. The protein is shown as a green cartoon with interacting side chains highlighted as colour-coded sticks (C, green; N, blue; O, red). The sequence-specific interaction between Asn150 and T18 on the methylated A7–T18 base pair is also depicted. The unmethylated and methylated DNA strands are shown as pale yellow and orange sticks, respectively, with O and N atoms shown in red and blue for clarity. Hydrogen bonds are depicted as dashed yellow lines and the electron-density map (contoured at 1σ) as a blue mesh. This figure was prepared using PyMOL [105].*

2.5 Discussion

Extensive oligonucleotide modification yielded SeqAΔ(41–59)-A25R–DNA crystals that diffracted X-rays to 3.0 Å resolution. Three modifications of the DNA sequence were key to obtaining these cocrystals: (i) setting the GATC spacing to nine base pairs, (ii) fixing the duplex length to 20 base pairs and (iii) the inclusion of dinucleotide overhangs. As revealed from the molecular-replacement solution, all three modifications improved the packing of the SeqAΔ(41–59)-A25R–DNA complex. On the other hand, the CG snap-fastener did not have an effect on the diffraction quality of the crystals, probably because crystal packing was not mediated by groove–backbone interactions between neighbouring DNA molecules.

Unexpectedly, the overhanging base pairs did not mediate head-to-tail interactions between symmetry-related DNA molecules. Instead, they fold away from the duplex axes and interact with neighbouring protein molecules. The single-stranded dinucleotide on the methylated strand folds back and interacts with the C-terminal domain of the SeqAΔ(41–59)-A25R molecule bound to its 5’ hemimethylated GATC site. While the 5’ guanine (G1) is hydrogen bonded to the side chains of Lys136 and Glu125 (Figure 2.3), adenine A2 does not interact with the protein, revealing why a single overhang could not support this interaction. We are currently assessing whether this interaction strengthens the protein–DNA complex or simply facilitates crystal packing. Conversely, the other end of the DNA duplex does not mediate such intimate contacts with the protein. The single-stranded dinucleotide on the unmethylated strand interacts with the N-terminal domain of a symmetry-related SeqAΔ(41–59)-A25R molecule. However, both T1 and C2 appear to be highly flexible as revealed by weak and

disconnected electron density. In conclusion, the presence of two-nucleotide overhangs was the most important feature to obtain diffraction-quality crystals of SeqAΔ(41–59)-A25R bound to hemimethylated DNA. Our results indicate that the length of single-stranded overhangs may be a powerful variable to consider in protein–DNA cocrystallization even when end-to-end stacking of neighbouring DNA molecules is not expected.

2.6 Acknowledgements

We thank the PXRR staff at the NSLS (Brookhaven National Laboratory) for assistance during data collection. We are grateful to Monica Pillon for help with DNA purification. This work was supported by the Canadian Institutes of Health Research (MOP 67189).

Chapter 3

Structural insights into the cooperative binding of SeqA to a tandem GATC repeat

The full citation is as follows:

Chung, YS., Brendler, T., Austin, S and Guarné, A. (2009) Structural insights into the cooperative binding of SeqA to a tandem GATC repeat. *Nucleic Acids Research*, **37(10)**, 3143-3152.

3.1 Author's Preface

The work presented in chapter 3 has been published in the peer-reviewed *Nucleic Acids Research* (Impact factor of 7.836) and appears in published format. This work describes the structural and functional characterization of the SeqA dimer, the minimal functional unit with DNA. I have solved the first crystal structure of the SeqA functional unit bound to hemimethylated DNA and propose a model to explain the DNA binding constraints of SeqA, based on the relative spacing between adjacent GATC sites. Furthermore, I propose a potential role of SeqA at the replication forks. I conducted all biochemical and structural analysis of SeqA bound to DNA and the electrophoretic mobility shift assays. Dr. Therese Brendler and Dr. Stuart Austin carried out the DNA-binding specificity assay and flow cytometry. The four of us analyzed the data and wrote the manuscript.

3.2 Abstract

SeqA is a negative regulator of DNA replication in *Escherichia coli* and related bacteria that functions by sequestering the origin of replication and facilitating its resetting after every initiation event. Inactivation of the *seqA* gene leads to unsynchronized rounds of replication, abnormal localization of nucleoids and increased negative superhelicity. Excess SeqA also disrupts replication synchrony and affects cell division. SeqA exerts its functions by binding clusters of transiently hemimethylated GATC sequences generated during replication. However, the molecular mechanisms that trigger formation and disassembly of such complex are unclear. We present here the crystal structure of a dimeric mutant of SeqA (SeqAΔ(41–59)-A25R) bound to tandem hemimethylated GATC sites. The structure delineates how SeqA forms a high-affinity complex with DNA and it suggests why SeqA only recognizes GATC sites at certain spacings. The SeqA–DNA complex also unveils additional protein–protein interaction surfaces that mediate the formation of higher ordered complexes upon binding to newly replicated DNA. Based on this data, we propose a model describing how SeqA interacts with newly replicated DNA within the origin of replication and at the replication forks.

3.3 Introduction

Initiation of replication is the cascade of events that causes unwinding of DNA at an origin of replication. In *Escherichia coli*, regulation of replication initiation ensures that the chromosome is replicated once, but only once, during the cell cycle. Three main processes control timing and synchrony of replication initiation: regulatory inhibition of the initiator protein DnaA (RIDA), titration of free DnaA and sequestration by SeqA [72,83,116]. The initiator protein DnaA forms two different complexes with the origin of replication (*oriC*). These are analogous to the eukaryotic origin recognition and the pre-priming complexes [117]. The origin recognition complex is formed when DnaA binds to the three high-affinity DnaA boxes within *oriC*. This complex persists throughout most of the cell cycle, whereas assembly of a pre-priming complex only occurs at the time of initiation of DNA synthesis and requires binding of DnaA to high and low-affinity recognition sites in an ATP-dependent fashion [61]. Upon initiation of replication, the SeqA protein forms a high-affinity complex with transiently hemimethylated GATC sites within *oriC* that partially overlap low-affinity DnaA boxes. This process, known as sequestration of *oriC*, represses the assembly of the pre-priming complex and, in turn, ensures that all origins are reset to form an origin recognition complex before a new round of replication starts [60,85]. SeqA sequestration is also important for survival of replication fork damage because it prevents convergence of forks upon DNA damage [118].

Fluorescently labeled SeqA proteins form visible foci in the cells. These appear to be clusters of SeqA bound to newly replicated DNA at the replication forks rather than at the origin of replication [91,101,119]. Formation of these foci depends on Dam methylation and ongoing DNA replication, but not on the presence of *oriC* [90,91,111].

This suggests that SeqA binding to newly replicated DNA also plays a role in organization of the chromosome. Indeed, *seqA null* strains exhibit increased negative superhelicity and abnormal localization of nucleoids [87,120], and mutation of the condensin-like protein MukB, a known participant in DNA segregation, has a mutually suppressive effect [121]. SeqA binding beyond the origin is also necessary for a full stringent response and cell cycle arrest [122]. Excess SeqA also interferes with nucleoid segregation, causing a delay in cell division and affecting topoisomerase IV activities [88,93]. These roles of SeqA have also been identified in bacteria bearing more complex genomes than *E. coli* such as *Vibrio cholerae* [123].

SeqA has two functional domains: an N-terminal oligomerization domain (residues 1–35), and a C-terminal DNA-binding domain (residues 64–181) [97]. The C-terminal domain of SeqA (SeqA-C), binds specifically to hemimethylated GATC sequences [97], but it can also recognize other hemimethylated sequences to a lesser extent [124]. The N-terminal domain mediates dimerization of SeqA. SeqA dimers further associate to form left-handed spiral linear polymers [98]. While SeqA dimer formation is sufficient to form a high-affinity complex with DNA, filament formation is required for proper function *in vivo*. A molecular model based on the crystal structures of the N- and C-terminal domains of SeqA suggests that the SeqA foci at the replication forks could be SeqA polymers restraining negative supercoils on newly replicated DNA [98]. In agreement with this, wild type SeqA restrains negative supercoils in DNA [102]. Interestingly, SeqA mutants with impaired ability to form filaments have been shown to introduce positive supercoils in DNA [102]. A wealth of information regarding how various mutations affect the activities of SeqA has been generated [98,102,125,126].

However, the molecular interactions that trigger the formation of a high-affinity complex of SeqA with newly replicated DNA have remained unclear.

In an effort to understand the topological constraints imposed by the SeqA dimer on newly replicated DNA, we have solved the crystal structure of a SeqA mutant unable to form oligomers larger than dimers bound to a DNA duplex containing two adjacent hemimethylated GATC sequences. The structure reveals how SeqA forms a high-affinity complex with newly replicated DNA and why it prefers to bind to GATC sites that are on the same face of the DNA helix. The organization of the linkers connecting the dimerization and DNA-binding domains suggests that SeqA uses a stepwise unwinding mechanism to bind pairs of GATC sequences with different spacings. Moreover, the association of SeqA–DNA complexes in the crystal unveils additional protein–protein interaction surfaces within the DNA-binding domain of SeqA that do not play a role in replication synchrony but could affect chromosome organization at the replication fork. Based on these findings, we propose a model to explain the independent roles of SeqA at *oriC* and the replication forks.

3.4 Materials and Methods

3.4.1 Crystallization and structure determination

SeqAΔ(41–59)-A25R was over-produced and purified as previously described [127]. Oligonucleotides $^{5'}$GAGTCG(mA)TCGGCGGG(mA)TCCTTA$^{3'}$ and $^{5'}$TCTAAGGATCCCGCCGAT CGAC$^{3'}$ were annealed to yield a 20-bp duplex encompassing two hemimethylated GATC sites. Crystallization and data collection details of SeqAΔ(41–59)-A25R bound to DNA

were described elsewhere [127]. The initial phases were determined by molecular replacement using Phaser [115] and the structures of the N- and C-terminal domains of SeqA as models (PDB ID: 1XRX and 1LRR). The structure was refined using standard protocols in Refmac and Phenix.refine [128,129].

3.4.2 SeqA mutants

All SeqA mutants were derived from a pET-11a plasmid encoding wild-type SeqA (pSS1/pAG8013, Table A.1) using QuikChange site-directed mutagenesis kit (Stratagene). Sequences of the mutants were verified by DNA sequencing (MOBIX Laboratory at McMaster). Mutant SeqA proteins were over-produced and purified as described elsewhere [98].

3.4.3 Electromobility shift assays

The same oligonucleotides used for crystallization were annealed and used for electromobility shift assays. Constant amounts of DNA duplex (80 nM) were incubated with increasing amounts of protein (15–250 nM) and incubated 15 min at room temperature followed by 30 min at 4°C. Samples were resolved on 10% native TBE gels and stained with SYBR Green (1:10 000) (Cambrex, Inc.). Bands were quantified using the ImageJ software (http://rsbweb.nih.gov/ij/index.html). Experiments were run in triplicate and the K_d was estimated from the average plots.

3.4.4 DNA-binding specificity assays

The randomly chosen 72-bp sequence and the design of hemimethylated DNA duplexes with two GATC sites at various spacings used for SeqA protein binding were as described

earlier [100,101]. The preparation and radioactive labeling of hemimethylated duplex DNA and the conditions for the SeqA binding electrophoretic mobility shift assay (EMSA) were done as described previously [84,100]. Unless otherwise stated, SeqA protein binding assays were performed using 0.2 μM of either wild-type or mutant SeqA protein.

3.4.5 Flow cytometry assays

Host strain BL21DE3/pLysS was made *ΔseqA::tet* by P1 transduction with lysate from MM294 *ΔseqA::tet* (a kind gift from Dr. Kleckner). This strain was supplemented with pET-11a derivatives encoding: SeqA (pSS1/pAG8013, Table A.1), SeqA-A25R (pAG8015, Table A.1), SeqAΔ(41–59) (pAG8023, Table A.1), SeqAΔ(41–59)-A25R (pAG8033, Table A.1), SeqA-R70S/R73S (pAG8270, Table A.1), SeqA-A25R/R70S/R73S (pAG8268, Table A.1) and SeqAΔ(41–59)-A25R/R70S/R73S (pAG8269, Table A.1). In each case, the average number of origins per cell was determined by the flow cytometry 'run-off' method with modifications [130]. Overnight cultures were grown in the absence or presence of 25 μM IPTG at 37°C in M63 minimal media with the appropriate antibiotics. The overnight cultures were diluted to an OD_{600} of 0.02 and grown to an OD_{600} ~0.1 prior to incubation for 3 h with rifampicin (200 μg/ml) and cephalexin (36 μg/ml). After fixing with 77% ethanol, cells were analyzed in a Bryte SH flow cytometer (Biorad) using WinBryte software (Figure 3.1) and in an Apogee A40-Mini FCM flow cytometer (Apogee Flow Systems) using Apogee Histogram Software version 1.94 (Figure 3.4D).

3.5 Results and Discussion

3.5.1 DNA binding plasticity of SeqA

We had previously shown that the SeqA-A25R mutant disrupts the ability of SeqA to form filaments and hence causes loss of replication synchrony [98]. However, replication synchrony of the SeqA-A25R mutant could be restored by protein over-expression suggesting that filament formation aids the assembly of a sequestered origin complex, but that the complex can still form in the presence of a high local concentration of SeqA. For this study, a double mutant of SeqA lacking residues 41–59 within the linker connecting the two functional domains and carrying the A25R point mutation was generated [127]. Similarly to SeqA-A25R, the SeqAΔ(41–59)-A25R double mutant was stable at low ionic strength and eluted from a size-exclusion column at a volume consistent with the formation of a dimer (Figure 3.1A). Deletion of residues 41–59 abolished the ability of SeqA to bind hemimethylated GATC sequences separated by more than one helical turn (Figure 3.1B). This restricted DNA binding favoured crystallization of SeqAΔ(41–59)-A25R bound to a tandem GATC repeat. The plasmid encoding SeqAΔ(41–59) restored replication synchrony of the *ΔseqA::tet* strain similarly to wild-type SeqA (Figure 3.1C), suggesting that filament formation can compensate the DNA-binding defects caused by the deletion of the linker region. However, the SeqAΔ(41–59)-A25R double mutant behaved similarly to SeqA-A25R [98] and only restored replication synchrony upon protein overexpression as seen by the even number of chromosome equivalents (Figure 3.1C).

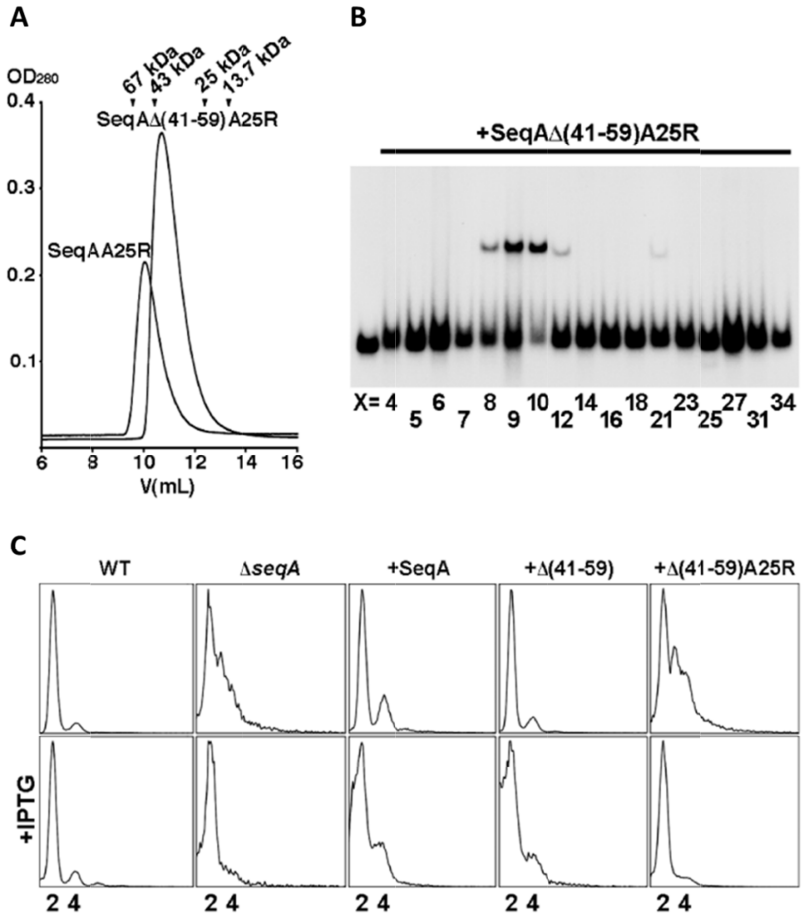

Figure 3.1. Characterization of the SeqAΔ(41–59)-A25R. ***A****. Elution profiles of SeqA-A25R (20,315 Da) and SeqA⊿(41–59)-A25R (18,513 Da) over a Superdex-75 size exclusion chromatography column (GE Healthcare). Elution volumes of albumin (67 kDa), ovoalbumin (43 kDa), chymotrypsinogen A (25 kDa) and ribonuclease A (13.7 kDa) are indicated.* ***B****. Electrophoretic mobility shift assay of SeqA⊿(41–59)-A25R with DNAs containing two hemimethylated GATC sequences separated by a variable number of base pairs (X). The left-most lane contains an equimolar mixture of DNAs with 5, 7, 12, 21, 25 and 34 base pairs between the two GATC sites but no protein.*

Continued on pg. 53

Continued from pg. 52

C*. From left to right, flow cytometry profiles of a wild-type strain, the ΔseqA::tet strain and the ΔseqA::tet strain transformed with pET-11a plasmids encoding wild-type SeqA, SeqAΔ(41–59), and SeqAΔ(41–59)-A25R. Wild-type SeqA and SeqAΔ(41–59) restore replication synchrony, however synchrony is lost upon protein overexpression. Conversely, SeqAΔ(41–59)-A25R only restores replication synchrony upon protein overexpression by addition of 25 μM IPTG.*

3.5.2 Structure determination of SeqAΔ(41–59)-A25R

SeqAΔ(41–59)-A25R was crystallized in complex with a hemimethylated duplex containing two GATC sequences separated by 9 bp [127]. The structure was solved by molecular replacement using the structures of the N- and C-terminal domains of SeqA (PDB ID: 1XRX and 1LRR) and refined using standard protocols in Refmac and Phenix.refine [128,129]. The asymmetric unit contains two identical protein–DNA complexes related by a 2-fold axis. The final model comprises two copies of protomer A (residues 1–40/60–181), two copies of protomer B (residues 1–35 and 60–181), two copies of the hemimethylated DNA duplex (with the exception of nucleotide Cyt2 from the unmethylated strands), thirty two water molecules and four 2-methyl-2,4-pentanediol (MPD) molecules (Table 3.1 and Figure 3.2). Over 97% of the residues lie in the most favoured regions of the Ramachandran plot, and none in disallowed regions.

Table 3.1. Refinement statistics

	SeqAΔ(41–59)-A25R
Refinement	
Resolution (Å)	35–3
No. of reflections (work/test)	30,285/1533
R_{work}/R_{free}	21.9/24.5
No. atoms	
Protein	5142
DNA	1758
MPD	4
Water	32
B-factors	
Protein/DNA	68.52
Solvent	46.59
R.m.s d	
Bond lengths (Å)	0.004
Bond angles (°)	0.877

3.5.3 Organization of SeqA functional domains

The SeqA monomer is organized into two domains, an N-terminal oligomerization domain (residues 1–33) and a C-terminal DNA-binding domain (residues 65–181) joined by a flexible linker (residues 34–64). Each SeqA monomer includes four β-strands (βN1 at the N-terminus, and βC2, βC3 and βC4 that define the small anti-parallel β-sheet within the C-terminal domain), and nine α-helices (αA and αB at the N-terminus and αC, αC1, αD, αE, αF, αG and αH forming the C-terminal domain) (Figure 3.2A and D). Formation of the SeqAΔ(41–59)-A25R dimer is mediated exclusively by the N-terminal domain, whereas the two DNA-binding domains of the dimer are unrelated to one another (Figure 3.2B and C). Notably, the A25R mutation did not change the extended conformation of the αA–αB loop (residues 18–24), leaving the side chain of isoleucine

21 completely exposed to solvent (Figure 3.2B). Minor changes on the tracing of the main chain at loop αA–αB were attributed to crystal-packing environment.

Although SeqA can recognize pairs of hemimethylated GATC sites with the methyl groups on the same or opposite DNA strands *in vitro* [100], newly replicated GATC sequences have all methyl groups on the template strand. Binding in this configuration forces the SeqA dimer to re-arrange its DNA-binding domains to recognize tandem hemimethylated GATC sites. Indeed, while the two N-terminal domains in the SeqA dimer are related by a 2-fold axis, the symmetry of SeqAΔ(41–59)-A25R is lost at the flexible linker joining the N- and C-terminal domains (Figure 3.2B and C). This linker mediates the 180°-rotation of one of the DNA-binding domains required to recognize a tandem GATC repeat (Figure 3.2C).

In protomer A, the linker is completely ordered and mainly helical, with helix αB encompassing residues Arg25 to Ser39 (Figure 3.2A–C). However, the electron density of this linker in protomer B was very weak. Consequently, residues Ser36 to Gln40 were not included in the final model (Figure 3.2B). Lys34 is the pivotal point that re-orients the C-terminal domains towards the target DNA. While in protomer A Lys34 is part of helix αB, Lys34 exchanges the orientations of its main and side chains in protomer B, breaking the 2-fold symmetry (Figures 3.2C and 3.3D). Note that the two protomers of the dimer cannot adopt the conformation seen in protomer A concurrently as this would cause steric hindrance at residues Ala37–Gln40 (Figure 3.2C). Therefore, Lys34 may be an intrinsically flexible point in the SeqA protein even in the absence of DNA.

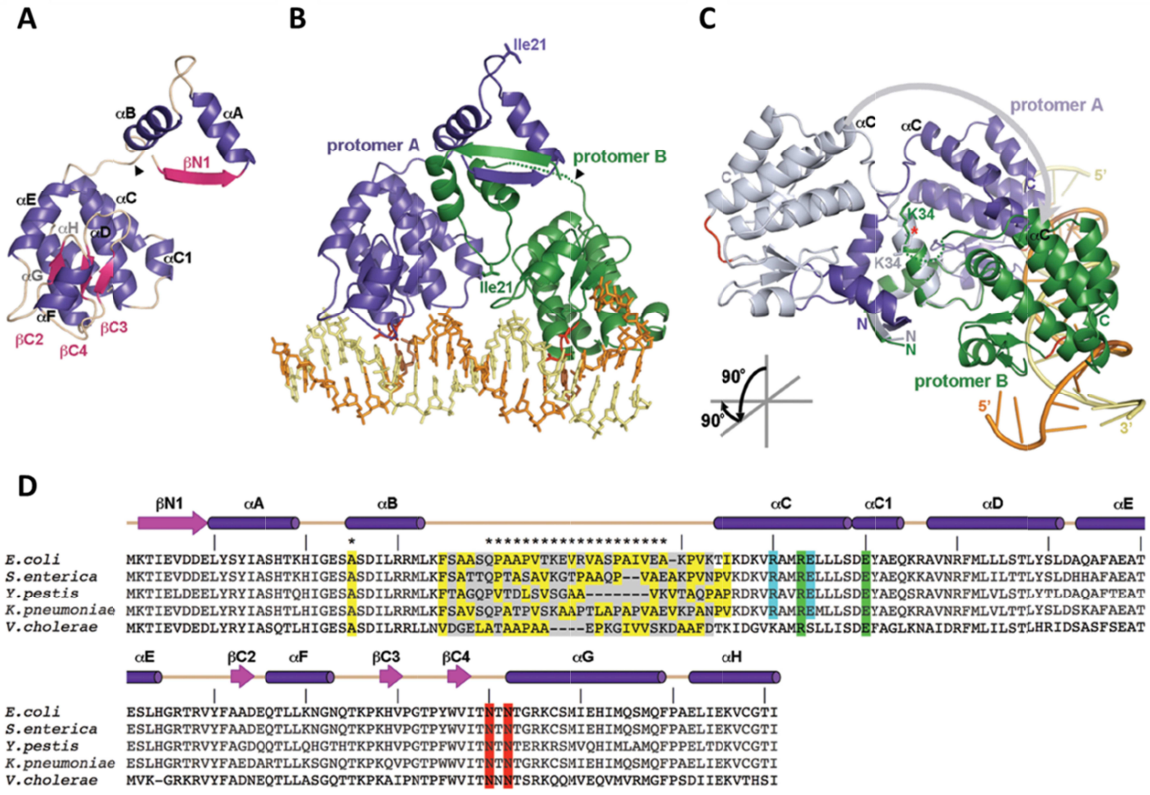

Figure 3.2. Structure of SeqAΔ(41–59)-A25R bound to a hemimethylated GATC repeat.

***A**. Ribbon diagram of the SeqAΔ(41–59)-A25R monomer with helices in purple and strands in pink. The junction between residues Gln40 and Lys60 is indicated with a black arrowhead. **B**. SeqAΔ(41–59)-A25R dimer bound to DNA. Protomer A (encompassing residues 1–40/60–181) is shown in purple and protomer B (encompassing residues 1–35/60–181) is shown in green. The disordered linker in protomer B is shown as a green dotted line. DNA binding loops are shown in red with the side chains of residues Asn150 and Asn152 as red sticks. Ile21 at the tip of the αA–αB loops are shown as sticks. The hemimethylated DNA is shown in orange (methylated strand) and light yellow (unmethylated strand) with the methylated adenine in brown.*

Continued on pg. 57

Continued from pg. 56

***C**. Ribbon diagram depicting how the C-terminal domain breaks the two-fold symmetry of the N-terminal domain. The SeqAΔ(41–59)-A25R dimer is shown in purple (protomer A) and green (protomer B). Protomer A superimposed onto the N-terminus of protomer B is shown in light grey. The first α-helix on the C-terminal domain (αC) is labeled for reference. The grey arrow indicates the rotation of the C-terminal domain around Lys34 (red asterisk and sticks). The cartoon illustrates the transformation from (**B** to **C**). **D**. Sequence alignment of SeqA from Escherichia coli K-12, Salmonella enterica serovar Typhi str. CT18, Yersinia pestis CO92, Klebsiella pneumoniae MGH 78578 and Vibrio cholerae 623-39 (top to bottom). Secondary structure motifs from SeqAΔ(41–59)-A25R are shown. The deletion and point mutations in SeqAΔ(41–59)-A25R are marked with asterisks. Ala25 and hydrophobic residues within the linker region are highlighted in yellow. The variable linker between the two functional domains is shadowed in grey. DNA binding residues are highlighted in red and those involved in reciprocal salt-bridges between neighbour C-terminal domains in green and blue. Panels (**A–C**) were prepared using PyMol [105].*

3.5.4 DNA conformation

The 22-bp long oligonucleotides used for crystallization formed a 20-bp DNA duplex with two overhanging nucleotides at each 5'-end (Figure 3.3A). The overhanging dinucleotide on the methylated strand projects away from the duplex helical axis and interacts with the C-terminal domain of protomer B (Figure 3.2B and C). The overhanging dinucleotide on the unmethylated strand is more flexible and only the 5' thymine, which stacks on top of adenine 22 from the methylated strand, could be fitted on the electron density maps (Figure 3.3A). The duplex portion of the DNA molecule adopted an overall B-DNA conformation with local distortions around the methylated mA–T base pair (Table 3.2). The N^6-methylated A (mA) maintains both canonical hydrogen bonds with its paired T, but the propeller twist between the bases is -19° for both methylated sites rather than the average -11.4° [131,132]. This distortion is also present in the structure of the isolated C-terminal domain (SeqA-C) bound to a single hemimethylated GATC sequence, but not in the structures of free DNA encompassing hemi or unmethylated GATC sites. Similarly to SeqA, the mismatch repair protein MutH specifically recognizes hemimethylated GATC sequences. Binding of MutH to un- and hemimethylated GATC sites also imposes significant distortions on the DNA [133], however the nature of these distortions is different to those imposed by SeqAΔ(41–59)-A25R binding to DNA.

Other significant distortions (namely roll and inclination) were also seen at the mA–T$\rightarrow$T–A base-pair junction (Table 3.2). However, they are not discussed further because similar distortions are found in the NMR structures of free DNA encompassing a single GATC site [134], and hence cannot be caused by protein binding. Interestingly, a recent structural study on a fully methylated GATC sequence reveals that methylation of the second strand compensates the base-opening distortions introduced by the first

methyl group [135], suggesting that distortions at a hemimethylated GATC sequence are both a requisite and a consequence of specific binding of SeqA and MutH. Interestingly, the DNA duplex is significantly overwound at the base-pair junctions between G–C→mA–T and mA–T→T–A. This overwinding is partially compensated by the underwinding at the T–A→C–G junction of the GATC site (Table 3.2). It is tempting to speculate that duplex overwinding is the mechanism used by dimeric SeqA to introduce positive supercoils into DNA [102]. However, the minor overwinding detected in our structure (11° over 19 bp steps) could also be related to the overall DNA sequence. In fact, the G–C→mA–T base-pair junction is also significantly overwound in the crystal structures of MutH bound to DNA, as well as, the NMR structures of free unmethylated and hemimethylated DNA (Table 3.2). Conversely, the high-resolution crystal structure of a hemimethylated GATC site does not show this distortion [136], suggesting that the GATC sequence is intrinsically dynamic and that protein binding may stabilize a specific DNA conformation.

Table 3.2. DNA parameters at the hemimethylated GATC sites

Structure	PDB ID	GATC	GATC methylation	Propeller mAT	Opening mAT	Roll mA→T	Inclination mA→T	Helical Twist G→mA	Helical Twist mA→T	Helical Twist T→C
SeqAΔ(41–59)-A25R/DNA	3FMT	2	Hemi-	−17.0	−5.7	−14.3	−21.1	43.0	40.6	29.8
				−19.3	−5.9	−13.6	−21.2	45.2	38.4	33.3
SeqA-C/DNA	1LRR	1	Hemi-	−21.2	−9.7	−10.3	−17.0	46.1	35.9	31.5
MutH/DNA	2AOQ	1	Unmeth-	−6.5	0.5	0.9	1.7	40.1	31.7	20.0
MutH/DNA	2AOR	2	Hemi-	−6.3	−0.7	1.8	3.3	40.4	31.8	20.3
				−5.8	0.6	−0.7	−1.2	40.5	32.0	20.5
GATC duplex	1OPQ	1	Unmeth-	11	−8.7	−12.3	−19.4	38.1	37.7	32.6
GATC duplex	1OQ2	1	Hemi-	−8.3	4.1	−7.2	−18.5	48.6	23.0	38.0
	1DA3	2	Hemi-	−9.3	−3.7	4.2	8.0	36.5	30.5	32.5
				−11.5	−0.1	−3.3	−5.9	36.3	33.0	35.6
Average B-DNA [132]				−11.4	0.6	0.6	2.1		36.5	

3.5.5 DNA–SeqAΔ(41–59)-A25R interactions

Although it had been predicted that the dimerization domain of SeqA would introduce restraints in the complex forcing the interaction with the 4 bp of the GATC site [97,124], the structure of SeqAΔ(41–59)-A25R bound to DNA exhibited the same sequence-specific interactions seen on the structures of the isolated C-terminal domain bound to a single GATC site (Figure 3.3A and B). Hence, the presence of the N-terminal domain does not seem to modulate the specificity of the interaction with hemimethylated GATC sites. However, the binding of a second GATC site significantly increases the stability of the DNA–SeqAΔ(41–59)-A25R complex (K_d ~90 nM), suggesting that the binding cooperativity of the SeqA dimer might be related to fulfilling its valence (Figure 3.3C).

Protomer B binds the GATC sequence very similarly to the previous structures; however, protomer A is only bound loosely to its GATC site (Figure 3.3B). Conceivably, the tighter binding of protomer B could be due to the additional interactions with the overhanging $^{5'}$GA dinucleotide. However, the isolated C-terminal domain of SeqA (SeqA-C) achieves a similar interaction with the GATC site in the absence of any additional contacts with DNA [97,124]. The linker region, connecting the oligomerization and DNA-binding domains, is ordered and mostly helical (αB) in protomer A, while it is flexible and unstructured in protomer B (Figures 3.2 and 3.3D). Unwinding of a helical turn on αB may allow the apparent tighter DNA binding displayed by protomer B. Conversely, the longer αB helix in protomer A may prevent its tight binding with DNA. Interestingly, the last helical turn on αB (Ser36–Gln40) from protomer A no longer maintains the hydrogen bonding pattern expected for an α-helix, suggesting that this helix is also unwinding to reach its GATC site (Figures 3.3D). Therefore, the αB of protomer A in the

SeqAΔ(41–59)-A25R crystal structure may represent an intermediate state on the formation of a high-affinity complex with DNA.

The linker region (residues Lys34–Lys63) is chiefly hydrophobic, with 18 residues out of 30 being Ala, Ile, Phe, Pro or Val. While the length and sequence of the linker varies among different organisms, its hydrophobic character is conserved (Figure 3.2D). Some of these residues would be shielded if the linker region forms an amphipathic α-helix running along the surface of the dimer. The presence of several proline residues within the linker would aid on kinking the helix along the surface. Indeed, the C-terminal domains of the SeqAΔ(41–59)-A25R dimer fold against one side of the dimerization domain (Figure 3.2C) rather than adopting the extended conformation predicted previously [98]. This organization only allows interactions with GATC sites that reside approximately on the same face of the DNA and at a very restricted spacing (Figure 3.1B). Assuming that the full-length SeqA dimer adopts a similar organization, binding of GATC repeats separated by more than one helical turn could be mediated by further stepwise unwindings of the helical linker.

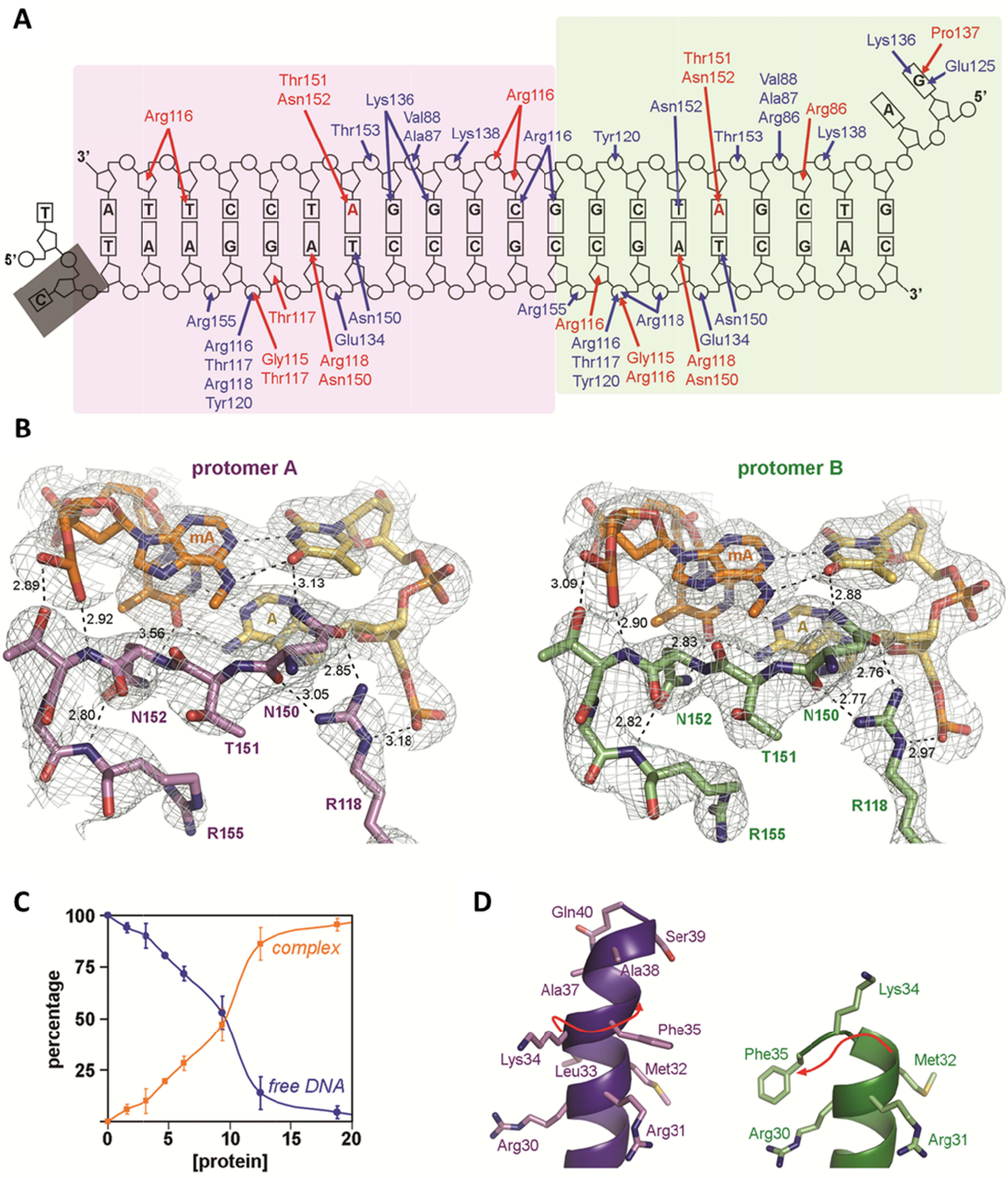

Figure 3.2. Recognition of hemimethylated DNA by SeqAΔ(41–59)-A25R.

Legend on pg. 64

Figure 3.3. Recognition of hemimethylated DNA by SeqAΔ(41–59)-A25R. ***A****. Diagram of the protein–DNA interactions. The purple and green shadow boxes indicate the SeqAΔ(41–59)-A25R protomer that interacts with each GATC site with hydrogen bonds shown in blue and van der Waals interactions in red. The methylated adenines are labeled in red and the disordered Cyt2 is shadowed with a grey box.* ***B****. Detail of the interaction between SeqAΔ(41–59)-A25R and the methylated and unmethylated A–T base pairs. The refined model is shown as sticks with protomer A shown in purple (left panel), protomer B in green (right panel), the methylated DNA strand in orange and the unmethylated DNA strand in yellow with the* $2F_o–F_c$ *electron density maps contoured at* 1σ*. Hydrogen bonds are shown as black dashed lines with distances labeled.* ***C****. Electrophoretic mobility shift assays of the oligonucleotide used for crystallization (80 nM) when incubated with increasing quantities of SeqAΔ(41–59)-A25R (nM).* ***D****. Detail of helix* α*B in the two protomers of the dimer. Red arrows indicate the path of the main chain on each protomer. Panels (****B*** *and* ***D****) were prepared using PyMol [105].*

3.5.6 Alternate surfaces facilitate DNA organization beyond *oriC*

The two SeqA–DNA complexes in the asymmetric unit interact through a reciprocal network of hydrogen bonds and hydrophobic interactions between residues Glu74, Asp79 and Leu77 from protomer A on one dimer and Arg70, Arg73 and Leu77 from protomer B on the adjacent dimer (Figure 3.4A). Arg70 forms a bi-dentate salt bridge with Glu74, while Arg73 is hydrogen-bonded to Asp79 (Figure 3.4A). This interaction was also found in the structure of SeqA-C encompassing residues 50–181 bound to DNA (Figure 3.4B), but not in the structures of shorter SeqA-C mutants lacking Arg70 [124].

This interface is relatively small, only ~400 $Å^2$ of surface is excluded by the pair-wise interactions. However, the residues involved in this interaction are well-conserved (Figure 3.2D) and their presence in several crystal structures suggests that this surface could contribute to the multimerization properties of SeqA. Indeed, a previous study had already shown that mutation of Lys66 and Arg70 did not affect DNA binding by SeqA but abrogated protein aggregation *in vitro* and foci formation *in vivo* [125]. Furthermore, conservation of Arg70 and Arg73 is correlated with conservation of Glu74 and Asp79, respectively even in most divergent species such as *V. cholerae* (Figure 3.2D). Interestingly, protein–protein association through αC seems to be promoted by DNA binding since SeqA-A25R and SeqAΔ(41–59)-A25R do not form higher order species in solution (Figure 3.1A and [98]).

To further explore the role of this surface on the function of SeqA, we mutated residues Arg70 and Arg73 and analyzed the ability of the SeqA-R70S/R73S mutant to restore synchrony of replication using flow cytometry. Although the R70S/R73S double mutation did not affect synchrony of replication (Figure 3.4D), *ΔseqA* cells transformed with plasmids encoding SeqA-R70S/R73S or SeqAΔ(41–59)-A25R/R70S/R73S grew much slower than *ΔseqA* cells transformed with plasmids encoding other SeqA mutants, reinforcing the idea that Arg70 and Arg73 are important for the function of SeqA *in vivo* as it had been previously proposed [125]. Therefore, we concluded that the SeqA interaction mediated by Arg70 and Arg73 is not required to sequester or reset *oriC*. However, it may contribute to additional functions of SeqA in fork management or chromosome segregation.

Conceivably, these weak interactions between αC helices from adjacent SeqA dimers may assist in compacting the SeqA–DNA filament (Figure 3.5). Interactions

through αC may cross-link dimers, either bound to different DNA duplexes (intermolecular cross-link), or bound to the same DNA duplex (intramolecular cross-link). We favour the latter because previous studies of complexes with duplex containing multiple GATC sites do not show the presence of intermolecular cross-linked complexes [103].

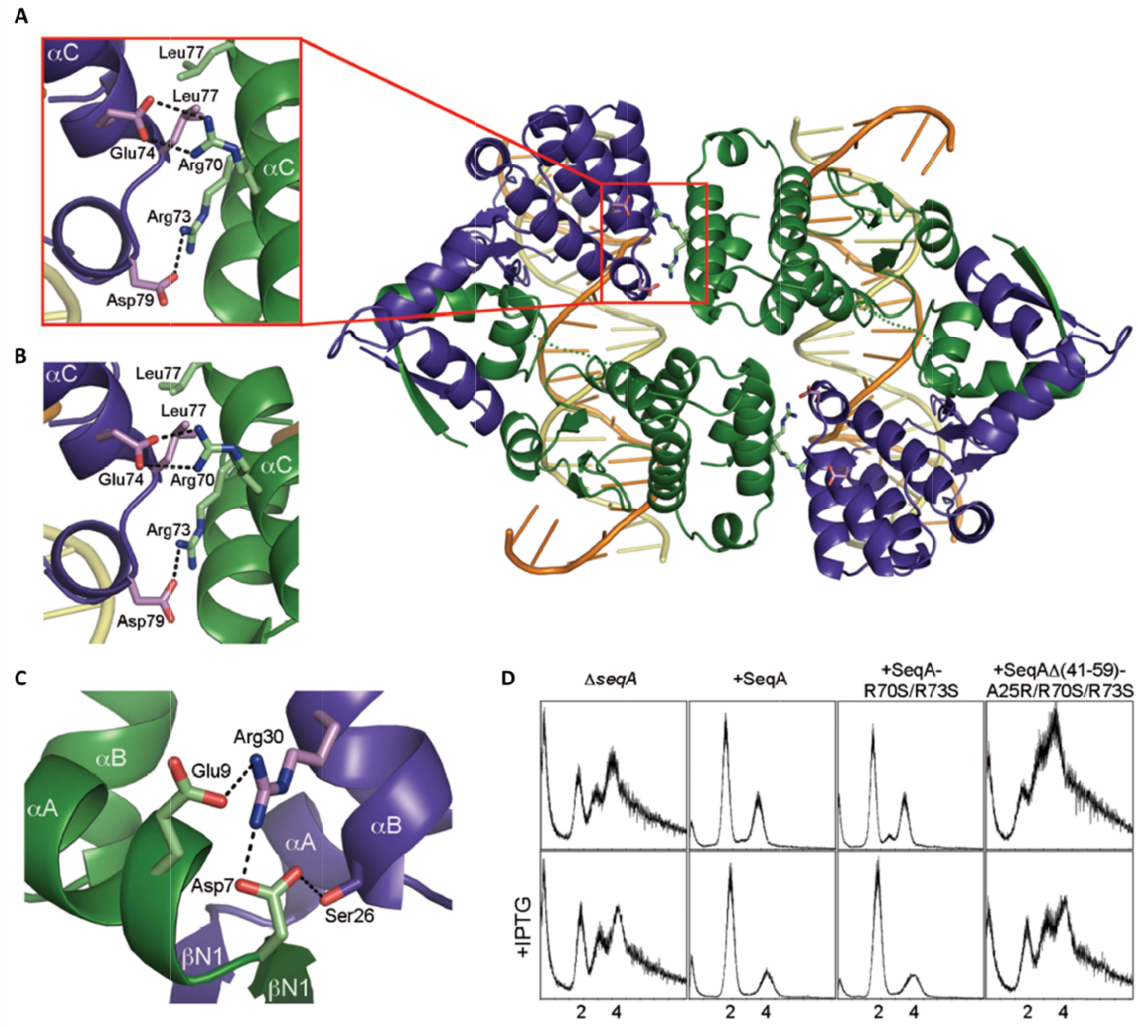

Figure 3.3. Additional aggregation surfaces in SeqA. ***A****. Ribbon diagram depicting the protein–protein interaction surfaces in the DNA–SeqAΔ(41–59)-A25R complex relating the two complexes of the asymmetric unit. Relevant residues involved in polar interactions are depicted as colour-coded sticks. A detailed view of the interaction surface between the adjacent SeqAΔ(41–59)-A25R dimers with hydrogen bonds depicted as dashed lines is shown on the left-hand side.*

Continued on pg. 68

Continued from pg. 67

***B**. Detail of the interface between the two SeqA-C molecules (PDB ID: 1LRR and [97]). **C**. Network of hydrogen bonds between Asp7, Glu9 and Arg30. **D**. Flow cytometry profiles of the ΔseqA::tet strain and the ΔseqA::tet strain transformed with pET-11a plasmids encoding wild-type SeqA, SeqA-R70S/R73S and SeqAΔ(41–59)-A25R/R70S/R73S in the absence (top) or presence of 25 μM IPTG (bottom). Panels (**A**–**C**) were prepared using PyMol [105].*

3.5.7 Distinct roles of SeqA at *oriC* and the replication forks

Apart from residues Asn150 and Asn152 in the DNA-binding loop [97,124], other key residues have been identified for the proper function of SeqA. Mutants that disrupt filament formation (A25R, T18E or I21R) retain binding to pairs of hemimethylated sites (Figure 3.5 and [98]), but lose cooperative binding to additional GATC sites [125]. SeqA mutants that are unable to form filaments cause replication asynchrony [98]. However, synchrony can be restored by protein overexpression, suggesting that the role of SeqA oligomerization is to facilitate saturation of linked sites rather than to remodel the origin of replication (Figure 3.5A).

Mutants affecting SeqA oligomerization may have a more deleterious effect at the replication forks where the GATC sites are spaced further apart and GATC-binding of few protomers along the SeqA filament might be the only mechanism to form a high-affinity complex with DNA (Figure 3.5 and [98]). Thus, the SeqA–DNA complex will likely weaken as the forks enter regions of the chromosome with low GATC content. Indeed, the original study by Campbell and Kleckner already revealed that the rate of adenine

methylation by dam methylase was inversely proportional to the distance from *oriC* with the exception of regions containing GATC clusters like the *recB* gene [137]. In this context, additional weak interactions between neighbouring DNA-binding sites mediated by helix αC may aid in extending the half-life of the SeqA–DNA complex (Figure 3.5B) and, in turn, assist in organizing newly replicated DNA.

Additional residues within the dimerization domain of SeqA also modulate its interaction with DNA. SeqA-D7K or SeqA-E9K do not bind pairs of GATC sites but can form larger complexes when more than two GATC sites are present (26). Close inspection of the SeqAΔ(41–59)-A25R structure reveals that Asp7 stabilizes the dimerization domain in two ways (Figure 3.4C). On one hand, it caps and hence stabilizes α-helix αA. On the other hand, it interacts with the side chain of Arg30 from the neighboring protomer, stabilizing the dimerization interface. Similarly, Glu9 is also hydrogen-bonded to Arg30. Mutation of Glu9 to an amino acid bearing an opposite charge (Glu→Arg) disrupts replication synchrony, but mutations removing the net charge of the side chain do not [98]. Conceivably, introducing a positive charge (E9R or E9K) in close vicinity to Arg30 may be more disruptive than simply removing the side chain of Glu9 (E9A), thus explaining the different phenotypes observed in these mutants [98,126].

In the absence of the Asp7 or Glu9 anchors, the SeqA dimer probably loosens up, causing the loss of detectable DNA binding by the SeqA dimer (Figures 3.4C and 3.5A). However, filament formation would presumably compensate for the weaker dimerization surface, since these mutants display enhanced DNA binding when more than two GATC sites are present [126], suggesting once again that the SeqA filament can compensate the defects of the dimer. Mutations affecting the stability of the SeqA dimer

are likely to have a significant impact in *oriC* sequestration where filament formation plays a mere supportive role (Figure 3.5A).

In conclusion, the structure of SeqAΔ(41–59)-A25R bound to a pair of hemimethylated GATC sites reveals how SeqA forms a high affinity complex with DNA and starts to unravel how different interaction surfaces contribute either to origin sequestration or DNA aggregation at the replication forks.

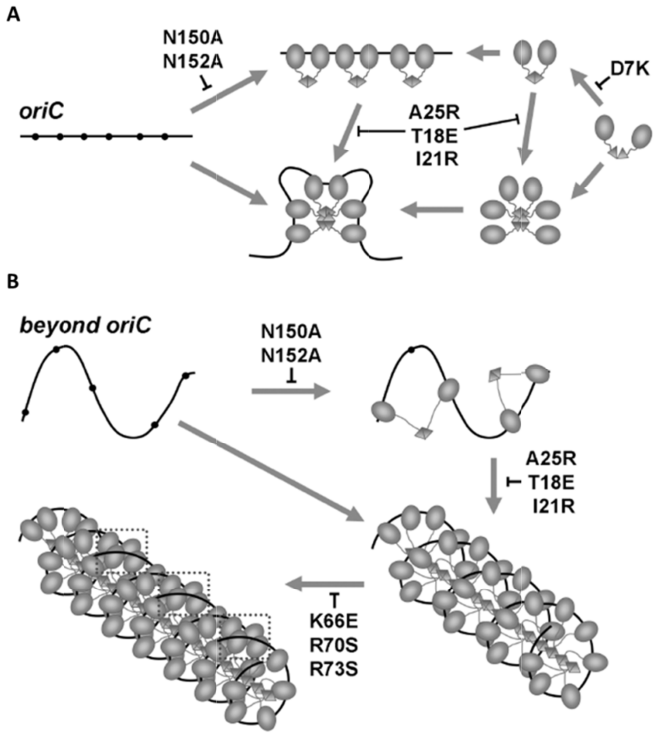

Figure 3.4. Modes of operation of SeqA at *oriC* and the replication forks. ***A****. Reduced spacing between GATC sites (shown as black dots) at oriC permits sequestration with only few SeqA molecules. Sequestration of the origin is abolished by lack of DNA binding (SeqA-N150A/N152A) and weakened by lack of SeqA oligomerization (SeqA-A25R, SeqA-T18E, SeqA-I21R).* ***B****. SeqA works similarly at the replication forks. However, formation of a left handed SeqA–DNA filament at the forks compensates for the increased spacing between GATC sites. SeqA–DNA filaments are further stabilized by protein–protein interactions through their DNA-binding domains (marked with boxes on the bottom left panel). Mutations in this surface (SeqA-K66E/R70E) affect SeqA induced DNA aggregation. Mutations destabilizing the dimerization domain (SeqA-D7K, SeqA-E9A) affect the functions of the SeqA dimer but not those of the SeqA filament and, hence, they likely have a more deleterious effect at the oriC.*

3.6 Coordinates

Atomic coordinates and structure factors have been deposited with the Protein Data Bank (accession code 3FMT).

3.7 Acknowledgements

We thank the PXRR staff at the NSLS (Brookhaven National Laboratory) for assistance during data collection and Monica Pillon for help with DNA purification.

3.8 Funding

Canadian Institutes of Health Research (MOP-67189 to A.G.); Intramural Research Program of the National Institutes of Health (to S.A.) in part. Funding for open access charge: Canadian Institutes of Health Research.

Chapter 4

Investigating the aggregational properties of *V. Cholerae* SeqA in the absence of the oligomerization domain

The work in this chapter is shown as a manuscript in preparation. The references from the manuscript are combined with the references from the thesis.

Authors: Chung, YS., Gupta, R.S. and Guarné, A.

4.1 Author's Preface

The work presented in chapter 4 is presented in the most recent edition. This work describes the structural and functional characterization of SeqA from *Vibrio cholerae* in the absence of the N-terminal oligomerization domain. I have shown that *V. cholerae* SeqA is structurally and functionally similar to *E. coli* SeqA. Dr. Gupta carried out the initial sequence analysis to identify SeqA homologues with truncations. I conducted all the work described in this chapter, prepared the figures and wrote the manuscript.

4.2 Abstract

SeqA is a negative regulator of DNA replication initiation in *Escherichia coli* and other bacterial species within the γ-proteobacteria class. At the *oriC*, SeqA binds to hemimethylated GATC sequences and sequesters the origin from new rounds of replication initiations. Beyond the *oriC*, SeqA binds to hemimethylated GATC sequences behind the replication forks forming large nucleoprotein complexes in a cooperative manner and forms aggregates that play a role in chromosome organization. Recent studies have shown that SeqA may play a similar role in the more complex organism, *Vibrio cholerae*. A blast search for SeqA homologues in *V. cholerae* have revealed that some strains of *V. cholerae* encode truncated versions of SeqA, lacking an N-terminal oligomerization domain responsible for cooperative binding. To examine if the conserved mechanism of protein–protein interaction at the replication forks can compensate for the lack of an N-terminal domain, we carried out structural and functional characterizations of *V. cholerae* SeqA. The crystal structure shows an identical structural organization to that of *E. coli* SeqA C-terminal domain. While the protein–protein interaction surface is intact within the crystal structure, *V. cholerae* SeqA does not show any aggregation on the DNA. Hence, our results suggest that the cooperative binding through the N-terminal domain may be necessary to achieve high-affinity binding.

4.3 Introduction

The faithful transmission of genetic material from one generation to the next is critical in the proliferation of all living organisms and requires the coordination and regulation of chromosome replication. In *Escherichia coli*, one regulatory mechanism of DNA replication initiation is through origin sequestration that plays a critical role in synchronizing DNA replication and ensuring that DNA replication occurs once and only once per cell cycle [58]. Origin sequestration prevents further rounds of replication initiation through the specific binding of SeqA to eleven hemimethylated GATC sequences within the *oriC* for approximately one third of the cell cycle [59]. These GATC sequences are normally fully methylated for replication initiation, but become hemimethylated as a result of semi-conservative DNA replication where the newly synthesized strand is unmethylated.

SeqA also binds to the hemimethylated GATC sequences that are generated with the passage of the replication forks in a cooperative manner, forming a large nucleoprotein complex that can be visualized as discrete foci [89]. *SeqA null* mutants display abnormal localization of nucleoids and increased negative superhelicity while cells with excess SeqA display delays in nucleoid segregation and cell division indicating that SeqA binding at the replication forks plays a role in chromosome organization as well as in proper segregation of the chromosomes [83,86,87,88]. The phenotypes seen with excess SeqA is also observed in the more complex organism, *Vibrio cholerae*, one of the few bacterial species that contains a multipartite genome [123].

V. cholerae contains two chromosomes, chromosome I (2.96 Mbp) and chromosome II (1.07 Mbp) each with a distinct *oriC* that is under replication control by

different initiation proteins [138]. However, replication initiation from both chromosomes is coordinated with the cell cycle indicating the presence of a global regulatory system that controls the timing of replication of both chromosomes [139,140]. Indeed, SeqA that regulates the timing of replication initiation in *E. coli* has shown to play a similar role in origin sequestration of both chromosomes in *V. cholerae* [123]. However, the roles of SeqA beyond the origin in chromosome organization have yet to be studied in *V. cholerae*.

Functional and structural characterizations of *E. coli* SeqA have shown two functionally distinct domains, the C-terminal DNA-binding domain and an N-terminal oligomerization domain, and the two domains are connected by a flexible linker region [97,98]. SeqA C-terminus interacts with the methylated adenine through van der Waals contacts and hydrogen bonds through the major groove of the DNA, which is stabilized by additional binding of the surrounding minor groove [97]. Besides DNA binding, SeqA C-terminus also contains a protein–protein interaction surface that mediates the formation of higher-ordered complexes upon binding to DNA [141]. The N-terminal domain forms a dimer which further oligomerizes through reciprocal hydrophobic interactions to form a filament [98]. Mutants with impaired filament formation display asynchronous DNA replication, but the resulting SeqA dimer is able to bind DNA similar to wild-type SeqA [98]. Synchrony of replication can be restored with overexpression of the SeqA dimer, suggesting that high local concentration of SeqA is sufficient for origin sequestration and does not require oligomerization. However, oligomerization is essential for SeqA function outside of the *oriC*, as SeqA dimers lose cooperative binding to additional GATC sequences and are unable to form foci *in vivo* [102]. The foci formation is also disrupted in SeqA mutants that are incapable of protein–protein

interactions through the C-terminal domain, suggesting that this interaction surface may induce aggregation of SeqA–DNA complexes [125].

Sequence alignment of SeqA homologues in *V. cholerae* demonstrated the presence of multiple *V. cholerae* strains that either lack parts of, or the entire N-terminal oligomerization domain based on the genomic sequence information. It is not clear if truncated forms of SeqA are still able to oligomerize in the absence of the N-terminal domain or have obtained alternate modes of oligomerization. To examine if the protein–protein interaction surface on the C-terminal domain contributes to the aggregation of SeqA on DNA, we have characterized a *V. cholerae* SeqA that lacks the N-terminal domain completely and show that this truncated SeqA is capable of binding DNA, but does not show DNA dependent aggregation.

4.4 Materials and Methods

4.4.1 Homology search and sequence alignment

Homologues of SeqA were found by doing a BLAST search with BLASTP (http://blast.ncbi.nlm.nih.gov) with the *E. coli* SeqA sequence (ACCESSION NO. AAA19855) using the default parameters. The BLAST search of SeqA in multiple strains of *V. cholerae* was carried out using TBLASTN (searching translated nucleotide using a protein query) searching the *Vibrio* Whole Genome Shotgun sequencing projects with the *E. coli* SeqA sequence (ACCESSION NO. AAA19855). Alignments of 16S ribosomal RNA and SeqA were carried out using CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2) [142]. On the basis of parameters obtained by using the CLUSTALW2

method, phylogenetic tree was drawn by using TREEVIEW software (http://taxonomy.zoology. gla.ac.uk/rod/treeview.html).

4.4.2 SeqA Mutants

SeqA gene from *V. cholerae* NCTC 8457 was synthesized from GenScript (pAG8305, Table A.1) (NJ, USA) and further cloned into a pET11a plasmid (pAG8306, Table A.1). This construct was insoluble and although the addition of an N-terminal Histidine tag using a pET15b plasmid (pAG8330, Table A.1) rendered the protein soluble, the protein was not cleavable by thrombin. Hence, an unremovable tag using the *E. coli* sequence ^{64}I/MKDKVRA^{71} was attached to the N-terminus of the *V. cholerae* SeqA in the pET11a plasmid (pAG8352, Table A.1). The *V. cholerae* SeqA utilized for crystallization had the MKDKVRA attachment and *V. cholerae* SeqA variants utilized for electrophoretic mobility shift assays all contained a single base mutation in the MKDKVRA attachment sequence to MKDKVSA to avoid any artificial effects it may have on aggregation. All mutants were generated by overlap PCR or using the QuikChange site-directed mutagenesis kit (Stratagene). Sequences of the constructs were verified by DNA sequencing (MOBIX Laboratory at McMaster). *V. cholerae* SeqA (pAG8352, Table A.1), *V. cholerae* SeqA* (pAG8526, Table A.1) and *V. cholerae* SeqA R2S I6D* (pAG8486, Table A.1) were overexpressed in BL21 (DE3), and *V. cholerae* SeqA* N78A N80A (pAG8527, Table A.1) was overexpressed in BL21 Star (DE3) (Invitrogen). All *V. cholerae* SeqA was purified as described elsewhere with the following modifications: the heparin column was loaded at 0.1M NaCl and run over a gradient of 0.1–0.7M NaCl and the MonoS ion exchange column was loaded at 0.1M NaCl and run over a gradient of 0.15–0.35M NaCl [127]. Purified *V. cholerae* SeqA was concentrated to 3mg ml^{-1} for crystallization. *E. coli* SeqA C-terminal domain was purified as described previously [97].

4.4.3 Crystallization, data collection and structure determination

V. cholerae SeqA crystals were grown in 30% PEG 4000 using the hanging-drop vapor-diffusion method and cryoprotected with the addition of 10% glycerol prior to freezing in liquid nitrogen. X-ray diffraction data from flash-frozen crystals diffracting to 1.5 Å was collected at wavelength of 1.1 Å at a temperature of 100 K at beamline X25C at the National Synchrotron Light Source (Brookhaven National Laboratory). The X-ray data was processed using HKL2000 [114]. The initial phases were determined by molecular replacement with PHASER using the C-terminal domain of *E. coli* SeqAΔ(41–59)-A25R (PDB Id: 3FMT) with non-identical residues changed to Alanine to remove any bias [115]. Structure refinement was carried out using PHENIX.refine and COOT [115,143]. All figures of the structure were prepared using PyMol [105]. The asymmetric unit contains two identical *V. cholerae* SeqA monomers. The final model consists of chain A and B (residues 1–41 and residues 46–109) and 238 water molecules (Table 1). Over 99% of residues lie in the favoured region of the Ramachandran plot, with no residues in the disallowed region.

4.4.4 Electrophoretic mobility shift assays

Oligonucleotides (YS1: $^{5'}$TTAAAAAGAAGATCTATTTATTTAGAGATCTGTTCTATTGTGATCTCT TATTAGGATCGCACTGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGG$^{3'}$ and YS2: $^{5'}$CCAGGTTGTTGATCTTAAAAGCCGGATCCTTGTTATCCACAGGGCAGTGCGATCCTAATA AGAGATCACAATAGAACAGATCTCTAAATAAATAGATCTTCTTTTTAA$^{3'}$) were synthesized from the KECK Oligonucleotide Synthesis Facility (Yale University) and purified as previously described [127]. YS1 contains six methylated adenines which are underlined and highlighted in red. Purified oligonucleotides were annealed at a 1:1 ratio and further purified over 4.5% non-denaturing PAGE to remove any tertiary structures that form

from annealing the DNA. The rest of the purification steps were carried out as previously described [127]. The subsequent dsYS1YS2 were radioactively labeled with T4 polynucleotide kinase (New England Biolab) and [γ-^{32}P] ATP. Electrophoretic mobility shift assay reactions were prepared by incubating 5 nM of ^{32}P-end-labeled hemimethylated DNA duplex YS1YS2 with increasing amounts of protein (20, 80, 320 and 1280 nM) at 30°C for 30 minutes in binding buffer (20 mM Tris pH 7.5, 10 mM Mg Acetate, 0.1 mM EDTA, 0.1% Igepal CA-630, 1 mM DTT, 5% Glycerol) with 0.1 μg/μl poly(dI-dC) (Sigma). After incubation, 1 μl of 6X loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol) was added and the samples were resolved on 4.5% native TBE gels at 100V in 45 mM Tris-borate and 1 mM EDTA buffer. After electrophoresis, the gels were dried and visualized by autoradiography.

4.5 Results

4.5.1 Diversity of SeqA in γ-proteobacteria

Previous phylogenetic studies have shown that SeqA is conserved in the Pasteruellaceae, Alteromonadaceae, Aeromonadaceae, Enterobacteriaceae and Vibrioneceae families within the γ-proteobacteria class. A search for SeqA homologues using the basic local alignment search tool (BLAST) with the *E. coli* SeqA amino acid sequence resulted in more than 350 hits. We have carried out sequence alignments of SeqA from nine different bacterial strains representing different bacterial families at least once (Figure 4.1A). SeqA proteins within its bacterial family display high sequence homology; however, there is greater sequence variation between different bacterial families,

especially in the linker region connecting the N- and C-terminus. Despite the sequence divergence, key residues that are important for the function of the protein are conserved across different bacterial species.

In *E. coli*, residues Asn150 and Asn152 mediate DNA binding of SeqA with the GATC sequence, and Arg118 further stabilizes this interaction through binding at the minor groove of the DNA [97]. All three of these DNA binding residues are conserved across all bacterial species. In addition, residues Thr18, Ile21 and Ala25 that are essential for oligomerization are all conserved as well. Asp7 and Glu9 are involved in stabilization of the SeqA dimer and Glu9 also plays a partial role in this stabilization of dimers, but only the negatively charged nature of the residue is conserved. Lastly, a partial conservation of the protein–protein interface reported in two previous studies is also present. Interaction between Arg70–Glu74 and Arg73–Asp79 has been reported to form a protein–protein interaction surface that helps to aggregate SeqA–DNA complexes. The Arg70–Glu74 and Arg73–Asp79 residues are completely conserved within the Enterobacteriaceae family, only the Arg73–Asp79 pair is conserved within the Vibrioneceae and neither pairs are conserved in Pasteruellaceae, Alteromonadaceae and Aeromonadaceae. The patterns of phylogenetic trees of small ribosomal RNA and SeqA homologues are similar to each other, suggesting the co-divergence of SeqA along with the evolution of the strains (Figure 4.1B–C). Despite the sequence divergence of SeqA, the changes are limited to similarly charged residues, suggesting the preservation of SeqA function.

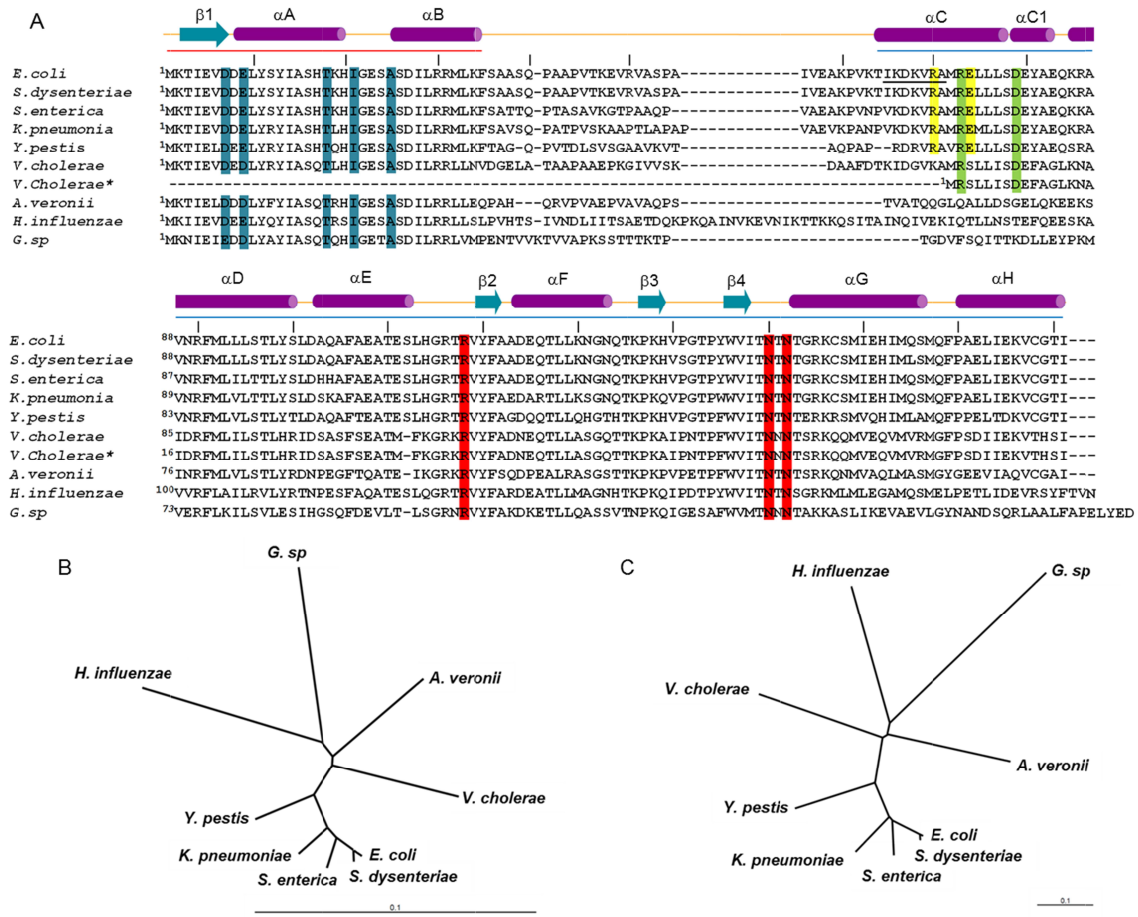

Figure 4.1. SeqA homologues in γ-proteobacteria. ***A****. Sequence alignment of SeqA from Escherichia coli K-12, Shigella dysenteriae Sd197, Salmonella enterica serovar Typhi str. CT18, Klebsiella pneumonia MGH 78578, Yersinia pestis KIM10, Vibrio cholerae N16961, Vibrio cholerae NCTC 8457 (V. cholerae*), Aeromonas veronii B565, Haemophilus influenzae R3021, Glaciecola species 4H-3-7+YE-5 (top to bottom). Secondary structure motifs based on the structure of SeqA dimer (PDB ID: 3FMT) are shown with helices in purple, strands in teal and loops in orange. The red and blue line below the alignment represents the N- and C-terminus, respectively.*

Continued on pg. 84

Continued from pg. 83

The black ticks above the sequence alignment correspond to every 10 amino acid residues. DNA binding residues are highlighted in red and residues involved in filament formation are highlighted in light blue. Residues involved in reciprocal salt-bridges and hydrogen bonds are shown in yellow and green. The underlined region of helix αC in the E. coli K-12 sequence represents the residues that were added to the VcSeqA for solubility. ***B****. A phylogenetic tree of the small ribosomal RNA from bacterial strains used in the sequence alignment.* ***C****. A phylogenetic tree of SeqA from bacterial strains used in the sequence alignment. Alignments of nucleotide or amino acid sequences were performed using the ClustalW2 software and phylogenetic trees were drawn with the TreeView program. The scale bar represents 0.1 substitutions per nucleotide site.*

4.5.2 Construction of soluble *Vibrio cholerae* SeqA

The BLAST search for SeqA homologues brought our attention to multiple strains of *V. cholerae* that contain SeqA with truncations to the N-terminal domain. Based on previous comparative genomic studies on the evolution and lineage of *V. cholerae* strains, the SeqA truncations do not show any apparent correlation to the evolution of *V. cholerae* strains [144]. Hence, it is not clear if this variation in length occurred through the evolution of *V. cholerae* strains or if the truncations are a spontaneous occurrence. We selected SeqA from *V. cholerae* NCTC 8457 strain that only contains the C-terminal domain in an attempt to understand if truncated SeqA retains its ability to form aggregates on DNA in the absence of the oligomerization domain (Figure 4.1A). Despite lacking the N-terminal domain, the C-terminal domain sequences are identical to SeqA

from *V. cholerae* N16961, the most commonly used strain for studies involving *V. cholerae*.

V. cholerae SeqA was insoluble and could not be purified. Based on secondary structure predictions, SeqA from *V. cholerae* NCTC 8457 and *E. coli* display similar organization of secondary structures. This suggested that *V. cholerae* SeqA C-terminal domain begins midway through the first helix, αC, likely exposing hydrophobic residues on the nearby α helices to the solvent. Hence, we generated a construct that contained a non-removable tag using sequences ^{64}I/MKDKVRA^{71} from *E. coli* SeqA. This construct rendered soluble protein and will be referred to as VcSeqA from here onwards.

4.5.3 Structure of *V. cholerae* SeqA

We have solved the crystal structure of *V. cholerae* SeqA using molecular replacement (Table 1). Each VcSeqA polypeptide chain consists of three β-strands (β2, β3 and β4) and seven α-helices (αC, αC1, αD, αE, αF, αG and αH) organized with an identical fold to that of *E. coli* SeqA C-terminus (EcSeqA) bound to DNA (Figure 4.2A). The structure of VcSeqA can be superimposed onto the structure of EcSeqA C-terminus bound to a single GATC site and EcSeqA dimer bound to two GATC sites with an r.m.s deviation of 0.673 and 0.874 Å for 667 and 684 atoms, respectively. As predicted from the secondary structure prediction of VcSeqA, helix αC forms a full helix with the addition of the residues MKDKVRA and shields the hydrophobic side chains of residues on helices αH and αD.

The DNA binding residues Asn78 and Asn80 are in an identical conformation to the corresponding DNA binding residues Asn150 and Asn152 of EcSeqA. On the other hand, residues 42–45 are disordered in both chains of the asymmetric unit. In the two

crystal structures of EcSeqA bound to DNA, this region forms a loop that binds the minor groove of the surrounding hemimethylated GATC sequence further stabilizing the protein–DNA interaction. Hence, it is likely disordered in our structure due to increased flexibility of the loop in the absence of DNA. The two monomers in the asymmetric unit are related by the non-crystallographic two-fold axis, forming a small interface, where ~200 $Å^2$ of the surface is excluded by hydrogen bonds between residues on helix αF (Figure 4.2B).

Table 4.1. Data collection and refinement statistics

Data Collection[a]	
Space Group	$P2_1\ 2_1\ 21$
Cell dimensions	
a, b, c (Å)	29.008, 55.029, 135.534
Resolution (Å)	50–1.46 (1.49–1.46)
Completeness (%)	99.6 (98.3)
R_{merge}	8.2 (46.2)
I/σ(I)	21.26 (4.40)
Redundancy	6.9 (4.2)
Refinement	
Resolution (Å)	35–1.5
No. reflections (work/test)	35,671 /1,797
R_{work} / R_{free}	15.6 / 20.4
No. atoms	
Protein	1,867
Water	238
B-factors	
Protein/DNA	17.59
Solvent	30.45
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.068

[a] Numbers in parenthesis indicate the highest resolution shell

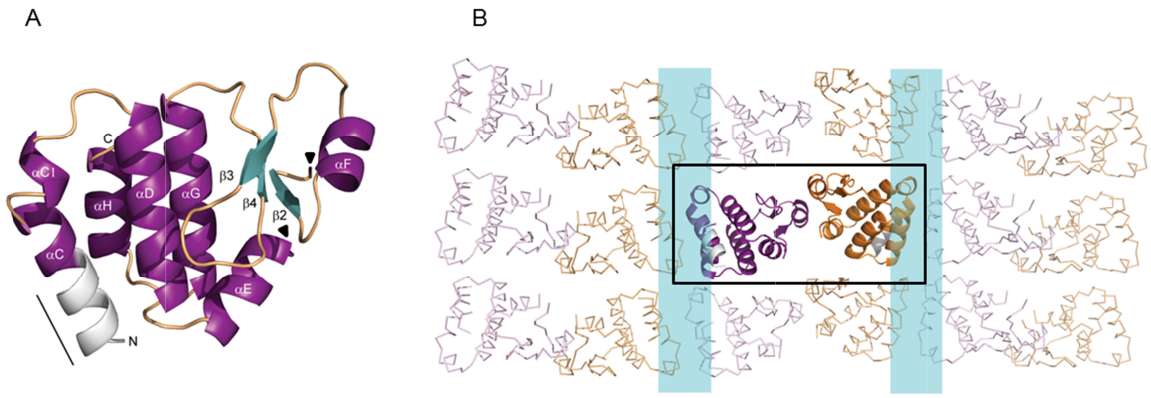

Figure 4.2. Structure of VcSeqA. ***A****. Ribbon diagrams of the Ribbon diagram of the VcSeqA monomer with helices in purple, strands in blue and loops in light orange. Black arrowheads show the disordered DNA binding loop (residues 42–45). The expression tag using the E. coli MKDKVRA sequences are shown on helix αC in white.* ***B****. The crystal packing of VcSeqA. There are two monomers in the asymmetric unit, shown as cartoons in purple and orange within the black box. Each monomer interacts with symmetry mates forming an aggregation interface. This surface is highlighted with a blue box. The symmetry mates are shown as ribbons in light purple and light orange.*

4.5.4 Protein–protein interactions

In the VcSeqA structure, the two monomers in the asymmetric unit interact with its symmetry mates through reciprocal polar and hydrophobic interactions between Arg2 and Leu6 from one monomer and Asp8 and Leu6 from a symmetry mate. This interface was first identified in the crystal structure of EcSeqA C-terminus bound to a single GATC site and was also observed in the structure of EcSeqA SeqA dimer bound to two GATC sites (Figure 4.3A, PDB ID: 1LRR, [97] and Figure 4.3C, PDB ID: 3FMT [141]). However, the

protein–protein interaction surface of these two EcSeqA contains an additional interaction pair that forms a bi-dentate salt bridge between an arginine and glutamic acid. Despite the absence of one of the interaction pairs, the protein–protein interaction surface in the VcSeqA structure still excludes a surface of ~320 $Å^2$ (Figure 4.3B). This is a smaller surface but still comparable to the excluded surface of ~430 $Å^2$ and ~400 $Å^2$ in the crystal structures of EcSeqA C-terminus bound to DNA and EcSeqA dimer bound to DNA with two GATC sequences, respectively. Hence, lacking one of the interaction pair that forms the bi-dentate salt bridge does not seem to affect the formation of the interface.

The presence of this interface in multiple crystal structures suggests that this surface could contribute to aggregation of SeqA–DNA complexes. Indeed, a study by Kang and co-workers have shown that in *E. coli*, mutations of residues Lys66 and Arg70 in SeqA protein results in mutants that maintain cooperative binding to DNA but are defective in DNA dependent aggregation *in vitro* as well as foci formation *in vivo* [125].

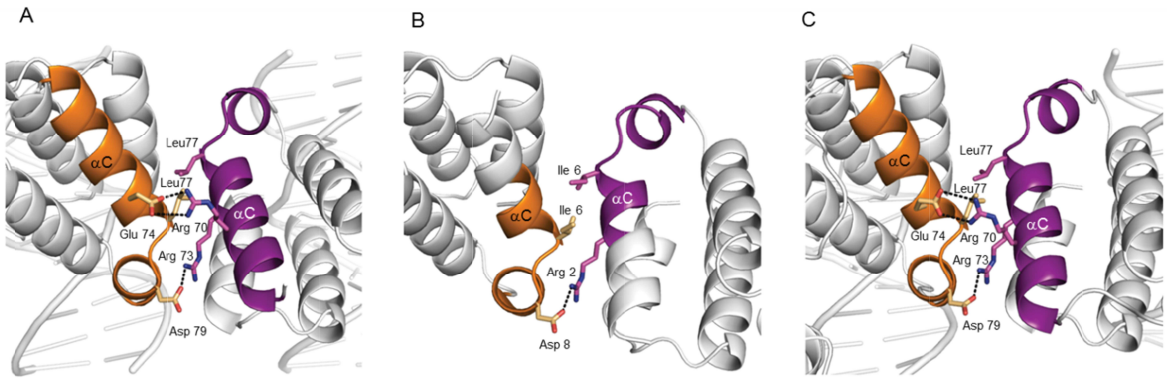

Figure 4.3. Additional aggregation interface of SeqA. ***A****. EcSeqA C-terminus bound to a single hemimethylated GATC site (PDB ID: 1LRR, [97]) shown as a cartoon in white with the interface of one monomer in purple and that of the other monomer in orange. Residues that form salt-bridges or hydrogen bonds are shown as sticks.* ***B****. VcSeqA in the absence of DNA (PDB ID:3SK7) still maintains the aggregation interface. The E. coli sequence expression tag on helix αC is shown in white and the remaining helix αC is in purple for one monomer and in orange for the other monomer of the symmetry mate.* ***C****. EcSeqA dimer bound to two hemimethylated GATC sites (PDB ID:3FMT, [141]) with the same colour scheme as in* ***A****.*

4.5.5 DNA binding characterization of VcSeqA surface mutants

Given the recurrence of the protein–protein interaction surface in multiple crystal structures of SeqA, we contemplated the possibility that this surface could compensate for the absence of the N-terminal oligomerization domain. Hence, we compared the DNA binding of EcSeqA with VcSeqA using a DNA aggregation assay previously described by Kang and co-workers [125]. For our crystallization study, we used a VcSeqA construct

with the N-terminal tag ^{64}MKDKVRA^{71} from the EcSeqA sequence. However, Arg70 from the *E. coli* tag sequence does not exist naturally in VcSeqA and was mutated to Serine generating VcSeqA* to function as wild-type. The ability of VcSeqA* to shift and form aggregates on DNA was compared to that of VcSeqA* R2S I6D, an aggregation interface mutant, *E. coli* SeqA C-terminus and VcSeqA* N78A N80A DNA binding mutant as a negative control.

All SeqA variants were homogeneous after purification and eluted from the size exclusion column as a monomer (Figure 4.4A–B). For characterization of the aggregation ability of these constructs, we carried out electrophoretic mobility shift assays using a 108 bp hemimethylated duplex containing six GATC sites (Figure 4.5A). As expected, the DNA binding mutant (VcSeqA* N78A N80A) did not show a shift even with higher concentrations of protein. However, VcSeqA*, VcSeqA* R2S I6D and EcSeqA C-terminus all bound and shifted DNA to a similar extent. Previously, sedimentation equilibrium analysis of EcSeqA C-terminus determined that it binds to a single hemimethylated GATC site with a K_d of approximately 7 μM [97]. Hence, it is likely that VcSeqA binds DNA with similar affinity as EcSeqA C-terminus. Unexpectedly, substantial increases in the concentration of VcSeqA* failed to induce aggregation as previously reported in *E. coli* by Kang and co-workers, and shifted DNA to a similar extent as the protein–protein interface mutant VcSeqA* R2S I6D. Given our results, VcSeqA* does not seem to be able to form aggregates on DNA as initially expected.

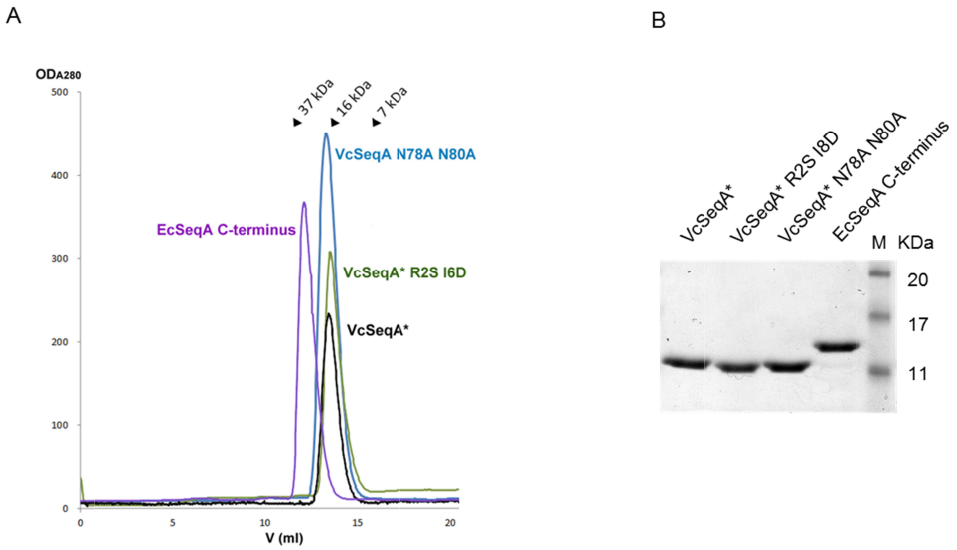

Figure 4.4. VcSeqA is a monomer in solution. ***A****. Elution profiles of VcSeqA variants over a Superdex-75 size exclusion chromatography column (GE Healthcare). VcSeqA* (13,169 Da) is in black, VcSeqA* R2S I6D (13,102 Da) in green, VcSeqA* N78A N80A (13,083 Da) in blue and EcSeqA C-terminus (14,897 Da) in purple.* ***B****. SDS-PAGE of the constructs in* ***A****. M represents the protein molecular weight marker.*

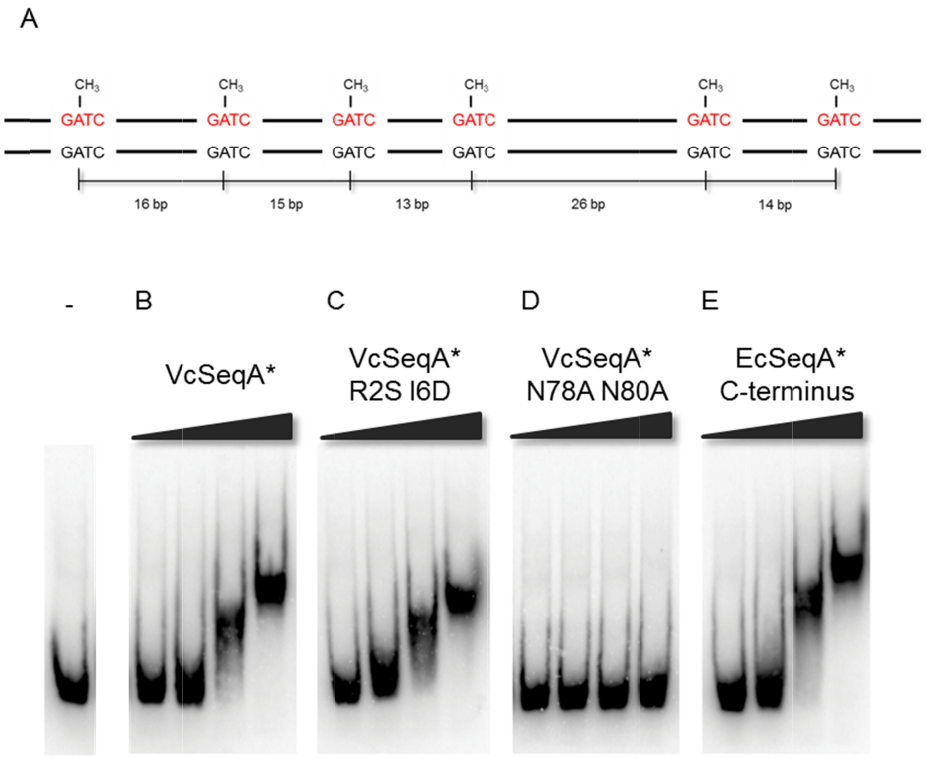

Figure 4.5. Electrophoretic mobility shift assays of SeqA with hemimethylated DNA. *A*. *A schematic illustrating the DNA substrate with six GATC sites. The methylated GATC sites are shown in red with a methyl group. The spacing between two GATC sites is also shown below the duplex. (**B**–**D**) 5 nM of YS1YS2 was incubated with 20, 80, 320, 1280 nM of SeqA variants indicated. The - lane indicates DNA alone.*

4.6 Discussion

DNA replication is a complex process that requires precise coordination of multiple proteins to ensure the passage of genetic information to progeny cells. The mechanism and regulation of DNA replication has been extensively studied in the model bacterium *E. coli* that contains a single chromosome, establishing the paradigm in prokaryotic

replication. However, the discovery of bacteria with multipartite genomes have led to the realization that a broader use of model prokaryotic organisms may be necessary for a more comprehensive understanding of DNA replication coordination with the cell cycle. The presence of multiple chromosomes has been demonstrated in a wide range of bacterial species such as *Agrobacterium tumefaciens, Brucella melitensis, Deinococcus radiodurans* and in *Vibrio cholerae.* The reason for the presence of multiple chromosomes in certain bacteria remains largely unknown, but many of these bacterial species with multiple chromosomes interact with hosts [145]. Despite the increasing numbers of sequenced multi-partite genomes, there is limited data on how multiple chromosomes are replicated and regulated.

Recent studies of replication initiation regulation in *V. cholerae* have shown that SeqA negatively regulates replication initiation through origin sequestration of both chromosomes similar to its functions in *E. coli*. However, replication of the two *V. cholerae* chromosomes are timed to achieve termination synchrony rather than initiation synchrony, where approximately two-thirds of chromosome I is replicated before replication initiation of the smaller chromosome II [140]. Furthermore, the two chromosomes segregate through chromosome specific partitioning (Par) proteins [146]. This suggests a requirement for a more stringent regulation in the coordination of chromosome replication and segregation and raises the possibility that SeqA may play additional roles in *V. cholerae* given its distinct roles in origin sequestration and chromosome organization in *E. coli*.

Studies of EcSeqA SeqA have shown that its function at the replication fork is dependent on the ability of SeqA to form a filament through the N-terminal oligomerization domain. Hence, the discovery of multiple *V. cholerae* strains that

contain SeqA with N-terminal truncations was unexpected and led us to examine any unique interactions that may compensate for the lack of an oligomerization domain. In our work, we have shown the structural and functional similarities between VcSeqA and EcSeqA C-terminus, suggesting the conservation of its function in different bacterial species within the γ-proteobacteria class. In addition, VcSeqA did not form any aggregates on DNA, similar to EcSeqA C-terminus. This suggests that the aggregation of SeqA on DNA may only be apparent once SeqA binds to DNA in a cooperative manner through the N-terminal domain.

During preparations for this manuscript, we amplified and sequenced the seqA gene from *V. cholerae* NCTC 8457 genomic DNA, which revealed a full length VcSeqA protein containing an N-terminal domain. This was unexpected, and is likely a result of incomplete annotations of the whole shotgun genome sequences. While our analysis of VcSeqA only pertains to the C-terminal domain, we have confirmed that the SeqA C-terminal domains from *V. cholerae* and *E. coli* in deed share structural and functional similarity, allowing us to delineate the parallel functions of SeqA in both bacteria. However, functions of full length VcSeqA remain to be validated to investigate any additional roles of SeqA in *V. cholerae*.

4.7 Coordinates

Atomic coordinates and structure factors of the *V. cholerae* SeqA have been deposited in the Protein Data Bank (accession code 3SK7).

4.8 Acknowledgements

We thank the PXRR staff at the NSLS (Brookhaven National Laboratory) and Dr. Murray Junop for assistance during data collection. We also thank Dr. Ali, Dr. Li and Dr. Brendler for their advice on DNA binding assays, and Lindsay Matthews for critical reading of the manuscript.

4.9 Funding

This work was supported by the Canadian Institutes of Health Research [MOP-67189 to A.G.]. YS. C. was supported in part by a McMaster Internal Prestige Scholarship.

Chapter 5

Conclusions and Final Remarks

This dissertation examines the functional role of SeqA in origin sequestration and chromosome organization. Despite the extensive data available on SeqA, prior to this work, the exact mechanism through which SeqA carries out its function was unknown. The work described in this thesis reveals how SeqA forms a high-affinity complex with hemimethylated DNA for its dual functions at the origin and at the replication forks.

To address our questions, we chose X-ray crystallography for investigating the intermolecular interactions of a functional SeqA dimer with DNA. Although X-ray crystallography can provide a high resolution structure of the protein of interest, the one most important limitation is the necessity of a crystal. However, not all crystals are suitable for data collection and phase determination. Therefore, one of the most challenging bottlenecks in the structure-determination pathway is obtaining ordered crystals of good diffraction quality. The addition of extra variables or substrates to the crystallization trials can exacerbate this already challenging process, requiring extra variables to consider for crystal optimization as discussed in chapter 2.

In general, there are three important factors to consider for crystallization of proteins in complex with DNA, which are homogeneity of both protein and DNA components, molecular flexibility of the protein and the design of the oligonucleotide substrate. All of these factors can affect the packing of the protein–DNA molecules within the crystal, affecting its resolution limit.

Homogeneity of the protein and its conformation can influence and determine the crystallization potential, as both impurities and conformational heterogeneity impedes crystal growth through disruption of orderly packing. Since many proteins are composed of many functional domains with flexible regions, it is common practice to remove as many flexible parts, retaining the smallest functional domains to facilitate crystal growth. Initial crystallization of SeqA is an example of this, where the N- and C-terminal domains were crystallized independently due to the flexibility of the linker region that connects the two domains. However, the structures of the minimal domains may not always represent the biologically relevant state, as in the case of *E. coli* SeqA, whose minimal functional unit is a dimer. Hence, it is imperative that the protein construct used in crystallization represents a functional unit that is biologically relevant.

SeqA is intrinsically flexible with a linker region (residues 36–63) which influences the conformational flexibility of the protein impeding crystallization. Characterization of various SeqA constructs with varying linker region deletions have shown that this region is not required for DNA binding, but adds plasticity to the DNA binding. Longer deletions of the linker region result in weaker binding for sites separated by more than two helical turns or abolishment in binding for sites separated by more than three helical turns (Figure A.1). Despite retaining its DNA binding ability, the deletion length affects the crystallization potential of the SeqA–DNA complex as crystals were more readily

obtained with SeqA constructs lacking residues 41–59. Based on the DNA binding of this protein, the spacing between the two hemimethylated GATC sequences were selected at 9 bp which showed maximal binding.

For protein–DNA cocrystallization, oligonucleotides can have an exceptionally large effect on the packing of the complex depending on the relative sizes of protein and DNA. This is due to the interactions between DNA molecules that generate crystal packing. DNA is able to pack through sequence specific groove-backbone interactions and can impose the overall geometry of packing [113,147]. These effects are amplified when the protein and DNA are similar in size, as the DNA will contribute a similar surface area for mediating crystal packing. Generally, lengths of 10 bp which corresponds to a full turn of a helix are used as a starting point for protein–DNA cocrystallization as they can stack in a head-to-tail manner to form a continuous B-DNA within the crystal lattice [148]. Furthermore, it is observed that systematic screening of DNA with variations in lengths drastically changes the crystallization conditions and crystal quality [149]. In our work, 20 bp was the most optimal DNA length that gave rise to high resolution diffraction quality crystals.

Another variable to consider is the ends of the DNA where unpaired overhanging bases can be added. Overhanging bases can mediate and stabilize complementary interactions of the overhanging bases that can form a pseudocontinuous helix in a head-to-tail stacking or form a triplex structure by base pairing between three bases [53,150]. Overhangs have even shown to form stabilizing interactions with the protein side chains, which in turn improves the packing of the protein–DNA complex, thereby increasing the resolution limit of the crystals [151,152]. In our structure of SeqA–DNA, the dinucleotide overhangs also mediate stabilizing interactions with the protein side chain rather than

forming a linear DNA through head-to-tail stacking. Only the first overhanging base partakes in this interaction, explaining why crystal of high resolution diffracting quality were not obtained when utilizing DNA with only a single overhang (Table 2.1 and Figure 2.3).

Our structure of the SeqA–DNA complex described in chapter 3 reveals a novel conformation of the minimal functional unit bound to two hemimethylated GATC sequences. Although the N- and C-terminal domains are identical to the structures of the two individual domains solved previously, the presence of a shortened linker region mediates the specific organization of the two C-terminals resulting in the loss of symmetry. This allows the specific recognition of the tandem hemimethylated GATC sequences by SeqA. The loss of symmetry is mediated by the 180° rotation of the C-terminal domain of one protomer through the localized unwinding of its linker region while the linker region of the other protomer exhibits a helical organization. It is conceivable that full-length SeqA dimer would assume a similar helical organization of the linker region, which can unwind in a stepwise manner to mediate binding of GATC sequences that are separated by more than one helical turn (Figure A.1).

Our work has also identified a key protein interaction surface that separates its roles at the origin and at the replication forks. SeqA dimer is the functional unit and exhibits asynchronous replication which can be restored with an increase in the local concentration of the protein. Hence, SeqA dimer is sufficient for origin sequestration and does not require filament formation. However, at the replication forks, the oligomerization of SeqA is essential for its function. In *E. coli*, the GATC sequences are more sparse outside of the *oriC* requiring the oligomerization of SeqA for its function at the replication forks. SeqA oligomerization forms a left-handed filament along the DNA

compensating for the increased spacing between the GATC sequences. The protein–protein interaction surface stabilizes and extends the half-life of the SeqA–DNA filament at the replication forks.

Another research contribution of my thesis was in validating the conservation of SeqA functions in other bacterial species, specifically in *V. cholerae* described in chapter 4. The presence of two chromosomes makes *V. cholerae* an interesting system to study the regulation of replication initiation and segregation. Furthermore, the presence of multiple *V. cholerae* strains with SeqA homologues that have truncations in the N-terminal domain was surprising as well as intriguing as the N-terminal oligomerization domain is essential for SeqA function in *E. coli*. Structural investigation of a truncated SeqA that only contains the C-terminal domain reveals that it is identical to the C-terminal domain of *E. coli* SeqA with a r.m.s deviation of 0.705 Å. The SeqA C-terminus can be superimposed onto the SeqA dimer C-terminus with a r.m.s deviation of 0.301 Å with 100% sequence identity. Considering that *V. cholerae* SeqA has a sequence identity of 57% to *E. coli* SeqA, the two structures are superimposed with a low r.m.s deviation. *V. cholerae* SeqA is functionally similar to *E. coli* SeqA as it can bind DNA with similar affinity, but does not exhibit aggregation in the absence of the N-terminal domain. This suggests that aggregation can only be observed in the presence of the oligomerization domain. *V. cholerae* SeqA also contains the protein–protein interaction surface that mediates the formation of higher ordered structures, although the significance of this cannot be confirmed in the absence of DNA.

The lack of any difference in the binding of *V. cholerae* SeqA to DNA in the absence of the N-terminal domain was intriguing as we anticipated an alternate mode of protein–protein interaction. This led us to examine the genomic DNA of multiple *V.*

cholerae strains that contain truncated SeqA. Most of the *V. cholerae* genome has been sequenced through whole shotgun sequencing and many protein entry are through conceptual translation without direct sequencing. It was interesting to find that in all the strains that possessed truncated SeqA, the SeqA gene was placed at the end of a 'contig', a set of overlapping DNA segments derived from the genome. Genomic amplification of the *V. cholerae seqA* gene revealed that SeqA from *V. cholerae* NCTC 8457 does contain the N-terminal domain.

Despite the added advantage of having access to a plethora of genomic sequences with increasing genomic sequencing projects, the user should always exercise caution as the rate of sequencing is much faster than the annotation process and the sequence in the National Centre for Biotechnology Information (NCBI) may not always accurately represent the correct genomic sequence.

5.1 Future Research Directions

The structure of SeqA dimer in complex with DNA reveals the interaction of the functional dimer with its cognate binding sequence and makes a critical connection in the structural organization of the protein to its functional ability to recognize GATC sequences that are separated by 1–3 helical turns. Based on the work, the linker region was suggested to adopt a helical organization, which can be unwound in a step-wise manner to mediate SeqA binding of GATC sequences that are further apart. However, our structure only represents one snapshot of SeqA binding to DNA. To provide an additional snapshot of the SeqA–DNA complex, another SeqA dimeric variant with a longer linker region should be utilized for structural studies. Since a single turn of a helix corresponds to 3.6 amino acid residues, a dimeric variant with eight extra residues may be ideal to generate two extra helical turns in the helical organization of the linker region. Furthermore, the effect of the linker region on the topology of the substrate DNA should be closely examined to explore the roles of SeqA in chromosome organization.

Beyond *E. coli*, *V. cholerae* SeqA structure is the first reported within the γ-proteobacteria class. Increasing numbers of studies indicate a similar role of SeqA in origin sequestration and chromosome organization in other bacterial species outside of *E. coli*. In addition, several studies have also suggested the role of SeqA in pathogenicity. Two independent studies have shown that *seqA*$^-$ strains are attenuated for virulence in *Salmonella enterica* serovar *Typhimurium* by the oral route and show decreased invasion of the host [106,107]. Lack of SeqA in *S. enterica* results in similar phenotypic defects as in *E. coli*, however, SeqA seems to regulate a small fraction of genes that are *Salmonella* specific [106]. However, the existence of SeqA regulated genes might reflect

physiological changes occurring in response to a *seqA* mutation rather than a direct role of SeqA in transcriptional control. SeqA is not essential in *E. coli* and there have been conflicting results on the dispensability of SeqA in *V. cholerae.* In *E. coli*, cells that lack SeqA or have excess SeqA are viable, but display replication asynchrony and segregation defects. This suggests that SeqA may decrease the overall physical fitness of the cell, thereby affecting its virulence. In fact, a *dam*$^-$ mutant exhibits similar attenuation of virulence in *S. enterica*, suggesting that virulence is regulated through the methylation state of DNA [153,154]. The role of SeqA in other pathogenic bacteria such as *V. cholerae*, *Haemophilus influenzae* and *Yersinia pestis* remain unchartered and future work may broaden the scope of research by stepping out from the model organism *E. coli*.

Another avenue of research as an extension of the research described in this thesis would be to understand the potential roles of SeqA in the transposition pathway that specifically recognizes DNA replication. Transposons are a class of mobile genetic elements that are highly prevalent and move between non-homologous regions of DNA. Hence, transposons can have a profound effect on the genomic diversity. Tn7 is a bacterial transposon with unusual target specificity, resulting in transposition events with an orientation bias that is consistent with the direction of DNA replication. This target specific transposition is mediated through a transposon-encoded DNA-binding protein, TnsE, which identifies targets sites by recognizing a structure or complex associated with DNA replication [155].

Recent genetic studies on the transposition events of Tn7 in *dam*$^-$ and *seqA*$^-$ strains have shown that transposition is stimulated by about three-fold in a *dam*$^-$ strain and about ten-fold in a *seqA*$^-$ strain (personal communication with Dr. Joseph Peters).

Stimulation of transposition does not seem to be a result of defective regulation of replication initiation as stimulation of transposition is not observed in *dnaA* or *hda* knockout strains. Interestingly, a similar level of transposition stimulation is seen in a *recA*$^-$ strain. RecA is a protein that plays a central role in homologous recombination for DNA repair and maintenance. Hence, the combination of these results indicate that SeqA could be directly inhibiting transposition events into specific chromosomal regions, or be involved in another pathway that allows specific targeting by Tn7. Dr. Peters laboratory has recently shown the direct interaction between the DNA-binding protein, TnsE and SeqA as well as SeqA specific protection of TnsE from proteolytic degradation by trypsin. Hence, further work should be carried out to probe the potential roles of SeqA in the regulation of Tn7 transposition events. Although the additional functions of SeqA remain to be validated, it reinforces the idea that SeqA is truly a master regulator.

Appendix

Appendix 1

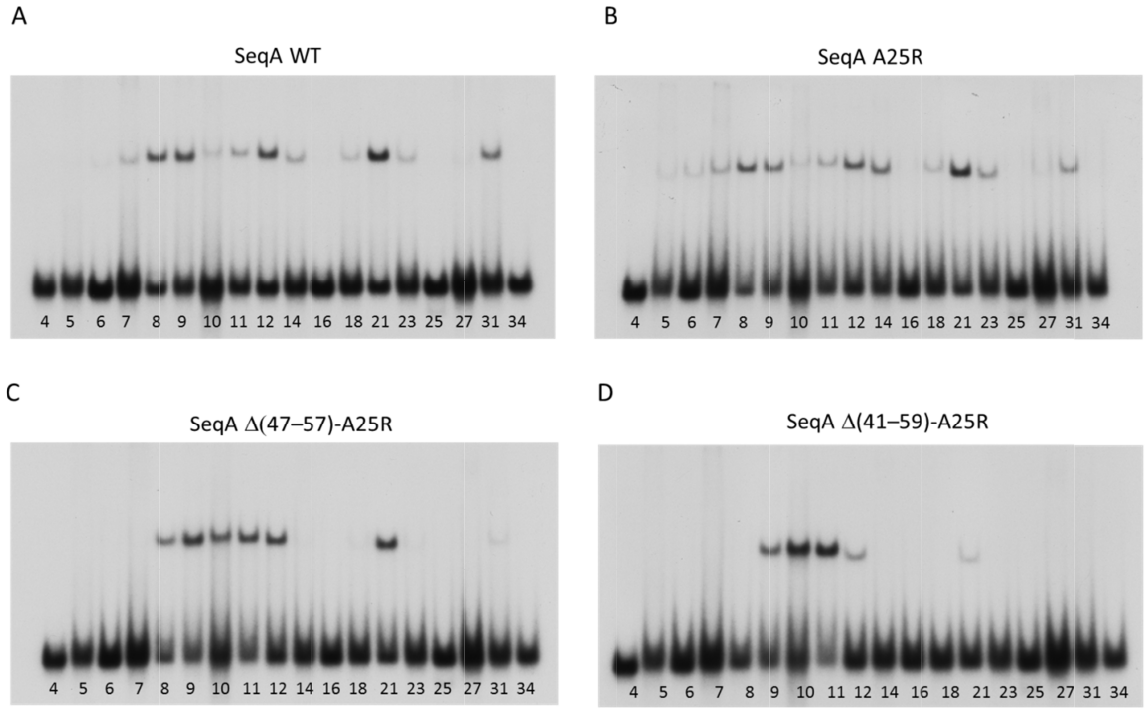

Figure A.1. EMSA of SeqA wild-type, A25R, Δ(47–57)-A25R and Δ(41–59)-A25R. *The numbers below the bands represent the spacing between two hemimethylated GATC sequences. (**A**–**D**) EMSA with SeqA variants indicated. SeqA WT and A25R bind DNA in a similar manner with maximal binding when two hemimethylated GATC sequences are separated by 1, 2 or 3 helical turns. Deletion of residues 47–57 results in weaker binding for two GATC sequences separated by more than two helical turns. Deletion of residues 41–59 results in weaker binding for two GATC sequences separated by two helical turns, and binding is abolished with three helical turn separation.*

Appendix 2

Table A.1. List of all SeqA constructs

Plasmid No.	SeqA construct	Putative defects
8013	WT	None
8014	I21R	Filamentation
8015	A25R	Filamentation
8016	T18E	Filamentation
8017	E9A	Dimerization
8018	E9R	Dimerization
8019	N150A/N152A	DNA binding
8020	Δ(47–57)	None
8021	Δ(45–59)	None
8022	Δ(43–59)	None
8023	Δ(41–59)	None
8025	Δ(47–57)-I21R	Filamentation
8026	Δ(47–57)-A25R	Filamentation
8032	Δ(43–59)-A25R	Filamentation
8033	Δ(41–59)-A25R	Filamentation
8041	Δ(39–59)	None
8042	Δ(37–61)	None
8043	R31A	Dimerization
8064	Δ(37–61)-I21R	Filamentation
8065	Δ(39–59)-I21R	Filamentation
8072	Δ(39–59)-A25R	Filamentation
8195	R31S	Dimerization
8197	Δ(45–59)-T18E	Filamentation
8204	R31S	Dimerization
8205	Δ(41–59)-T18E	Filamentation
8210	Δ(43–59)-T18E	Filamentation
8211	Δ(41–59)-T18E	Filamentation
8214	Δ(43–59)-R31S	Dimerization
8215	Δ(41–59)-R31S	Dimerization
8268	A25R R70S R73S	Filamentation, CTD–CTD[a]
8269	Δ(41–59)-A25R R70S R73S	Filamentation, CTD–CTD
8270	WT R70S R73S	CTD–CTD

Table continued on pg. 107

Table continued from pg. 106

Plasmid No.	SeqA construct	Putative defects
8305	$SeqA_{Vibrio}$	None
8306	$SeqA_{Vibrio}$	None
8330	$SeqA_{Vibrio}$	None
8342	TEV-$SeqA_{Vibrio}$	None
8352	M-$SeqA_{Vibrio}$ [b]	None
8353	HIS-$SeqA_{Vibrio}$	None
8376	Δ(45–59)-A25R	Filamentation
8377	Δ(37–61)-A25R	Filamentation
8445	SeqA with a poly-Glycine linker region	None
8484	SeqA A25R R73S	Filamentation, CTD–CTD
8486	M*-$SeqA_{Vibrio}$ R2S I6D [c]	CTD–CTD
8487	SeqA A25R R70S R73S I77D	Filamentation, CTD–CTD
8488	SeqA R70S R73S I77D	Filamentation, CTD–CTD
8526	M*-$SeqA_{Vibrio}$	None
8527	M*-$SeqA_{Vibrio}$ N78A N80A	DNA binding
8528	M*-$SeqA_{Vibrio}$ R2S I6D N78A N80A	DNA binding, CTD–CTD
8536	M*-$SeqA_{Vibrio}$ with *E. coli* DNA loop (110–117)	None
8537	M*-$SeqA_{Vibrio}$ R73S I77D with *E. coli* DNA loop (110–117)	None

[a] CTD–CTD refers to protein–protein interaction through the C-terminal domains

[b] M refers to the ^{64}MKDKVRA^{71} sequence from *E. coli*

[c] M* refers to the ^{64}MKDKVRA^{71} sequence from *E. coli* with a mutation in R70S

Appendix 3

Table A.2. Sequence of all oligonucleotides used for SeqA–DNA cocrystallization

	DNA Sequence[a]	Length/spacing (overhang)
D4-1 D4-2	5′TAAGGATCCAAGCACCGATCGACT3′ 3′ATTCCTAGGTTCGTGGCTAGCTGA5′	24/12 (0)
D4-3 D4-4	5′GGTAAGGATCCAAGCACCGATCGACTGG3′ 3′CCATTCCTAGGTTCGTGGCTAGCTGACC5′	28/12 (0)
D4-5 D4-6	5′GAGTCGATCGGTGCGATCCTTAG3′ 3′CTCAGCTAGCCACGCTAGGAATC5′	23/9 (0)
D4-7 D4-8	5′GAGTCGATCGGTGCTGATCCTTAG3′ 3′CTCAGCTAGCCACGACTAGGAATC5′	24/10 (0)
D4-9 D4-10	5′AGTCGATCGGTGCGATCCTTA3′ 3′TCAGCTAGCCACGCTAGGAAT5′	21/9 (0)
D4-11 D4-12	5′GAGTCGATCGGTGCTGATCCTTA3′ 3′CTCAGCTAGCCACGACTAGGAAT5′	23/10 (0)
D4-13 D4-14	5′GAGTCGATCGGTGCGATCCTTA3′ 3′TCAGCTAGCCACGCTAGGAATC5′	21/9 (1)
D4-15 D4-16	5′GAGTCGATCGGCGGGATCCTTA3′ 3′TCAGCTAGCCGCCCTAGGAATC5′	21/9 (1)
D4-17 D4-18	5′GAGTCGATCGGCGGGATCCTTA3′ 3′CAGCTAGCCGCCCTAGGAATCT5′	20/9 (2)
D4-19 D4-20	5′GAGTGATCCGGCGGGATCCTT3′ 3′CACTAGGCCGCCCTAGGAACT5′	21/10 (2)
D4-21 D4-22	5′GAGTCGATCGGCGGGATCCTT3′ 3′CAGCTAGCCGCCCTAGGAACT5′	21/9 (2)
D4-23 D4-24	5′GTCGATCGGCGGGATCCTTA3′ 3′CAGCTAGCCGCCCTAGGAAT5′	20/9 (0)
D4-17 D4-24	5′GAGTCGATCGGCGGGATCCTTA3′ 3′CAGCTAGCCGCCCTAGGAAT5′	20/9 (2 on one end)
D4-25 D4-26	5′GAGTCGATCGGCGGTGATCCTTA3′ 3′CAGCTAGCCGCCACTAGGAAT5′	21/10 (2 on one end)

[a]Methylated GATC sequences are shown in orange. CG snap-fasteners are shaded in grey.

References

1. Dahm R (2005) Friedrich Miescher and the discovery of DNA. Developmental biology 278: 274-288.
2. Griffith F (1928) The Significance of Pneumococcal Types. The Journal of hygiene 27: 113-159.
3. Avery OT, Macleod CM, McCarty M (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. The Journal of experimental medicine 79: 137-158.
4. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171: 737-738.
5. Meselson M, Stahl FW (1958) The Replication of DNA in Escherichia Coli. Proceedings of the National Academy of Sciences of the United States of America 44: 671-682.
6. Cairns J (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. Journal of molecular biology 6: 208-213.
7. Prescott DM, Kuempel PL (1972) Bidirectional replication of the chromosome in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 69: 2842-2845.
8. Fangman WL, Brewer BJ (1992) A question of time: replication origins of eukaryotic chromosomes. Cell 71: 363-366.
9. Edenberg HJ, Huberman JA (1975) Eukaryotic chromosome replication. Annual review of genetics 9: 245-284.
10. Stinchcomb DT, Struhl K, Davis RW (1979) Isolation and characterisation of a yeast chromosomal replicator. Nature 282: 39-43.
11. Marahrens Y, Stillman B (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science 255: 817-823.
12. Bogan JA, Natale DA, Depamphilis ML (2000) Initiation of eukaryotic DNA replication: conservative or liberal? Journal of cellular physiology 184: 139-150.
13. Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 357: 128-134.
14. Liang C, Weinreich M, Stillman B (1995) ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. Cell 81: 667-676.
15. Whittaker AJ, Royzman I, Orr-Weaver TL (2000) Drosophila double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. Genes & development 14: 1765-1776.
16. Tanaka T, Knapp D, Nasmyth K (1997) Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. Cell 90: 649-660.

17. Weinreich M, Liang C, Stillman B (1999) The Cdc6p nucleotide-binding motif is required for loading mcm proteins onto chromatin. Proceedings of the National Academy of Sciences of the United States of America 96: 441-446.
18. Stillman B (1996) Cell cycle control of DNA replication. Science 274: 1659-1664.
19. Spiesser TW, Klipp E, Barberis M (2009) A model for the spatiotemporal organization of DNA replication in Saccharomyces cerevisiae. Molecular genetics and genomics : MGG 282: 25-35.
20. Goldar A, Marsolier-Kergoat MC, Hyrien O (2009) Universal temporal profile of replication origin activation in eukaryotes. PloS one 4: e5899.
21. Diller JD, Raghuraman MK (1994) Eukaryotic replication origins: control in space and time. Trends in biochemical sciences 19: 320-325.
22. Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. FEMS Microbiol Rev 26: 355-374.
23. McGarry KC, Ryan VT, Grimwade JE, Leonard AC (2004) Two discriminatory binding sites in the Escherichia coli replication origin are required for DNA strand opening by initiator DnaA-ATP. Proceedings of the National Academy of Sciences of the United States of America 101: 2811-2816.
24. Erzberger JP, Pirruccello MM, Berger JM (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. Embo J 21: 4763-4773.
25. Marszalek J, Kaguni JM (1994) DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. The Journal of biological chemistry 269: 4883-4890.
26. Kawakami H, Katayama T (2010) DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. Biochemistry and cell biology = Biochimie et biologie cellulaire 88: 49-62.
27. Mott ML, Erzberger JP, Coons MM, Berger JM (2008) Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. Cell 135: 623-634.
28. Messer W, Blaesing F, Jakimowicz D, Krause M, Majka J, et al. (2001) Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. Biochimie 83: 5-12.
29. Weigel C, Seitz H (2002) Strand-specific loading of DnaB helicase by DnaA to a substrate mimicking unwound oriC. Molecular microbiology 46: 1149-1156.
30. Fang L, Davey MJ, O'Donnell M (1999) Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. Mol Cell 4: 541-553.
31. Johnson A, O'Donnell M (2005) Cellular DNA replicases: components and dynamics at the replication fork. Annual review of biochemistry 74: 283-315.

32. Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 9: 27-43.
33. Sekimizu K, Bramhill D, Kornberg A (1987) ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. Cell 50: 259-265.
34. Messer W, Blaesing F, Majka J, Nardmann J, Schaper S, et al. (1999) Functional domains of DnaA proteins. Biochimie 81: 819-825.
35. Ishida T, Akimitsu N, Kashioka T, Hatano M, Kubota T, et al. (2004) DiaA, a novel DnaA-binding protein, ensures the timely initiation of E. coli chromosome replication. J Biol Chem.
36. Abe Y, Jo T, Matsuda Y, Matsunaga C, Katayama T, et al. (2007) Structure and function of DnaA N-terminal domains: specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on oriC. The Journal of biological chemistry 282: 17816-17827.
37. Nozaki S, Ogawa T (2008) Determination of the minimum domain II size of Escherichia coli DnaA protein essential for cell viability. Microbiology 154: 3379-3384.
38. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO journal 1: 945-951.
39. Erzberger JP, Mott ML, Berger JM (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. Nature structural & molecular biology 13: 676-683.
40. Roth A, Messer W (1998) High-affinity binding sites for the initiator protein DnaA on the chromosome of Escherichia coli. Molecular microbiology 28: 395-401.
41. Fuller RS, Funnell BE, Kornberg A (1984) The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. Cell 38: 889-900.
42. Bramhill D, Kornberg A (1988) Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome. Cell 52: 743-755.
43. Miller DT, Grimwade JE, Betteridge T, Rozgaja T, Torgue JJ, et al. (2009) Bacterial origin recognition complexes direct assembly of higher-order DnaA oligomeric structures. Proceedings of the National Academy of Sciences of the United States of America 106: 18479-18484.
44. Cunningham EL, Berger JM (2005) Unraveling the early steps of prokaryotic replication. Current opinion in structural biology 15: 68-76.
45. Grimwade JE, Torgue JJ, McGarry KC, Rozgaja T, Enloe ST, et al. (2007) Mutational analysis reveals Escherichia coli oriC interacts with both DnaA-ATP and DnaA-ADP during pre-RC assembly. Molecular microbiology 66: 428-439.
46. Leonard AC, Grimwade JE (2011) Regulation of DnaA Assembly and Activity: Taking Directions from the Genome. Annual review of microbiology.

47. Speck C, Messer W (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. Embo J 20: 1469-1476.
48. Kawakami H, Keyamura K, Katayama T (2005) Formation of an ATP-DnaA-specific initiation complex requires DnaA Arginine 285, a conserved motif in the AAA+ protein family. The Journal of biological chemistry 280: 27420-27430.
49. Hansen FG, Christensen BB, Atlung T (2007) Sequence characteristics required for cooperative binding and efficient in vivo titration of the replication initiator protein DnaA in E. coli. Journal of molecular biology 367: 942-952.
50. Kowalski D, Eddy MJ (1989) The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. The EMBO journal 8: 4335-4344.
51. Bonnefoy E, Rouviere-Yaniv J (1991) HU and IHF, two homologous histone-like proteins of Escherichia coli, form different protein-DNA complexes with short DNA fragments. The EMBO journal 10: 687-696.
52. Roth A, Urmoneit B, Messer W (1994) Functions of histone-like proteins in the initiation of DNA replication at oriC of Escherichia coli. Biochimie 76: 917-923.
53. Rice PA, Yang S, Mizuuchi K, Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell 87: 1295-1306.
54. Gille H, Egan JB, Roth A, Messer W (1991) The FIS protein binds and bends the origin of chromosomal DNA replication, oriC, of Escherichia coli. Nucleic acids research 19: 4167-4172.
55. Filutowicz M, Ross W, Wild J, Gourse RL (1992) Involvement of Fis protein in replication of the Escherichia coli chromosome. Journal of bacteriology 174: 398-407.
56. Zyskind JW, Smith DW (1986) The bacterial origin of replication, oriC. Cell 46: 489-490.
57. Sanchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, et al. (2010) Dynamic distribution of seqa protein across the chromosome of escherichia coli K-12. mBio 1.
58. Campbell JL, Kleckner N (1990) E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. Cell 62: 967-979.
59. Slater S, Wold S, Lu M, Boye E, Skarstad K, et al. (1995) 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.
60. Nievera C, Torgue JJ, Grimwade JE, Leonard AC (2006) SeqA blocking of DnaA-oriC interactions ensures staged assembly of the E. coli pre-RC. Mol Cell 24: 581-592.
61. Ryan VT, Grimwade JE, Camara JE, Crooke E, Leonard AC (2004) Escherichia coli prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. Mol Microbiol 51: 1347-1359.

62. Keyamura K, Fujikawa N, Ishida T, Ozaki S, Su'etsugu M, et al. (2007) The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. Genes & development 21: 2083-2099.
63. Katayama T, Kubota T, Kurokawa K, Crooke E, Sekimizu K (1998) The initiator function of DnaA protein is negatively regulated by the sliding clamp of the E. coli chromosomal replicase. Cell 94: 61-71.
64. Camara JE, Breier AM, Brendler T, Austin S, Cozzarelli NR, et al. (2005) Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of Escherichia coli DNA replication. EMBO Rep 6: 736-741.
65. Kato J, Katayama T (2001) Hda, a novel DnaA-related protein, regulates the replication cycle in Escherichia coli. The EMBO journal 20: 4253-4262.
66. Kurz M, Dalrymple B, Wijffels G, Kongsuwan K (2004) Interaction of the sliding clamp beta-subunit and Hda, a DnaA-related protein. Journal of bacteriology 186: 3508-3515.
67. Davey MJ, Jeruzalmi D, Kuriyan J, O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. Nature reviews Molecular cell biology 3: 826-835.
68. Ogura T, Whiteheart SW, Wilkinson AJ (2004) Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. Journal of structural biology 146: 106-112.
69. Su'etsugu M, Takata M, Kubota T, Matsuda Y, 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 to cells : devoted to molecular & cellular mechanisms 9: 509-522.
70. Su'etsugu M, Nakamura K, Keyamura K, Kudo Y, Katayama T (2008) Hda monomerization by ADP binding promotes replicase clamp-mediated DnaA-ATP hydrolysis. The Journal of biological chemistry 283: 36118-36131.
71. Kitagawa R, Mitsuki H, Okazaki T, Ogawa T (1996) A novel DnaA protein-binding site at 94.7 min on the Escherichia coli chromosome. Mol Microbiol 19: 1137-1147.
72. Kitagawa R, Ozaki T, Moriya S, Ogawa T (1998) Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for Escherichia coli DnaA protein. Genes Dev 12: 3032-3043.
73. Nozaki S, Yamada Y, Ogawa T (2009) Initiator titration complex formed at datA with the aid of IHF regulates replication timing in Escherichia coli. Genes to cells : devoted to molecular & cellular mechanisms 14: 329-341.
74. Atlung T, Clausen ES, Hansen FG (1985) Autoregulation of the dnaA gene of Escherichia coli K12. Molecular & general genetics : MGG 200: 442-450.
75. Braun RE, O'Day K, Wright A (1985) Autoregulation of the DNA replication gene dnaA in E. coli K-12. Cell 40: 159-169.
76. Speck C, Weigel C, Messer W (1999) ATP- and ADP-dnaA protein, a molecular switch in gene regulation. Embo J 18: 6169-6176.

77. Lee YS, Hwang DS (1997) Occlusion of RNA polymerase by oligomerization of DnaA protein over the dnaA promoter of Escherichia coli. The Journal of biological chemistry 272: 83-88.
78. Fujimitsu K, Katayama T (2004) Reactivation of DnaA by DNA sequence-specific nucleotide exchange in vitro. Biochemical and biophysical research communications 322: 411-419.
79. Fujimitsu K, Senriuchi T, Katayama T (2009) Specific genomic sequences of E. coli promote replicational initiation by directly reactivating ADP-DnaA. Genes & development 23: 1221-1233.
80. Skarstad K, Lueder G, Lurz R, Speck C, Messer W (2000) The Escherichia coli SeqA protein binds specifically and co-operatively to two sites in hemimethylated and fully methylated oriC. Mol Microbiol 36: 1319-1326.
81. Riber L, Lobner-Olesen A (2005) Coordinated replication and sequestration of oriC and dnaA are required for maintaining controlled once-per-cell-cycle initiation in Escherichia coli. Journal of bacteriology 187: 5605-5613.
82. Russell DW, Zinder ND (1987) Hemimethylation prevents DNA replication in E. coli. Cell 50: 1071-1079.
83. Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in E. coli. Cell 77: 413-426.
84. Brendler T, Abeles A, Austin S (1995) A protein that binds to the P1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. Embo J 14: 4083-4089.
85. Boye E, Stokke T, Kleckner N, Skarstad K (1996) Coordinating DNA replication initiation with cell growth: differential roles for DnaA and SeqA proteins. Proc Natl Acad Sci U S A 93: 12206-12211.
86. von Freiesleben U, Krekling MA, Hansen FG, Lobner-Olesen A (2000) The eclipse period of Escherichia coli. Embo J 19: 6240-6248.
87. Weitao T, Nordstrom K, Dasgupta S (2000) Escherichia coli cell cycle control genes affect chromosome superhelicity. EMBO Rep 1: 494-499.
88. Bach T, Krekling MA, Skarstad K (2003) Excess SeqA prolongs sequestration of oriC and delays nucleoid segregation and cell division. Embo J 22: 315-323.
89. Hiraga S, Ichinose C, Niki H, Yamazoe M (1998) Cell cycle-dependent duplication and bidirectional migration of SeqA-associated DNA-protein complexes in E. coli. Mol Cell 1: 381-387.
90. Onogi T, Niki H, Yamazoe M, Hiraga S (1999) The assembly and migration of SeqA-Gfp fusion in living cells of Escherichia coli. Mol Microbiol 31: 1775-1782.
91. Hiraga S, Ichinose C, Onogi T, Niki H, Yamazoe M (2000) Bidirectional migration of SeqA-bound hemimethylated DNA clusters and pairing of oriC copies in Escherichia coli. Genes Cells 5: 327-341.
92. Fossum S, Crooke E, Skarstad K (2007) Organization of sister origins and replisomes during multifork DNA replication in Escherichia coli. Embo J 26: 4514-4522.

93. Kang S, Han JS, Park JH, Skarstad K, Hwang DS (2003) SeqA protein stimulates the relaxing and decatenating activities of topoisomerase IV. J Biol Chem.
94. Drlica K, Zhao X (1997) DNA gyrase, topoisomerase IV, and the 4-quinolones. Microbiology and molecular biology reviews : MMBR 61: 377-392.
95. Espeli O, Lee C, Marians KJ (2003) A physical and functional interaction between Escherichia coli FtsK and topoisomerase IV. The Journal of biological chemistry 278: 44639-44644.
96. Madabhushi R, Marians KJ (2009) Actin homolog MreB affects chromosome segregation by regulating topoisomerase IV in Escherichia coli. Molecular cell 33: 171-180.
97. Guarné A, Zhao Q, Ghirlando R, Yang W (2002) Insights into negative modulation of E. coli replication initiation from the structure of SeqA-hemimethylated DNA complex. Nat Struct Biol 9: 839-843.
98. Guarné A, Brendler T, Zhao Q, Ghirlando R, Austin S, et al. (2005) Crystal structure of a SeqA-N filament: implications for DNA replication and chromosome organization. Embo J 24: 1502-1511.
99. Fujikawa N, Kurumizaka H, Yamazoe M, Hiraga S, Yokoyama S (2003) Identification of functional domains of the Escherichia coli SeqA protein. Biochem Biophys Res Commun 300: 699-705.
100. Brendler T, Austin S (1999) Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. Embo J 18: 2304-2310.
101. Brendler T, Sawitzke J, Sergueev K, Austin S (2000) A case for sliding SeqA tracts at anchored replication forks during Escherichia coli chromosome replication and segregation. Embo J 19: 6249-6258.
102. Odsbu I, Klungsoyr HK, Fossum S, Skarstad K (2005) Specific N-terminal Interactions of the E. coli SeqA protein are required to form multimers that restrain negative supercoils and form foci. submitted to J Biol Chem.
103. Han JS, Kang S, Lee H, Kim HK, Hwang DS (2003) Sequential binding of SeqA to paired hemi-methylated GATC sequences mediates formation of higher-order complexes. J Biol Chem 278: 34983-34989.
104. Han JS, Kang S, Kim SH, Ko MJ, Hwang DS (2004) Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. J Biol Chem 279: 30236-30243.
105. DeLano WL (2002) The PyMOL Molecular Graphic Systems. DeLano Scientific, Palo Alto, CA, USA.
106. Prieto AI, Jakomin M, Segura I, Pucciarelli MG, Ramos-Morales F, et al. (2007) The GATC-binding protein SeqA is required for bile resistance and virulence in Salmonella enterica serovar typhimurium. J Bacteriol 189: 8496-8502.
107. Chatti A, Daghfous D, Landoulsi A (2007) Effect of seqA mutation on Salmonella typhimurium virulence. J Infect 54: e241-245.

108. Barras F, Marinus MG (1989) The great GATC: DNA methylation in E. coli. Trends Genet 5: 139-143.
109. Modrich P, Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annual review of biochemistry 65: 101-133.
110. Hiraga S (2000) Dynamic localization of bacterial and plasmid chromosomes. Annu Rev Genet 34: 21-59.
111. Niki H, Hiraga S (1998) Polar localization of the replication origin and terminus in Escherichia coli nucleoids during chromosome partitioning. Genes Dev 12: 1036-1045.
112. Bach T, Skarstad K (2004) Re-replication from non-sequesterable origins generates three-nucleoid cells which divide asymmetrically. Mol Microbiol 51: 1589-1600.
113. Timsit Y, Moras D (1994) DNA self-fitting: the double helix directs the geometry of its supramolecular assembly. The EMBO journal 13: 2737-2746.
114. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276: 307-326.
115. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. Journal of applied crystallography 40: 658-674.
116. Camara JE, Skarstad K, Crooke E (2003) Controlled initiation of chromosomal replication in Escherichia coli requires functional Hda protein. J Bacteriol 185: 3244-3248.
117. Leonard AC, Grimwade JE (2005) Building a bacterial orisome: emergence of new regulatory features for replication origin unwinding. Molecular microbiology 55: 978-985.
118. Sutera VA, Jr., Lovett ST (2006) The role of replication initiation control in promoting survival of replication fork damage. Mol Microbiol 60: 229-239.
119. Yamazoe M, Adachi S, Kanaya S, Ohsumi K, Hiraga S (2005) Sequential binding of SeqA protein to nascent DNA segments at replication forks in synchronized cultures of Escherichia coli. Mol Microbiol 55: 289-298.
120. Bahloul A, Meury J, Kern R, Garwood J, Guha S, et al. (1996) Co-ordination between membrane oriC sequestration factors and a chromosome partitioning protein, TolC (MukA). Mol Microbiol 22: 275-282.
121. Weitao T, Nordstrom K, Dasgupta S (1999) Mutual suppression of mukB and seqA phenotypes might arise from their opposing influences on the Escherichia coli nucleoid structure. Mol Microbiol 34: 157-168.
122. Ferullo DJ, Lovett ST (2008) The stringent response and cell cycle arrest in Escherichia coli. PLoS Genet 4: e1000300.
123. Saint-Dic D, Kehrl J, Frushour B, Kahng LS (2008) Excess SeqA leads to replication arrest and a cell division defect in Vibrio cholerae. J Bacteriol 190: 5870-5878.
124. Fujikawa N, Kurumizaka H, Nureki O, Tanaka Y, Yamazoe M, et al. (2004) Structural and biochemical analyses of hemimethylated DNA binding by the SeqA protein. Nucleic Acids Res 32: 82-92.

125. Kang S, Han JS, Kim SH, Park JH, Hwang DS (2007) Aggregation of SeqA protein requires positively charged amino acids in the hinge region. Biochem Biophys Res Commun 360: 63-69.
126. Kang S, Han JS, Kim KP, Yang HY, Lee KY, et al. (2005) Dimeric configuration of SeqA protein bound to a pair of hemi-methylated GATC sequences. Nucleic Acids Res 33: 1524-1531.
127. Chung YS, Guarne A (2008) Crystallization and preliminary X-ray diffraction analysis of SeqA bound to a pair of hemimethylated GATC sites. Acta Crystallogr Sect F Struct Biol Cryst Commun 64: 567-571.
128. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods in enzymology 374: 300-321.
129. Afonine PV, Grosse-Kunstleve RW, Adams PD (2005) The Phenix refinement framework. CCP4 Newsletter 42: Contribution 8.
130. Skarstad K, Bernander R, Boye E (1995) Analysis of DNA replication in vivo by flow cytometry. Methods Enzymol 262: 604-613.
131. Lu XJ, Olson WK (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nature protocols 3: 1213-1227.
132. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, et al. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. Journal of molecular biology 313: 229-237.
133. Lee JY, Chang J, Joseph N, Ghirlando R, Rao DN, et al. (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. Molecular cell 20: 155-166.
134. Bae SH, Cheong HK, Cheong C, Kang S, Hwang DS, et al. (2003) Structure and dynamics of hemimethylated GATC sites: Implications for DNA-SeqA recognition. J Biol Chem.
135. Bang J, Bae SH, Park CJ, Lee JH, Choi BS (2008) Structural and dynamics study of DNA dodecamer duplexes that contain un-, hemi-, or fully methylated GATC sites. Journal of the American Chemical Society 130: 17688-17696.
136. Baikalov I, Grzeskowiak K, Yanagi K, Quintana J, Dickerson RE (1993) The crystal structure of the trigonal decamer C-G-A-T-C-G-6meA-T-C-G: a B-DNA helix with 10.6 base-pairs per turn. Journal of molecular biology 231: 768-784.
137. Campbell JL, Kleckner N (1988) The rate of Dam-mediated DNA adenine methylation in Escherichia coli. Gene 74: 189-190.
138. Egan ES, Waldor MK (2003) Distinct replication requirements for the two Vibrio cholerae chromosomes. Cell 114: 521-530.
139. Egan ES, Lobner-Olesen A, Waldor MK (2004) Synchronous replication initiation of the two Vibrio cholerae chromosomes. Current biology : CB 14: R501-502.

140. Rasmussen T, Jensen RB, Skovgaard O (2007) The two chromosomes of Vibrio cholerae are initiated at different time points in the cell cycle. The EMBO journal 26: 3124-3131.
141. Chung YS, Brendler T, Austin S, Guarne A (2009) Structural insights into the cooperative binding of SeqA to a tandem GATC repeat. Nucleic Acids Res.
142. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.
143. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography 60: 2126-2132.
144. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, et al. (2009) Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. Proceedings of the National Academy of Sciences of the United States of America 106: 15442-15447.
145. Egan ES, Fogel MA, Waldor MK (2005) Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. Molecular microbiology 56: 1129-1138.
146. Fogel MA, Waldor MK (2005) Distinct segregation dynamics of the two Vibrio cholerae chromosomes. Molecular microbiology 55: 125-136.
147. Timsit Y, Moras D (1995) Self-fitting and self-modifying properties of the B-DNA molecule. Journal of molecular biology 251: 629-647.
148. Anderson J, Ptashne M, Harrison SC (1984) Cocrystals of the DNA-binding domain of phage 434 repressor and a synthetic phage 434 operator. Proceedings of the National Academy of Sciences of the United States of America 81: 1307-1311.
149. Jordan SR, Whitcombe TV, Berg JM, Pabo CO (1985) Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. Science 230: 1383-1385.
150. Schultz SC, Shields GC, Steitz TA (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 253: 1001-1007.
151. Yang W, Steitz TA (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell 82: 193-207.
152. Obmolova G, Ban C, Hsieh P, Yang W (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature 407: 703-710.
153. Chatti A, Landoulsi A (2008) GATC sites might regulate virulence of Salmonella typhimurium. Foodborne pathogens and disease 5: 555-557.
154. Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ (1999) An essential role for DNA adenine methylation in bacterial virulence. Science 284: 967-970.
155. Peters JE, Craig NL (2001) Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. Genes & development 15: 737-747.