



Published in final edited form as:

J Mol Biol. 2020 January 17; 432(2): 343–357. doi:10.1016/j.jmb.2019.08.019.

The solution structures and interaction of SinR and SinI: Elucidating the mechanism of action of the master regulator switch for biofilm formation in *Bacillus subtilis*

Morgan E Milton¹, G Logan Draughn², Benjamin G Bobay^{2,3}, Sean D Stowe², Andrew L Olson², Erik A Feldmann², Richele J Thompson¹, Katherine H Myers², Michael T Santoro², Daniel B Kearns⁴, John Cavanagh^{1,*}

¹Department of Biochemistry and Molecular Biology, The Brody School of Medicine, East Carolina University, Greenville, NC 27834, USA

²Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA

³Duke University NMR Center, Duke University, Durham, NC 27710, USA

⁴Department of Biology, Indiana University Bloomington, Bloomington, IN 47405, USA

Abstract

Bacteria have developed numerous protection strategies to ensure survival in harsh environments, with perhaps the most robust method being the formation of a protective biofilm. In biofilms, bacterial cells are embedded within a matrix that is composed of a complex mixture of polysaccharides, proteins and DNA. The Gram-positive bacterium *Bacillus subtilis* has become a model organism for studying regulatory networks directing biofilm formation. The phenotypic transition from a planktonic to biofilm state is regulated by the activity of the transcriptional repressor, SinR, and its inactivation by its primary antagonist, SinI. In this work, we present the first full-length structural model of tetrameric SinR using a hybrid approach combining high-resolution solution NMR, chemical crosslinking, mass spectrometry, and molecular docking. We also present the solution NMR structure of the antagonist SinI dimer, and probe the mechanism behind the SinR-SinI interaction using a combination of biochemical and biophysical techniques. As a result of these findings, we propose that SinI utilizes a residue replacement mechanism to block SinR multimerization, resulting in diminished DNA binding and concomitant decreased

*Corresponding author - John Cavanagh; cavanaghj19@ecu.edu.

Author Contributions S.D.S., G.L.D., and J.C. conceived the study. E.A.F., G.L.D., and M.E.M. designed the experimental approaches. B.G.B. and E.A.F. designed the computational approaches and performed the computational work. S.D.S., A.L.O., and G.L.D. performed the NMR-structure experiments. E.A.F., M.E.M., G.L.D., K.H.M., and M.T.S. performed the biochemical experiments. D.B.K. provided recombinant DNA materials. R.J.T. assisted in labelled sample preparation. J.C. was responsible for project design and correspondence. M.E.M. wrote the manuscript.

Accession numbers The structures of SinRN, SinRC, and SinI have been deposited to the Protein Data Bank with PDB entry IDs 5TN0 (SinRN), 5TN2 (SinRC), and 5TMX (SinI).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

repressor activity. Finally, we provide an evidence-based mechanism that confirms how disruption of the SinR tetramer by SinI regulates gene expression.

Keywords

Biofilms; *Bacillus subtilis*; transcriptional regulation; NMR solution structure

Introduction

With antimicrobial resistance escalating rapidly worldwide, understanding how bacteria regulate environmental stress responses is of increasing importance [1]. One of the ways that bacteria protect themselves from stressful environmental conditions is to form biofilms [2]. *Bacillus subtilis* is one of the most well studied bacterial species, being used as a paradigm in numerous genetic and biochemical investigations. It is a spore-forming, Gram-positive, non-pathogenic bacterium that has served as a model organism for studying the regulation of biofilm formation for over a decade [3]. *B. subtilis* has also been found to have a variety of health benefits as a probiotic [4–6]. Unfortunately, it also has a dark side, with recent findings that suggest it is not always beneficial or innocuous. Strains highly resistant to biocides and cleaning agents have been identified in hospitals [7]. This resistance is attributed to the production of especially thick biofilm matrices [8]. Pathogenic bacteria such as *Staphylococcus aureus* hide in *B. subtilis* biofilms, taking shelter from extensive disinfecting techniques [9]. *B. subtilis* is also able to transfer antibiotic resistance genes to serious human health threats such as *Clostridium difficile* [10]. With these deleterious behaviors coming to light, new interest has emerged in what was once thought to be a harmless bacterium.

Significant advances have been made toward our understanding of the critical components involved in the transition from planktonic to biofilm states [1,11–13]. The decision to execute this change in *B. subtilis* is governed predominantly by the *sin* operon [14–16]. The *sin* operon contains two critical gene products, SinR and SinI. SinR is a constitutively expressed transcriptional regulator, and has been identified as a master regulator of biofilm formation [17–19]. DNA binding by SinR leads to repression of the *epsA-O* and *tapA-sipW-tasA* operons [14,15], both of which are involved in the production of biofilm matrix polysaccharides and proteins. Expression of SinR inhibits transcription of these biofilm genes by physically blocking RNA polymerase access to its binding site. SinI is an antagonist of SinR [15,20]. Expression of SinI is triggered by environmental cues promoting biofilm formation [21]. The pivotal role SinR plays in this critical regulatory mechanism makes it the cornerstone of biofilm regulation in *B. subtilis*, and an important target for biofilm manipulation.

A significant amount of structural and biochemical work has been performed on SinR and SinI. SinR is a tetrameric protein composed of two domains, an N-terminal DNA-binding domain (SinRN) and a C-terminal oligomerization domain (SinRC) [22–24]. The first structure solved was that of the SinR-SinI heterodimer which revealed that the SinRC oligomerization domain interacts with the core of SinI to form a four-helix bundle [25].

Formation of this complex adversely affects SinR DNA binding [24,26]. The SinRN domain crystal structure was subsequently solved in its unbound state [24]. A structure of SinR bound to DNA was also elucidated, though no density was seen for the SinRC domain [26]. At this time there are no structures of either the tetrameric full-length SinR or full-length SinI apo-protein. Further, the mechanism by which SinI disrupts the SinR-DNA complex is not fully understood. Based on the currently available data, two models have been proposed for the structure of tetrameric SinR. In one model, the SinR tetramer is formed from a dimer of dimers with the SinRN DNA-binding domains of one dimer pair positioned on the same side of the tetramer [24](Fig S1a). The second model is formed by a dimer of dimers with the SinRN DNA-binding domains of one dimer pair residing on opposite sides of the tetramer [26](Fig S1b). To this point, it has been unclear as to which of these two models is correct.

Here, we address two major unanswered questions. First, how does SinR interact with DNA to repress biofilm formation genes, and second, how does SinI inhibit that function? To do this, we employ a hybrid approach that combines high-resolution NMR spectroscopy with chemical crosslinking and molecular modeling. By doing so, we provide for the first time solution structural models of both unbound full-length tetrameric SinR and unbound SinI. Next, we experimentally confirm SinR tetramer formation, and unravel the mechanism through which SinI inhibits SinR activity. Finally, we provide biochemical evidence for a proposed unusual binding mode which involves SinR bending DNA, and we present a model showing how DNA binding regulates biofilm gene expression.

Results and Discussion

Oligomeric state determination

Previous studies suggest that SinR is a tetramer in solution, with SinRC (residues 69–111) being solely responsible for oligomerization, and the DNA-binding domain SinRN (residues 1–69) being monomeric. It is also known that SinI is in monomer/dimer equilibrium, and that the SinR-SinI complex is a heterodimer. Using size exclusion chromatography (SEC), we showed that our construct of full-length SinR is a tetramer, our SinRN DNA-binding domain construct is a monomer, and our SinRC oligomerization domain construct is a tetramer (Fig. S2). These results agree with previous observations [23,24]. In a matrix such as those used in SEC and SEC multi-angle light scattering (SEC-MALS), SinI adopts a clear dimeric state [24] (Fig. S2). In solution, SinI is in a rapidly reversible monomer/dimer equilibrium as shown through analytical ultra-centrifugation (AUC) [23] (further discussed in Fig. 5c). We repeated SinI sizing using SEC and AUC and confirmed that our SinI construct has the same oligomeric state as previously found.

Structural characterization of SinR

Since prior crystallographic efforts were unsuccessful in obtaining structural information on full-length tetrameric SinR [25], we turned to NMR spectroscopy. SinR HSQC NMR spectra display two distinct sets of peaks with different intensities, indicative of differential motions between the two domains. Therefore, as we describe in detail below, we used individual SinRN and SinRC truncation constructs for initial structural characterizations. Once NMR

structures of each domain were determined, a full-length model was generated using distance constraints from multiple chemical crosslinking experiments and ESI-MS/MS analysis. This technique has proven to be successful for determining structures of highly dynamic proteins, especially those with independently moving domains [27,28].

Structure of the SinR DNA-binding domain, SinRN

The structure of monomeric SinRN is composed of five α -helices (Fig. 1a). Overlaying our solution structure with previous crystal structures results in a slight variation in root mean square deviations (RMSDs). The RMSDs for our structure and the original structure of SinR bound to SinI (PDB ID 1B0N [25]), the apo SinR DNA-binding domain (PDB ID 3QQ6 [24]), and the DNA bound SinR DNA-binding domain (PDB ID 3ZKC [26]) are 1.073, 0.871, and 1.005 Å, respectively. Differences in the RMSDs are due to a minor shift in the linker between helices α 3 and α 4 across all the structures and a minor shift in helix α 5 when comparing the DNA bound structure of SinRN (PDB ID 3ZKC [26]) with our solution structure (Fig. 1b). These results confirm that the structure of SinRN does not alter significantly between when it is free in solution, bound to DNA, or bound to SinI. Importantly from a functional standpoint, this indicates that SinI does not appear to hinder SinR by inducing a significant conformational change in the DNA-binding domain.

Structure of the SinR oligomerization domain, SinRC

Tetrameric SinRC is formed by a dimer of dimers through helical hooks (Fig. 2a). Each dimer is composed of a four-helix bundle created from interlocking helical hooks. Oligomer interactions position the N-terminal portions of this domain away from the core tetramer interface. NMR distance restraints confirm the tetramer interface previously identified [26]. This interface is comprised of helix α 7 residues from each monomeric unit (Fig. 2b). Adjacent helices are joined through salt bridges formed by Glu97 and Arg105 as well as pi-stacking from Tyr101 and Trp104. All four residues from each chain are involved in the tetramer interaction. Alignment of our SinRC solution structure with the crystal structure (PDB ID 2YAL[24]) results in an RMSD of 1.491 Å. A key difference between these two structures is the angle between the helices. In the NMR structure, helix α 6 adopts a slightly more curved conformation in solution, bringing Glu97 and Arg105 into closer proximity. Adaptation of this conformation most likely aids in formation of the tetramer interface salt bridges.

Full-length structural model of tetrameric SinR

Using a data-driven molecular modeling approach, we generated a full-length model of tetrameric SinR. This method has been used previously to successfully develop models of other highly flexible proteins [27,28]. Chemical crosslinking was used to stabilize and isolate monomeric, dimeric, and tetrameric SinR for subsequent analysis by mass spectrometry (Fig. S3). Two crosslinkers with different linker lengths, BS3 (bis(sulfosuccinimidyl) suberate) and DSG (disuccinimidyl glutarate), were found to produce inter-domain crosslinks between lysine residues [27]. Of the thirteen lysine residues in SinR, six belong to SinRN and six to SinRC, with the remaining lysine in the flexible linker region (K65). The lysine residues of both SinRN and SinRC are dispersed throughout the domains. Between the two reagents, DGS and BS3, there were three observed protein

species corresponding to monomeric, dimeric, and tetrameric SinR as seen through SDS-PAGE gel (Fig. S3a). Treatment with DSG provided six intramolecular crosslinks (Fig. S3c). One of these crosslinks (K23 to K28) was between residues within the C-terminus, while the other five correspond to C- to N-terminal crosslinks of a single SinR monomer (K23 to K93, K28 to K93, K28 to K103, K28 to K106, and K28 to K109). Treatment with BS3 provided five intramolecular and one intermolecular crosslink (Fig. S3c). Three of the intramolecular crosslinks correspond to intra-domain linkages (K7 to K23, K7 to K28, and K23 to K28). The other two intramolecular crosslinks connect the C- and N-terminal domains (K28 to K93 and K28 to K109). A single intermolecular crosslink was observed for BS3, connecting K28 from separate molecules of SinR. The valid crosslinking results were used as distance restraints in subsequent molecular docking procedures (see methods). Our resulting model of tetrameric SinR is reminiscent of those previously proposed by Colledge et al. (Fig. S1a) [24]. The SinRN DNA-binding domains project radially from the core helical bundles that compose the oligomerization domain (Fig. 3a). The oligomerization domains remain locked in place while the DNA-binding domains can sample a great deal of conformational space as demonstrated by our molecular dynamic simulations. (Fig. 3b). This domain mobility further supports the ability of SinR to recognize a wide array of variably-spaced DNA binding sites.

Structure of SinI

We determined the solution NMR structure of SinI. This represents the first structure of dimeric SinI. Unlike the well-ordered SinR protein, SinI is a partially disordered protein with a centralized helical hook domain (residues 19–45). Two helical hooks come together to form a four-helix bundle that is the core of the SinI dimer (Fig. 4). A combination of intermolecular hydrogen bonds and hydrophobic interactions hold the dimer together. These interactions include the hydrogen bonds formed between Asp12 and Lys23, Glu14 and Lys40, Glu14 and Arg33, Glu17 and Arg33, and Glu21 and Lys40. Hydrophobic interactions involve Trp15, Lys18, Ile27, Ile23, and Pro29. The dimerization region is flanked by a region of 19 disordered amino acids. The specific role of this disordered section is unknown.

SinR-SinI interactions

The SinI-SinR interaction is essentially irreversible, creating a dead-end complex [23] as established by the very low binding affinity and very slow dissociation rate [26]. The formation of this dead-end inhibitory complex is proposed to be necessary for sequestering free SinR thereby ensuring biofilm formation [23]. A mechanism for the interaction between SinR and SinI was first proposed in 1992 by Smith and co-workers [20]. Strong sequence similarity between SinI and the oligomerization domain of SinR suggested that SinI displaces a SinR monomer from its homotetramer complex [20]. The SinR dimer interface and the SinI dimer interface are highly homologous in sequence and structure (Fig. 5a,b).

Despite sequence and structural similarities, the four key residues involved in SinR tetramer formation (Fig. 2b) are significantly different in SinI. Residues involved in hydrogen bonding, Glu97 and Arg105 in SinR, are replaced with Lys34 and Ser42 in SinI, respectively. Aromatic residues Tyr101 and Trp104 from SinR are substituted for Leu38 and Lys41, respectively (Fig. 5a,b). To probe how SinI disrupts SinR tetramer formation, we developed a SinI mutant that would reestablish the SinR oligomerization residues. Our SinI

mutant, SinI^{quad}, contains four point mutations necessary to mimic the SinR tetramer interface: K34E, L38Y, K41W, and S42R. We employed AUC to evaluate oligomerization of the individual proteins. For both SinI and SinI^{quad}, the distribution of sedimentation coefficients (c(s)) indicates the presence of monomeric (1.2 S) and homodimeric (1.8 S) states (Fig. 5c, Fig. S4). Notably, SinI^{quad} has a greater population of dimers than SinI. This may be due to the K37E mutation which eliminates possible hydrogen bonding at this site. Finally, a uniform distribution of an apparent homotetramer (3.4 S) was observed for SinR. Next, we investigated complex formation. Equal molar ratios of SinR and SinI resulted in a complete disruption in the SinR homotetramer with a broad 1.7 S distribution (Fig. 5c). Adding SinI^{quad} to SinR also completely disrupted the SinR homotetramer. A small population corresponding to a SinR-SinI^{quad} heterodimer is present, but the c(s) distribution is dominated by the presence of a 2.7 S species, strongly suggestive of an apparent heterotetrameric complex of SinR₂-SinI^{quad}₂. Thus, we demonstrate that the mutations of SinI to mimic the tetramer interface of SinR confirm that E97, Y101, W104, and R105 are intimately involved in tetramer formation. Based on these results, we conclude and confirm that SinI inhibits SinR oligomerization, first by displacing one subunit from a SinR dimer [23,24,26], and then by preventing tetramer formation through the elimination of key hydrogen bond and pi-stacking interactions. Since SinI does not interact with DNA [20], disruption of the SinR tetramer must play a critical role in affecting SinR DNA binding activity.

Protein-DNA interactions

The ability for SinR to bind DNA is vital to its role as the master biofilm regulator. Numerous genes have been identified as being involved in biofilm formation in *B. subtilis* [12,29–31]. Identification of DNA sequences that are bound by SinR has been important in understanding regulation of these genes. SinR binding sequences have been observed in a variety of number and orientations such as in the *espA-O* and *tapA-sipW-tasA* operons (Fig. 6a)[14,15]. Interestingly, SinR binding sites are often too far upstream of the RNA polymerase –10 and –35 binding elements to directly compete with binding [15]. This suggests that SinR does not simply regulate gene expression by binding to DNA and blocking transcription [32].

Our SinR construct binds an inverted repeat substrate of the *tapA-sipW-tasA* promoter with an apparent K_d value of 249 ± 38 nM (Fig. 6b). This value is comparable to previously determined K_d values for inverted repeat substrates [24,26]. These results also indicate that SinR does not interact with the fluorescent dye on the DNA substrate. Fitting our data to a nonlinear regression results in a Hill coefficient of 1.012, corresponding to noncooperative binding.

Our structural model of full-length tetrameric SinR suggests that the SinRC oligomerization domain and the SinRN DNA-binding domain have somewhat independent motion (Fig. 3b). This is supported, as alluded to above, by the crystal structure of the SinR-DNA complex, where the SinRN DNA binding domain is clearly delineated, but the SinRC domain is absent due to lack of electron density [26]. Consequently, it is plausible that SinRN by itself would be able to bind DNA to some degree. With that in mind, we performed NMR chemical shift

perturbation studies where DNA was titrated into a ^{15}N -labeled SinRN sample. Chemical shift changes show that SinRN does indeed bind DNA, and the generally small perturbations confirm that SinRN conformational changes appear minor (Fig. S5). This supports the SinR-DNA structure determined by Newman et al. [26]. Also corroborating that work, the NMR data suggest that SinRN helices α_2 and α_3 are involved in DNA binding. Interestingly, chemical shift changes are also observed for residues 40–45, which comprise the linker between helices α_3 and α_4 , as well as the end of helix α_4 . These perturbations could result from interactions from adjacent DNA-binding domains as previously proposed by Colledge et al. [24]. These interactions may further stabilize SinR bound to DNA, locking the repressor in the correct position to block RNA polymerase. Certainly, these NMR data demonstrate that an isolated SinRN monomer can bind to DNA.

SinI adversely affects SinR DNA binding in a concentration-dependent manner [20,23,25], and is unable to disrupt a pre-formed SinR-DNA complex [20,24,26]. Using fluorescence anisotropy, we demonstrated that our SinI construct impedes SinR from binding to an *epsA-O* promoter-based DNA substrate in a dose-dependent manner (Fig. S6). DNA binding is reduced significantly, but not completely, even in conditions where there is a 10-fold excess of SinI compared to SinR. Previous studies [20] suggest that SinI fully inhibits SinR's ability to bind DNA but this gel-based assay is not as sensitive as the fluorescence anisotropy approach reported here. Regardless, the ability of SinI to drastically affect SinR DNA binding has functional consequences.

The question as to the precise mechanism by which SinI affects SinR DNA binding and subsequently function has not yet been established. Studies here and by others reveal that the formation of a SinR-SinI complex does *not* induce a structural change in the SinR DNA-binding domain [24,26], and we have shown that a single SinR DNA-binding domain is also able to interact with DNA. Consequently, the answer must lie not in the innate ability of the SinRN DNA-binding domain to recognize and interact with DNA, but more likely in the capacity of SinI to disrupt SinR tetramer formation and the multiple SinR recognition sites within the promoter regions of the operons that SinR regulates.

Regulation of biofilm formation through DNA bending

The presence of multiple SinR binding motifs around the promoter regions has long been perplexing. A mechanism for SinR binding to these multiple binding sites was first presented by Chu et al. in 2005. They proposed that the DNA could form a loop, allowing the upstream sites to stabilize the SinR tetramer [14]. With their model of full-length SinR (Fig. S1a), Colledge et al. noted that the length of SinR and the distance between the binding sites could accommodate complete tetramer binding through the creation of a loop in the DNA [24]. Newman et al. presented the first visual representation of SinR bending the *tapA-sipW-tasA* promoter by 180° to facilitate binding at the two inverted repeat sites [26]. By creating a similar model, we show that our full-length structure of SinR can comfortably bind two sets of inverted repeats from the *tapA-sipW-tasA* promoter by bending the DNA. Our full-length structure and molecular dynamics data allow us to more accurately explore the SinR DNA-bending mechanism.

The size and shape of the SinR tetramer will dictate the distance between the inverted repeats on either side of the DNA loop. Too short of a distance creates a loop that will put too much tension on the DNA, making it unfavorable to bend. The *tapA-sipW-tasA* promoter contains about 80 base pairs separating the inverted repeats and is able to take on a 180° bended conformation [26] (Fig. 7a). The *espA-O* promoter has approximately 100 base pairs between SinR binding sites, suggesting it also should easily be able to take on a loop conformation [24]. The two previous models for SinR have significantly different shapes and sizes. In the first model, the oligomerization domain is quite elongated and spans ~100 Å between DNA-binding domains of the adjoined SinR dimers (Fig. S1a and S7a). The distance between DNA-binding domains in the Newman et al. model was based on the crystal lattice [26]. This crystal packing spaced the DNA ~ 60 Å apart in the crystal lattice, leaving minimal room for the oligomerization domain (Fig. S1b and S7b). It turns out that the flexible nature of SinR allows it to accommodate a range of DNA loop dimensions. Our molecular dynamics model shows that free movement of the SinR DNA-binding domains about the SinRC oligomerization domains would allow SinR to have two binding orientations, one dimer per inverted repeat (Fig. S7a) or one subunit from each dimer per inverted repeat (Fig. S7b). These two orientations explain the significant differences between the two previously proposed models, and how SinR can bind sites only 60 Å apart.

To test if SinR could physically bend DNA *in vitro*, we developed a fluorescence-based assay. A DNA substrate designed from the *tapA-sipW-tasA* promoter sequence was engineered to contain an internal fluorophore and a quencher (Fig. 7b). Bending of the DNA brings the fluorophore into close proximity to the quencher, resulting in decreased fluorescence signal. As SinR was titrated into a constant concentration of DNA, the fluorescent signal decreased (Fig. 7c). No SinR controls and the previous determination that SinR does not interact with the fluorescein fluorophore indicate that SinR is bending the DNA in solution. Incubating SinR with increasing concentrations of SinI prior to exposure to DNA restores the fluorescent signal, though not fully (Fig. 7d). The inability to completely restore the fluorescent signal and the larger error bars may be due to the fact that the interaction between SinR, SinI, and DNA do not appear to follow standard Michaelis-Menton kinetics (Fig. S5). More extensive kinetics studies would need to be conducted to determine this mechanism. Our data indicate that the observed change in fluorescence is due to SinR binding and bending the DNA. These results provide the first experimental evidence to support the DNA bending model for transcriptional regulation by SinR.

By combining structural and biochemical data from our work and that of several other groups, we can now build a model for how SinR represses transcription, and how SinI regulates SinR activity through the *tapA-sipW-tasA* promoter (Fig. 7e). SinR binds multiple recognition sequences within the promoter region, creating a loop in the DNA. Binding of all four DNA-binding domains stabilizes the SinR tetramer, while the loop creates a physical barrier, preventing RNA polymerase from binding to the *tapA-sipW-tasA* promoter. Expression of SinI produces dead-end SinR-SinI complexes. The sequestration of SinR by SinI leaves the *tapA-sipW-tasA* promoter regions in a linear state, allowing RNA polymerase to carry out transcription. While this SinR DNA-bending mechanism very likely occurs for other SinR-binding promoters such as the *espA-O* promoter (Fig. 6A), it is not clear if SinR

could physically block RNA polymerase binding due to the distance between the polymerase start site and the SinR binding motif (58 bp, ~190Å for the *espA-O* promoter). In such a case, the binding of DNA by SinR could block other regulatory elements such as sigma factors from binding the DNA and preventing recruitment of the RNA polymerase. Further work would help determine the exact mechanism of inhibition for promoters that contain SinR DNA-binding sites are not in proximity to the RNA binding site.

Conclusion

Over the years, extensive research has been performed on SinR and SinI. While much was learned, the story was far from complete. Here, we have summarized prior work from several groups and filled in many of the gaps with our new structural and biochemical data. By combining solution NMR with chemical crosslinking, mass spectrometry, and molecular modeling, we provide the first structures of both full-length tetrameric SinR and dimeric SinI in solution. Twenty years after the first crystal structure of SinR bound to SinI gave us hints at the regulatory mechanism, we finally have a complete structural picture of these two essential biofilm regulators. Our molecular dynamics studies highlight the highly flexible nature of both SinR and SinI, and this appears to play a critical role in the ability of SinR to recognize a wide range of DNA binding sites. By probing the SinR tetramer interface with our SinI^{quad} mutant, we were able to determine the molecular mechanism by which SinI inhibits SinR. Binding of SinI not only displaces a SinR molecule, but also disrupts tetramerization with other SinR molecules. Furthermore, we provide the first biochemical evidence to support the proposed SinR DNA-bending model. The bending of a DNA promoter is likely the key mechanism for gene repression by SinR, and is facilitated by the formation of the flexible SinR tetramer. SinI disrupts the SinR tetramers, inhibiting SinR from effectively binding DNA. Since derepression of SinR by SinI induces a biofilm state, targeting the SinR-SinI protein interaction could be a viable therapeutic intervention point. Inhibition of the SinR-SinI interaction would impede biofilm formation, making *B. subtilis* more vulnerable and unable to harbor dangerous pathogens.

Methods

Cloning, expression, and protein purification

The plasmid containing the full length *sinR* was provided by Professor Richard Losick at Harvard University. Generation of the SinR, SinRN (residues 1–69), and SinRC (residues 69–111) constructs was described previously [33]. *sinI* was amplified and cloned into pET28a containing a thrombin-cleavable N-terminal His₆ affinity tag. A mutant version of SinI (SinI^{quad}) was generated using the QuickChange II site-directed mutagenesis kit (Agilent Technologies) to make the following mutants: K34E, L38Y, K41W, and S42R. The SinI^{quad} construct remained in the pET-28a expression system with a thrombin-cleavable N-terminal His₆ affinity tag and all mutations were confirmed by sequencing (Eton Bioscience).

All SinR constructs were expressed as previously described [33]. Samples of SinI and SinI^{quad} were prepared from 1 L cultures grown at 34°C to OD₆₀₀ of ~0.7 before a six-hour induction using 1 mM IPTG before final harvest. Uniformly labelled ¹³C/¹⁵N samples of

SinI, SinRN and SinRC were prepared using a previously described method [33]. SinR, SinRN, and SinRC were purified based on previous methods [33]. Over expressed SinI and SinI^{quad} were resuspended in lysis buffer containing 50 mM Tris-HCl pH 8.5, 300 mM NaCl, and 0.02 % NaN₃, sonicated to homogeneity, and clarified by centrifugation at 15,000 g. The supernatants were immediately passed over a pre-equilibrated column containing Ni-NTA agarose (QIAGEN) and washed with 50 mL lysis buffer and 50 mL of lysis buffer containing 10 mM imidazole. Protein was eluted into 10 mL fractions using lysis buffer with an imidazole gradient from 0 to 300 mM. All SinI/SinI^{quad}/SinRC samples were cleaved overnight using 200 units of thrombin (Merck Millipore) overnight and assessed for successful removal of His₆ affinity tag by SDS-PAGE before final dialysis and concentration.

Size Exclusion Chromatography

A Waters Acquity H-Class UPLC was utilized to carry out analytical size-exclusion liquid chromatography (SELC). Prior to injection, each construct was equilibrated into an identical mobile phase buffer containing 20 mM MES pH 6.0, 200 mM NaCl, and 0.02% NaN₃. Individual samples were diluted to a final concentration of 250 μM and 25 μL were applied to a pre-equilibrated BEH-125 SELC column (Waters) using an autosampler under a constant 0.2 mL/min flow rate. Absorbance at 280 nm was continuously measured and the data processed and analyzed using Empower 3.0 software (Waters). Retention times were recorded for each species and plotted against a standard curve generated using a mixture of Gel Filtration Standards (BioRad) to obtain an apparent molecular weight for each sample. Molecular weights of proteins complexes were evaluated using an identical method.

Analytical Ultracentrifugation

Sedimentation velocity experiments were carried out at 45,000 rpm on an Optima XL-A analytical ultracentrifuge using a An-60 Ti four-space analytical centrifuge rotor (Beckman Coulter) and Epon charcoal-filled centerpieces. Absorbance measurements were recorded at 280 nm every 8 minutes with 0.03 cm spacing. Velocity profiles were processed and analyzed with SedFit (Peter Schuck, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD). Sedimentation coefficients were corrected for temperature and buffer composition using Sednterp (<http://sednterp.unh.edu>). All experiments were carried out at 20 °C in buffer containing 20 mM MES pH 7.0, 300 mM NaCl, and 0.02% w/v sodium azide.

NMR Spectroscopy

Backbone, sidechain, and aromatic chemical shift assignments for SinRN and SinRC were completed previously [33] and can be accessed at the BioMagResBank (BMRB) under accession numbers 30193 and 30194, respectively. The NMR experiments for SinI were performed at 298 K on either a Varian Inova 600 MHz or a Bruker Avance 700 MHz, both equipped with cryoprobes. Backbone chemical shifts were assigned in a sequential manner from the following experiments: [¹⁵N-¹H] HSQC, [¹⁵N-¹H] TROSY-HSQC, HNCOC, HN(CA)CO, CBCA(CO)NH, HNCACB, and C(CO)NH (15 and 18 ms). Sidechain proton chemical shifts were assigned using the following experiments: HBHA(CO)NH, H(CCO)NH (12 and 18 ms), ¹⁵N-TOCSY-HSQC (50 and 75 ms), and ¹⁵N-NOESY-HSQC

(80 ms). Data were processed using NMRPipe [34] and analyzed using NMRView [35]. Dihedral angles and secondary structure predictions were calculated using the program TALOS+ [36] and C_{α} chemical shift index deviations [37], and confirmed using ^1H - ^{15}N heteronuclear NOE experiment with a 5 s relaxation delay (Table S1). The chemical shift assignments for SinI can be accessed through the BMRB under accession number 30192.

Distance restraints were obtained for all proteins from ^{15}N -NOESY-HSQC (75 and 100 ms), ^{13}C -NOESY-HSQC (100 and 120 ms), and ^{13}C aromatic NOESY experiments (80 ms). Identification of residues involved in the multimerization interface was accomplished through a ^{13}C -filtered, ^{12}C -edited NOESY-HSQC (100 ms). $^1\text{D}_{\text{NH}}$ residual dipolar couplings (RDC) were measured for both SinRC and SinI using radially compressed 6.0 mm to 4.2 mm polyacrylamide gels with acrylamide concentrations of 4% and 8% w/v, respectively, and data were collected from the IPAP-HSQC experiment and analyzed using NMRView [35].

Protein Structure Calculation and Validation

For all structures, initial models were calculated using NOE-derived distance restraints with CYANA v.3.1 (noeassign) [38–40] along with TALOS+ [36] predicted backbone dihedral angle restraints. For SinI and SinRC, both of which are multimeric models, noncrystallographic symmetry restraints were utilized to keep the C_{α} atoms of the monomers superimposable, and distance symmetry potential was used to ensure that the relative orientations of all C_{α} atoms within respective monomers were symmetrical. SinRC was originally calculated as a dimeric protein with removal of known tetramer-forming NOEs. The final step in the CYANA calculation of SinRC was a manual run composed of the upper-limit and .aco files of the dimer, duplication of the dimeric files, and noncrystallographic symmetry restraints for generation of the tetrameric structure. The final 100 CYANA conformers with the lowest target function values were subjected to restrained energy minimization in implicit solvent in AMBER [41]. The lowest 10 energy structures as calculated through AMBER were then analyzed, minimized for further violations, and ultimately submitted to the PDB (Table S1). Molecules were visualized and aligned with PyMOL. The structure was further analyzed by PSVS validation suite to confirm stereochemical quality of a protein structure [34,35]. The solution structures for SinRN, SinRC, and SinI were deposited in the RCSB Protein Data Bank (PDB) and can be accessed under PDB IDs 5TN0, 5TN2, and 5TMX, respectively.

NMR Structure Statistics

The N-terminal domain, SinRN, structures were determined with a total of 1274 nuclear Overhauser enhancements (NOEs) composed of 226 intra-residue, 231 sequential, 416 medium range, and 401 long range NOEs. 40 and 42 Phi/Psi angles were used for structure calculations. This exhibited 19.6 restraints and 5.81 long-range constraints per residue. The final ten structures resulted in no NOE or dihedral violations per structure at 0.2 Å and 5°, respectively. The structural ensemble has an average C_{α} RMSD of 0.20 ± 0.10 Å. PSVS (Protein Structure Validation Server), MolProbity, WHATCHECK, and PROCHECK were used to analyze the ensemble that was subsequently determined to be of good quality.

Ramachandran analysis confirmed that 99.7% of residues lie within the additionally allowed or better conformational space.

For SinRC, the structures were determined with a total of 989 NOEs per monomer, of which 210 were intra-residue, 197 were sequential, 242 were medium range, and 340 were long range. Finally, 50 Phi/Psi angles were used. This resulted in 22.15 restraints per residue with 7.23 long-range constraints per residue. The final 10 structures had zero NOE and 2 dihedral violations per structure at 0.2Å and 5°, respectively. The structural ensemble showed an average C^α RMSD of 0.24 ± 0.02 Å for secondary structure (backbone atoms) and of 0.65 ± 0.03 Å for secondary structure (heavy atoms; i.e., all non-hydrogen atoms). PSVS, MolProbity, WHATCHECK, and PROCHECK were used to analyze the ensemble that was subsequently determined to be of good quality. Ramachandran analysis confirmed that 100% of residues lie within the additionally allowed or better conformational space.

The solution structure of SinI was characterized using solution NMR spectroscopy, representing the first independent structure solved for the SinI dimer. Approximately 95% of the assignable backbone residues, 95% of the sidechain carbons, and 85% of proton side-chain assignments were completed (the missing assignments reside in the C-terminal residues 57–61, most likely due to conformational flexibility of this region). The structures were determined with a total of 297 NOEs per monomer, of which 56 were intra-residue, 83 were sequential, 77 were medium range, and 81 were long range. Finally, 21 Phi/Psi angles were used. This resulted in a total of 11.77 restraints per residue over ordered residues with 7.77 long-range constraints per residue. The final 10 structures had zero NOE violations above 0.2 Å and zero dihedral violations >5°. The structural ensemble displays an average C^α root-mean-square deviation (RMSD) of 0.35 ± 0.06 Å for secondary structure (backbone atoms) and of 0.90 ± 0.06 Å for secondary structure (heavy atoms; i.e., all non-hydrogen atoms). PSVS, MolProbity, WHATCHECK, and PROCHECK were used to analyze the ensemble that was subsequently determined to be of good quality. Ramachandran analysis confirmed that 94.8% of residues lie within the additionally allowed or better conformational space.

Chemical Crosslinking and Mass Spectrometry

Equivalent volumes of serially diluted amine-to-amine crosslinkers BS3 and DSG (Thermo Scientific) were incubated in 100 µL reaction volumes containing a fixed 100 µM concentration of purified full-length SinR. The reactions took place under ambient conditions and proceeded for 0.5 h before being quenched using 50 mM Tris. The reaction mixtures were then separated by 1D-SDS-PAGE using a single-percentage 12% tricine gel. Following separation, the molecular mass region corresponding to a SinR tetramer (~52 kDa) (Fig. S3) was excised and subjected to an in-gel reduction, iodoacetamide alkylation, and trypsin digestion as previously described [42]. Extracted peptides were lyophilized to dryness and then resuspended in 20 µL of 2% acetonitrile and 0.1% formic acid prior to liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis. Chromatographic separation was performed on a Waters NanoAcquity UPLC equipped with a 1.7 µm BEH130 C₁₈ 75 µm ID × 250 mm reversed-phase column. The mobile phase consisted of (a) 0.1% formic acid in water and (b) 0.1% formic acid in acetonitrile. Following a 4 µL injection,

peptides were trapped for 5 min on a 5- μ m Symmetry C₁₈ 180 μ m ID \times 20 mm column at 5 μ L/min in 99.9% A. The analytical column was then switched in-line and a linear elution gradient of 5% B to 40% B was performed over 60 min at 400 nL/min. The analytical column was connected to a fused silica PicoTip emitter (New Objective, Cambridge, MA) with a 10- μ m tip orifice and coupled to a Waters Synapt G2 Quadrupole Time-of-Flight mass spectrometer through an electrospray interface. The instrument was operated in data-dependent mode of acquisition with precursor mass spectrometry scans from m/z 50–2000 and the top three most abundant precursor ions being subjected to MS/MS fragmentation. For all experiments, charge-dependent CID energy settings were employed, and a 120-s dynamic exclusion was employed for previously fragmented precursor ions.

Raw LC-MS/MS data files were processed in Mascot distiller (Matrix Science) and then submitted to independent Mascot database searches (Matrix Science) against a SwissProt (taxonomy *B. subtilis*) database (4290 forward sequences, updated December 2012) appended with the reverse sequence of all of the forward entries. Search tolerances were 10 ppm for precursor ions and 0.04 Da for product ions using trypsin specificity with up to two missed cleavages. Carbamidomethylation (+57.0214 Da on C) was set as a fixed modification, whereas oxidation (+15.9949 Da on M) and hydrolyzed DSG (+114.031694 Da on K) or hydrolyzed BS3 (+156.0786 Da on K) were considered as variable modifications. All searched spectra were imported into Scaffold (Proteome Software), and protein confidence thresholds were set using a Bayesian statistical algorithm based on the PeptideProphet and ProteinProphet algorithms that yielded a peptide and protein false discovery rate of <1% [43,44]. To identify crosslinked species, we generated Mascot distiller. MGF files were submitted to MassMatrix (v 2.4.2, February 2012) searches against a forward/reverse SwissProt database (taxonomy *B. subtilis*) of SinR [45]. Search mass tolerances and modifications were as described for Mascot searches, with the “advanced search” option enabled to allow for interpeptide or intrapeptide crosslinking of DSG (+96.0211 Da) and BS3 (+138.0681 Da). Specificity of the BS3 and DSG cross-linkers was initially confined to lysine–lysine. A secondary search was then performed for each crosslinking reagent replacing lysine with glycine to allow for the mapping of crosslinked sites to the protein N-terminal primary amine. Trypsin rules were set to not allow cleavage at a crosslinked modified residue and only one crosslink per peptide pair was allowed. A peptide match within MassMatrix was only considered if peptide scoring thresholds were above that required for a matching probability less than p-value of <0.05. All crosslinked MS/MS spectra were manually inspected for adequate fragment ion coverage.

Crosslink Derived Molecular Docking

HADDOCK in combination with chemical cross-linking mass spectrometry was used to generate structural models for the SinR tetramer. Default HADDOCK parameters were used throughout the docking procedure [46] with the following exceptions: non-crystallographic and C2 symmetry restraints were used for each monomer. Semi-flexible residues were defined automatically via HADDOCK. Fully flexible residues were set to accommodate the mass spectrometry crosslinking data, BS3 has a 11.4-Å spacer arm and DSG has a 7.4-Å spacer arm. For defining the appropriate HADDOCK flexibility parameters, the following criteria were taken into account: the large distances of the crosslinking moieties, the

dynamic nature of the protein, the symmetry of the tetramer, and the need to retain the identified multimeric state of both SinRN and SinRC at all concentrations. Lastly, ambiguous restraints were not included. One thousand structures were generated for the first iteration (rigid docking), 200 were generated for the second iteration (semi-flexible docking), and 200 lowest energy structures were water refined. The C α RMSD values of the complexes were calculated using ProFit (Martin, A.C.R., <http://www.bioinf.org.uk/software/profit/>). A cluster analysis was performed on the final docking solutions using a minimum cluster size of 4. The RMSD cutoff for clustering was manually determined to be 5.5 Å (lower than the default 7.5 Å). The RMSD matrix was calculated over the backbone atoms of the interface residues of the proteins. The lowest energy structure from the molecular docking within the highest populated cluster was further analyzed by PSVS to confirm stereochemical quality of a protein structure [47].

NMR Chemical Shift Perturbation

A series of ¹⁵N-HSQC experiments on a Bruker Avance 700 MHz equipped with a cryoprobe were performed at 298 K in conjunction with titrations of unlabeled DNA. ¹⁵N-SinRN:DNA titrations were carried out until a molar ratio of 1:4 protein to DNA was achieved. The oligonucleotide, 5'-TTTGTTCTCTAAAGAGAACTTA, and its complement (Integrated DNA Technologies) were annealed at 1:1 in Sin NMR buffer by heating to 95°C for 10 minutes and slowly cooled to room temperature. The NMR data was processed using NMRPipe and analyzed using NMRView.

Fluorescence Polarization DNA-binding Assay

The DNA binding affinities of SinR were determined using a fluorescence polarization assay performed on a Synergy H1 Hybrid Multi-Mode microplate reader (BioTek). All assays were performed at ambient temperature in 20 mM Tris pH 7.9, 100 mM NaCl, 2 mM MgCl₂, 1 mM DTT, and 0.1 mg mL⁻¹ BSA with constant 20 nM fluorescein labeled DNA substrate. The DNA substrate was based on the *espA-O* promoter consensus sequence [14] labeled with 6-carboxyfluorescein (6-FAM). The 22-mer contained the following sequence: 5'-6-FAM-TTTGTTCTCTAAAGAGAACTTA (the inverted repeat DNA binding site is centrally positioned and underlined). Experiments were performed with 24 μM, final, SinR followed by a series of seven 2-fold serial dilutions. The reactions were initiated by the addition of protein dilutions to 10 nM, final, DNA 22-mer in a final reaction volume of 100 μL in a black/opaque 96-well microplate (BrandTech). Fluorescent substrates were excited at 482 nm and polarization data were collected at 528 nm. Data were analyzed and K_d values were determined using DynaFit software (BioKin Ltd.).

Fluorescence Intensity DNA-binding Assays

The ability of SinR to bend DNA was determined using a fluorescence intensity assay performed on a Synergy H1 Hybrid Multi-Mode microplate reader (BioTek). All assays were performed at ambient temperature in 20 mM Tris pH 7.9, 100 mM NaCl, 2 mM MgCl₂, 1 mM DTT, and 0.1 mg mL⁻¹ BSA. The 90 bp DNA substrate was based on the *tapA-sipW-tasA* promoter. One strand of the DNA contained a dT labeled with Internal Fluorescein (5'-GAAAGTTCTCTAAAGAGAAACA/iFluorT/TATAAATAATAAGGGAAGTGCAGTAAATTA

GAGGAAAATCATGTATTGTTCTCTAAAGAGAACTTAG), while the other strand contained a dT labeled with Internal BHQ-1 (5'-CTAAGTTCTCTTTAGAGACAATA/iBHQ-1dT/
CATGATTTTCCTCTAATTTACTGCACTTCCCTTATTATTTATAATGTTCTCTTTAGAGAACTTTC). Dye and quencher were tethered to the 5 position of the thymine ring with a 6-carbon linker arm, ~15 Å long. Experiments with SinR alone were performed with 24 µM, final, SinR followed by a series of seven 2-fold serial dilutions. The reactions were initiated by the addition of protein to the 10 µM, final, DNA resulting in a final reaction volume of 100 µL. For studies containing SinI, 72 µM final SinI was serial diluted 2-fold, seven times. The dilutions were incubated with 24 µM final SinR. The protein mixture was then added to 10 nM final DNA. Reactions were carried out in black/opaque 96-well microplates (BrandTech). Fluorescent substrates were excited at 482 nm and emission data were collected at 528 nm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grant R01 GM055769 (J.C.) and the V Foundation for Cancer Research (J.C.). We would like to thank Richard Losick of the Department of Microbiology and Molecular Genetics at Harvard Medical School for providing us with the full-length *sinR* construct. We would also like to thank Roy R. Hantgan of Wake Forest University School of Medicine for use of the Molecular Partners Core Facility to collect AUC data and for thoughtful discussion of the project. Additionally, we would like to thank the members of the Duke Center for Genomic and Computational Biology at Duke University Medical Center for technical advice with crosslinking experiments and critical analysis of MS/MS data. Finally, we would like to thank AJ Milton for critically reviewing the manuscript.

References

- [1]. Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R, Sticking together: building a biofilm the *Bacillus subtilis* way, *Nat. Rev. Microbiol.* 11 (2013) 157–168. doi:10.1038/nrmicro2960. [PubMed: 23353768]
- [2]. Jefferson KK, What drives bacteria to produce a biofilm?, *FEMS Microbiol. Lett.* 236 (2004) 163–173. doi:10.1016/j.femsle.2004.06.005. [PubMed: 15251193]
- [3]. Lemon KP, Earl AM, Vlamakis HC, Aguilar C, Kolter R, Biofilm development with an emphasis on *Bacillus subtilis*, *Curr. Top. Microbiol. Immunol.* 322 (2008) 1–16. [PubMed: 18453269]
- [4]. Lefevre M, Racedo SM, Ripert G, Housez B, Cazaubiel M, Maudet C, Jüsten P, Marteau P, Urdaci MC, Probiotic strain *Bacillus subtilis* CU1 stimulates immune system of elderly during common infectious disease period: a randomized, double-blind placebo-controlled study, *Immun. Ageing A* 12 (2015). doi:10.1186/s12979-015-0051-y.
- [5]. Ayala FR, Bauman C, Cogliati S, Leñini C, Bartolini M, Grau R, Microbial flora, probiotics, *Bacillus subtilis* and the search for a long and healthy human longevity, *Microb. Cell.* 4 (n.d.) 133–136. doi:10.15698/mic2017.04.569.
- [6]. Duc LH, Hong HA, Barbosa TM, Henriques AO, Cutting SM, Characterization of *Bacillus* Probiotics Available for Human Use, *Appl. Environ. Microbiol.* 70 (2004) 2161–2171. doi:10.1128/AEM.70.4.2161-2171.2004. [PubMed: 15066809]
- [7]. Martin DJH, Denyer SP, McDonnell G, Maillard J-Y, Resistance and cross-resistance to oxidising agents of bacterial isolates from endoscope washer disinfectors, *J. Hosp. Infect.* 69 (2008) 377–383. doi:10.1016/j.jhin.2008.04.010. [PubMed: 18602194]

- [8]. Bridier A, Coq DL, Dubois-Brissonnet F, Thomas V, Aymerich S, Briandet R, The Spatial Architecture of *Bacillus subtilis* Biofilms Deciphered Using a Surface-Associated Model and In Situ Imaging, *PLOS ONE*. 6 (2011) e16177. doi:10.1371/journal.pone.0016177. [PubMed: 21267464]
- [9]. Bridier A, del M Sanchez-Vizueté P, Coq DL, Aymerich S, Meylheuc T, Maillard J-Y, Thomas V, Dubois-Brissonnet F, Briandet R, Biofilms of a *Bacillus subtilis* Hospital Isolate Protect *Staphylococcus aureus* from Biocide Action, *PLOS ONE*. 7 (2012) e44506. doi:10.1371/journal.pone.0044506. [PubMed: 22973457]
- [10]. Mullany P, Wilks M, Lamb I, Clayton C, Wren B, Tabaqchali S, Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*, *Microbiology*. 136 (1990) 1343–1349. doi:10.1099/00221287-136-7-1343.
- [11]. Lopez D, Vlamakis H, Kolter R, Generation of multiple cell types in *Bacillus subtilis*, *FEMS Microbiol. Rev.* 33 (2009) 152–163. doi:10.1111/j.1574-6976.2008.00148.x. [PubMed: 19054118]
- [12]. Cairns LS, Hobley L, Stanley-Wall NR, Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms, *Mol. Microbiol.* 93 (2014) 587–598. doi:10.1111/mmi.12697. [PubMed: 24988880]
- [13]. López D, Kolter R, Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*, *FEMS Microbiol. Rev.* 34 (2010) 134–149. doi:10.1111/j.1574-6976.2009.00199.x. [PubMed: 20030732]
- [14]. Chu F, Kearns DB, Branda SS, Kolter R, Losick R, Targets of the master regulator of biofilm formation in *Bacillus subtilis*, *Mol. Microbiol.* 59 (2006) 1216–1228. doi:10.1111/j.1365-2958.2005.05019.x. [PubMed: 16430695]
- [15]. Kearns DB, Chu F, Branda SS, Kolter R, Losick R, A master regulator for biofilm formation by *Bacillus subtilis*, *Mol. Microbiol.* 55 (2005) 739–749. doi:10.1111/j.1365-2958.2004.04440.x. [PubMed: 15661000]
- [16]. Chai Y, Kolter R, Losick R, Paralogous antirepressors acting on the master regulator for biofilm formation in *Bacillus subtilis*: Paralogous antirepressors acting on the master regulator, *Mol. Microbiol.* 74 (2009) 876–887. doi:10.1111/j.1365-2958.2009.06900.x. [PubMed: 19788541]
- [17]. Gaur NK, Dubnau E, Smith I, Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies, *J. Bacteriol.* 168 (1986) 860–869. [PubMed: 3096962]
- [18]. Mandic-Mulec I, Gaur N, Bai U, Smith I, Sin, a stage-specific repressor of cellular differentiation., *J. Bacteriol.* 174 (1992) 3561–3569. [PubMed: 1592811]
- [19]. Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R, Fruiting body formation by *Bacillus subtilis*, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 11621–11626. doi:10.1073/pnas.191384198. [PubMed: 11572999]
- [20]. Bai U, Mandic-Mulec I, Smith I, SinI modulates the activity of SinR, a developmental switch protein of *Bacillus subtilis*, by protein-protein interaction, *Genes Dev.* 7 (1993) 139–148. [PubMed: 8422983]
- [21]. Gaur NK, Cabane K, Smith I, Structure and expression of the *Bacillus subtilis* sin operon., *J. Bacteriol.* 170 (1988) 1046–1053. [PubMed: 3125149]
- [22]. Gaur NK, Oppenheim J, Smith I, The *Bacillus subtilis* sin gene, a regulator of alternate developmental processes, codes for a DNA-binding protein, *J. Bacteriol.* 173 (1991) 678–686. [PubMed: 1898931]
- [23]. Scott DJ, Leejeerajumnean S, Brannigan JA, Lewis RJ, Wilkinson AJ, Hoggett JG, Quaternary re-arrangement analysed by spectral enhancement: the interaction of a sporulation repressor with its antagonist, *J. Mol. Biol.* 293 (1999) 997–1004. doi:10.1006/jmbi.1999.3221. [PubMed: 10547280]
- [24]. Colledge VL, Fogg MJ, Levdikov VM, Leech A, Dodson EJ, Wilkinson AJ, Structure and Organisation of SinR, the Master Regulator of Biofilm Formation in *Bacillus subtilis*, *J. Mol. Biol.* 411 (2011) 597–613. doi:10.1016/j.jmb.2011.06.004. [PubMed: 21708175]
- [25]. Lewis RJ, Brannigan JA, Offen WA, Smith I, Wilkinson AJ, An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex, *J. Mol. Biol.* 283 (1998) 907–912. doi:10.1006/jmbi.1998.2163. [PubMed: 9799632]

- [26]. Newman JA, Rodrigues C, Lewis RJ, Molecular basis of the activity of SinR protein, the master regulator of biofilm formation in *Bacillus subtilis*, *J. Biol. Chem.* 288 (2013) 10766–10778. doi:10.1074/jbc.M113.455592. [PubMed: 23430750]
- [27]. Olson AL, Liu F, Tucker AT, Goshe MB, Cavanagh J, Chemical Crosslinking and LC/MS Analysis to Determine Protein Domain Orientation: Application to AbrB, *Biochem. Biophys. Res. Commun.* 431 (2013) 253–257. doi:10.1016/j.bbrc.2012.12.124. [PubMed: 23313475]
- [28]. Logan Draughn G, Milton ME, Feldmann EA, Bobay BG, Roth BM, Olson AL, Thompson RJ, Actis LA, Davies C, Cavanagh J, The structure of the biofilm-controlling response regulator BfmR from *Acinetobacter baumannii* reveals details of its DNA-binding mechanism, *J. Mol. Biol.* (n.d.) doi:10.1016/j.jmb.2018.02.002.
- [29]. Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R, Fruiting body formation by *Bacillus subtilis*, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 11621–11626. doi:10.1073/pnas.191384198. [PubMed: 11572999]
- [30]. Hamon MA, Lazazzera BA, The sporulation transcription factor Spo0A is required for biofilm development in *Bacillus subtilis*, *Mol. Microbiol.* 42 (2001) 1199–1209. [PubMed: 11886552]
- [31]. Hamon MA, Stanley NR, Britton RA, Grossman AD, Lazazzera BA, Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*, *Mol. Microbiol.* 52 (2004) 847–860. doi:10.1111/j.1365-2958.2004.04023.x. [PubMed: 15101989]
- [32]. Winkelman JT, Bree AC, Bate AR, Eichenberger P, Gourse RL, Kearns DB, RemA is a DNA-binding protein that activates biofilm matrix gene expression in *Bacillus subtilis*, *Mol. Microbiol.* 88 (2013). doi:10.1111/mmi.12235.
- [33]. Stowe SD, Olson AL, Losick R, Cavanagh J, Chemical shift assignments and secondary structure prediction of the master biofilm regulator, SinR, from *Bacillus subtilis*, *Biomol. NMR Assign* 8 (2014) 155–158. doi:10.1007/s12104-013-9473-7. [PubMed: 23475644]
- [34]. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR.* 6 (1995) 277–293. [PubMed: 8520220]
- [35]. Johnson BA, Blevins RA, NMR View: A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR.* 4 (1994) 603–614. doi:10.1007/BF00404272. [PubMed: 22911360]
- [36]. Shen Y, Delaglio F, Cornilescu G, Bax A, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR.* 44 (2009) 213–223. doi:10.1007/s10858-009-9333-z. [PubMed: 19548092]
- [37]. Wishart DS, Sykes BD, The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data, *J. Biomol. NMR.* 4 (1994) 171–180. [PubMed: 8019132]
- [38]. Güntert P, Mumenthaler C, Wüthrich K, Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol.* 273 (1997) 283–298. doi:10.1006/jmbi.1997.1284. [PubMed: 9367762]
- [39]. Güntert P, Braun W, Wüthrich K, Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA, *J. Mol. Biol.* 217 (1991) 517–530. [PubMed: 1847217]
- [40]. Herrmann T, Güntert P, Wüthrich K, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J. Mol. Biol.* 319 (2002) 209–227. [PubMed: 12051947]
- [41]. Case DA, Betz R, Cerutti DS, Cheatham T, Darden T, Duke R, Giese TJ, Gohlke H, Götz A, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee T-S, LeGrand S, Li P, Lin C, Luchko T, Kollman PA, Amber 2016, University of California, San Francisco., 2016. doi:10.13140/RG.2.2.27958.70729.
- [42]. Wilm M, Shevchenko A, Houthaave T, Breit S, Schweigerer L, Fotsis T, Mann M, Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry, *Nature.* 379 (1996) 466–469. doi:10.1038/379466a0. [PubMed: 8559255]

- [43]. Keller A, Nesvizhskii AI, Kolker E, Aebersold R, Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74 (2002) 5383–5392. [PubMed: 12403597]
- [44]. Nesvizhskii AI, Keller A, Kolker E, Aebersold R, A statistical model for identifying proteins by tandem mass spectrometry, *Anal. Chem.* 75 (2003) 4646–4658. [PubMed: 14632076]
- [45]. Xu H, Freitas MA, A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data, *BMC Bioinformatics.* 8 (2007) 133. doi:10.1186/1471-2105-8-133. [PubMed: 17448237]
- [46]. Dominguez C, Boelens R, Bonvin AMJJ, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, *J. Am. Chem. Soc.* 125 (2003) 1731–1737. doi:10.1021/ja026939x. [PubMed: 12580598]
- [47]. Bhattacharya A, Tejero R, Montelione GT, Evaluating protein structures determined by structural genomics consortia, *Proteins.* 66 (2007) 778–795. doi:10.1002/prot.21165. [PubMed: 17186527]

Highlights

- Full-length structural model of tetrameric SinR and dimeric SinI
- SinI inhibits SinR tetramerization by blocking multimerization
- SinR bends DNA to inhibit RNA polymerase

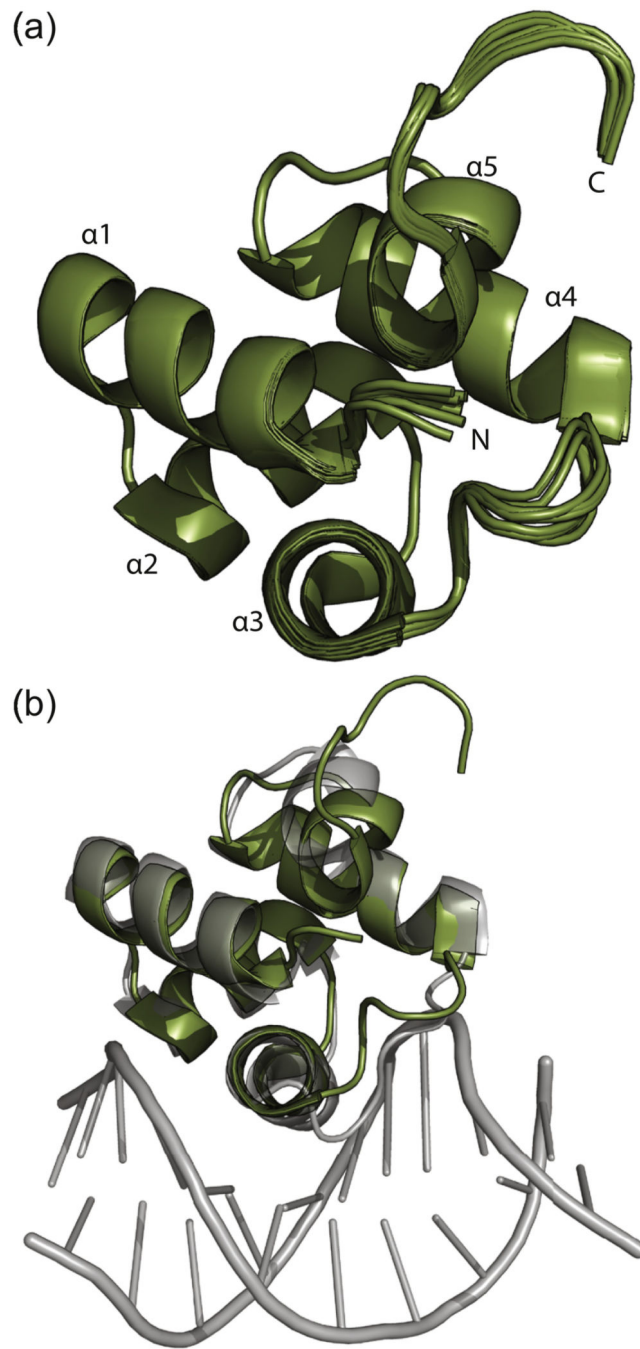


Fig. 1. Solution NMR structure of SinRN, the DNA-binding domain of SinR. (a) NMR structure ensemble of SinRN with helices labeled. Ensembles contain the ten lowest energy models as a result of SinRN structure calculations. (b) SinRN (green) overlaid with DNA bound crystal structure (gray) shows the nearly identical conformation.

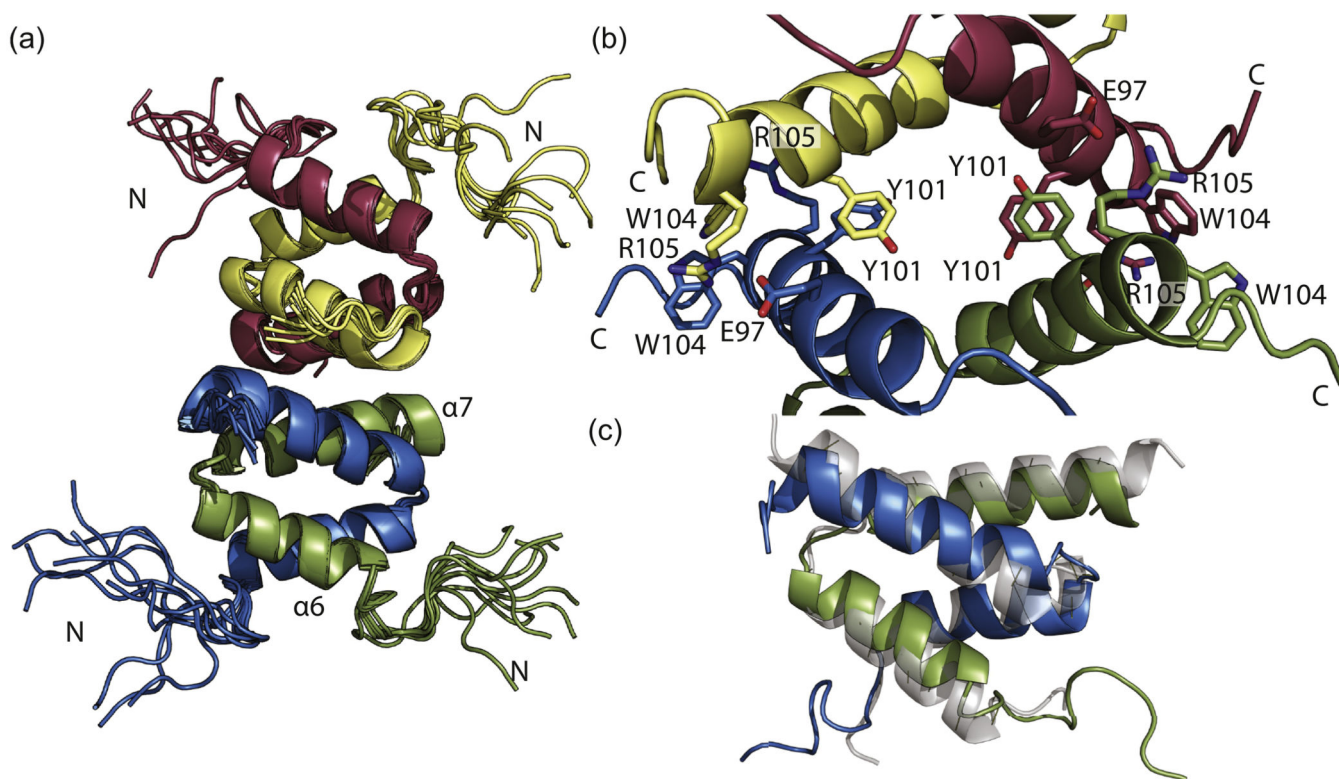


Fig. 2. Solution NMR structure of SinRC. (a) NMR structure ensemble of SinRC with helices of one monomer (green) labeled. Ensembles contain the ten lowest energy models as a result of SinRC structure calculations. (b) Closeup of tetramer interface with each monomer. Key residues involved in tetramer formation are labeled. (c) SinRC (green and blue) overlaid with PDB ID 2YAL shows slight difference in the angle of dimer helices.

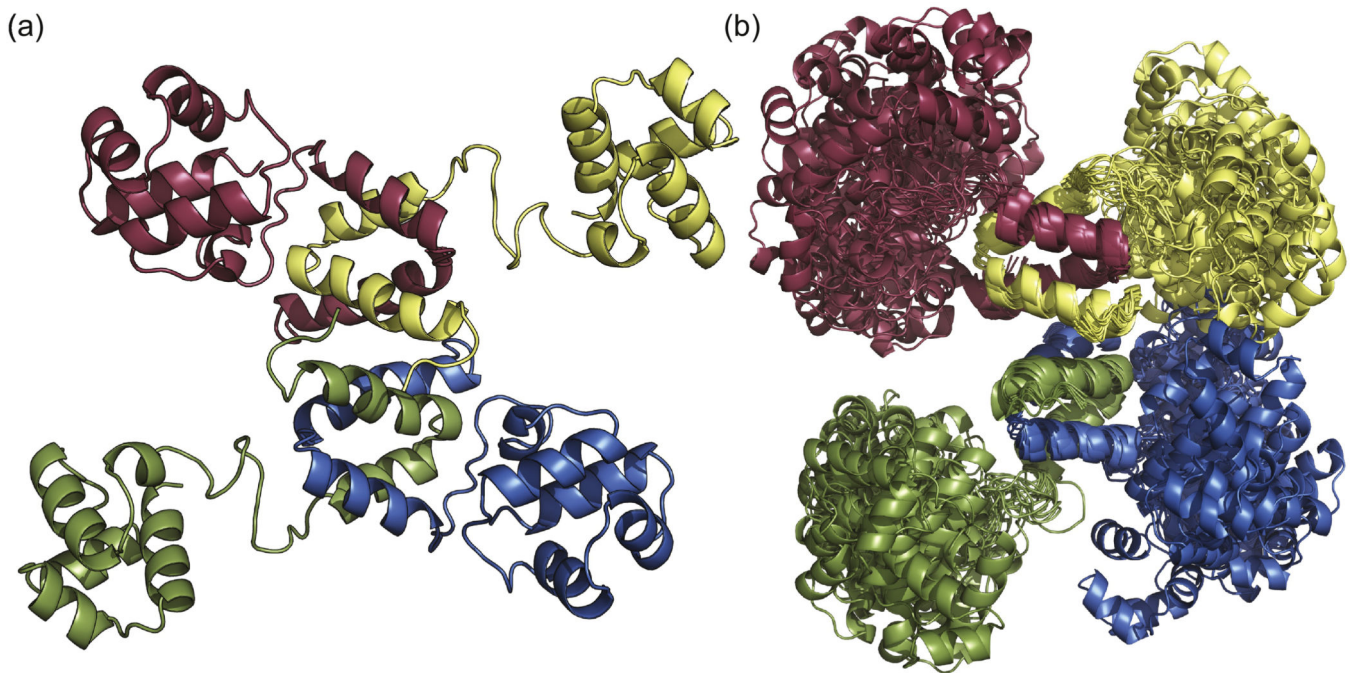


Fig. 3. Structure of full-length SinR tetramer. (a) Representative model of SinR derived using a hybrid approach of NMR, chemical crosslinking, mass spectrometry, and molecular docking. (b) Ensemble of 20 molecular dynamics solutions from the 200 lowest energy cluster HADDOCK results. Note the relative stability of the SinRC tetramerization core compared to the conformational freedom within the SinRN DNA-binding domains.

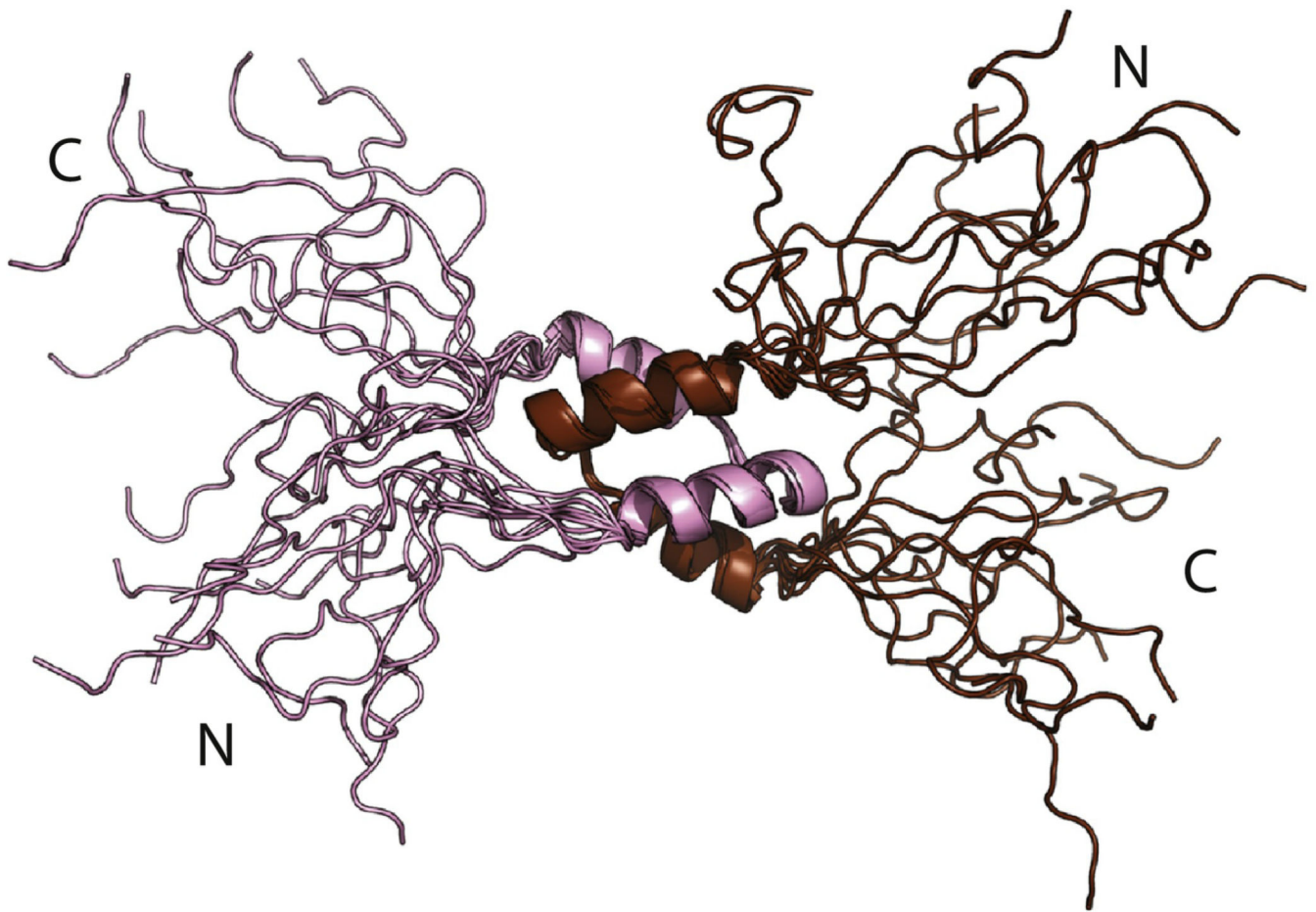


Fig. 4. Solution NMR structure of SinI. (a) NMR structure ensemble of dimeric SinI with one monomer in pink and the second in brown. Ensemble contains the ten lowest energy models as a result of SinI structure calculations. (b) Model of SinR-SinI complex using the full-length SinR (green) and SinI (pink) structures. Model was generated by superimposing the two structures on the original SinR-SinI crystal structure, PDB ID 1B0N [25].

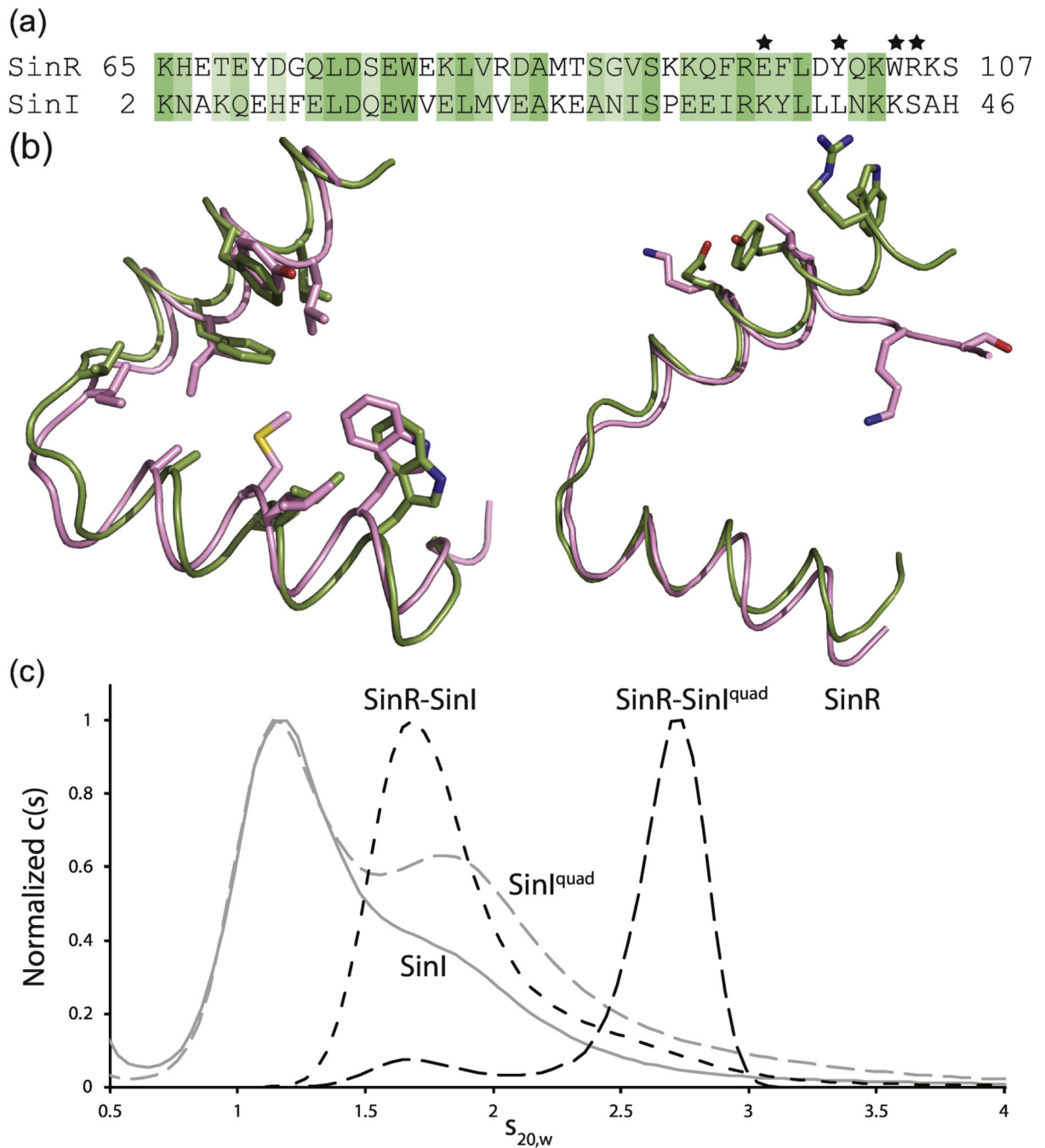


Fig. 5. Examination of the SinR-SinI interaction. (a) Sequence alignment of the helical hook regions of SinR and SinI. Residues are colored based on sequence similarity from Clustal Omega. Stars denote the residues involved in formation of the SinR tetramer interface (Fig. 2b). These residues in SinI were mutated to the corresponding residue of SinR to generate the SinI^{quad} mutant. (b) Structural alignment of SinR (green) and SinI (pink) helical hooks. Left, conserved residues line the dimerization interface. Right, SinR residues necessary for tetramer formation are significantly different in SinI. (c) AUC analysis of SinR (solid black

line), SinI (solid grey line), SinI^{quad} (grey dashed line), SinR-SinI complex (black dot-dashed line), and SinR-SinI^{quad} (black dashed line).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

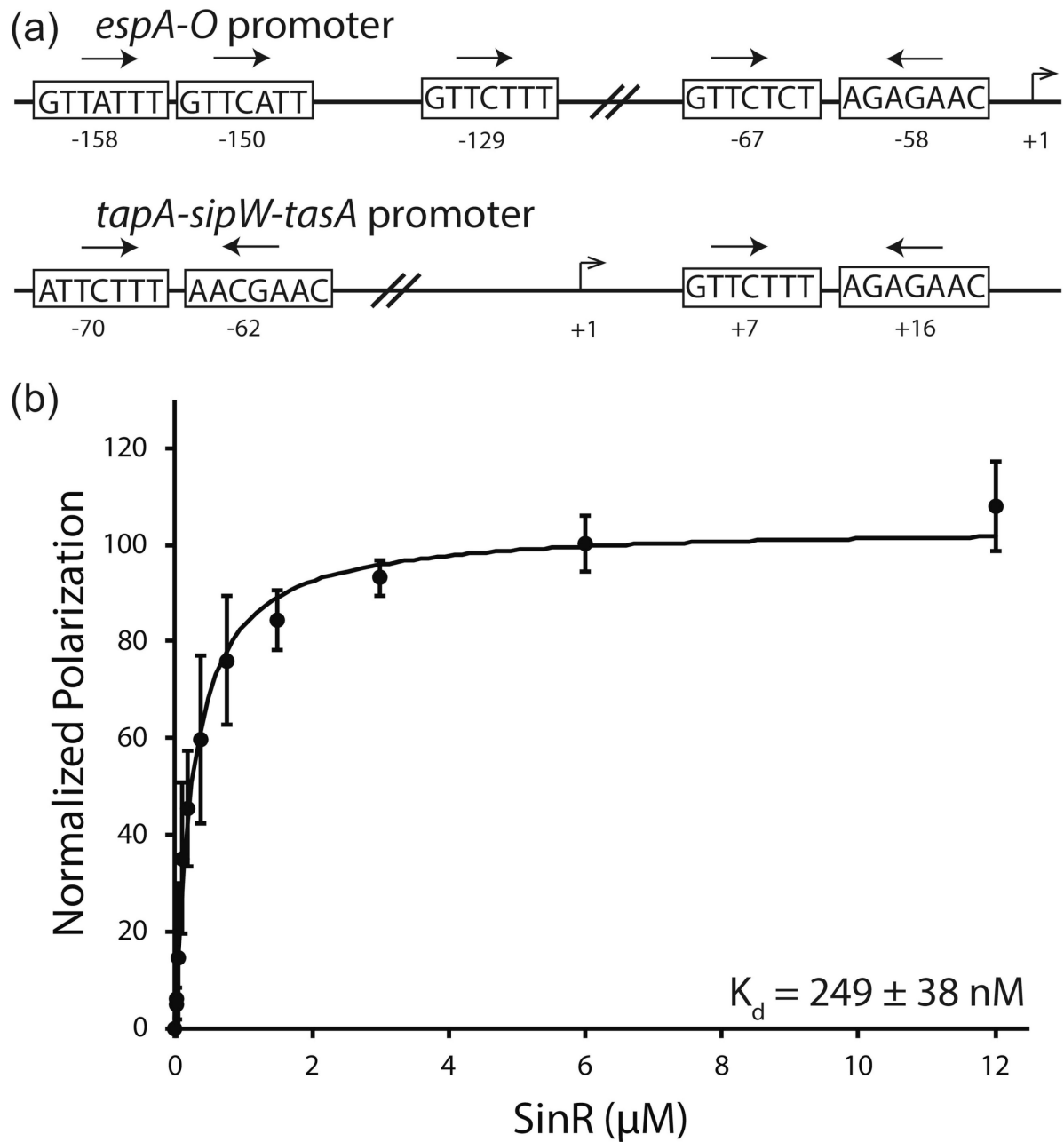
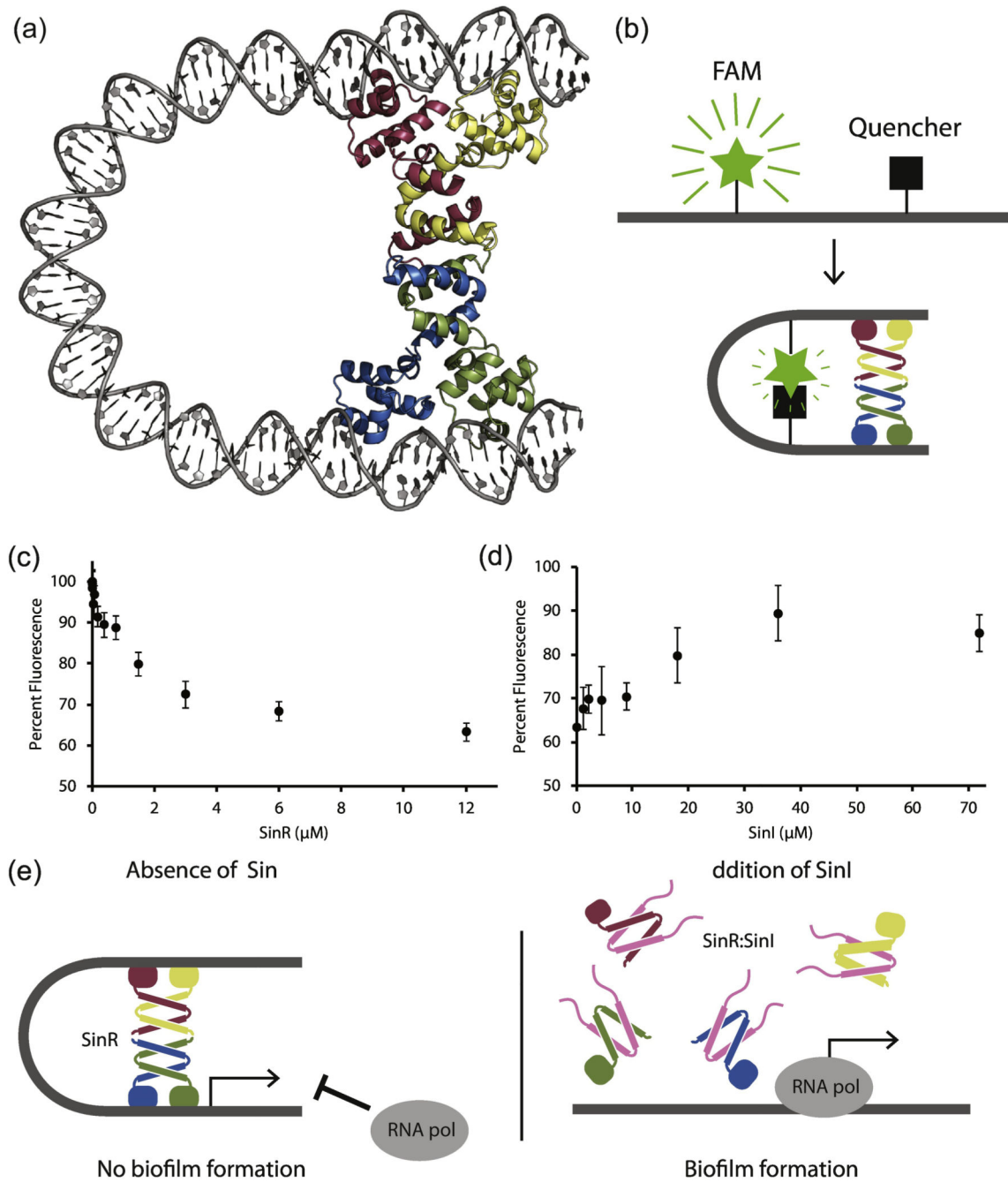


Fig. 6. Examination of the SinR-DNA interaction. (a) Arrangement of the SinR binding motifs in *epsA-O* and *tapA-sipW-tasA* promoters of *B. subtilis*. SinR binding sequences are shown in boxes. Arrows denote the orientation of the binding sequence. The number of base pairs between binding sequences is noted between the boxes. Transcription start site is labeled at +1 with relative locations of binding sites labeled below each box. (b) Fluorescent anisotropy analysis of SinR binding to a 6-FAM 5'-labelled DNA substrate based on the +7 and +16 *tapA-sipW-tasA* promoter sequences.

**Fig. 7.**

SinR bends DNA. (a) Fluorescently labeled substrate based on the *tapA-sipW-tasA* promoter. 6-FAM (green star) gives off a fluorescent signal when DNA is linear. The bending of DNA by SinR brings the 6-FAM into proximity to the Black Hole Quencher-1 (black square) and reduces the fluorescent signal. (b) Model of the SinR tetramer bound to the *tapA-sipW-tasA* promoter. (c) SinR bends the DNA. Fluorescent signal decreases with increasing concentrations of SinR. Assay was repeated in triplicate with three technical replicates per experiment. (d) SinI inhibits SinR from bending the DNA. Incubating SinR with SinI in increasing concentrations prevents the loss of fluorescent signal. Assay was

repeated in triplicate with three technical replicates per experiment. (e) Proposed mechanism for gene regulation by SinR and SinI using the DNA-bending model. SinR bends DNA, blocking RNA polymerase from binding to the promoter. The introduction of SinI forms SinR-SinI heterodimers that cannot effectively bind and bend the DNA. Released from the physical restraint, the DNA is now exposed to RNA polymerase which can initiate transcription.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript