Published in final edited form as:

Biochemistry. 2017 February 21; 56(7): 932–943. doi:10.1021/acs.biochem.6b01273.

Identification of a ligand-binding site on the *Staphylococcus aureus* DnaG primase C-terminal domain

Jonathan Catazaro[†], Jessica Periago[†], Matthew D. Shortridge, Bradley Worley, Andrew Kirchner, Robert Powers^{*}, and Mark A. Griep^{*}

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304

Abstract

The interface between the DnaG primase C-terminal domain (CTD) and the N-terminal domain of DnaB helicase is essential for bacterial DNA replication because it allows coordinated priming of DNA synthesis at the replication fork while the DNA is being unwound. Since these two proteins are conserved in all bacteria and distinct from those in eukaryotes, their interface is an attractive antibiotic target. To learn more about this interface, we determined the solution structure and dynamics of the DnaG primase CTD from *Staphylococcus aureus*, a medically important bacterial species. It has two subdomains like the primase CTD structures from other bacteria in which the first six helices create the C1 subdomain and the last two helices, 7 and 8, create the C2 subdomain. Comparison with the known primase CTD structures shows there are two biologically relevant conformations – an open conformation that likely binds to DnaB helicase and a closed conformation that does not. The *S. aureus* primase CTD is in the closed conformation, but NMR dynamic studies indicates there is considerable movement in the linker between the two subdomains and that N564 is the most dynamic residue within the linker. A high-throughput NMR ligand-affinity screen identified potential binding compounds, among which were acycloguanosine and myricetin. Although the affinity for these compounds and adenosine was in the millimolar range, all three bind to a common pocket that is only present on the closed conformation of CTD. This binding pocket is at the opposite end of helices 6 and 7 from N564, the key hinge residue. The identification of this binding pocket should allow the development of stronger-binding ligands that can prevent formation of the CTD open conformation that binds to DnaB helicase.

Graphical Abstract

^{*}Co-corresponding authors: Robert Powers (722 Hamilton Hall; 402-472-3039; rpowers3@unl.edu), Mark A. Griep (736 Hamilton Hall; 402-472-3429; mgriep1@unl.edu), University of Nebraska-Lincoln, Department of Chemistry, Lincoln, NE 68588-0304.

[†]Equal contribution

Supporting Information.

This work is accompanied by Supplementary Online Material containing: 1) sequence alignment for primase from eight bacteria; 2) pairwise structure comparison root mean square deviation (RMSD) matrices for the eight primase C-terminal domain structures; 3) list of the residues in the disallowed regions of the Ramachandran plots for the 20 NMR models; and 4) alignment of *G. stearothermophilus* and *S. aureus* DnaC helicase N-terminal domains.

Keywords

DnaG primase; DnaB helicase; Acycloguanosine; Adenosine; Myricetin; NMR Structure; Structure Dynamics

The initiation of DNA synthesis in living organisms and many large viruses results from the coordinated interaction of two enzymes - primase and helicase.^{1, 2} In bacteria, DnaG primase is the specialized DNA-dependent RNA polymerase that synthesizes short oligoribonucleotide polymers called primers. DNA polymerase, an enzyme that lacks the ability to initiate chain synthesis, elongates the primers. DnaG primase is stimulated by the homohexameric DnaB helicase. (Note: The name DnaB helicase is used here for the protein that is actually named DnaC helicase in *Staphylococcus aureus* and *Bacillus subtilis* because, in all other bacteria, DnaB is the name of the replicative helicase and DnaC is the name of the helicase-loading enzyme). DnaB helicase dissociates the two strands of duplex DNA during DNA replication while hydrolyzing ATP, and travels processively in the 5'–3' direction along the single-stranded lagging template toward the replication fork, communicating allosterically with the multi-subunit replicative DNA polymerase. This action keeps it in proximity of the replication fork and ensures the primers are synthesized on the exposed single-stranded DNA nearest to the replication fork. Since primase activity is weak, the stimulation by DnaB causes primase to synthesize primers only at the replication fork when and where they are needed.^{3, 4} Finally, the DnaG-DnaB interaction is an attractive antibiotic target because it is conserved in bacteria, essential for DNA replication, and is distinctly different from that of viruses, archaea, and eukaryotes.^{5–7}

DnaG primase is composed of three functional domains: the N-terminal zinc-binding domain (ZBD) responsible for DNA binding specificity; the RNA polymerase domain (RPD) responsible for enzymatic activity; and the C-terminal domain (CTD) responsible for the interaction with the DnaB helicase.^{8, 9} The 110-residue ZBD contains the highest percentage of conserved residues and structural conservation, whereas the 140-residue CTD has the least sequence conservation and the most structural diversity (Table S1). Of these domains, the ZBD is unique to bacterial primases, the magnesium-binding residues of the

RPD shares some similarity to the magnesium-binding residues of topoisomerases and some DNA-hydrolyzing enzymes, while the CTD is similar only to the N-terminal domain of DnaB helicases, the domain to which it binds.^{10–13} Currently, no high-resolution structure has been solved for a full, intact bacterial primase, but structures are available for all three individual domains. The primase CTD is the domain with the most experimental structures.^{6, 14–17} There are X-ray crystal structures and NMR solution structures (Table 1) from *S. aureus* (2LZN), *E. coli* (2HAJ, 1T3W), *G. stearothermophilus* (1Z8S, 2R6A), and *H. pylori* (4EHS). In addition, the ZBD-RPD domain pair from *Aquifex aeolicus* has been experimentally determined.¹⁰ It has been proposed that the conformations of the two *E. coli* structures in 1T3W may not be biologically relevant due to crystallization conditions and packing effects.¹⁶

To develop the DnaG-DnaB interface as an antibiotic target, we determined the solution structure of the DnaG CTD from *S. aureus*. This organism was chosen because there is an urgent need for new antibiotic targets due to the rapid rise in antibiotic resistance.¹⁸ A comparison to known structures from other bacteria indicated that the *S. aureus* CTD was composed of two subdomains. The CTD from *Firmicutes* and *Proteobacteria* may adopt similar conformations when bound to helicase, but may adopt different free conformations. A study of the structural dynamics of the free *S. aureus* CTD confirmed that the junction between the two subdomains is the focal point for the greatest movement. Also, one of the subdomains is significantly more ordered than the other. The structure was used to identify three small compounds that bound to the close conformation, which is likely to reduce the primase-helicase interaction.

Experimental Procedures

Chemicals.

The dimethyl sulfoxide-d₆ (99.9% D) and deuterium oxide (99.9% D) were obtained from Aldrich (Milwaukee, WI). The 3-(trimethylsilyl)propionic-2,2,3,3-d₄ acid sodium salt (98% D) was purchased from Cambridge Isotopes (Andover, MA). The potassium phosphate dibasic salt (anhydrous, 99.1% pure) and monobasic salt (crystal, 99.8% pure) were purchased from Mallinckrodt (Phillipsburg, NJ). All compounds used for screening were obtained as described earlier.¹⁹ Briefly, the compound library is composed of 437 known biologically active compounds distributed across 113 mixtures with 3–4 compounds in each mixture.

Proteins.

Uniformly ¹³C, ¹⁵N-labeled *S. aureus* DnaG primase CTD and uniformly ¹⁵N-labeled *S. aureus* DnaG primase CTD were designed and purified by Nature Technologies (Lincoln, NE). The details of protein expression and purification were previously described.²⁰ The vector added 20 N-terminal residues: an HN metal affinity tag (MGHNHNHNHNHNGG) followed by a protease-sensitive DDDD sequence.²⁰

NMR data analysis, structure calculations and refinement.

The NMR spectra for obtaining the protein backbone assignments were collected at 298 K on a five-channel 600-MHz Bruker Avance spectrometer equipped with a 5-mm TXI probe. The NMR spectra for the protein side chain assignments were collected at the Rocky Mountain Regional 900-MHz NMR Facility on a four-channel 900-MHz Varian INOVA spectrometer equipped with a 5-mm HCN probe. The nearly complete *S. aureus* DnaG primase CTD NMR resonance assignments have been previously reported.²⁰ Distance constraints were obtained from 3D ¹⁵N-edited NOESY and 3D ¹³C-edited NOESY that were collected at 900 MHz.²¹ Hydrogen bond constraints were determined using the (CLEANEX-PM)-FHSQC experiment.²² All torsion angle constraints were obtained by chemical shift analysis using the TALOS²³ software program, and measured coupling constants from an HNHA experiment.²⁴

For the backbone NMR assignment experiments, ¹³C,¹⁵N-labeled protein was concentrated to 1.2 mM in a 95% H₂O/5% D₂O buffered solution of 100 mM NaCl, 25 mM potassium phosphate at pH 6.64 (uncorrected) using an Amicon ultra centricon (MW cutoff 10,000 Da). Long-term protein stability was enhanced by adding 50 mM arginine and 50 mM glutamine. For the side chain experiments, ¹³C,¹⁵N-labeled protein was concentrated to 1.4 mM in the same buffer. All multidimensional experiments were processed using NMRPipe²⁵ and analyzed using PIPP²⁶ and CCPNMR.²⁷

Nuclear Overhauser Effect (NOE) assignments were obtained by using 3D ¹⁵N-edited NOESY and 3D ¹³C-edited NOESY. NOE intensities were sorted visually into four classes: strong (1.8–2.5 Å), medium (1.8–3.0 Å), weak (1.8–4.0 Å), and very weak (3.0–5.0 Å). Upper limits for distances involving methyl protons and non-stereospecifically assigned methylene protons were corrected appropriately for center averaging. Initial NOE assignments were completed by the program Autostructure.²⁸ Despite the high magnetic field of 900 MHz, the extent of cross-peak overlap was great enough to warrant manual refinement of the NOE assignments.

Hydrogen bond constraints were determined using the (CLEANEX-PM)-FHSQC experiment, which identifies amide residues with fast water exchange rates.²² Hydrogen bond constraints were assigned to amides in secondary structure regions for any 2D ¹H-¹⁵N HSQC peaks lacking a corresponding peak in the (CLEANEX-PM)-FHSQC spectrum. The carboxyl oxygen to amide nitrogen hydrogen bond distances were set at 2.8 Å while the carbonyl oxygen to amide proton distances were set at 1.8 Å. All carboxyl groups within 2.5 Å of slowly exchanging amide groups were constrained to be involved in a hydrogen bond.

The structures were refined using the hybrid distance geometry dynamic-simulated annealing method²⁹ using the XPLOR-NIH program^{30, 31} adapted to incorporate pseudopotentials for ³J(HN-Hα) coupling constants, secondary ¹³Cα/¹³Cβ chemical shift constraints, and a conformational database potential.^{32–36} The force constant for the conformational database was kept relatively low (0.5–1.0 kcal/mol) throughout the simulation to allow the experimental distance and torsion angle constraints to predominately influence the resulting structures. The force constant for the NOE and dihedral constraints were 30 times and 10 times stronger than the force constants used for the conformational

database.³⁷ All peptide bonds were constrained to be planar and trans. There were no hydrogen-bonding or electrostatic empirical potential energy terms in the target function. A total of 1000 structures were calculated and the 20 lowest energy structures were selected to become the ensemble of *S. aureus* DnaG primase CTD structures. Each of these twenty structures were subjected to further energy minimization using explicit water with Crystallography and NMR system (CNS) version 1.2.³⁸ Explicit water solvation was accounted for with Lennard-Jones and electrostatic potentials using a modification of the procedure and force field of Nilges.^{39, 40}

Protein backbone dynamics.

The NMR experiments for protein dynamics analysis were collected on a Bruker 500 MHz Avance spectrometer (Billerica, MA) equipped with a triple resonance, Z-axis gradient cryoprobe. The sample was uniformly ¹⁵N-labeled *S. aureus* DnaG primase CTD concentrated to 1.2 mM in a 95% H₂O/5% D₂O buffered solution of 100 mM NaCl, 50 mM arginine, 50 mM glutamine, 25 mM potassium phosphate, pH 6.64 (uncorrected) using an Amicon ultra centricon (MW cutoff 10 000 Da).

The previously described^{41–43} experiments included a 2D ¹⁵N-¹H HSQC experiment (*hsqct1etf3gpsi*) designed to measure T₁ relaxation rates with delay times of 0.0 ms, 5.39 ms, 53.92 ms, 134.80 ms, 269.60 ms, 404.40 ms, 539.20 ms, 674.00 ms and 1078.40 ms, a 2D ¹⁵N-¹H HSQC experiment (*hsqct2etf3gpsi*) designed to measure T₂ relaxation rates with delay times of 0.0 ms, 17.6 ms, 35.2 ms, 52.8 ms, 70.4 ms, 105.6 ms, 123.2 ms, 140.8 ms, 158.4 ms, 176.0 ms, and a 2D ¹⁵N-¹H HSQC experiment (*hsqcnof3gpsi*) designed to measure NOE changes.

The relaxation rates (T₁, T₂) for each amino acid within the structure were calculated by fitting the intensity of each peak to the intensity decay curve (eq 1) where I_t is the intensity of each peak at the delay time *t*, I₀ is the initial steady state intensity.

$$I_t = I_0 \exp\left(-\frac{t}{T_{1,2}}\right) \quad \text{eq 1}$$

The NOE values were determined by the ratio of peak intensity between the saturated (I_{sat}) and unsaturated (I_{unsat}) spectra.

$$NOE = I_{sat}/I_{unsat} \quad \text{eq 2}$$

The T₁ rates, T₂ rates, and NOE ratios were determined from their fits to their respective equations using Kaleidagraph (Synergy Software, Reading, PA) and exported to the FAST-MODEL FREE program⁴⁴ to calculate an overall correlation time (τ_m); and per residue order parameters (S²), internal motion (τ_e), and chemical exchange (R_{ex}) using the Lipari-Szabo model-free method.⁴⁵ Some of the NOE ratios were above the theoretical maximum of 1.0 and were set to 1.0 to run the FAST-MODEL FREE program.

Primase CTD and C1 subdomain similarity phylograms.

An all-versus-all root-mean-square deviation (RMSD) matrix table (Table S2) was generated using the MatchMaker tool in UCSF Chimera.⁴⁶ All chains from X-ray structures were included in the structural comparison. Only the best representative structure from the NMR ensemble, identified by XPLOR as having the lowest RMSD relative to the mean, was included in the comparison (See Table S3).

The computed RMSD matrices were then inserted into the Splitstree4⁴⁷ program to generate the unrooted phylogenetic trees depicted in Figure 3. The trees were generated using the Neighbor-join algorithm. The pairwise RMSDs were calculated between the complete C-terminal domains (Figure 3A) as well as between the aligned C1 subdomains (Figure 3C).

Primase CTD sequence similarity.

The sequences for the *E. coli*, *G. stearothermophilus*, *H. pylori*, and *S. aureus* primase C-terminal domains were acquired from the Protein Data Bank.⁴⁸ Three different sequence alignment programs, Muscle, T-Coffee, and Clustal omega,^{49–51} were used to determine a sequence similarity between the four primase CTD proteins. The results obtained from the three sequence alignment programs were identical and were converted into the Phylip format for analysis using Splitstree4.⁴⁷ The unrooted phylogenetic tree (Figure 3B) was created using the Neighbor-join algorithm and equal angle method.

NMR ligand affinity assays.

Sample preparation and experimental parameters for the NMR ligand affinity screen were performed as described previously.⁵² Briefly, each ligand mixture (113 total) was screened using a 1D ¹H NMR spectrum with excitation sculpting.⁵³ Each NMR sample contained 100 μM ligand and 25 μM protein in a 99.99% D₂O buffered solution of 20 mM d₁₉-bis-Tris at pH 7.0 (uncorrected) with 2% DMSO-d₆ to maintain ligand solubility and 11.1 μM 3-(trimethylsilyl)propionic-2,2,3,3-d₄ acid sodium salt as a chemical shift reference. All 1D ¹H NMR spectra were processed with the ACD/1D NMR manager v. 12.0 (Advanced Chemistry Development, Inc., Toronto, Ontario). Each 1D ¹H NMR spectrum was compared to the corresponding free ligand mixture reference spectrum and visually analyzed to identify binding ligands based on a decrease in the intensity of ligand NMR resonances. A 2D ¹H-¹⁵N HSQC spectrum was collected for each ligand with a positive response from the 1D ¹H NMR line-broadening screen using 500 μM of ligand and 100 μM of protein under the same buffer conditions as the 1D ¹H NMR screen except for 95% H₂O/5% D₂O. A single ligand-free 2D ¹H-¹⁵N HSQC spectrum was collected as a reference, where a binding event was confirmed based on a clustering of surface residues that incurred a chemical shift change one standard deviation above the average.

Protein-ligand titration and docking.

DnaG primase CTD HSQC was collected at a protein concentration of 1 mM and was titrated with adenosine (0.9, 1.8, 2.6, 3.3, and 4.0 mM), acycloguanosine (0.7, 1.9, 3.5, and 5.2 mM), and myricetin (0.9, 1.8, 2.6, 3.3, and 4.0 mM). All NMR data was collected at 25 °C on a 700 MHz Bruker Avance III spectrometer equipped with a 5 mm QCI-P probe with cryogenically cooled carbon and proton channels. The 2D ¹H-¹⁵N HSQC protein

titration data was processed and analyzed in NMRPipe.²⁵ Briefly, the spectral titration data for each ligand was automatically peak picked and adjusted manually within NMRDraw. The ^1H - ^{15}N and ^{15}N chemical shifts from the selected peaks were then fit to a K_a (1/Kd) according to a previously published method.⁵⁴ Subsequently, the peaks that showed significant binding were manually assigned to residues. AutoDock was then used to individually dock the three ligands into a primase CTD binding site defined by the observed chemical shift perturbations. Chimera was used to visualize the lowest energy docked complexes.⁴⁶

Results

To fight antibiotic-resistant infections, there is a need to identify new antibiotic targets. The bacterial primase C-terminal domain is an especially promising target because eukaryotic primases do not have a cognate sequence or structure and because its critical function is to interact with DnaB helicase to limit DNA replication to the replication fork. Among a number of factors, evaluating the “drugability” of a protein target requires a fundamental understanding of its structure-function relationship.⁵⁵ As a first step towards this goal, we determined the three-dimensional structure of the primase CTD from the pathogen *S. aureus*.

S. aureus primase CTD structure determination.

Backbone and side chain resonance assignments were completed using standard triple resonance NMR experiments.²⁰ In summary, the backbone resonance assignment was 93% complete with 133 amino acids out of the 143 primase residues assigned unambiguously in the 2D ^1H - ^{15}N HSQC spectrum. Nearly all the peaks were uniformly-shaped and separated from the others, which indicates that each residue is in a unique environment and that *S. aureus* primase CTD exists in a single conformation. Unassigned residues include M1-H13, D19, E470, H479, L480, M481, T500, R536, E537, E543, P551, and Y552. Primase CTD residues are numbered relative to the complete DnaG primase sequence, while the N-terminal purification tag is simply numbered from residue 1 to 20. Most of the unassigned residues were located in the purification tag or within highly solvent exposed regions. The structure consists of eight α -helices. Residues H479, L480, and M481 were in a turn region between helix 1 and 2. Residue T500 was in an unstructured loop region between helix 2 and helix 3, whereas residues E543, P551 and Y552 were in an unstructured loop region between helix 5 and helix 6. The only unassigned residues that appear to be localized within a secondary structure were E470, R536, E537, and E543. E470 is the second residue of helix 1, while R536 and E537 are in the middle of helix 5, and E543 is near the end of helix 5. An exhaustive analysis of the NMR data set was unable to yield assignments for these residues, suggesting the end of helices 1 and 5 may undergo partial unfolding leading to chemical shift exchange broadening for these residues.

The solution structure of the *S. aureus* primase CTD was calculated using 1823 distance constraints, 280 dihedral constraints, 256 $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ carbon chemical shift constraints, and 82 $^3J_{\text{NH}\alpha}$ coupling constant constraints, among others (Table 2). In the initial phase, 1000 structures were calculated from 10 individual sets of 100 structures each using XPLOR-NIH^{30, 31} as described previously.⁵⁶ Each set of structures was started from a

randomly generated seed. The lowest energy structures were consolidated to generate a set of 20 low energy structures that were further refined within a virtual water environment using the RECOORD recalculated coordinates⁴⁰ as implemented in CNS.³⁸

The resulting *S. aureus* primase CTD structures (Figure 1) form a self-consistent set as determined by a range of statistical analyses (Table 2). The root mean square deviation (RMSD) for all 1823 experimental distance constraints is $0.040 (\pm 0.023)$ Å, which implies a good agreement between the structural ensemble and the constraints. None of the distance constraints have a violation that exceeds 0.247 Å. There are also low deviations from the constraints for the ¹³Ca chemical shifts, ¹³Cβ chemical shifts, and the ³J(HN-Hα) coupling constants, indicating consistent conformations for both the backbone and side chains. Similarly, the deviations from idealized covalent geometries are consistent with good quality structures.

The quality of the *S. aureus* primase CTD NMR structure was assessed using two suites of programs (Table 3). PROCHECK⁵⁷ indicates the average minimized structure has an overall G-Factor of -0.12 ± 0.04 with no bad contacts, consistent with a good quality structure. The Protein Structure Validation Software suite of programs (PSVS)⁵⁸ gave a Verify3D score of -4.58 ± 0.71 , which is within the typical range. The Molprobit module indicates that the *S. aureus* primase CTD structure has a very good Z-score (-2.50) compared to the average Z-score (-10.74) for all NMR structures in the Protein Data Bank. Finally, a Ramachandran plot of backbone dihedral angles for all non-glycine residues indicates that $83.4 \pm 3.2\%$ lie within the “most favored” region and only $0.7 \pm 0.6\%$ lie within “disallowed” regions. That is, one residue in 14 of the 20 structures is found in the disallowed region but never in a consistent location within the structure (Table S4). These residues tend to have a lower number of constraints, which results in a low penalty for unusual conformations during energy minimization.

The final step in the analysis was to compare the 20 energy-minimized structures with the mean structure (Table 4). Within the C1 subdomain comprising the first two-thirds of the residues, the average root-mean square deviation (RMSD) of the 20 lowest energy structures about the mean coordinate position is 1.54 ± 0.19 Å for backbone atoms and 2.37 ± 0.17 Å for all heavy atoms. When only secondary structure elements are used for alignment, the mean coordinate position deviation is 1.18 ± 0.14 Å for backbone atoms and 1.93 ± 0.16 Å for all heavy atoms. These deviations indicate the backbone structure is well defined. The side chain residues have fewer constraints and correspondingly higher structural disorder.

***S. aureus* primase CTD structure.**

The 20 energy-minimized *S. aureus* primase CTD structures (Figure 1) are deposited in the Protein Database with the identification code 2LZN. The structure is composed of 8 helices arranged into two subdomains as is the primase CTD NMR structure from *G. stearothermophilus*.⁶ In our *S. aureus* primase CTD structure, the first six helices create the C1 subdomain and encompass residues 467 to 565 (Figure 1A). The last two helices, 7 and 8, create the C2 subdomain encompassing residues 572 to 603. When the 20 structures are overlaid using the alignment of the C1 subdomain backbone residues as a guide (Figure 1B), the C2 subdomain structures do not superimpose well. The converse overlay based on the

alignment of the C2 subdomain shows the same effect (Figure 1C). The poor superimposition arises because of the very low number of structural restraints between the subdomains. However, the linker between the subdomains is constrained by 11 NOEs from the C1 subdomain, 77 NOEs within the linker, and 17 NOEs from the C2 subdomain. On the other hand, the loop region between the subdomains lacks sequential NH-NH NOEs in the ^1H - ^{15}N HSQC-edited NOESY and the loop residues G567, Q568 and E569 exhibit exchange peaks in the CLEANEX experiment.²² These residues are likely undergoing rapid exchange with the solvent, a feature indicative of exposed residues that lack protection from hydrogen bonds found in secondary structures. As we shall see below, the dynamics analysis indicates that the subdomains behave as two independent domains.

The surface charge distribution shows the exposed surface is uniformly but weakly electrically positive (Figure 2). The buried interface between the subdomains is electrically negative, as reflected in the moderately large Verify 3D Z-score. This repulsion may destabilize the interaction between the C1 and C2 subdomains, causing them to prefer a more extended conformation. In addition, there are some partially buried electrically negative residues in the C1 subdomain. The overall charge distribution is unlike the uniformly negative CTD from *G. stearothermophilus*,⁶ suggesting that it is not a universal feature of primase CTDs.

Primase CTD conformations.

Every enzyme family has a distinct sequence/structure relationship linked to the evolutionary pressures on the sequence and structure within each organism.⁵⁹ To determine how the *S. aureus* primase CTD structure relates to other known primase CTD structures (Table 1), the MatchMaker tool in UCSF Chimera⁴⁶ was used to calculate pairwise RMSDs for the eight known primase CTD structures (Table S2). The resulting unrooted phylogram (Figure 3A) shows three main branches that correlate with distinct conformations (Figure 4). The first branch (conformation A) includes our NMR structure from *S. aureus* (2LZN-14) and the *G. stearothermophilus* NMR structure (1Z8S-5). The second branch (conformation B) includes the *E. coli* NMR structure (2HAJ-19), the *G. stearothermophilus* crystal structure (2R6A-C), and both *H. pylori* crystal structures (4EHS-A and B). The third branch (conformation C) includes both *E. coli* crystal structures (1T3W-A and B). Conformations A and B are more similar to each other than they are to conformation C. Furthermore, it has been proposed¹⁶ that conformation C may not be biologically relevant because the crystals were formed at low pH (4.6), which resulted in a packing effect of swapped dimer conformation.

Since the unrooted phylogram for the full length structure did not split according to sequence phylogeny, we created unrooted phylograms based on CTD sequence similarity (Figure 3B) and the primase CTD C1 subdomain structure (Figure 3C). From the sequence similarity phylogram, we observed two distinct branches corresponding to the expected phylogenetic split between the *Firmicutes* (*G. stearothermophilus*, *S. aureus*) and the *Proteobacteria* (*E. coli*, *H. pylori*). The same phylogenetic split was observed for the C1 subdomain structures, showing that there is a structure-sequence phylogenetic split for this protein.

Primase CTD flexibility.

An NMR dynamics analysis of the *S. aureus* primase CTD structure was conducted to complement the structural analysis. The NMR relaxation parameters T_1 and T_2 , and the relative ratio of NOE enhancement were measured on a per residue basis and analyzed using the Fast-Model free program⁴⁴ to calculate order parameters (S^2) and chemical exchange (R_{ex}) values (Figure 5). In general, the T_1 values for the C1 subdomain residues were higher than the T_1 values for the C2 subdomain residues, indicating that the two subdomains exhibit different dynamics. A similar observation was made for the T_2 values, except that the T_2 values for the C1 subdomain were lower than the T_2 values for C2.

It is important to note that Fast-Model free did not converge on a result when using the entire primase CTD structure. This is consistent with the observation that the C1 and C2 subdomains behave as two separate domains and are dynamically independent. A convergent result was only obtained when the NMR dynamics data of the C1 and C2 subdomains were modeled separately. The C1 subdomain included residues 463–572 and the C2 subdomain included residues 565–605. The linker region, corresponding to residues 565–572, was included in the dynamics analysis for both subdomains.

The order parameters (S^2) for the C1 subdomain residues covered a large range (0.240–1.000) whereas the C2 subdomain residues were close to 1 (Figure 5D). The average S^2 values were 0.79 ± 0.02 and 0.97 ± 0.03 for the C1 and C2 domains, respectively. In general, an S^2 value approaching 1.0 indicates a highly ordered residue with limited flexibility, while a lower S^2 value indicates greater flexibility. This implies that the C1 subdomain has a greater overall flexibility relative to C2. Within the primase CTD structure there are two stretches of three or more residues with S^2 values less than 0.75: residues 488 to 498; residues 555 to 557 and residues 561 to 563. Residues 488 to 498 lie in helix 2 and the linker between helix 2 and 3. Residues 555 to 557 and 561 to 563 lie in helix 6 (Table S5). The flexibility of the linker between helix 2 and 3 is consistent with the variable lengths for helix 3 observed in the structures from other organisms. It is possible that the reduction in the flexibility of helix 3 is part of an allosteric system that provides an entropic energy change when helix 6 and 8 bind to the helicase NTD.⁶⁰ The flexibility of the second half of helix 6 shows how the unstructured linker between helix 6 and 7 may act as a nucleus for unraveling helix 6. Surprisingly, the linker region corresponding to residues 565–572 was observed to be dynamic, indicating that the point of flexibility between the two domains is located at the end of helix 6.

To identify the hinge between the two subdomains, we searched for a residue with a low order parameter and fast chemical exchange. N564 had one of the lowest order parameters (0.658) and the fastest chemical exchange (17.7 Hz), indicating a high degree of motion on both the ps-ns and ms timescales. It is the final residue of helix 6 and is located between a very rigid C2 subdomain and a flexible C1 domain, potentially implicating N564 as a key residue for a conformational change.

In addition to per-residue order parameters and chemical exchange rates, an overall correlation time (τ_m) was calculated for each subdomain. The correlation times for the C1 and C2 subdomains were 7.2 ns and 6.4 ns, respectively. The predicted value for a hydrated

protein with a molecular weight of 17.2 kDa, corresponding to the intact primase CTD structure, is 7.2 ns, where $\tau_m \approx MW/2400$.⁶¹ The predicted tumbling times were 5.45 ns and 2.1 ns for C1 (13.1 kDa) and C2 (5.0 kDa). The observed τ_m values for both subdomains are substantially higher than predicted and share a greater degree of similarity than the predicted values. This indicates the C1 and C2 subdomains are both tumbling at approximately the same rate as the intact protein. The model that emerges is one in which the two subdomains can adopt a wide range of relative orientations, with N564 as a pivot point, but share an overall tumbling rate.

Identification of ligands that bind to *S. aureus* primase CTD.

A high-throughput NMR ligand affinity screen of the *S. aureus* primase CTD was undertaken to identify potential binding compounds. Out of the 423 compounds tested, a total of 12 compounds were shown to bind *S. aureus* primase CTD using a 1D ¹H NMR line-broadening screen: acycloguanosine, 3-aminopropionitrile fumarate, chelerythrine chloride, didecylidimethylammonium bromide, 5,5-diphenylhydantoin, L-histidine, (±)-α-lipoamide, 1-methylimidazole, mitoxantrone dihydrochloride, myricetin, (±)-propranolol hydrochloride, sodium creatine phosphate, and sodium DL-lactate. Of these compounds, acycloguanosine and myricetin were further analyzed for binding with primase CTD from *S. aureus*. Furthermore, adenosine was selected for binding studies due to its structural similarity to acycloguanosine. Quercetin, luteolin and kaempferol were also selected for their structural similarity to myricetin, but their low solubility prevented further study.

A 2D ¹⁵N-¹H HSQC spectrum was collected for free primase and for the primase-ligand complex using acycloguanosine, adenosine, or myricetin. All three compounds were shown to bind primase CTD based on protein chemical shift perturbations (CSPs) of Y552, N564, T570, L574, E580, I584, G585, L589, and Q590 (Figure 6A). None of the three compounds reached CSP saturation at the highest concentration tested (4.0 mM), indicating that their K_D 's were greater than 4 mM. Acycloguanosine, adenosine and myricetin exhibited CSPs above one standard deviation from the average of all non-binding site residues. Specifically, acycloguanosine, adenosine and myricetin caused CSPs of 0.35 ± 0.09 ppm, 0.16 ± 0.29 ppm and 0.07 ± 0.02 ppm, respectively, for the binding site residues.

When the nine residues incurring CSPs were mapped onto the structure of primase CTD, eight of the residues were located adjacent to one another on helix 6 or 7 (Figure 6B). Since these residues are only adjacent to one another in the protein's closed conformation, the proposed ligand-binding site further validates the accuracy of the primase CTD NMR structure. The residue furthest from the other binding site residues was N564, which was previously identified as a key residue for a C1-C2 conformation change (Figure 5). When adenosine was docked into the groove between helices 6 and 7 (Figure 6B), the adenine moiety filled a hydrophobic pocket while its ribose was exposed to solvent and made specific interactions with three residues: Y552, I584, and G585 (Figure 6C). All three compounds docked into the same primase CTD pocket and adopted a uniform binding conformation. The hydrophobic fused rings were buried deep into the pocket and the hydrophilic moieties interacted with the same three residues at the opening of the pocket

(Figure 6D). Since these compounds bind to the closed conformation of primase CTD, they are likely to act as an inhibitor and reduce the helicase-primase interaction.

Discussion

We report the first primase CTD structure from a medically relevant, mesophilic *Firmicutes* bacterium. Its structure shares many features with the known thermophilic *Firmicutes* and mesophilic proteobacterial primase CTD. It has an N-terminal 6-helix bundle called the C1 subdomain connected by linker residues to a C-terminal helix-turn-helix called the C2 subdomain. The C1 subdomains from different species share the same overall fold, although there are differences in the lengths of several helices such that *S. aureus* and *G. stearothermophilus* are more similar to each other than to *E. coli* or *H. pylori*. The overall resolution of the *S. aureus* structure was low due to significant peak overlap in the NMR spectra, even at a field of 900 MHz. The high leucine and α -helical content of the CTD, coupled with a high degree of dynamic motion, led to peak broadening such that only 56% of side chain H γ and 76% H δ could be unambiguously assigned. The number of long range (>5 Å) NOEs was also lower than anticipated (140) for a protein of 19.6 kDa. Nevertheless, the overall rmsd of the 20 lowest energy structures was reasonable at 1.54 ($\pm$ 0.19) Å and the ensemble of structures lies in acceptable ranges of common structure validation programs (all Z-scores above -5.0).

When a representative structure for *S. aureus* primase CTD was compared to the x-ray and NMR CTD structures from other organisms, three conformational classes were observed (Figure 3A, 4). Two of the conformations are likely related to the function of primase. The conformation of the *E. coli* 1T3W structure is probably not biologically relevant given the low pH of its crystallization and some strong interface packing effects.¹⁶ Of all the structures, the *G. stearothermophilus* crystal structure 2R6A is the only one bound to helicase.¹⁵ This suggests that the structures of *H. pylori* (4EHS) and *E. coli* (2HAJ) are also helicase-bound forms. Since conformation B contains members from both *Firmicutes* and *Proteobacteria*, it indicates that the helicase-bound form of primase CTD is not phylum-dependent. Since the solution structures of *G. stearothermophilus* and *S. aureus* primase CTD are in conformation A, it appears to be the conformation adopted when primase dissociates from helicase. In the case of *S. aureus*, it has also been possible to use NMR dynamics to identify N564 as the key hinge residue that allows helix 6 to bend into helices 6 and 7.

Three small molecules, acycloguanosine, adenosine, and myricetin, were discovered to bind a common site on the closed conformation of primase CTD. The ligand-binding pocket is only created when helices 6 and 7 are adjacent to one another. Among the nine residues with CSPs that define this ligand-binding site, N564 exhibited the largest CSP and was also identified as a key residue in a conformational change between the open and closed form of primase CTD. As a result, this ligand-binding site is a potential target for the further development of small molecule inhibitors that may “lock” primase CTD in its closed form and prevent its interaction with the NTD of the helicase. In fact, two of the compounds shown to bind primase CTD have known inhibitory effects on DNA replication proteins, although their known mode of actions are different from the one identified here.

Acycloguanosine is a known inhibitor of the primase-helicase interaction in herpes simplex virus and acts as a chain terminator.^{62, 63} Additionally, myricetin inhibits bacterial helicases with an IC₅₀ of 10 μM⁶⁴ by competing with the ATPase active sites.

Since the *S. aureus* CTD structure shares many features with the *G. stearothermophilus* solution structure, we used the *G. stearothermophilus* DnaB/primase CTD co-crystal structure¹⁵ to identify potential interface residues. The co-crystal structure of the *G. stearothermophilus* primase CTD C2 subdomain (Figure 7A,C) contains five nonpolar residues (F577, L578, A581, A584, and I588) in the final helix that could make contact with DnaB NTD1 (L67, A72, A75, L80 and V86) (Table S6). Following this idea, we created a homology model of the *S. aureus* primase CTD and DnaB NTD (Figure 7B, D) and found there were conserved but not identical residues in the equivalent positions. The *S. aureus* CTD C2 subdomain residues were V598, L594, K591, E588 and V587, and the *S. aureus* DnaB NTD1 residues were V67, D72, S75, L80, and P86. It has been previously shown that a C-terminal fragment containing the last three helices of the *G. stearothermophilus* C2 subdomain⁶ exhibited sub-nanomolar binding affinity to DnaB, but was not capable of stimulating the ATPase activity of DnaB. It could be that DnaG and DnaB from these two organisms interact using spatially similar, hydrophobic, but differently sized residues. The DnaG and DnaB structures may have coevolved to preserve the primase CTD interaction in a manner that led to species-specificity.

The observation that the primase CTD-helicase structural interface is species-specific is consistent with previous studies showing that *S. aureus* helicase will only stimulate primer synthesis when incubated with the cognate primase.⁶ Stimulation of DnaB ATPase activity by primase binding has been shown to involve the C1 subdomain Y548 in *G. stearothermophilus*⁶⁵ and F534 in the *H. pylori*⁶⁶ C1 subdomain. Mutation of these residues to an alanine reduced binding affinity and ATPase stimulation. An examination of the *G. stearothermophilus* primase CTD structure shows that Y548 and the adjacent residue D547 in helix 6 of the C1 subdomain make contact with P97, T98, and N101 in the DnaB NTD2 (Figure 7C, Table S5). The structural equivalent residues in *S. aureus* C1 subdomain are D558 and Y559 and within the *S. aureus* DnaB NTD2 are also P97, T98, and N101 (Figure 7D). Since the residues are identical between *G. stearothermophilus* and *S. aureus*, it indicates the mechanism of ATPase stimulation is highly conserved.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

This work was supported by grants from the UNL/UNMC Collaborative Research Fund and the Layman Foundation Fund to M. G. and from the National Institute of Allergy and Infectious Diseases (R21AI081154), Nebraska Tobacco Settlement Biomedical Research Development Funds, Nebraska EPSCoR, the Maude Hammond Fling Faculty Research Fellowship, and a Nebraska Research Council Interdisciplinary Research Grant to R. P. The research was performed in facilities renovated with support from the NIH (RR015468-01). The authors would like to thank the Rocky Mountain Regional 900 MHz NMR Facility for contributing NMR data to this project.

References

- (1). Griep MA, and Periago J (2017) Primase, In Reference Module in Life Sciences pp 1–6, Elsevier, New York.
- (2). Frick DN, and Richardson CC (2001) DNA primases, *Annu. Rev. Biochem* 70, 39–80. [PubMed: 11395402]
- (3). Johnson SK, Bhattacharyya S, and Griep MA (2000) DnaB helicase stimulates primer synthesis activity on short oligonucleotide templates, *Biochemistry* 39, 736–744. [PubMed: 10651639]
- (4). Chintakayala K, Larson MA, Grainger WH, Scott DJ, Griep MA, Hinrichs SH, and Soultanas P (2007) Domain swapping reveals that the C- and N-terminal domains of DnaG and DnaB, respectively, are functional homologues, *Mol. Microbiol* 63, 1629–1639. [PubMed: 17367384]
- (5). Keck JL, and Berger JM (2001) Primus inter pares (first among equals), *Nat. Struct. Biol* 8, 2–4. [PubMed: 11135655]
- (6). Syson K, Thirlway J, Hounslow AM, Soultanas P, and Waltho JP (2005) Solution structure of the helicase-interaction domain of the primase DnaG: a model for helicase activation, *Structure* 13, 609–616. [PubMed: 15837199]
- (7). Koepsell SA, Larson MA, Griep MA, and Hinrichs SH (2006) *Staphylococcus aureus* helicase but not *Escherichia coli* helicase stimulates *S. aureus* primase activity and maintains initiation specificity, *J. Bacteriol* 188, 4673–4680. [PubMed: 16788176]
- (8). Urlacher TM, and Griep MA (1995) Magnesium acetate induces a conformational change in *Escherichia coli* primase, *Biochemistry* 34, 16708–16714. [PubMed: 8527445]
- (9). Tougu K, and Marians KJ (1996) The extreme C terminus of primase is required for interaction with DnaB at the replication fork, *J. Biol. Chem* 271, 21391–21397. [PubMed: 8702920]
- (10). Corn JE, Pease PJ, Hura GL, and Berger JM (2005) Crosstalk between primase subunits can act to regulate primer synthesis in trans, *Mol. Cell* 20, 391–401. [PubMed: 16285921]
- (11). Podobnik M, McInerney P, O'Donnell M, and Kuriyan J (2000) A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases, *J. Mol. Biol* 300, 353–362. [PubMed: 10873470]
- (12). Aravind L, Leippe DD, and Koonin EV (1998) Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins, *Nucleic Acids Res.* 26, 4205–4213. [PubMed: 9722641]
- (13). Soultanas P (2005) The bacterial helicase-primase interaction: a common structural/functional module, *Structure* 13, 839–844. [PubMed: 15939015]
- (14). Pan H, and Wigley DB (2000) Structure of the zinc-binding domain of *Bacillus stearothermophilus* DNA primase, *Structure* 8, 231–239. [PubMed: 10745010]
- (15). Bailey S, Eliason WK, and Steitz TA (2007) Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase, *Science* 318, 459–463. [PubMed: 17947583]
- (16). Su XC, Schaeffer PM, Loscha KV, Gan PH, Dixon NE, and Otting G (2006) Monomeric solution structure of the helicase-binding domain of *Escherichia coli* DnaG primase, *FEBS J* 273, 4997–5009. [PubMed: 17010164]
- (17). Oakley AJ, Loscha KV, Schaeffer PM, Liepinsh E, Pintacuda G, Wilce MC, Otting G, and Dixon NE (2005) Crystal and solution structures of the helicase-binding domain of *Escherichia coli* primase, *J. Biol. Chem* 280, 11495–11504. [PubMed: 15649896]
- (18). Larson E (2007) Community factors in the development of antibiotic resistance, *Annu. Rev. Public Health* 28, 435–447. [PubMed: 17094768]
- (19). Mercier KA, Germer K, and Powers R (2006) Design and characterization of a functional library for NMR screening against novel protein targets, *Comb. Chem. High Throughput Screen* 9, 515–534. [PubMed: 16925512]
- (20). Shortridge MD, Griep MA, and Powers R (2012) (1)H, (1)3C, and (1)5N NMR assignments for the helicase interaction domain of *Staphylococcus aureus* DnaG primase, *Biomol. NMR Assign* 6, 35–38. [PubMed: 21644056]

- (21). Sattler M, Schleucher J, and Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.
- (22). Hwang TL, Mori S, Shaka AJ, and vanZijl PCM (1997) Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs, *J. Am. Chem. Soc.* 119, 6203–6204.
- (23). Shen Y, Delaglio F, Cornilescu G, and Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR* 44, 213–223. [PubMed: 19548092]
- (24). Vuister GW, Wang AC, and Bax A (1993) Measurement of 3-Bond Nitrogen Carbon-J Couplings in Proteins Uniformly Enriched in N-15 and C-13, *J. Am. Chem. Soc.* 115, 5334–5335.
- (25). Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, and Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6, 277–293. [PubMed: 8520220]
- (26). Garrett DS, Powers R, Gronenborn AM, and Clore GM (1991) A Common-Sense Approach to Peak Picking in 2-Dimensional, 3-Dimensional, and 4-Dimensional Spectra Using Automatic Computer-Analysis of Contour Diagrams, *J. Magn. Res* 95, 214–220.
- (27). Fogh R, Ionides J, Ulrich E, Boucher W, Vranken W, Linge JP, Habeck M, Rieping W, Bhat TN, Westbrook J, Henrick K, Gilliland G, Berman H, Thornton J, Nilges M, Markley J, and Laue E (2002) The CCPN project: an interim report on a data model for the NMR community, *Nat. Struct. Biol* 9, 416–418. [PubMed: 12032555]
- (28). Huang YJ, Tejero R, Powers R, and Montelione GT (2006) A topology-constrained distance network algorithm for protein structure determination from NOESY data, *Proteins* 62, 587–603. [PubMed: 16374783]
- (29). Clore GM, Appella E, Yamada M, Matsushima K, and Gronenborn AM (1990) Three-dimensional structure of interleukin 8 in solution, *Biochemistry* 29, 1689–1696. [PubMed: 2184886]
- (30). Schwieters CD, Kuszewski JJ, Tjandra N, and Clore GM (2003) The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson* 160, 65–73. [PubMed: 12565051]
- (31). Schwieters CD, Kuszewski JJ, and Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination, *Prog. Nucl. Magn. Reson. Spectrosc* 48, 47–62.
- (32). Garrett DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW, Gronenborn AM, and Clore GM (1994) The impact of direct refinement against three-bond HN-C alpha H coupling constants on protein structure determination by NMR, *J. Magn. Reson., Ser. B* 104, 99–103. [PubMed: 8025816]
- (33). Kuszewski J, Qin J, Gronenborn AM, and Clore GM (1995) The impact of direct refinement against ¹³C alpha and ¹³C beta chemical shifts on protein structure determination by NMR, *J. Magn. Reson., Ser B* 106, 92–96. [PubMed: 7850178]
- (34). Kuszewski J, Gronenborn AM, and Clore GM (1996) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases, *Protein Sci.* 5, 1067–1080. [PubMed: 8762138]
- (35). Kuszewski J, Gronenborn AM, and Clore GM (1997) Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids, *J. Magn. Reson.* 125, 171–177. [PubMed: 9245376]
- (36). Kuszewski J, and Clore GM (2000) Sources of and solutions to problems in the refinement of protein NMR structures against torsion angle potentials of mean force, *J. Magn. Reson* 146, 249–254. [PubMed: 11001840]
- (37). Powers R, Mirkovic N, Goldsmith-Fischman S, Acton TB, Chiang Y, Huang YJ, Ma L, Rajan PK, Cort JR, Kennedy MA, Liu J, Rost B, Honig B, Murray D, and Montelione GT (2005) Solution structure of *Archaeoglobus fulgidis* peptidyl-tRNA hydrolase (Pth2) provides evidence for an extensive conserved family of Pth2 enzymes in archaea, bacteria, and eukaryotes, *Protein Sci.* 14, 2849–2861. [PubMed: 16251366]
- (38). Brunger AT (2007) Version 1.2 of the Crystallography and NMR system, *Nat. Protoc* 2, 2728–2733. [PubMed: 18007608]

- (39). Linge JP, and Nilges M (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation, *J. Biomol. NMR* 13, 51–59. [PubMed: 10905826]
- (40). Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, and Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank, *Proteins* 59, 662–672. [PubMed: 15822098]
- (41). Kay LE, Torchia DA, and Bax A (1989) Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry* 28, 8972–8979. [PubMed: 2690953]
- (42). Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, and Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation, *Biochemistry* 33, 5984–6003. [PubMed: 7514039]
- (43). Mandel AM, Akke M, and Palmer AG 3rd. (1995) Backbone dynamics of *Escherichia coli* ribonuclease HI: correlations with structure and function in an active enzyme, *J. Mol. Biol* 246, 144–163. [PubMed: 7531772]
- (44). Cole R, and Loria JP (2003) FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data, *J. Biomol. NMR* 26, 203–213. [PubMed: 12766418]
- (45). Lipari G, and Szabo A (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, *J. Am. Chem. Soc* 104, 4546–4559.
- (46). Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J. Comput. Chem* 25, 1605–1612. [PubMed: 15264254]
- (47). Huson DH, and Bryant D (2006) Application of phylogenetic networks in evolutionary studies, *Mol. Biol. Evol* 23, 254–267. [PubMed: 16221896]
- (48). Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, and Bourne PE (2000) The Protein Data Bank, *Nucleic Acids Res.* 28, 235–242. [PubMed: 10592235]
- (49). Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32, 1792–1797. [PubMed: 15034147]
- (50). Notredame C, Higgins DG, and Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol* 302, 205–217. [PubMed: 10964570]
- (51). Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, and Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol* 7, 539. [PubMed: 21988835]
- (52). Mercier KA, Baran M, Ramanathan V, Revesz P, Xiao R, Montelione GT, and Powers R (2006) FAST-NMR: functional annotation screening technology using NMR spectroscopy, *J. Am. Chem. Soc* 128, 15292–15299. [PubMed: 17117882]
- (53). Hwang TL, and Shaka AJ (1995) Water Suppression That Works - Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients, *J. Magn. Reson., Ser. A* 112, 275–279.
- (54). Johnson PE, Tomme P, Joshi MD, and McIntosh LP (1996) Interaction of Soluble Cellooligosaccharides with the N-Terminal Cellulose-Binding Domain of *Cellulomonas fimi* CenC. 2. NMR and Ultraviolet Absorption Spectroscopy, *Biochemistry* 35, 13895–13906. [PubMed: 8909286]
- (55). Barril X (2013) Druggability predictions: methods, limitations, and applications, *WIREs Comput. Mol. Sci* 3, 327–338.
- (56). Halouska S, Zhou Y, Becker DF, and Powers R (2009) Solution structure of the *Pseudomonas putida* protein PpPutA45 and its DNA complex, *Proteins* 75, 12–27. [PubMed: 18767154]
- (57). Laskowski RA, Macarthur MW, Moss DS, and Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures, *J. Appl. Crystallogr* 26, 283–291.
- (58). Bhattacharya A, Tejero R, and Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia, *Proteins* 66, 778–795. [PubMed: 17186527]

- (59). Shortridge MD, Triplet T, Revesz P, Griep MA, and Powers R (2011) Bacterial protein structures reveal phylum dependent divergence, *Comput. Biol. Chem* 35, 24–33. [PubMed: 21315656]
- (60). Motlagh HN, Wrabl JO, Li J, and Hilser VJ (2014) The ensemble nature of allostery, *Nature* 508, 331–339. [PubMed: 24740064]
- (61). Cantor CR, and Schimmel PR (1980) *Biophysical Chemistry Part II: Techniques for the study of biological structure and function*, p 461, W. H. Freeman and Co., San Francisco.
- (62). Crute JJ, Grygon CA, Hargrave KD, Simoneau B, Faucher AM, Bolger G, Kibler P, Liuzzi M, and Cordingley MG (2002) Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease, *Nat. Med* 8, 386–391. [PubMed: 11927945]
- (63). Kleymann G, Fischer R, Betz UA, Hendrix M, Bender W, Schneider U, Handke G, Eckenberg P, Hewlett G, Pevzner V, Baumeister J, Weber O, Henninger K, Keldenich J, Jensen A, Kolb J, Bach U, Popp A, Maben J, Frappa I, Haebich D, Lockhoff O, and Rubsamen-Waigmann H (2002) New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease, *Nat. Med* 8, 392–398. [PubMed: 11927946]
- (64). Griep MA, Blood S, Larson MA, Koepsell SA, and Hinrichs SH (2007) Myricetin inhibits *Escherichia coli* DnaB helicase but not primase, *Bioorg. Med. Chem* 15, 7203–7208. [PubMed: 17851081]
- (65). Chintakayala K, Larson MA, Griep MA, Hinrichs SH, and Soultanas P (2008) Conserved residues of the C-terminal p16 domain of primase are involved in modulating the activity of the bacterial primosome, *Mol. Microbiol* 68, 360–371. [PubMed: 18366438]
- (66). Abdul Rehman SA, Verma V, Mazumder M, Dhar SK, and Gourinath S (2013) Crystal structure and mode of helicase binding of the C-terminal domain of primase from *Helicobacter pylori*, *J. Bacteriol* 195, 2826–2838. [PubMed: 23585534]
- (67). Clore GM, Nilges M, Sukumaran DK, Brunger AT, Karplus M, and Gronenborn AM (1986) The three-dimensional structure of alpha1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics, *EMBO J.* 5, 2729–2735. [PubMed: 16453716]
- (68). Nilges M, Gronenborn AM, Brunger AT, and Clore GM (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2, *Protein Eng.* 2, 27–38. [PubMed: 2855369]
- (69). Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, and Karplus M (1983) Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations, *J. Comput. Chem* 4, 187–217.

Figure 1. Primase CTD ensemble overlay.

(A) A ribbon diagram of the average water refined structure. The C1 subdomain is composed of helices 1–6 and the C2 subdomain is composed of helices 7–8. All structures were generated with Visual Molecular Dynamics (VMD-XPLOR) and are colored according to the secondary structure: red, α -helix; green, loop. (B) An overlay of the backbone trace of the 20 lowest-energy, water-refined structures aligned with residues 468–566 from the C1 subdomain. (C) An overlay of the backbone trace of the 20 lowest-energy structures aligned with residues 570–603 from the C-terminal C2 subdomain.

Figure 2. Electrostatic surface potential.

Surface charge distribution of the C-terminal domain showing (A) the face toward the RNA polymerase domain (F463 is the N-terminal residue of this domain) with space-filled C2 subdomain on the left and the C1 subdomain on the right and (B) the solvent-exposed face (180° rotation around the x-axis relative to the structures in panel A).

Figure 3. Unrooted phylograms.

Representation of the primase CTD structure (A), sequence similarity (B) and primase CTD C1 subdomain structure (C). All trees were generated with SplitsTree4. Each tree contains the following species: *S. aureus* primase CTD (this study), *G. stearothermophilus* primase CTD, *H. pylori* primase CTD, and *E. coli* primase CTD.

Figure 4. Primase CTD conformation classes.

The species structures were categorized in three conformations from Figure 3A based on compactness: A compact, B intermediate, C extended.

Figure 5. Dynamics of *S. aureus* primase CTD.

The NMR relaxation parameters T_1 (A) and T_2 (B), NOE enhancements (C), S^2 order parameter (D), and R_{ex} chemical exchange rate are plotted per residue. The C1 subdomain included residues 463–572 and the C2 subdomain included residues 565–605. The linker region corresponding to residues 565–572 was included in the dynamics analysis for both subdomains.

Figure 6. Adenosine binding site identification.

When *S. aureus* primase CTD (1 mM) was titrated with up to 4.0 mM adenosine, nine residues exhibited chemical shift perturbations in the 2D ^1H - ^{13}C HSQC spectrum of primase CTD. A detail from the overlay of the 2D ^1H - ^{13}C HSQC spectra is colored according to the addition of 0 mM (blue), 2.6 mM (purple) and 4.0 mM (red) of adenosine (A). Expanded view of a surface rendition of the primase CTD NMR structure is superimposed onto a ribbon structure to highlight secondary structure elements. Adenosine (blue tessellated structure) was docked into the primase CTD region defined by the nine residues with significant chemical shift perturbations (colored purple), which are located between helices 6 and 7. The side chains for all the residues that interact with adenosine are drawn with licorice bonds and are the same side chains displayed in panel C. Residue N564 and α -helices 6 and 7 are labeled (B). An expanded view of the ligand-binding pocket (flipped 180 degree from panel B) represented as a ribbon structure with side chains interacting with adenosine shown as licorice bonds. The residues interacting with adenosine are labeled. The labels for residues with significant CSPs are colored purple (C). Superposition of the docked conformations of myricetin (green), adenosine (black), and acycloguanosine (orange) into the same expanded view of the primase CTD binding site from panel C (D).

Figure 7. Critical interface residues between primase and helicase.

The co-crystal structure of the complex from *G. stearothermophilus* involves two DnaB NTDs, one of which interacts with the C1 subdomain of the primase and the other with the C2 subdomain (A). The specific interface residues are shown in an expanded view (C). This was the template for the homology model of *S. aureus* primase CTD and DnaB NTDs (B), which is also shown in an expanded view (D).

Table 1:

Primase CTD structures used for analysis

Species	PDB ID – chain	Reference
<i>S. aureus</i>	2LZN – 14	Current paper
<i>G. stearothermophilus</i>	1Z8S – 5	6
<i>E. coli</i>	2HAJ – 19	16
<i>G. stearothermophilus</i>	2R6A – C	15
<i>H. pylori</i>	4EHS – A and 4EHS – B	66
<i>E. coli</i>	1T3W – A and 1T3W – B	17

Table 2:

Structure Statistics^a

Root mean square parameter	<SA>	σ	(SA) _r
Distance restraints (Å)			
All (1823)	0.040	0.023	0.022
Inter-residue sequential ($ i-j = 1$) (460)	0.056	0.021	0.030
Inter-residue short range ($1 < i-j \leq 5$) (444)	0.062	0.018	0.041
Inter-residue long-range ($ i-j > 5$) (140)	0.247	0.093	0.374
Intra-residue (663)	0.001	0.001	0.003
Backbone hydrogen bonds (116)	0.014	0.005	0.008
Dihedral restraints (deg) (280)	1.21	0.23	1.12
C α chemical shift restraints (ppm) (130)	1.05	0.05	1.08
C β chemical shift restraints (ppm) (126)	1.06	0.04	1.09
³ J _{NHα} coupling restraints (Hz) (82)	0.011	0.0005	0.011
F _{NOE} (kcal mol ⁻¹)	87	36	46
F _{torsion} (kcal mol ⁻¹)	28	20	21
F _{VDW} (kcal mol ⁻¹)	-509	31	-510
Idealized covalent geometry			
Bond length (Å) (2400)	0.0114	0.0005	0.0108
All angles (deg) (4302)	1.44	0.06	1.43
Improper torsion angles (deg) (1287)	1.95	0.23	1.84

^a <SA> is the average value from an all-versus-all comparison of the set of 20 annealed structures; σ is the standard deviation for the all-versus-all comparison; and (SA)_r is the value for the restrained minimized mean structure. The number of restraints for each parameter is given in parentheses. For backbone NH-CO hydrogen bonds, the two restraints are $r(\text{NH}\cdots\text{O}) = 1.5 - 2.3 \text{ \AA}$ and $r(\text{N}\cdots\text{O}) = 2.5 - 3.3 \text{ \AA}$. The values of the square-well NOE (F_{NOE}) and torsion angle (F_{torsional}) potentials⁶⁷ are calculated with force constants of $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ and $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, respectively. The value of the Lennard-Jones van der Waals term (F_{VDW})⁶⁸ is calculated with a force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-4}$ with the CHARMM⁶⁹ empirical energy function. The improper torsion angle restraints serve to maintain planarity and chirality

Table 3:Ensemble self-consistency^a

Parameter	<SA>	σ	(SA) _r
PROCHECK ^b			
Overall G-Factor	-0.12	0.04	-0.09
H-bond energy	0.82	0.06	0.8
Number bad contacts/100 residues	0.0		0.0
PSVS Z-scores ^c			
Verify3D	-4.58	0.71	-4.82
ProsaII (-ve)	-1.14	0.34	-1.20
Procheck (phi-psi)	-0.56	0.42	-0.28
Procheck (all)	-1.92	0.32	-2.07
MolProbity clash score	-2.50	0.97	-2.46
Ramachandran space ^d			
Most favored regions	83.4%	3.2	85.5%
Additional allowed regions	14.5%	3.0	13.0%
Generously allowed regions	1.43%	1.2	0.7%
Disallowed regions	0.7%	0.56	0.7%

^a <SA> is the average value from an all-versus-all comparison of the set of 20 annealed structures; σ is the standard deviation for the all-versus-all comparison; and (SA)_r is the value for the restrained minimized mean structure.

^b PROCHECK values are in comparison to the likelihood of finding each residue within the derived structure as compared to a database of residues found in similar environments within structures with resolution below 1.8 Å.

^c PSVS (Protein Structure Validation Suite) provides an analysis of five separate structural analysis programs, each of which measures a different parameter. The PSVS reports Z-scores for each analysis, where a Z-score is the number of standard deviations from the mean. The authors of PSVS suggest that Z-scores lower than -5 should be reevaluated.

^d Ramachandran space analysis from the PDB sum website uses a database of structures with resolution below 1.8 Å to decide whether a region is considered to be “most favored”, etc.

Table 4:Ensemble self-consistency of the C1 subdomain^a

Comparison	All residues	Backbone atoms	All heavy atoms
<u>All residues</u>			
<SA> vs. SA _{mean}		1.54 ± 0.19 Å	2.37 ± 0.17 Å
<SA> vs. (SA) _r		1.98 ± 0.28	3.05 ± 0.32
(SA) _r vs. SA _{mean}		1.26	1.94
<u>Secondary structures</u>			
<SA> vs. SA _{mean}		1.18 ± 0.14	1.93 ± 0.16
<SA> vs. (SA) _r		1.56 ± 0.22	2.55 ± 0.26
(SA) _r vs. SA _{mean}		1.03	1.66

^a <SA> is the average value from an all-versus-all comparison of the set of 20 annealed structures; (SA)_r is the value for the restrained minimized mean structure; and SA_{mean} is the Cartesian mean structure created from the geometric mean for each atom of all 20 structures. The SA_{mean} structure does not adhere to any of the restraints for bond lengths, angles, etc. The secondary structure residues are: 469–478 (α1), 484–493 (α2), 502–515 (α3), 524–528 (α4), 533–544 (α5), 552–564 (α6), 573–582 (α7), and 588–602 (α8). The secondary structure rmsd values were measured by aligning all secondary structure elements and calculating an average and standard deviation.