Supporting Methods

Construction of plasmids.

DNA primers used are listed in Supplementary Table S4.

Plasmid M13*oriC*MS9 was an M13mp18 derivative bearing a 431 bp chromosome-derived NcoI-NsiI fragment which carried *oriC*. M13*oriC*MS9-τ2l' and -τ2r' were constructed by PCR using M13*oriC*MS9 and primers listed in Table S4.

Plasmid pBR*oriC* was described previously (1). pBR*oriC*-derivatives bearing an altered length between R2 and C3 boxes were constructed by PCR using pBR*oriC* and primers listed in Table S4.

DUE unwinding (P1 nuclease) assay.

This assay was performed as described previously with minor modifications (1, 2) . Briefly, the ATP-bound DnaA and IHF (32 nM) were incubated at 38 °C for 3 min in buffer containing 100 mM KCl, 5 mM ATP or ADP, and M13*oriC*MS9 plasmid or its derivative (1.32 nM), followed by incubation with P1 nuclease. DNA was further incubated with EcoRI, followed by agarose gel electrophoresis analysis. When pBRoriC and its derivatives were used, DNA digested by AlwNI was analyzed (1).

DnaB helicase loading (form I*) assay.

This assay was performed as described previously with minor modifications (1, 2). Briefly, the form I of pBR*oriC* or its derivative was incubated at 30 °C for 15 min in buffer containing ATP, DnaA, IHF, DnaB, DnaC, SSB, and gyrase, followed by 0.65% agarose gel electrophoresis analysis and quantification of intensities of bands corresponding to the form I and form I*.

Pull-down assay using a biotin-tagged *oriC* fragment

This assay was performed as described previously (3). Briefly, DnaA, His-DnaB, DnaC, and a biotin-tagged *oriC* fragment were incubated on ice for 15 min in buffer (10 μL), followed by further incubation at 4 °C for 15 min in the presence of streptavidin-coated beads (Promega). Bound proteins were recovered and analyzed by SDS-10% PAGE and silver staining.

EMSA using the R2-C3 region.

This assay was basically performed as previously described (1). Briefly, DnaA was incubated at 4ºC for 10 min in buffer containing 80 mM KCl, 2 mM ATP or ADP, 0.2 μg/mL λDNA as a competitor, and 100 nM each DNA fragment, followed by incubation at 30ºC for 10 min. DNA complexes were analyzed by 6% polyacrylamide gel electrophoresis and Gelstar staining. DNA used was prepared by annealing of oligonucleotides (Table S4).

EMSA using the middle and right-half DOR.

This assay was performed as previously described (1). Briefly, in this method, chimeric DnaA (ChiDnaA) consisting

of *E. coli* DnaA (EcoDnaA) domain I-III and *Thermotoga maritima* DnaA domain IV is used. ChiDnaA as well as *T. matirima* DnaA specifically binds to *T. maritima* DnaA box sequence which is fundamentally different from *E. coli* DnaA box sequence. Thus, when a certain DnaA box within an *oriC* region is replaced with *T. maritima* DnaA box, ChiDnaA, but not EcoDnaA, binds to the site (1). As affinities of EcoDnaA for ATP and ADP are very high (i.e., Kd of 10-100 nM), ATP-EcoDnaA and ADP-EcoDnaA are extremely stable in solution and nucleotide exchange of ADP-DnaA does not substantially occur even in the presence of 2 mM ATP (4). In addition, ATPase activity of ATP-DnaA is negligible at 30ºC (5). Affinities of ChiDnaA for ATP and ADP are similar to those of EcoDnaA, and stabilities of ATP-ChiDnaA and ADP-ChiDnaA even in the presence of 2 mM ATP are also demonstrated (1). Based on these, in the present study, the nucleotide-free forms of EcoDnaA and chiDnaA were purified and then incubated at 0°C with ATP or ADP to yield nucleotide-bound forms as we described previously (1, 4, 6). The resultant ATP-EcoDnaA and ATP/ADP-ChiDnaA were incubated at 30ºC for 10 min in buffer containing 2 mM ATP, 60 mM KCl, 20 μg/mL λDNA as a competitor DNA, and 35 nM oriR2-R4 or R2R4-R2tma fragment. DNA fragments used were prepared by PCR amplification using pBRoriC and its derivatives as a template and pairs of primers, ori-2 and oriR2R4 or R2R4tmaR2 f (Table S4).

Protein model for coarse grained (CG) MD simulations

For coarse-grained (CG) molecular dynamics simulations, we used a software CafeMol (7). The protein model used is the atomic-interaction-based coarse-grained model (AICG2+) developed by Li et al (8). In this model, each amino acid is represented by a particle located at the Cα atom position. The energy function is the so-called structure-based model; the energy function is based on the three-dimensional native (or reference) structure of the protein. We did not represent ligands (ATPs) explicitly. Instead, by using the structure of ATP-bound form of DnaA as the reference structure, we model the ATP-bound form of DnaA.

The potential energy function is expressed as

$$V_{AICG2+} = V_{bond} + V_{loc}^{flp} + V_{angle} + V_{dihedral} + V_{contact} + V_{excluded}.$$

The first term V_{bond} is the potential energy function to restraint the virtual bond length. For an N-amino acid protein,

$$V_{bond}^{AICG2+} = \sum_{I=1}^{N-1} k_b (r^I - r_0^I)^2$$

where r^I is the virtual bond length between *I* th and *I+1* th residues and r_0^I is the corresponding bond length at the reference structure. For $k_b = 110.40$ kcal/mol/Å^2 is a constant tuned by Li et al using the fluctuation matching method. The second term V_{loc}^{flp} is the sequence-dependent flexible local potential developed by Terakawa and Takada (9) that represents the intrinsic local property of polypeptide. This is statistical local potential for the virtual bond angles and the virtual dihedrals. For an N-amino acid protein, this potential is given by

$$V_{loc}^{flp} = -\sum_{I=1}^{N-2} k_B T \ln\left(\frac{P_{ba}(\theta_I)}{sin\ \theta_I}\right) - \sum_{I=1}^{N-3} k_B T \ln\left(P_{dih}(\varphi_I)\right)$$

where k_B is Boltzmann constant and $T = 300.0$K is a temperature. θ_I is virtual bond angle defined by two consecutive virtual bonds and φ_I is virtual dihedral angle defined by three consecutive virtual bonds. $P_{ba}(\theta)$ and $P_{dih}(\varphi)$ are probability distributions of virtual bond angle and virtual dihedral angle, respectively (9), constructed for each type of

amino acids from a dataset of loop regions in 13598 protein structures in Protein Data Bank (http://www.rcsb.org/pdb/). The $P_{ba}(\theta)$ and $P_{dih}(\varphi)$ are histograms with the bin size of 10 degrees. For calculating the force and energy of virtual bond angle, continuous energy functions were obtained by cubic spline interpolation of the $P_{ba}(\theta)$. In order to avoid virtual bond angles going out of the range with sufficient samples, a linear potential is employed outside of the range of sufficient samples. For the case of dihedral angle, tabulated $P_{dih}(\varphi)$ is fit by the truncated Fourier series as

$$f_{dih}^{flp}(\varphi) = \sum_{m}^{3} ks_m \sin(m\varphi) + \sum_{n}^{3} kc_n \cos(n\varphi) + C$$

where ks_m, kc_n, and C are Fourier coefficient.

The flexible local potential is available in CafeMol (7) (www.cafemol.org). The parameters for flexible local potential are in cafemol/para/flexible_local.para. In "<<<< dihedral_angle", Fourier coefficients are listed in order of C, kc_1, ks_1, kc_2, ks_2, kc_3, ks_3. In "<<<< bond_angle_x", the median value of each bin of $P_{ba}(\theta)$ are listed. Corresponding $-k_BT \ln\left(\frac{P_{ba}(\theta)}{sin\theta}\right)$ are listed in "<<<< bond_angle_y".

The third and fourth terms V_{angle} and $V_{dihedral}$ are the structure-based local potentials. Explicitly, the potential energy functions are given by

$$V_{angle} = \sum_{J=I+2} \varepsilon_{loc}^{IJ} \exp\left(-\frac{\left(r^{IJ} - r_0^{IJ}\right)^2}{2W_{IJ}^2}\right)$$

$$V_{dihedral} = \sum_{J=I+3} \varepsilon_{loc}^{IJ} \exp\left(-\frac{\left(\varphi^{IJ} - \varphi_0^{IJ}\right)^2}{2W_{\varphi,IJ}^2}\right)$$

where ε_{loc}^{IJ} is position dependent parameter, r^{IJ} and φ^{IJ} are the distance between *I*-th and *J* th residues, and the dihedral angle formed by *I*-th, *I+1*-th, *I+2*-th, and *I+3*-th residues. r_0^{IJ} and φ_0^{IJ} are the corresponding parameters at the native structure. W_{IJ} and $W_{\varphi,IJ}$ are the width parameters of the Gaussian potentials.

The fifth term $V_{contact}$ is the structure-based non-local contact potential. When two residues are in contact at the native structure, an attractive potential function is employed. When two heavy atoms belonging to different residues (excluding pairs that are within 3 residues in sequence) are within 6.5 Å at the native structure, we consider these residues are in contact at the native structure. Explicitly, the potential energy function is given by

$$V_{contact} = \sum_{I>J+3}^{nat\ contact} \varepsilon_{nloc}^{IJ} \left[5\left(\frac{r_0^{IJ}}{r^{IJ}}\right)^{12} - 6\left(\frac{r_0^{IJ}}{r^{IJ}}\right)^{10}\right]$$

where ε_{nloc}^{IJ} is position dependent parameter. The default parameters for ε_{nloc}^{IJ} were tuned in Li et al. In order to enhance DnaA domain III oligomerization, ε_{nloc}^{IJ} of inter-DnaA domain III are set as the two times of the default values. $V_{excluded}$ represents the excluded volume effect for particle pairs which are not in contact at the native structure,

$$V_{excluded} = \sum_{I>J+3}^{non-native} \varepsilon_{ex} \left(\frac{C}{r^{IJ}}\right)^{12}$$

where $\varepsilon_{ex} =$ 0.6 kcal/mol and $C = 4.0$ Å are parameters.

In this work we used two types of DnaA domain III-IV models. One is the rigid model. In this model the potential energy function is given by V_{AICG2+} throughout all residues (K135-S467) for a homology-modeled reference structure (described below). The other model is the flexible model, in which both local and non-local structure-based potentials were removed for the boundary region between domains III-IV (L367-T375).

DNA model for CG MD simulation

We used the 3SPN.1 model as the coarse grained DNA model (10). In this model each nucleotide is represented by three particles (phosphate, sugar, base). The phosphate and sugar particle are placed at the center of mass of the corresponding atom groups. The base particle is placed at N1 position (adenine, guanine) or N3 position (cytosine, thymine). This model was empirically tuned to reproduce the sequence-dependent melting temperatures and the persistent length of DNA well. The potential energy function is given by

$$V_{DNA} = V_{bond}^{DNA} + V_{angle}^{DNA} + V_{dihedral}^{DNA} + V_{stuck}^{DNA} + V_{base}^{DNA} + V_{excluded}^{DNA} + V_{solv}^{DNA} + V_{ele} .$$

The potential energy functions for virtual bond length, bond angle, and dihedral angle are given by

$$V_{bond}^{DNA} = \sum_{I} [k_1^{DNA}(r^I - r_0^I)^2 + k_2^{DNA}(r^I - r_0^I)^4]$$

$$V_{angle}^{DNA} = \sum_I k_a^{DNA}(\theta^I - \theta_0^I)^2$$

$$V_{dihedral}^{DNA} = \sum_{I} k_\varphi^{DNA}[1 - \cos(\varphi^I - \varphi_0^I)]$$

where $k_1^{DNA} = 0.1839$ kcal/mol/Å^2, $k_2^{DNA} = 183.9$ kcal/mol/Å^4, $k_a^{DNA} = 128.73$ kcal/mol/rad^2, $k_\varphi^{DNA} = 5.1942$ kcal/mol are constants. r^I, θ^I, and φ^I are the *I*-th bond length, bond angle, and dihedral angle. r_0^I,θ_0^I,φ_0^I are the corresponding parameters in the canonical B-type duplex DNA structure.

Non-boded interaction are given by

$$V_{stuck}^{DNA} = \sum_{I<J} 4\varepsilon \left[\left(\frac{\sigma_{IJ}^{DNA}}{r^{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}^{DNA}}{r^{IJ}}\right)^{6}\right]$$

$$V_{base}^{DNA} = \sum_{I<J} 4\varepsilon_{bIJ}^{DNA} \left[5\left(\frac{\sigma_{bI}^{DNA}}{r^{IJ}}\right)^{12} - 6\left(\frac{\sigma_{bI}^{DNA}}{r^{IJ}}\right)^{10}\right]$$

$$V_{excluded}^{DNA} = \sum_{I<J} \begin{cases} 4\varepsilon \left[\left(\frac{\sigma_0^{DNA}}{r^{IJ}}\right)^{12} - \left(\frac{\sigma_0^{DNA}}{r^{IJ}}\right)^{6}\right] + \varepsilon \left(r^{IJ} < r_{coff}\right) \\ 0 \left(r^{IJ} \geq r_{coff}\right) \end{cases}$$

$$V_{solv}^{DNA} = \sum_{I<J} E_s^{DNA}\left[1 - e^{-\alpha(r^{IJ}-r_s^{DNA})}\right]^2$$

$$V_{ele} = \sum_{I<J} \frac{q_I q_J}{4\pi\varepsilon_0\varepsilon_k r^{IJ}} e^{-r^{IJ}/\kappa_D}$$

V_{stack}^{DNA} is the potential energy function for base stacking of natively-contacting particle pairs based on the canonical B-type DNA. V_{ele} is the Debye-Hückel type electrostatic interactions. q_I and q_J are charges, ε_0 is the electric constant, and ε_k=78.0 is dielectric constant. r^{IJ} is distance between *I* th particle and *J* th particle. κ_D is the Debye length given by

$$\kappa_D = \left(\frac{\varepsilon_0\varepsilon_k k_B T}{2N_A e^2 I}\right)^{1/2}$$

where N_A is Avogadro's constant, e is the elementary electric charge, and *I* is the ionic strength. The ionic strength *I* is defined as $I = 0.5\sum z_i^2 c_i$. c_i is molar density and $z_i = q_i/e$. In this work ionic condition is 200 mM NaCl (I = 0.2). In 3SPN.1 model, all phosphates have z= -1 charges.

Protein-DNA interaction in CG MD simulation

The potential energy function is given by,

$$\mathrm{V}_{protein-DNA} = \sum_{I\in protein, J\in DNA}^{native} \varepsilon^{IJ}\left[5\left(\frac{r_0^{IJ}}{r^{IJ}}\right)^{12} - 6\left(\frac{r_0^{IJ}}{r^{IJ}}\right)^{10}\right] + \mathrm{V}_{ele} + \mathrm{V}_{excluded}$$

The first term is the structure-based potential that represents specific protein-DNA interactions. This is similar to the non-local potential in the off-lattice Go models (11). Following our previous work (12), in DNA we used bases and sugars, but not phosphates for the structure-based potentials: We regarded that amino acid-phosphate interaction is well described by V_{ele} and $\mathrm{V}_{excluded}$. The parameter ε^{IJ} is a constant: For the IHF-DNA interaction ε^{IJ} is 1.2 kcal/mol. For DnaA domain IV-DNA we used three values, $\varepsilon_{strong} = 0.36$ kcal/mol for high-affinity DnaA boxes (R1 box, R4 box, and R2 box), $\varepsilon_{weak} = 0.315$ kcal/mol for the other DnaA boxes (C- boxes, I- boxes, τ2l box, and τ2r box), and $\varepsilon_{prepare} = 0.6$ kcal/mol for preparation (Details are described below). r^{IJ} is the distance between I-th and J-th particles and the parameter r_0^{IJ} is the corresponding value at the native structure. V_{ele} is the same as that in the DNA model. In proteins, we set all Arg and Lys with +1 charges and Asp, Glu with -1 charges.

Fully atomistic MD simulations

Fully atomistic MD simulation was conducted by GROMACS 4.5.5 (for the energy minimization of the native structures) (13–17) and GROMACS 5.1.1 (for refinement simulation of *E. coli* replication initiation complex) (13–20). The force field for proteins and nucleic acids is ffAMBER99SB with ParmBSC0 nucleic acid parameters (21–24). ATP topology was obtained at (http://redmine.gromacs.org/issues/721). The water model is TIP3P (25). We conducted simulations in the triclinic box with the periodic boundary condition. In the left-half *E. coli* replication initiation complex, there are DNA, proteins, 5 ATPs, 5 Mg^{2+} ions, 990 Na^+ ions, 785 Cl^- ions, and 199,080 waters. In the middle- right-half part, there are DNA, proteins, 6 ATPs, 6 Mg^{2+} ions, 1435 Na^+ ions, 1255 Cl^- ions, and 326,379 waters. In the DnaA domain III-

hexameric complex on middle–right-half *oriC*, there are DNA, proteins, 6 ATPs, 6 Mg^{2+} ions, 1088 Na^{+} ions, 922 Cl^{-} ions, and 235,717 waters. We neutralized our system and set ion concentration to 0.2 M NaCl by adding Na+ and Cl- ions. The particle-mesh Ewald (26) method was used for long range electrostatic interactions.

The energy minimization was conducted by the steepest descent minimization algorithm. For *E. coli* replication initiation complex we subsequently conducted 100ps NVT equilibration at 300K. We restrained the position of all heavy atoms in proteins, DNAs, and ligands (ATP, Mg^{2+}). We used an accurate and efficient leap-frog stochastic dynamics integrator for molecular move algorithm and as a thermostat (27). The inverse friction constant was 2.0 ps. After NVT equilibration, we conducted 1 ns NPT equilibration at 300K, 1 bar using the Parrinello-Rahman pressure coupling (28). The condition of position restraints of atoms and the integrator were same as those in the NVT equilibration. The production run was conducted at 300K, 1 bar for 10 ns. The positions of heavy atoms were not restrained in the production run. The other conditions were the same as that in the NPT equilibration. In NVT equilibrations, NPT equilibrations, and production runs, all bond lengths were constrained by the LINCS (29) except for waters. For waters we used the SETTLE (30).

Preparation of the structure of globular protein domains and protein-DNA complex for CG MD simulation

In this process, we used PyMOL for superposition of two molecules and for changing atoms (31). We started from homology modeling of *E. coli* DnaA domain III-IV native structure. First we modeled *E. coli* DnaA domain III by I-TASSER modeling server (32–34). We specified *T. maritima* DnaA domain III (PDB ID 2Z4S) as the template (35). Next we modelled *E. coli* DnaA domain III-IV by I-TASSER specifying modelled *E. coli* DnaA domain III (K135-L369) and *E. coli* DnaA domain IV crystal structure (PDB ID : 1J1V) (36). Using PyMOL, we modified the output model of ACP bound *E. coli* DnaA domain III-IV by changing carbon of ACP to oxygen and placing Mg^{2+} ion referring to PDB ID 3R8F. Finally we obtained ATP and Mg^{2+} bound *E. coli* DnaA domain III-IV.

Subsequently we prepared *E. coli* DnaA domain III-IV homo-oligomer. We utilized *A. aeolicus* DnaA domain III-IV tetramer with ssDNA (PDB ID 3R8F) (37) and replaced *A. aeolicus* DnaA with our ATP and Mg^{2+} bound *E. coli* DnaA model. In order to replace proteins, we superposed domain III of *E. coli* DnaA domain III-IV homology model on each *A. aeolicus* DnaA.

Finally we conducted the energy minimization of *E. coli* DnaA domain III-IV tetramer with ssDNA complex by GROMACS 4.5.5 (13–17). We defined one of inner DnaA structure in *E. coli* DnaA tetramer as the native (reference) ATP-DnaA domain III-IV structure for coarse-grained simulations and used the K135-L367 of inner-DnaA domain III dimer for the native (reference) DnaA domain III-domain III interactions.

Next we prepared *E. coli* DnaA domain IV-dsDNA complex structure of each DnaA box. The structure is based on *E. coli* DnaA domain IV-dsDNA co-crystal structure (PDB ID 1J1V) (36). First we superposed *E. coli* DnaA domain IV native structure (our homology model) on crystal structure and replaced DnaA domain IV. Next we changed DNA sequence by changing bases of the nucleic acid. We obtained the structure of bases (adenine, cytosine, guanine, thymine) by NAB in AmberTools 13, and fitted three atoms of the NAB-generated base on the three atoms in the crystal structure (38, 39). For adenine or guanine we used N9, C4, C8 atoms. For cytosine or thymine we used N1, C2, C6 atoms. We

subsequently deleted overlapping bases in the crystal structure. Finally we conducted energy minimization of the complex structures by GROMACS 4.5.5 (13–17). The complexes we obtained contain 13-basepair (bp) DNA. In coarse-grained simulations, we only used structural information of DnaA box (the 9 bp) – DnaA domain IV interaction.

At last we modeled native structure of *E. coli* IHF – dsDNA complex using *E. coli* IHF – DNA crystal structure (PDB ID 1IHF)(40). First we used I-TASSER modeling server specifying PDB ID 1IHF and supplemented missing residues and atoms. We superposed this structure and replaced the IHF protein. The original crystal structure contains DNA nick. We changed this nick into the continuous DNA structure referring the structure of chain C residue ID -38 to -34 and chain E residue ID 34 to 38. Then we changed DNA sequence into the IHF binding site in *oriC* in the same way as DnaA domain IV - dsDNA complex (40–42). We transformed Cd^{2+} ions into Mg^{2+} ions. Finally we conducted the energy minimization of the complex structures by GROMACS 4.5.5 (13–17). When we defined the natively contacting pairs between DNA and proteins, we used not only the contact information between IHF specific site and proteins but also the other contact information between DNA and proteins in this complex.

Detail of coarse grained (CG) MD simulations

We conducted Langevin dynamics simulations with friction coefficient $\gamma = 0.02$ and temperature $\mathrm{T} = 300\mathrm{K}$. The detail of molecular move algorithm is described in CafeMol manual (43).

A) Determining parameters

First we determined the parameter ε_{strong} and ε_{weak} by fitting the experimental value of dissociation constants. We simulated DnaA domain IV (L373-S467) and 21 base pair dsDNA (5'-TCCTTGTTATCCACAGGGCAG-3') (R1 box underlined) system in 100 Å capsule. When the geometric center of domain IV reach the capsule boundary, the pushing back force acts. For 21 base pair DNA, the sugar particle of the terminal thymine is anchored at the center of the capsule. We conducted 10^9-MD step simulations changing ε^{IJ} value of native contact interaction by 0.015 kcal/mol. We conducted 2 simulations for each ε value changing the random seed of Langevin dynamics. The coordinates and energy were recorded in every 10^4 MD steps.

From trajectories, we estimated the dissociation constant. For analysis, we discarded first 10^7 MD steps. When the total interaction energy between DnaA domain IV and DNA is below $-0.3\ \mathrm{kcal/mol}$, we regarded that the two molecules are bound. We defined the dissociate constant (K_d) as follows.

$$r_{bound} = N_{bound}/N_{total}$$

$$c_{bound} = \frac{r_{bound}}{N_A} \times \frac{1}{\frac{4}{3} \times \pi \times \left(R_{cap}\right)^3}$$

$$c_{unbound} = \frac{(1 - r_{bound})}{N_A} \times \frac{1}{\frac{4}{3} \times \pi \times \left(R_{cap}\right)^3}$$

$$K_d = \frac{{c_{unbound}}^2}{c_{bound}}$$

Where N_{bound} represents the number of snapshots in which DnaA domain IV binds to DNA. N_{total} = 9,9000 is the number of snapshots. N_A is Avogadro constant. R_{cap} = 100 Å is radius of capsule. We calculated K_d value for each trajectory and the result was summarized in Table S5. Dissociation constants for this system were experimentally measured by Schaper and Messer (44). For R1 and R4 box, K_d~1.0 nM and for weak boxes $K_d > 200$ nM, respectively. We found that the use of ε = 0.36 kcal/mol reproduce the K_d with the order of ~1 nM and ε = 0.315 kcal/mol reproduce the K_d with the order of ~100 nM. We defined $\varepsilon_{strong} = 0.36$ kcal/mol and used for interactions between DnaA domain IV and R1/R4/R2 box. We defined $\varepsilon_{weak} = 0.315$ kcal/mol and used for interactions between DnaA domain IV and all low affinity boxes(C- boxes, I- boxes, τ2l box, and τ2r box).

B) Preparing initial structures and production runs

- Simulation B1: The simulation of DnaA domain III-IV monomer in solution was conducted by Langevin dynamics for 1×10^8 MD steps. We simulated both the rigid model and the flexible model. For each model, we conducted 5 simulations individually changing random seeds. Initial structure was the native structure of DnaA domain III-IV.
- Simulation B2: The simulation of flexible model of DnaA domain III-IV with the DNA fragment of R1 box was also conducted for 1×10^8 MD steps. The DNA sequence was CCTGTGGATAACA. The initial structures were DnaA domain III-IV binding to the R1 box (underlined). We superposed DnaA domain III-IV on DnaA domain IV -DNA crystal structure (36) and obtained the initial structure. ε^{IJ} were set to ε_{strong}. We conducted 5 simulations individually changing random seeds. In all simulations we recorded the coordinates and energy of system every 10^4 MD steps.
- Simulation B3: For the simulations of left-half subcomplex, we included five DnaA domain III-IVs and IHF and 102-bp left-half *oriC* fragment (See Table S6).

  The initial structures for simulations were prepared in 2 steps. The first step is to bind DnaA domain III-IV on the DnaA box. We prepared IHF-bound left-half *oriC* DNA fragment which is generated by connecting coarse-grained B-type DNA fragments and coarse-grained IHF-DNA complex structure. The B-type DNA was generated by NAB(38, 39). We connected them by superposing the terminal basepair of one fragment on the other using PyMOL(31). Next we made DnaA domain III-IV bound on the R1 box. We arranged DnaA domain III-IV close to R1 box and induced DnaA binding on the R1 box by 5×10^6 MD steps simulations. We conducted 10 simulations with different random numbers in Langevin dynamics. ε^{IJ} were set 3.0 kcal/mol and DnaA was set to rigid model. After obtaining R1-bound DnaA-IHF-DNA complex, we next made DnaA binding to the R5M box in the same way as R1 box by 5 simulations. In the same way, we repeatedly conducted DnaA binding simulations in order of the τ2 (τ2l or τ2r), I1, and I2 by 5-10 simulations for each DnaA box. During simulations we fixed all Cα atoms of IHF proteins. In this step inter-DnaA domain III attractions were not applied. The second step is to prepare different initial structures. We conducted 10^7 MD steps simulations 10 times with different random forces. ε^{IJ} was $\varepsilon_{prepare}$ for DnaA box – DnaA domain IV. The DnaA domain III-IV model was either rigid or flexible for the left-half simulation. During the left-half simulations we fixed three amino acids of IHF (IHFα P65, IHFβ H16 and R62). In these simulations, inter-DnaA domain III attractions were not applied. The obtained final structures were used as the

initial structures of the production runs.

About the production runs for the complex formations, we conducted 1×10^8MD steps simulations 10 times with different random forces. We applied inter-DnaA domain III native contact potentials. ε^{IJ} were ε_{strong} for high affinity box (R1) and ε_{weak} for low affinity boxes (the others). Interacting DnaA domain III pairs was determined by our previous work (1). More specifically, in left-half simulations we set AICG2+ structure-based contact potentials to form

DnaA(R1)$_{\text{Arg-ATP}}$DnaA(R5M)$_{\text{Arg-ATP}}$DnaA(τ2)$_{\text{Arg-ATP}}$DnaA(I1)$_{\text{Arg-ATP}}$DnaA(I2)

where DnaA(R1)$_{\text{Arg-ATP}}$DnaA(R5M) means Arg finger side of DnaA(R1) binds ATP side of DnaA(R5M).

- Simulation B4: For middle–right-half subcomplex, we included six DnaA domain III-IVs and wild type right-half *oriC* fragment (Table S6).

  For the initial structure, we conducted simulations in the same way as the left-half and made DnaA's binding on DnaA boxes in order of R4, C1, I3, C2, C3, R2. For the second step, we only used the flexible model of DnaA. During the middle–right-half simulations we fixed R2-side terminal base pair. Other setups are the same as those in the left-half subcomplex.

  In the production runs, ε^{IJ} were ε_{strong} for high and middle affinity boxes (R4, and R2) and ε_{weak} for low affinity boxes (the others). In middle–right-half simulations we set AICG2+ structure-based contact potential to form

  DnaA(R4)$_{\text{Arg-ATP}}$DnaA(C1)$_{\text{Arg-ATP}}$DnaA(I3)$_{\text{Arg-ATP}}$DnaA(C2)$_{\text{Arg-ATP}}$DnaA(C3)$_{\text{Arg-ATP}}$DnaA(R2).

  The other setups are identical to those in Simulation B3. We conducted the right-half production runs with two settings. One is a setting with attractive interaction between DnaA on C3 box and DnaA on R2 box. In the other case, we did not include the attractive force between these two DnaA molecules.
- Simulation B5: We also conducted simulations of DnaA domain III-IV and mutant *oriC (*and IHF for left-half). For the left-half subcomplex, we simulated the left-half R1IBS-d5 mutant. This mutant *oriC* contains 5-bps deletion between R1 box and IHF box (see supplemental Table S6). Initial structures of production run were generated using initial structures for wildtype left-half production runs. Then we conducted the second stage of the initial structure preparation (10^7-MD step simulations) and obtained 10 different structures. Production runs were conducted with the same conditions as Simulation B3. Flexible model of DnaA was used. We conducted 1×10^8MD steps simulations 10 times with different random forces.
- Simulation B6: For the middle–right-half subcomplex, we simulated 5 *oriC* mutants (R2C3Δ5, R2C3Δ10, R2C3Δ18, R2C3+4, R2C3+10). By simulations, we tested whether DnaA domain III hexamer forms or not. Simulation settings were the same as those in the wildtype case. In the same way as wild-type, we conducted the right-half production runs with two settings (with or without attractive interaction between DnaA on C3 box and DnaA on R2 box).
- Simulation B7: Using the complex structure formed in Simulation B4 and B6, we investigated the dynamics of DnaA domain III oligomer on wild-type or mutant *oriC's*, For this, we conducted 2.0×10^8MD steps of wild-type or mutant right-half *oriC* – DnaA domain III-IV complex simulations. Initial structures were complex structures appeared in Simulation B4 and B6. We conducted 9 simulations changing random seed for Langevin dynamics for

each setting. We discarded the first 5×10^7MD steps in the analysis. Simulation conditions such as potentials and fixed particles were the same as Simulation B4.

- Simulation B8: For preparation of fully atomistic simulations, we also simulated the DnaA-IHF-minimal *oriC* complex including the left-half and the middle–right-half *oriC*. Initial structure was prepared by connecting the left-half complex and middle–right-half complex. Simulation was 5×10^5MD steps. The settings were the same as complex formation simulations. The coordinates were recorded in every 10^3 MD steps.
- Simulation B9: For preparation of fully atomistic simulations, we performed simulated annealing and made local structure of the complex near the reference structure. Simulations were 1×10^4 MD steps. By dropping the temperature by 1.495 K 200 times, the temperature decreased linearly from 300 K to1 K. The settings were the same as complex formation simulations. For the DnaA-IHF-minimal *oriC* complex, we selected a snapshot whose minimum Q_{box} among 11 DnaA box was the highest during simulation B8 as initial structure. We defined Q_{box} as the fraction of intermolecular native contacts of a DnaA domain IV- DnaA box. For the DnaA domain III-hexameric complex on the middle–right-half *oriC*, we selected a snapshot whose Q_{R2} (box is R2 box) is the highest among 19 snapshots (SI table. S3).

Preparation of *E. coli* replication initiation complex structure for atomistic MD simulation

We transformed coarse-grained (CG) representation of *E. coli* initiation complex into an atomistic model. For DNA we superposed fully atomistic nucleotides (dAMP, dGMP, dCMP, dTMP) on corresponding coarse-grained nucleotides. We superposed the center of mass of phosphate, center of mass of sugar and N1 (for adenine, guanine) or N5 (for cytosine, thymine) on the CG model. Fully atomistic models for nucleotides were generated by NAB as B-type DNA structure model (38, 39). For the 5' end nucleotide, we generated fully atomistic models of nucleotide with adjoining phosphate and superposed them on the sugar particle and base particle of the corresponding nucleotides and phosphate particle of neighboring nucleotide.

We used the one to one threading method of Phyre2 server and transformed the CG representation of DnaA domain III-IV into an atomistic model (45). When we fitted the atomistic DnaA domain III-IV on the CG DnaA pentamer model, there was no space sufficient for ATP in the ATP binding site. Based on the *E. coli* ATP-DnaA tetramer we prepared above, we prepared fully atomistic pentameric-, hexameric-, or monomeric- ATP-Mg^{2+}-DnaA K135-L366, which we call domain III core model. We superposed the domain III core model on the Phyre2 model. For each DnaA protomer we determined a first boundary amino acid, which is close to L366 on the amino acid sequence and the gap between the coordinates in Phyre2 model and that in the domain III core model is small. From K135 to the first boundary amino acid, we regarded the superposed domain III core model as the transferred model.

For the DnaA domain IV, we superposed DnaA domain IV (V374-S467) crystal structure (36) on the Phyre2 model and determined the second boundary amino acid in the same way as domain III. From the next amino acid of the first boundary amino acid to the second boundary amino acid, we regarded the Phyre2 model as the transferred atomistic model. For the others, we regarded the superposed crystal structure as the transferred atomistic model. Two selenomethionines in the crystal structure were changed into the methionines by PyMOL (31). When a part of the complex

overlapped sterically, we remodeled the sidechain using PyMOL(31) or SCWRL4 (46).

For IHF, we superposed the IHF native structure (prepared above) on the CG model and replaced the CG model with fully atomistic model.

Because the whole complex was too large, we divided the atomistic complex into left-half part (contains R1 box -I2 box part) and right-half part (contain R2 box-R4 box part) and conducted MD simulation for each halves (see supplemental table S6).

Analysis

The molecule structures were visualized by PyMOL(31) or VMD(47). Figures of molecules were generated by PyMOL.

In order to describe formation of the replication initiation complex, we defined $Q_{III\text{-}III}$, Q_{box} as the fraction of intermolecular native contacts of DnaA domain III-domain III, DnaA domain IV- DNA (DnaA box), respectively. When two heavy atoms belong to different molecules and they are closer than 6.5 Å at the native structure, we consider the CG particle pair which the heavy atoms belong to are in intermolecular native contact. When the distance of natively contacting CG particle pair was smaller than 1.2 times of the distance in the native structure in a snapshot, we considered the contact of the pair was formed in the snapshot. We considered the nucleoprotein complexes which meets the following two criteria as complete complex.

1) $Q_{III\text{-}III} > 0.99$ for all DnaA domain III - DnaA domain III
2) $Q_{box} > 0.1$ for all DnaA box

We also defined the integer numbers N_{pp} and N_{pd}. The N_{pp} represents the number of DnaA domain III-domain III pairs whose $Q_{III\text{-}III} > 0.99$. The N_{pd} represents the number of DnaA domain IV-DnaA box pairs whose $Q_{box} > 0.1$. The criteria of pentameric domain III complex in Simulation B4 with attractive interaction between DnaA on C3 box and DnaA on R2 box are as follows.

1) $Q_{III\text{-}III} > 0.99$ for DnaA domain III - DnaA domain III except for between DnaA on C3 box and DnaA on R2 box.
2) $Q_{III\text{-}III} = 0$ for between DnaA on C3 box and DnaA on R2 box.
3) $Q_{box} > 0.1$ for all DnaA box

For Figs. 3B, 5, 8, S1-S8, and S13, we analyzed only the snapshots which met the criteria for the complete complex or which met $Q_{box} > 0.1$ in Simulation B2.

In describing the orientation of DnaA domain IV on τ2 box, we defined $\theta_{DnaA\text{-}\tau2}$ ($0° < \theta_{DnaA\text{-}\tau2} < 180°$). This is an angle between two vectors. One is a vector with the DnaA L422 being the start point and A404 being the end point, and the other is a vector along the τ2l box. The latter starts from the geometric center of sugar and base particles of AG base pairs in τ2l box (5'-AGgatcacc-3') (shown in capital letters), and ends at the geometric center of sugar and base particles of CC base pairs in τ2 a box (underlined in sequence).

In Fig. S4, in determine the orientation of DnaA domain IV on each DnaA box, we defined $\theta_{DnaA\text{-}box}$ ($0° < \theta_{DnaA\text{-}box} <$

180°) by similar definition as $\theta_{DnaA\text{-}\tau 2}$. Box indicates each DnaA box. A vector along each DnaA box starts from the geometric center of sugar and base particles of the first 2 base pairs of each DnaA box, and ends at the geometric center of sugar and base particles of the last 2 base pairs. We determined the orientation of DnaA domain IV based on whether $\theta_{DnaA\text{-}box} < 90°$ or not.

In Fig. S8B, for all pairs of DnaA boxes, we analyzed conformations of bound DnaAs. We analyzed the trajectories in Simulation B3 and B4. For each pair of DnaA boxes, we first superimposed the I376-S467 of domain IV for all the snapshots on the native structure. Then, the distance between the geometric centers of two domains III is calculated. This calculation was performed for all the snapshot pairs and their average was calculated.

For measuring the rotation of the right-half DnaA pentamer by mutation in *oriC*, we defined $\varphi_{complex}$ ($0° < \varphi_{complex} < 180°$). We calculated this angle as follows. First we obtained the vector **r4** from each structure wherein the start point is the geometric center of Cα atom of I3-bound DnaA K135-L366 and end point is that of DnaA domain III on R4 box. In the same way as **r4**, we also obtained vector **c3**, where in the end point is the geometric center of Cα atom of DnaA domain III core on C3 box. Next we calculated the outer product of **r4** and **c3** for each structure.

We calculated the angle of two outer products. One is from a snapshot in target complex in Simulation B7. The other outer product is obtained from a snapshot in wild-type simulation of Simulation B7. We calculated the angles for all pairs of mutant-complex snapshots – wild-type-complex snapshots. Finally, we obtained $\varphi_{complex}$ as an average over all pairs. For control, we also calculated this value between wild-type snapshots and wild-type snapshots.

In Simulation B3, B4, B5, and B6, the representative structure for each subcomplex was chosen as follows. First, we selected a typical structure in each trajectory in the following way. In each trajectory, using all snapshots with the complete initiation complex we calculated pairwise RMSDs of the DnaA oligomer. Here, we used K135-L366 and I376-S467 of DnaA for the RMSD calculations because L367-T375 was set to flexible. For each snapshot, we averaged its RMSD to all the other snapshots obtaining the average RMSD from the snapshot. We chose the typical structure of each trajectory as the snapshot with the smallest average RMSD. Second, using thus obtained typical snapshots in all trajectories, we calculated pairwise RMSDs in the same way as above. We chose the typical structure of entire ensemble as the snapshot with the smallest average RMSD.

In Fig S4, the representative structure of each type of complex was chosen as follows. We calculated pairwise RMSDs of the K135-L366 and I376-S467 DnaA pentamer. For each snapshot, we averaged its RMSD to all the other snapshots obtaining the average RMSD from the snapshot. We chose the representative structure as the snapshot with the smallest average RMSD.

In Simulation B7, the representative structure for each subcomplex was chosen as follows. We calculated the geometrical center of DnaA K135-L366 of DnaA pentamer (DnaA binding to R4, C1, I3, C2, or C3 box). Then we calculated the mean coordinate of the geometric center of all subcomplexs. We chose the typical structure as the structure whose geometrical center of DnaA K135-L366 of DnaA pentamer is nearest to the mean coordinate.

Error bars of histograms express sample standard deviation of trajectories (Fig 3B, 5A, 8B, S3, and S6). In calculating the sample standard deviation, each trajectory was weighted by the number of snapshots in which DnaA proteins and DNA forms subcomplex (Fig 3B, 5A, S3, and S6). In Fig 8B, error bars of histograms express sample

standard deviation of 9 individual experiments.

After atomistic simulations, we connected the left-half subcomplex and the middle–right-half subcomplex. When we superposed two structures, we used PyMOL (31).

We made figures using gnuplot (48) and inkscape (49).

Supporting references

1. Noguchi Y, Sakiyama Y, Kawakami H, Katayama T (2015) The Arg fingers of key DnaA protomers are oriented inward within the replication origin *oriC* and stimulate DnaA subcomplexes in the initiation complex. *J Biol Chem* 290(33):20295–20312.
2. Keyamura K, et al. (2007) The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP-DnaA-specific initiation complexes. *Genes Dev* 21(16):2083–2099.
3. Keyamura K, Abe Y, Higashi M, Ueda T, Katayama T (2009) DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *J Biol Chem* 284(37):25038–25050.
4. Sekimizu K, Bramhill D, Kornberg A (1987) ATP activates DnaA protein in initiating replication of plasmids bearing the origin of the *Escherichia coli* chromosome. *Cell* 50(2):259–265.
5. Kasho K, Katayama T (2013) DnaA binding locus *datA* promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc Natl Acad Sci U S A* 110(3):936–941.
6. Nishida S, et al. (2002) A nucleotide switch in the *Escherichia coli* DnaA protein initiates chromosomal replication: evidence from a mutant DnaA protein defective in regulatory ATP hydrolysis *in vitro* and *in vivo*. *J Biol Chem* 277(17):14986–14995.
7. Kenzaki H, et al. (2011) CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J Chem Theory Comput* 7(6):1979–1989.
8. Li W, Terakawa T, Wang W, Takada S (2012) Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc Natl Acad Sci U S A* 109(44):17789–17794.
9. Terakawa T, Takada S (2011) Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys J* 101(6):1450–1458.
10. Sambriski EJ, Schwartz DC, de Pablo JJ (2009) A mesoscale model of DNA and its renaturation. *Biophys J* 96(5):1675–1690.
11. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298(5):937–953.
12. Kenzaki H, Takada S (2015) Partial unwrapping and histone tail dynamics in nucleosome revealed by coarse-grained molecular simulations. *PLOS Comput Biol* 11(8):1004443.
13. Pronk S, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *BIOINFORMATICS* 29(7):845–854.
14. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced,

and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447.

15. van der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26(16):1701–1718.
16. Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7(8):306–317.
17. Berendsen HJC, van der Spoel D, Vandrunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 91(1–3):43–56.
18. Pall S, Abraham MJ, Kutzner C, Hess B, Lindahl E (2015) Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *SOLVING SOFTWARE CHALLENGES FOR EXASCALE*, Lecture Notes in Computer Science., ed Markidis, S and Laure E, pp 3–27.
19. Abraham MJ, et al. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1:19–25.
20. Bondi A (1964) van der Waals Volumes and Radii. *J Phys Chem* 68(3):441-451.
21. Guy AT, Piggot TJ, Khalid S (2012) Single-stranded DNA within nanopores: Conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys J* 103(5):1028–1036.
22. Hornak V, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *PROTEINS-STRUCTURE Funct Bioinforma* 65(3):712–725.
23. Sorin EJ, Pande VS (2005) Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J* 88(4):2472–2493.
24. Perez A, et al. (2007) Refinenement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys J* 92(11):3817–3829.
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
26. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593.
27. Goga N, Rzepiela AJ, de Vries AH, Marrink SJ, Berendsen HJC (2012) Efficient algorithms for langevin and DPD dynamics. *J Chem Theory Comput* 8(10):3637–3649.
28. Parrinello M, Rahman A (1981) Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J Appl Phys* 52(12):7182–7190.
29. Hess B, Bekker H, Berendsen HJC, Fraaije J (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18(12):1463–1472.
30. Miyamoto S, Kollman PA (1992) SETTLE - An analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* 13(8):952–962.
31. Schrödinger LLC (2015) *The PyMOL Molecular Graphics System, Version 1.7*.
32. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
33. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738.
34. Yang J, et al. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12(1):7–8.

35. Ozaki S, et al. (2008) A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J Biol Chem* 283(13):8351–8362.
36. Fujikawa N, et al. (2003) Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res* 31(8):2077–2086.
37. Duderstadt KE, Chuang K, Berger JM (2011) DNA stretching by bacterial initiators promotes replication origin opening. *Nature* 478(7368):209-213.
38. Case DA, et al. (2012) *AMBER 13* (University of California, San Francisco).
39. Macke TJ, Case DA (1998) Modeling unusual nucleic acid structures. *MOLECULAR MODELING OF NUCLEIC ACIDS*, ACS SYMPOSIUM SERIES., ed Leontis, NB and SantaLucia, J, pp 379–393.
40. Rice PA, Yang SW, Mizuuchi K, Nash HA (1996) Crystal structure of an IHF-DNA complex: A protein-induced DNA u-turn. *Cell* 87(7):1295–1306.
41. Ozaki S, Katayama T (2009) DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62(2):71–82.
42. Craig NL, Nash HA (1984) *Escherichia coli* integration host factor binds to specific sites in DNA. *Cell* 39(3):707–716.
43. Kenzaki H, et al. (2013) *CafeMol 2.1 manual*.
44. Schaper S, Messer W (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* 270(29):17622–17626.
45. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858.
46. Krivov GG, Shapovalov M V, Dunbrack Jr. RL (2009) Improved prediction of protein side-chain conformations with SCWRL4. *PROTEINS-STRUCTURE Funct Bioinforma* 77(4):778–795.
47. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph Model* 14(1):33–38.
48. Williams T, et al. (2012) *Gnuplot 4.6: an interactive plotting program*.
49. The Inkscape Team (2003-2016) *INKSCAPE*.
50. Felczak MM, Simmons LA, Kaguni JM (2005) An essential tryptophan of *Escherichia coli* DnaA protein functions in oligomerization at the *E. coli* replication origin. *J Biol Chem* 280(26):24627–24633.
51. Arias-Palomo E, O'Shea VL, Hood I V, Berger JM (2013) The Bacterial DnaC Helicase Loader Is a DnaB Ring Breaker. *Cell* 153(2):438–448.
52. Abe Y, et al. (2007) Structure and function of DnaA N-terminal domains - Specific sites and mechanisms in inter-DnaA interaction and in DnaB helicase loading on *oriC*. *J Biol Chem* 282(24):17816–17827.
53. Kaguni JM (2011) Replication initiation at the *Escherichia coli* chromosomal origin. *Curr Opin Chem Biol* 15(5):606–613.

Table S1 DnaA box

	Sequence (5' - 3')
R1	**TTATCCACA**
R5M	**T**C**ATTCACA**
τ2 l	AGGA**TCAC**C
τ2 r	G**T**GA**TC**CTG
I1	**TTATAC**GGT
I2	C**T**GA**TC**C**CA**
R2	**TTATACACA**
C3	**TT**G**TTC**TTT
C2	A**TA**A**C**T**AC**C
I3	**TT**GA**TC**CA**A**
C1	**TT**CC**T**G**ACA**
R4	**TTATCCACA**

Identical bases to the 9-mer consensus are indicated by red.

Table S2 Statistics of left-half subcomplex formation simulations

τ2 alignment	**L367-T375**	Accumulative Time(10^4 MD steps)
τ2 l	Rigid	0
τ2 l	Flexible	36,418
τ2 r	Rigid	0
τ2 r	flexible	2,818

We recorded snapshots by10^4 MD step and summarized the number of snapshots which satisfy the complete complex criteria.

Table S3 Statistics of right-half subcomplex formation simulations

DNA fragment	DnaA origomerization	Trajecrtories	Accumulative Time (10^4 MD steps)
WT *oriC*	Pentamer	9	61,561
		(without attractive potential between DnaAs bound to C3 and R2 boxes)	
WT *oriC*	Pentamer	7	19,827
		(with attractive potential between DnaAs bound to C3 and R2 boxes)	
WT *oriC*	Hexamer	3	19
R2C3Δ5	Pentamer	10	68,707
R2C3Δ5	Hexamer	3	910
R2C3Δ10	Pentamer	10	68,630
R2C3Δ10	Hexamer	9	52,232
R2C3Δ18	Pentamer	10	69,070
R2C3Δ18	Hexamer	9	59,074
R2C3+4	Pentamer	10	66,484
R2C3+4	Hexamer	1	3,852
R2C3+10	Pentamer	9	56,725
R2C3+10	Hexamer	0	0

We recorded snapshots by 10^4 MD step and summarized the number of snapshots which satisfy the complete complex criteria.

Table S4 List of DNA primers

Name	Sequence
tau2sub-1	CGATCATTCACAGTTAATGATCCTTTC
tau2sub-2	GTGATGGTGGACCGTATAAGCTGGGATCAG
tau2sub-3	CTCATCCTGGACCGTATAAGCTGGGATCAG
oriC3-f	GTTGTTCTTTGGATAACTACCG
oriC3-f0	TTGTTCTTTGGATAACTACCG
oriR2-r	TTGTGTATAACCCCTCATTCTG
oriR2-r-15	TGAGTTGTGTATAACCCCTC
oriR2-r-10	GTTTTTGAGTTGTGTATAACC
oriR2C3+4-r	AGGGTGTTGTTCAGTTTTTGAG
oriR2C3+10-r	ACACCTAGGGTGTTGTTCAGTTTTTGAG
R2-C3 f	GGGTTATACACAACTCAAAAACTGAACAACAGTTGTTCTTTGGA
R2-C3 r	TCCAAAGAACAACTGTTGTTCAGTTTTTGAGTTGTGTATAACCC
R2R1-C3 f	GGGTTATCCACAGCTCAAAAACTGAACAACAGTTGTTCTTTGGA
R2R1-C3 r	TCCAAAGAACAACTGTTGTTCAGTTTTTGAGCTGTGGATAACCC
R2-C3delta5 f	GGGTTATACACAACTCAAAAACTGAACTTGTTCTTTGGA
R2-C3delta5 r	TCCAAAGAACAAGTTCAGTTTTTGAGTTGTGTATAACCC
R2-C3delta10 f	GGGTTATACACAACTCAAAAACTTGTTCTTTGGA
R2-C3delta10 r	TCCAAAGAACAAGTTTTTGAGTTGTGTATAACCC
R2-C3delta18 f	GGGTTATACACAAGTTGTTCTTTGGA
R2-C3delta18 r	TCCAAAGAACAACTTGTGTATAACCC
ori-2	CAAATAAGTATACAGATCGTGCG
oriR2R4	GAATGAGGGGTTATACACAACTC
R2R4tmaR2f	GAATGAGGGGAAACCTACCACC

Table S5 Dissociation constants in DnaA domain IV – 21 base pair DNA system

$\varepsilon^{IJ}_{DnaA-DNA\ interaction}$ (kcal/mol)	$k_d\ (M)$
0	3.98×10^{-4}
	4.18×10^{-4}
0.270	3.08×10^{-5}
	3.01×10^{-5}
0.285	6.15×10^{-6}
	1.16×10^{-5}
0.300	2.54×10^{-6}
	7.37×10^{-7}
0.315	1.51×10^{-7}
	4.08×10^{-7}
0.330	3.10×10^{-8}
	2.18×10^{-7}
0.345	2.98×10^{-9}
	2.49×10^{-8}
0.360	–
	1.27×10^{-8}
0.375	1.14×10^{-8}
	–
0.390	–
	–

Detail of simulations are described in supporting methods.

For each $\varepsilon^{IJ}_{DnaA-DNA\ interaction}$, we conducted 2 simulations.

When dissociation of protein from DNA was not observed, hyphen is specified.

Table S6 List of DNA fragments we used in simulations

	Sequence (5' - 3')
Left-half (R1-I2)	TGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGGAAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGG
	ATCAG
R1IBSd5	TGCCCTGTGGATAAC-----TCCGGCTTTTAAGATCAACAACCTGGAAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGG
	ATCAG
Middle—right-half WT*oriC*	AGGGGTTATACACAACTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGTAGA
R2C3A5	AGGGGTTATACACAACTCAAAAACTGAAC-----TTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGTAGA
R2C3A10	AGGGGTTATACACAACTCAAAAAC----------TTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGTAGA
R2C3A18	AGGGGTTATACACAA----------GTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGTAGA
R2C3+10	AGGGGTTATACACAACTCAAAAACTGAACAACACCCTAGGTGTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCAC
	AGTAGA
R2C3+4	AGGGGTTATACACAACTCAAAAACTGAACAACACCCTGTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGTAGA
Left-half (atomistic MD)	TGCCCTGTGGATAACAAGGATCCGGCTTTTAAGATCAACAACCTGGAAAGGATCATTAACTGTGAATGATCGGTGATCCTGGACCGTATAAGCTGGG
	ATCAGAATGAGGGGTTA
Middle—right-half (atomistic MD)	CAGAATGAGGGGTTATACACAACTCAAAAACTGAACAACAGTTGTTCTTTGGATAACTACCGGTTGATCCAAGCTTCCTGACAGAGTTATCCACAGT
	AGA

R2 boxes are red, C3 box are blue, insertion sequences between R2 box and C3 box are underlined, and deletion sequences are shown by hyphen, respectively.

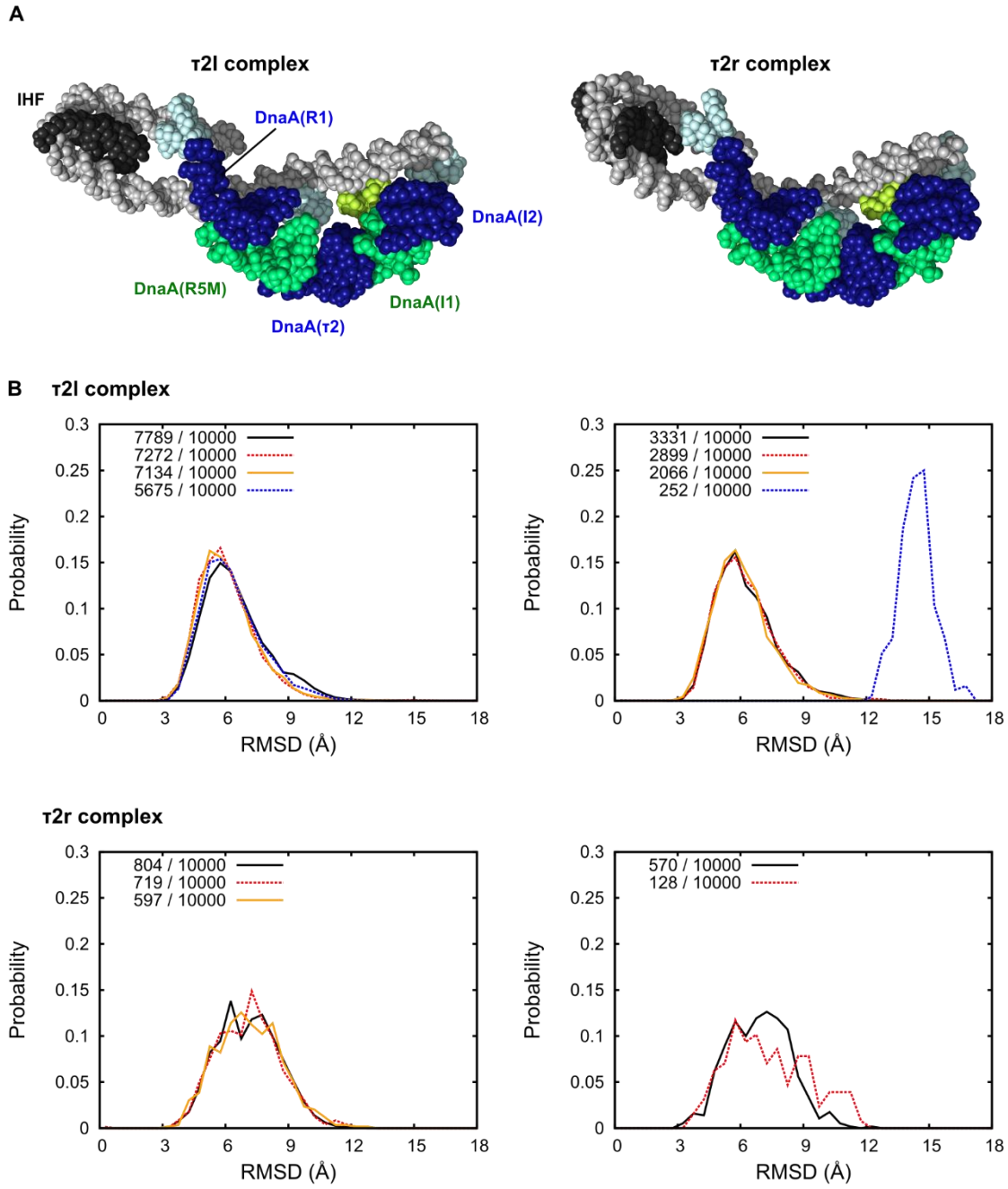


Figure S1. Structural variations in the left-half subcomplex. A) A representative structure of left-half subcomplexes. The structure of τ2l case (left) and τ2r case (right). DNA in grey. IHF in black. ATP-bound DnaA is drawn in alternative choices of two color schemes; one in dark blue (domain III) and light blue (domain IV), and the other in green (domain III) and yellowish green (domain IV). B) After superposing on representative structure, we calculated the RMSD of each left-half subcomplex and plotted its distribution by trajectory. We calculated RMSD only for the complex structures which satisfy the complex formation criteria. For RMSD calculation we used only K135-L366 and I376-S467 parts of DnaA. The top two panels are for the τ2l setting, while the bottom two panels are for the τ2r setting. Each curve represents

results from one trajectory that fulfilled the criteria of complex formation. Results for eight and five trajectories in the top and bottom lines, respectively, are divided into two panels merely for clarity. On the left top of the panels, the number of snapshots which satisfied the complex formation criteria was described beside the graph color. For each simulation, we output totally 10,000 snapshots.

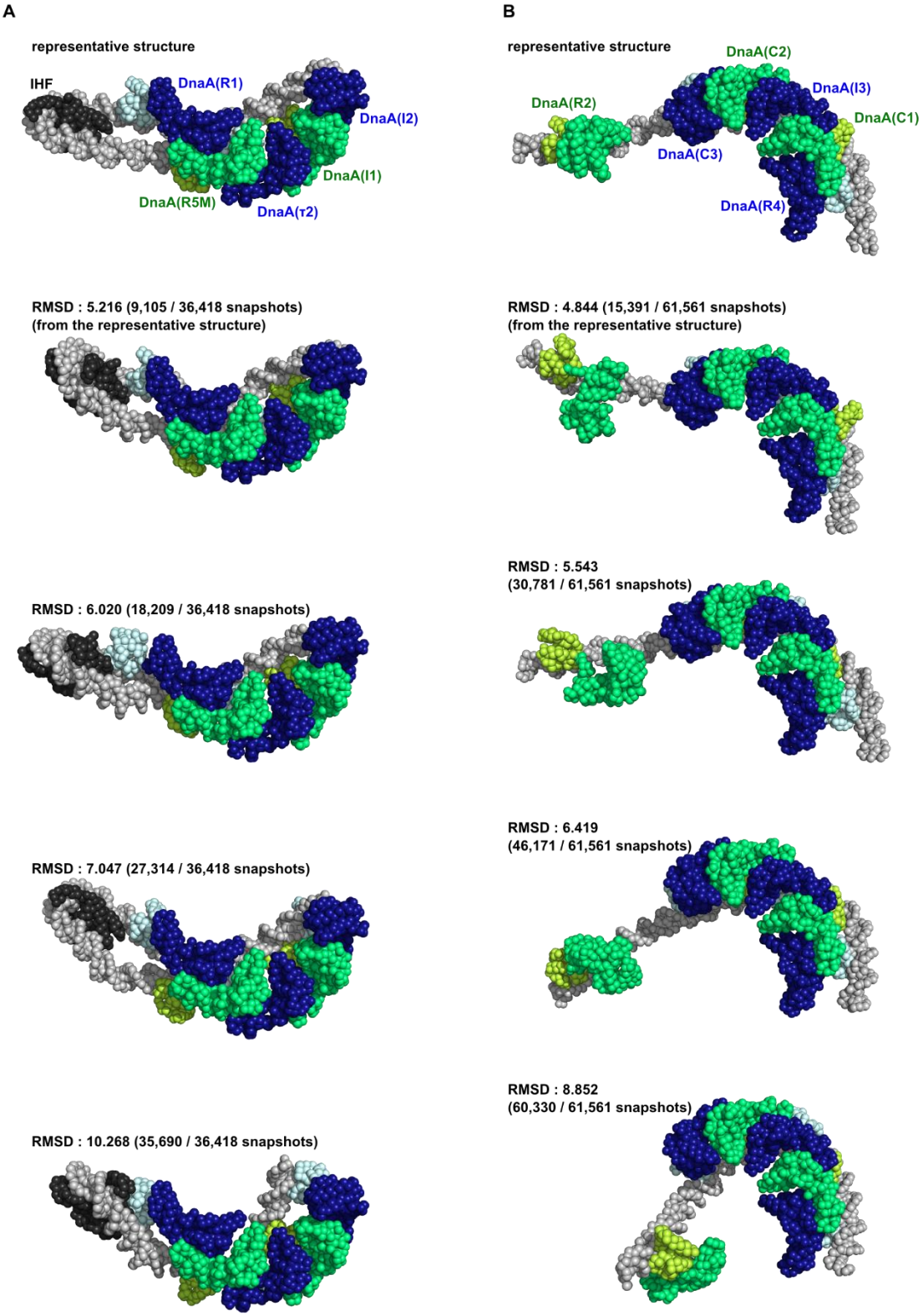


Fig. S2 Snapshots of left-half and middle–right-half subcomplexes. Several subcomplex structures satisfying the criteria of complex formation are shown. The left-half subcomplexes are shown in A, and the middle-right-half

subcomplexes are shown in B. The representative structures of subcomplexes (the same ones as Fig. 2) are shown at the top. For others, in addition to snapshots we described the RMSD of K135-L366 and I376-S467 parts of DnaA pentamer from the representative structure. We ordered snapshots by the similarity to the representative structure using this RMSD value, and the order of each snapshot is also shown. The models are spatially aligned by the K135-L366 and I376-S467 parts of DnaA pentamer. The color scheme is identical to Fig. S1. In simulations, all the structures of the left-half subcomplex and the middle–right-half subcomplexes were similar.

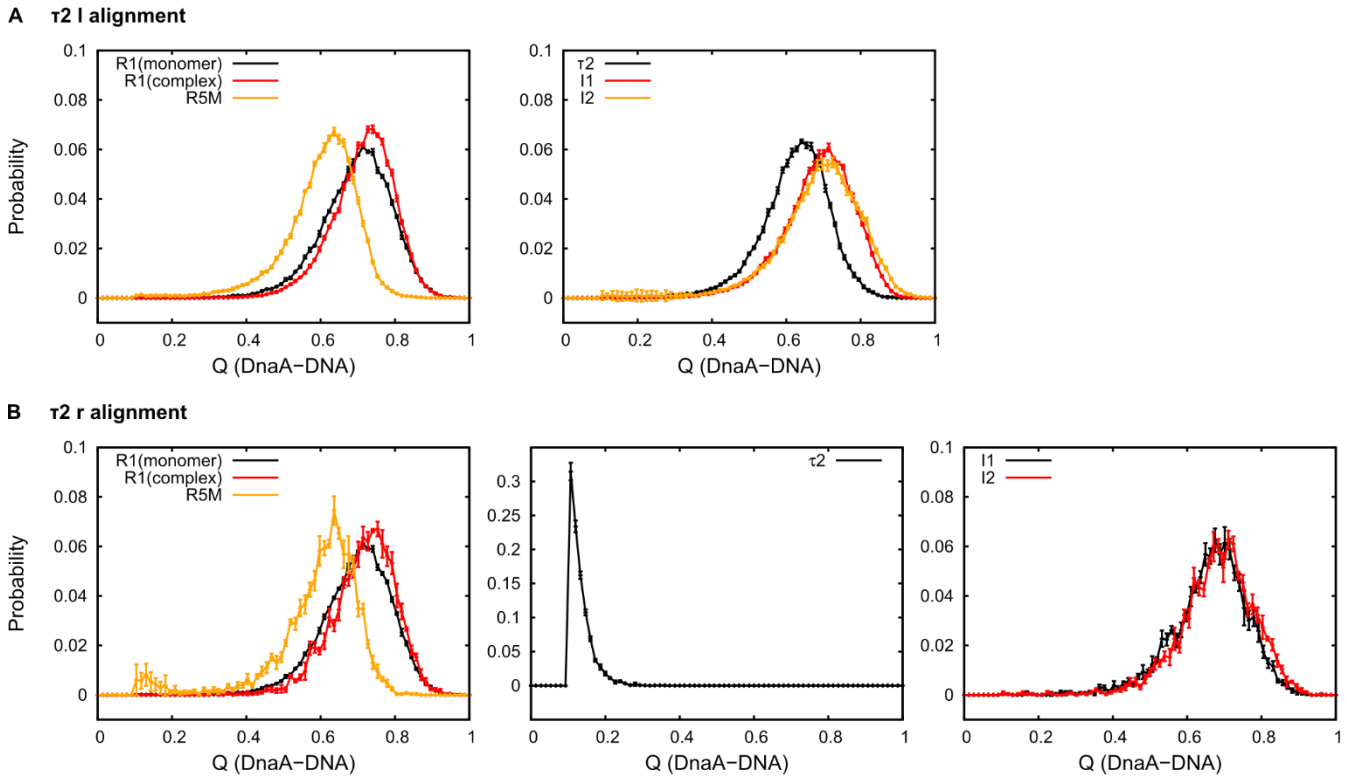


Figure S3. The *Q*-score distributions of DnaA domain IV-DNA in the formed DnaA subcomplex. *Q*-score distributions for the interface between each DnaA box and DnaA domain IV appeared in the left-half subcomplex are plotted. For comparison, *Q*-score distribution for the interface between DnaA domain IV and R1 box in the system of monomeric DnaA domain III-IV and 13 bp DNA is also shown (R1(monomer)). A) The left-half subcomplex with the set of τ2l alignment B) The left-half subcomplex with the set of τ2r alignment. Error bars express sample standard deviations.

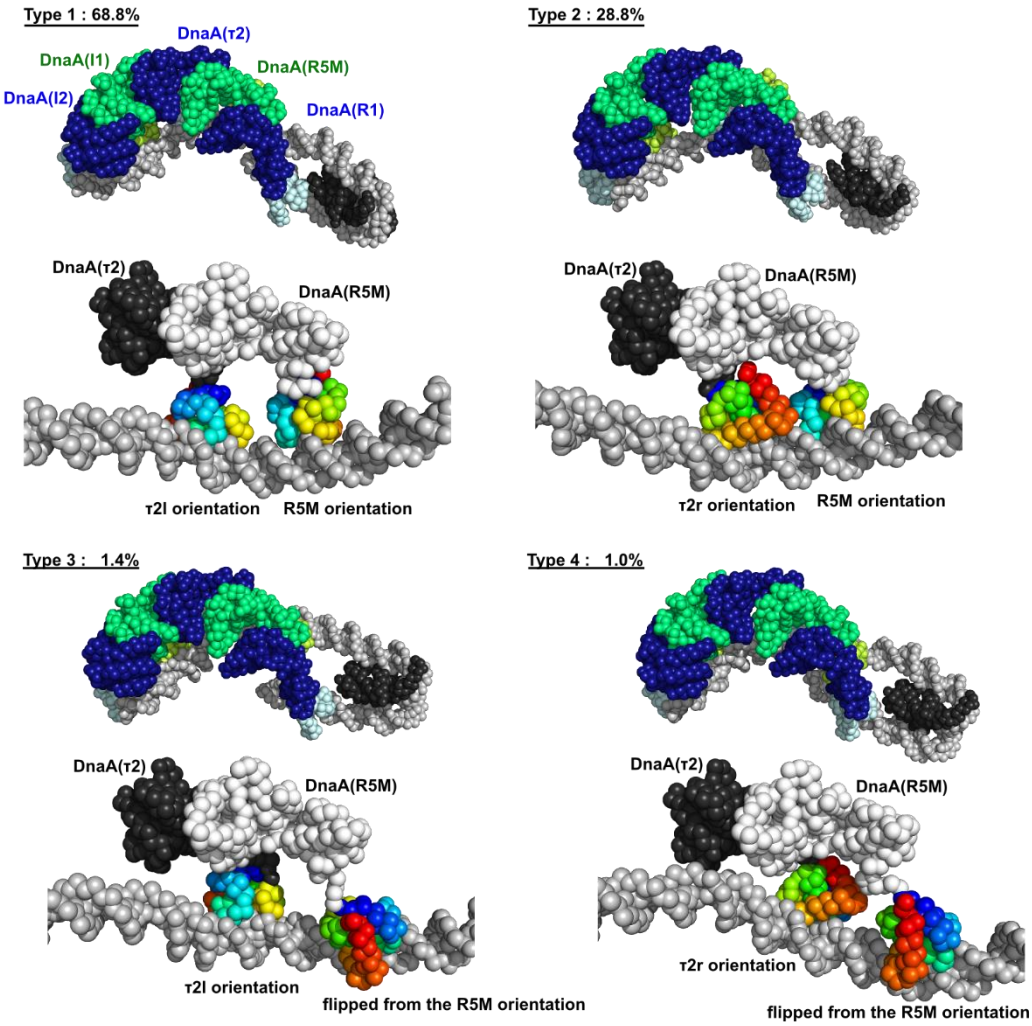


Fig S4 The domain IV flipping in complexes with τ2r setting. From the point of view of the DnaA domain IV orientation, we classified left-half subcomplexes with τ2r setting satisfying the criteria of complex formation into 4 types. **Type 1 :** The orientation of DnaA domain IV at R5M box is the same as reference structure (crystal structure (36)), and the orientation of DnaA domain IV at τ2 box is the same as the τ2l setting. This type was 68.8% of 2,818 snapshots. **Type 2 :** The orientation of DnaA domain IV at R5M box is the same as reference structure and the orientation of DnaA domain IV at τ2 box is the same as the τ2r setting. Type 2 was 28.8% of 2,818 snapshots. **Type 3 :** The DnaA domain IV at R5M box flipped from the reference structure, and the orientation of DnaA domain IV at τ2 box is the same as the τ2l setting. Type 3 was 1.4% of 2,818 snapshots. **Type 4 :** The DnaA domain IV at R5M box flipped from the reference structure, and the orientation of DnaA domain IV at τ2 box is the same as the τ2r setting. Type 4 was 1.0% of 2,818 snapshots.

The representative structures of the left-half subcomplex of each type are shown in the upper part. The color scheme is identical to Fig. S1. The structures around R5M-bound DnaA and τ2-bound DnaA are shown in the lower part. DNA is in gray, DnaA domain III of R5M-bound DnaA in white, DnaA domain III of τ2-bound DnaA in black. Each DnaA domain IV is colored by rainbow, from blue (V374) to red (S467). When R5M-bound DnaA domain IV flip or τ2-bound

DnaA domain IV has τ2r orientation, The C-terminal part of domain IV (colored orange to red) is seen on the front side. For other boxes (I-boxes, R1 box), τ2l-setting, or DnaA pentamer of WT middle–right-half subcomplex, the frequencies of flipping DnaA domain IV were less than 0.3%.

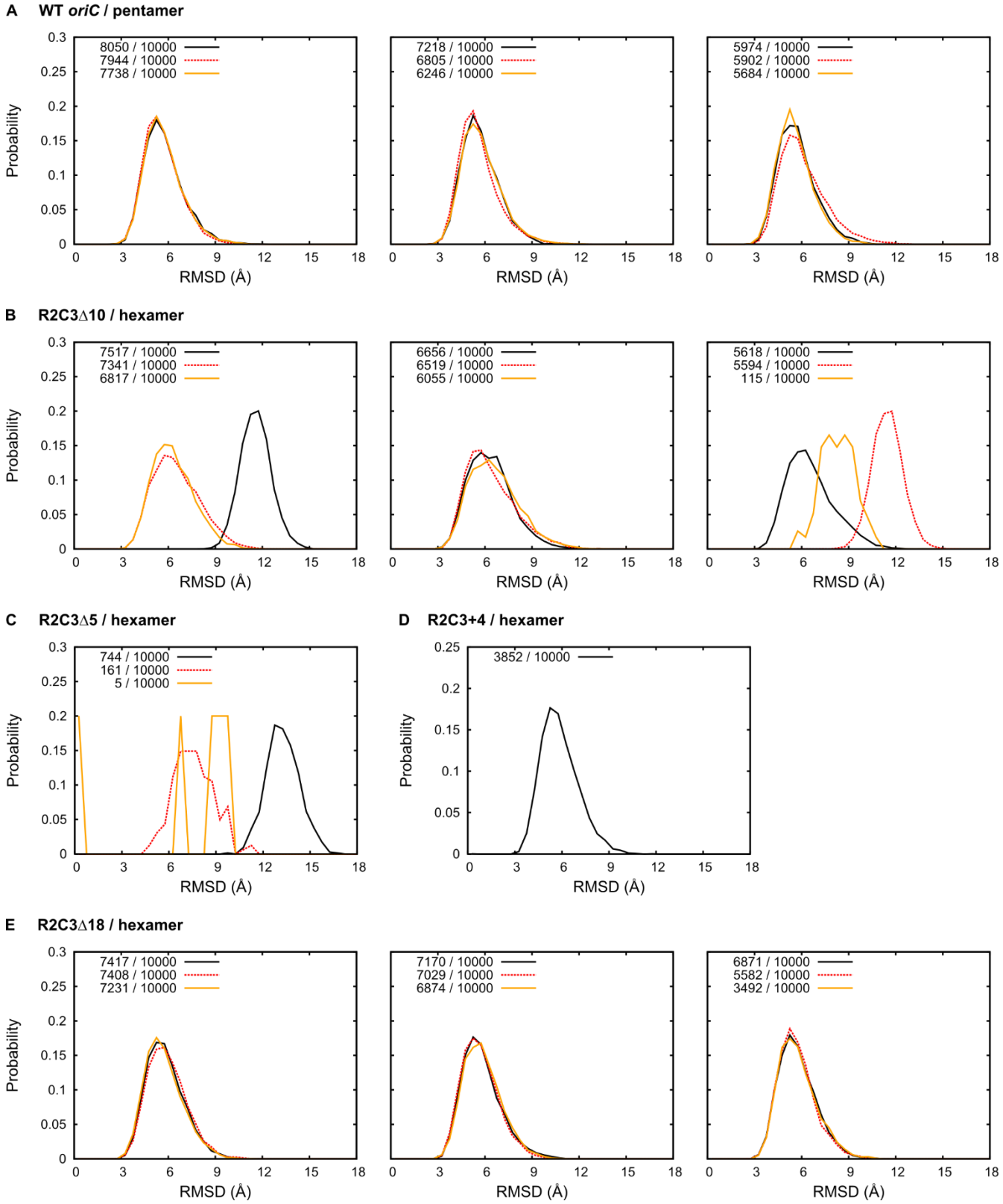


Figure S5. Structural variations in the middle-right-half subcomplex. A-E) Analysis of oligomeric DnaA domain III

formation simulations. After superposing on a representative structure, we calculated the RMSD of each right-half subcomplex and plotted its distribution by trajectory. Each curve represents one trajectory in which fulfilled the criteria of complex formation. For WT (A), R2C3Δ10 mutant (B), and R2C3Δ18 mutant (E), we divided trajectories into three panels merely for clarity. On the left top of the panels, the number of snapshots which satisfied the complex formation criteria was described beside the graph color. For each simulation, we output totally 10,000 snapshots. We calculated RMSD only for complex structure which satisfy complex formation criterion. For the RMSD calculation we used only K135-L366 and I376-S467 parts of DnaA.

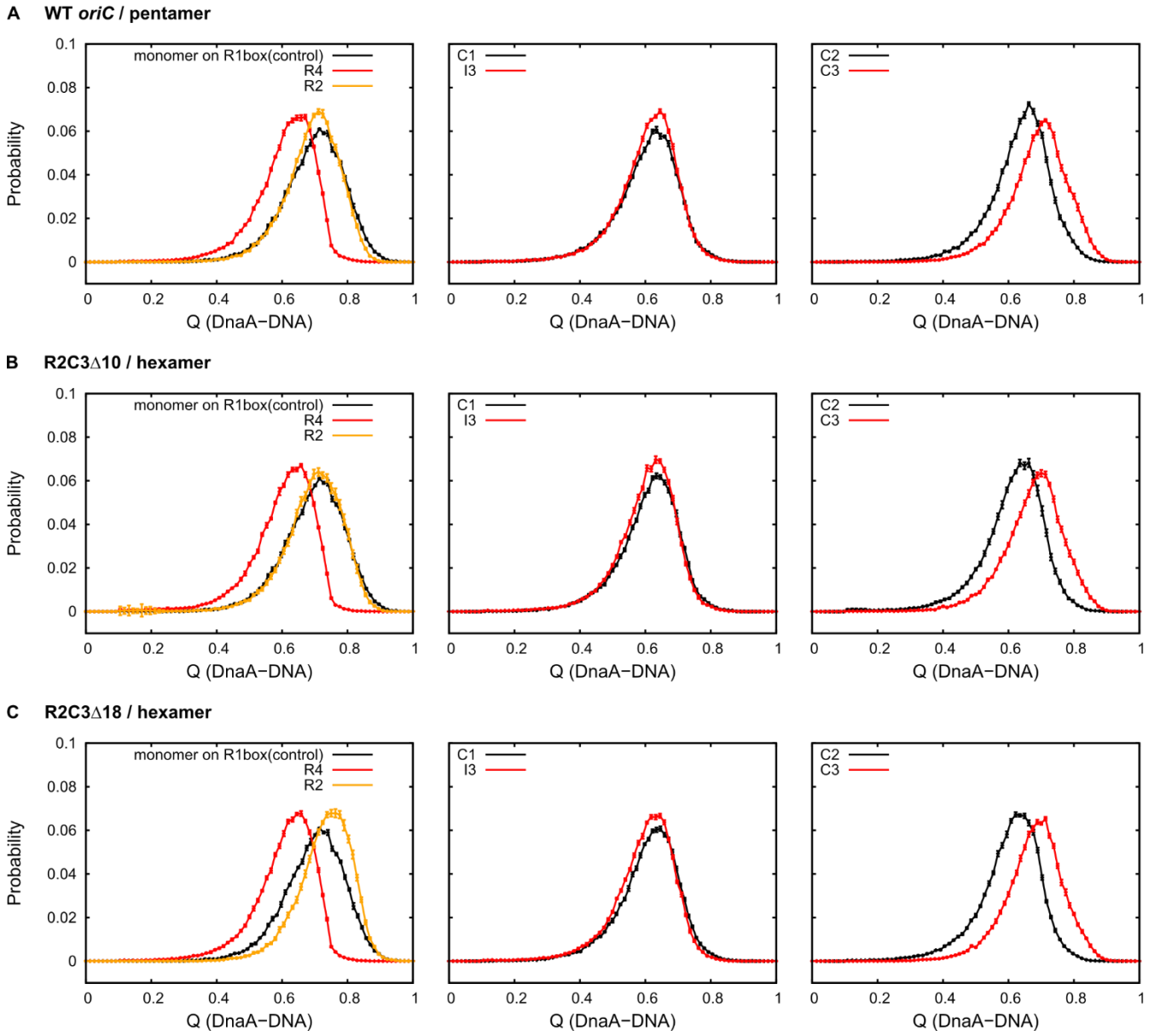


Figure S6 *Q*-score distribution for the interface between DnaA domain IV-DNA in *oriC* bearing altered R2-C3 spacing. The *Q*-score distributions in the middle–right-half subcomplexes formation simulations are analyzed. For comparison, we also plotted the *Q*-score distribution of DnaA domain IV and R1 box in the system of monomeric DnaA domain III-IV and 13 bp DNA (left black). Error bars express sample standard deviations.

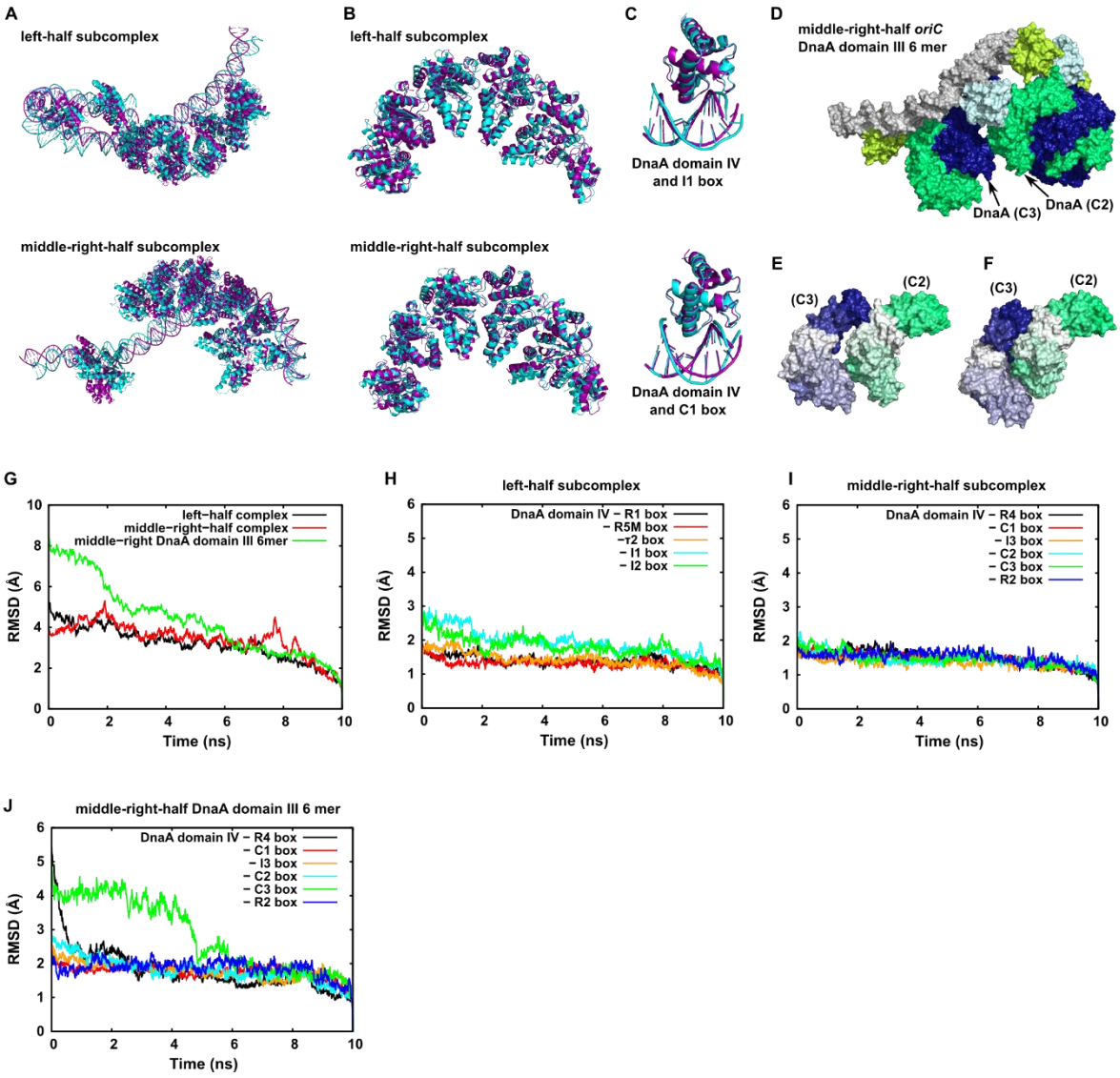


Figure S7. Structural change of each subcomplexes during 10 ns short atomistic MD simulations. A-C) Comparison between the structure before simulation (purple) and the structure after simulation (cyan). A) Entire left-half subcomplex (top), and middle–right-half subcomplex (bottom). Since DNA and hinge region between DnaA domain III and IV (L367-T375) are rather flexible, the entire structure of each subcomplex has changed to a certain extent. However the overall structures did not change much. B) The DnaA K135-L366 pentamer in the left-half subcomplex (top), and middle–right-half subcomplex (bottom). These contain DnaA domain III pentamer except the region we set flexible loop in the coarse-grained simulations. In the case of both left-half subcomplex and middle–right-half subcomplex, the overall structure almost unchanged. C) Complexes of DnaA I376-S467 and DnaA boxes. These are the DnaA domain IV-DnaA box complex except the region we set flexible loop in the coarse-grained simulations. The DnaA domain IV-I1 box is shown on top, and the DnaA domain IV-C1 box is shown at the bottom. In both cases, the overall structures did not

change. D, E) Results of the atomistic MD simulation of the DnaA domain III-haxameric complex on middle–right-half *oriC*. This is a control experiment done for an hypothetical assembly. This complex is unstable very rarely observed in coarse-grained simulations. D) Entire structure of the complex after simulation. DnaA bound to C3 box and DnaA bound to C2 box are indicated by allows. The two DnaA domain IIIs almost separated in the simulation. The color scheme is identical to Fig. S1. E) DnaA K135-L366 bound to C3 box and that bound to C2 box. This is the same structure as the structure in panel D. In the DnaA bound to C3 box, color gradually changes from white (K135) to blue (L 366). In the DnaA bound to C2 box, color gradually changes from white (K135) to green (L 366). F) DnaA K135-L366 bound to C3 box and that bound to C2 box in the middle–right-half subcomplex after 10 ns simulation. In this complex, inter-DnaA domain III interaction is maintained. G-J) The time series of RMSDs between the structure at each time and the structure at the end of the simulation. RMSD of DnaA K135-L366 pentamer is shown in (G). RMSDs of DnaA domain IV (I376-S467)-DnaA box complexes in the left-half subcomplex (H), in the middle–right-half subcomplex (I), and in the DnaA domain III-haxameric complex on the middle–right-half *oriC* (J) are shown. In panel G, the RMSD distributed around 4 Å for both the left-half subcomplex and the middle–right-half subcomplex, which is rather small for this size of complex. In contrast, the RMSD of the DnaA domain III-haxameric complex on middle–right-half *oriC* distributed around 7 Å in the first 2 ns. This result suggests the instability of the initial structure. In panels H and I, the RMSD distributes around 2 Å, which suggests that the structures are stable. On the other hand, in panel J, the structure of the DnaA domain IV-R4 or C3 box have changed more dynamically for the incorrect hexameric complex.

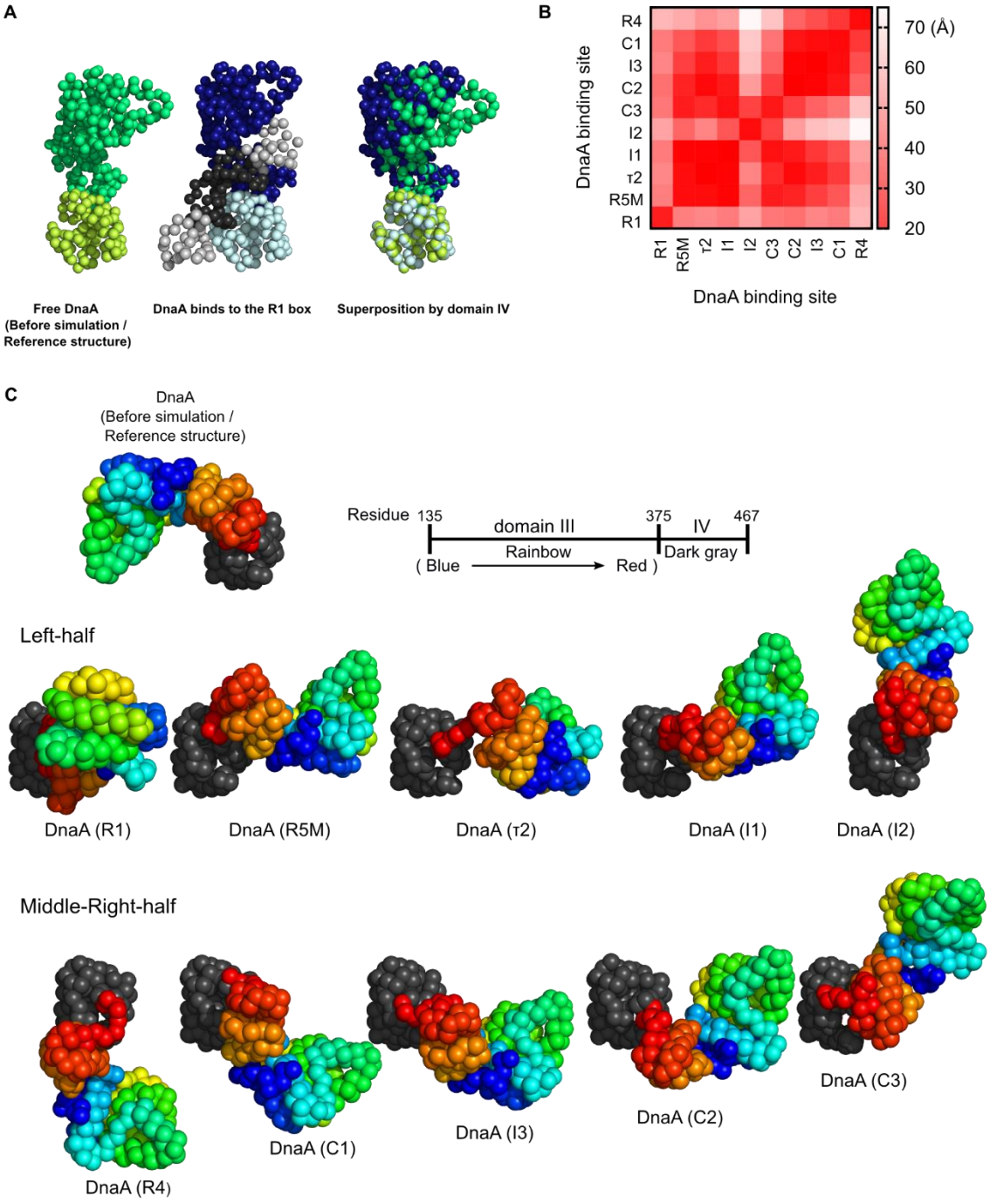


Figure S8. Comparison of conformations of each DnaA domain III-IV protomer in replication initiation complex. A) Inter-domain conformation change in DnaA domain III-IV upon binding to dsDNA. The left structure is reference (native) structure of DnaA domain III-IV. A dsDNA-bound structure of DnaA domain III-IV is shown at middle. By superposing domain IVs of two structures, we compared domain III orientation relative to the domain IV (right). DNA in gray and DnaA box (DNA) in black. DnaA domain III of reference structure in green and domain IV of reference structure in yellowish green. DnaA domain III of the dsDNA-bound structure in dark blue and domain IV of dsDNA-

bound structure in light blue. B) Pairwise comparison of conformations of DnaA molecules. For each pair, after the domain IV is superimposed, the average distance between the center of mass of two domains III is represented by darkness. C) Representative conformations of DnaA at individual boxes. The domain IV (grey) is spatially aligned for all images. The domain III is rainbow colored from blue (K135) to red (L 373) (blue-green-orange-red).

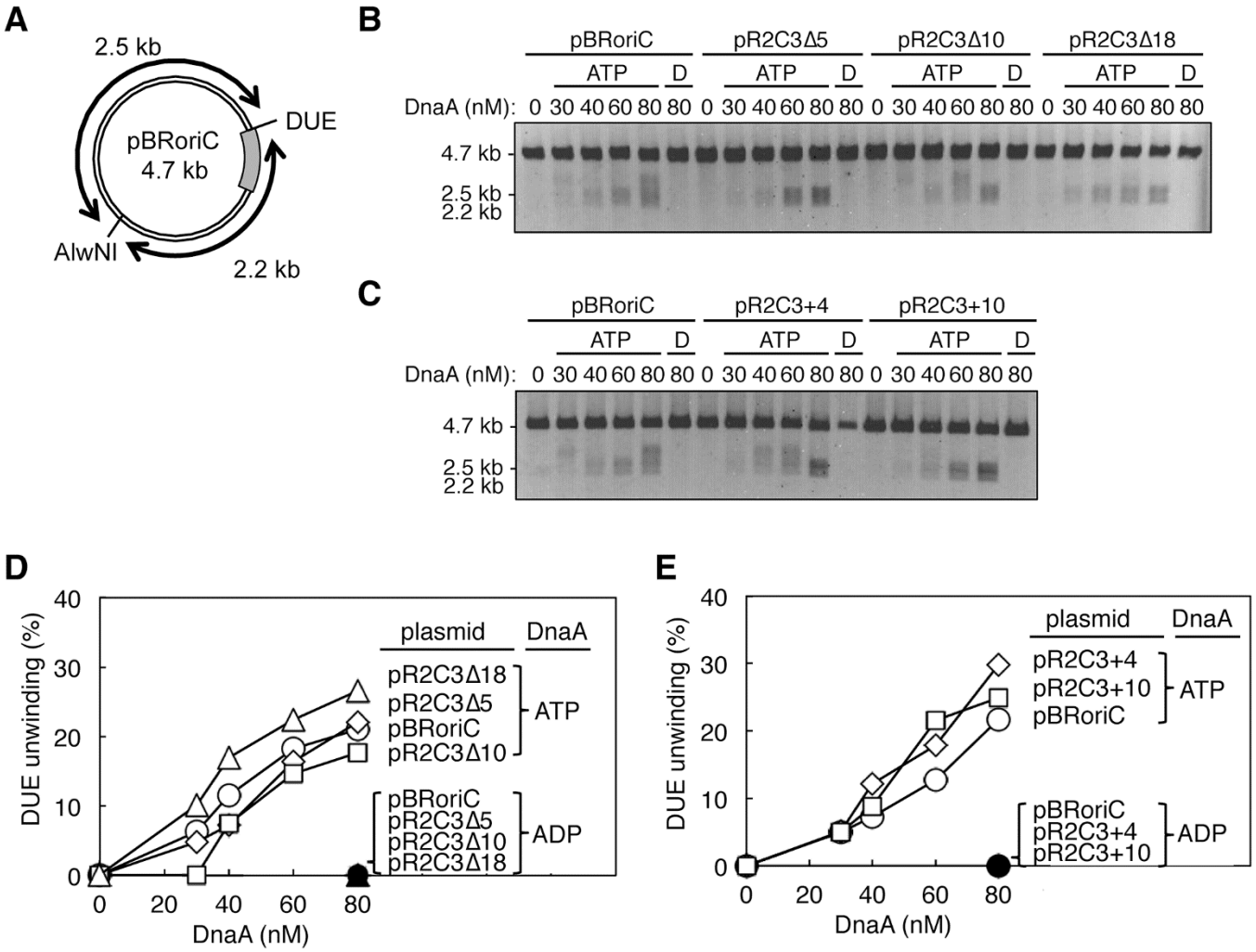


Figure S9. DUE unwinding activity of *oriC* derivatives A) Relevant structure of an *oriC* plasmid pBRoriC. B-D) DUE unwinding assay. This assay was performed as described for Figure 4, using *oriC* plasmids shown in Figure 7. The indicated amounts of ATP/ADP-DnaA were incubated with pBRoriC and its derivatives bearing deletions shown in Figure 7A, followed by digestion using P1 nuclease. The DNA samples were further digested with AlwNI and analyzed as described for Figure 4 (B and C). Percentages of P1 nuclease-digested *oriC* molecules per input DNA were quantified and are shown as 'DUE unwinding (%)'. ATP-DnaA and ADP-DnaA are respectively indicated by open and closed symbols (D and E).

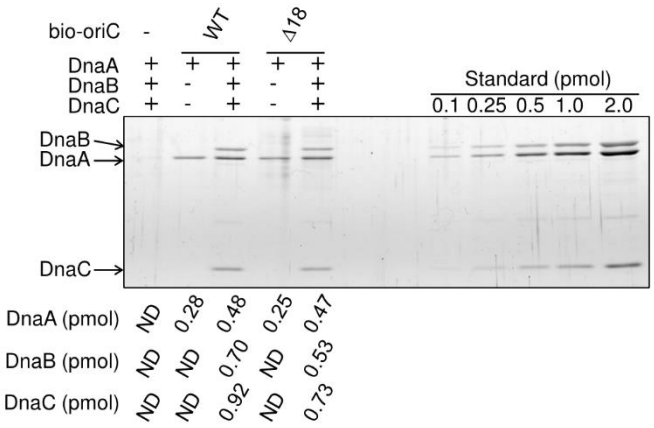
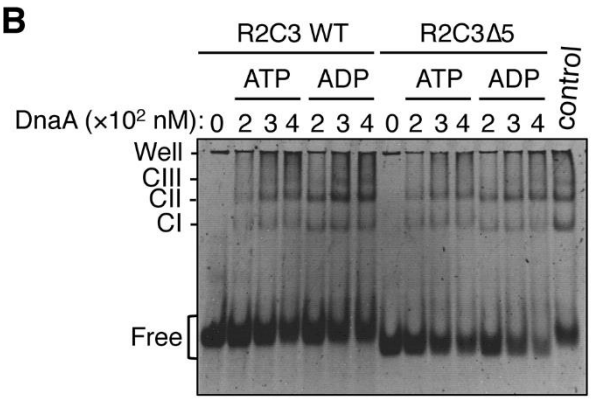


Figure S10. DnaB binding to the *oriC*-DnaA complex. DnaA (5 pmol) was incubated in the presence (+) or absence (-) of DnaB (5 pmol, as monomer), DnaC (5 pmol), and bio-tagged *oriC* DNA (100 fmol) bearing the wild-type sequence (WT) or the Δ18 mutation (Δ18). Proteins bound to the DNA were isolated using streptavidin beads and analyzed using SDS-PAGE and silver staining. Proteins for quantitative standard also were included for SDS-PAGE analysis. Band intensities of the gel image were quantified and the recovered amounts of proteins were deduced using the standards, and indicated below the gel image. Migration positions of each protein are indicated by arrows. ND, not determined.

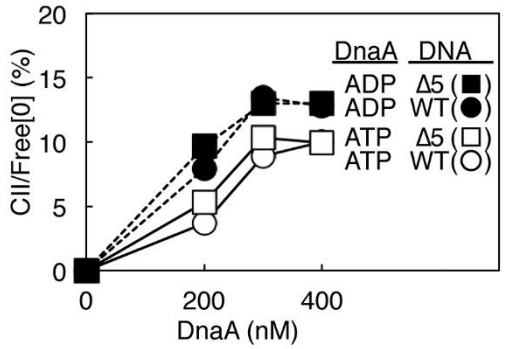
A

```
                    R2 box                                C3 box
R2C3 WT     GGGTTATACACAACTCAAAAACTGAACAACAGTTGTTCTTTGGA (44 bp)
R2-R1C3     GGGTTATCCACAGCTCAAAAACTGAACAACAGTTGTTCTTTGGA (44 bp)
R2C3Δ5      GGGTTATACACAACTCAAAAACTGAACTTGTTCTTTGGA      (39 bp)
R2C3Δ10     GGGTTATACACAACTCAAAAACTTGTTCTTTGGA           (34 bp)
R2C3Δ18     GGGTTATACACAAGTTGTTCTTTGGA                   (26 bp)
```

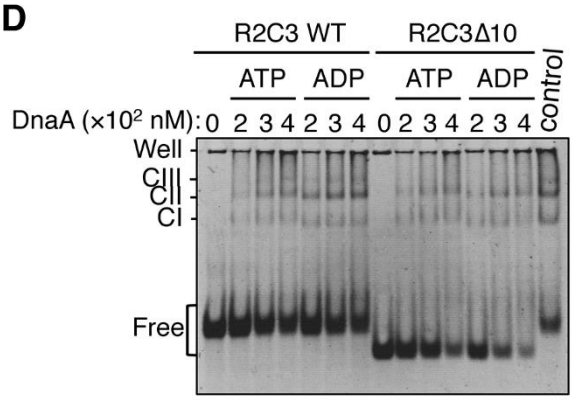
B



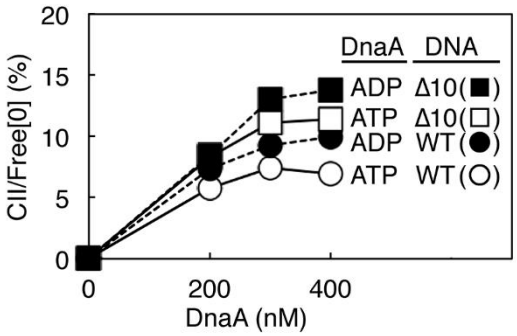
C



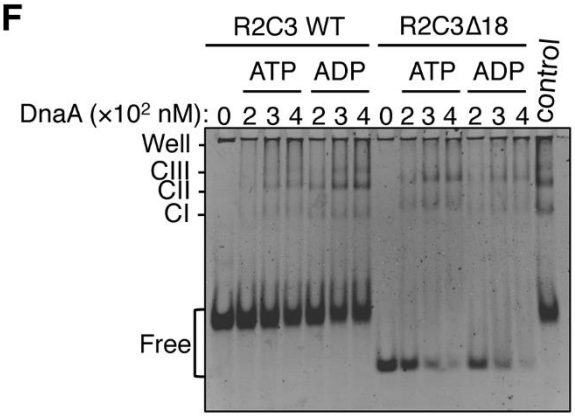
D



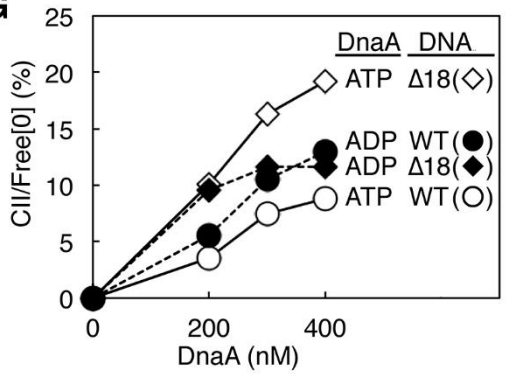
E



F



G



H

R2C3 frag.	CII/Free[0](%) at 400 nM DnaA	
	ATP	ADP
WT	10.8 ± 2.0	17.0 ± 4.0
Δ10	9.3 ± 0.6	11.3 ± 1.7
Δ18	19.3 ± 0.1	11.7 ± 0.1

Figure S11. Binding modes of DnaA to the *oriC* R2-C3 region A) Sequences and sizes of DNA fragments used. Sequences of the R2 and C3 boxes are highlighted by red and cyan letters, respectively. In the R2-R1C3 fragment, the R2 box was substituted with the R1 box sequence as shown by green letters and underlines. DNA used was prepared by annealing of oligonucleotides (Table S4). WT, wild-type B-G) EMSA. The indicated amounts of DnaA were incubated in buffer containing 100 nM DNA fragments, 0.2 μg/ mL λ phage DNA, and 2 mM ATP or ADP, followed by polyacrylamide gel electrophoresis and Gelstar staining. The R2-R1C3 fragment was used as a migration control (*control*). Well; gel well, CI; complex I, CII; complex II, CIII; complex III, Free; DnaA-unbound DNA. λ phage DNA used as a competitor remained in the gel well (B, D, and F). Percentages of CII per free DNA in the absence of DnaA (indicated as Free [0]), were quantified (C, E, and G). The mean values and the errors of CII/Free[0](%) at 400 nM DnaA were shown (n = 2) (H).

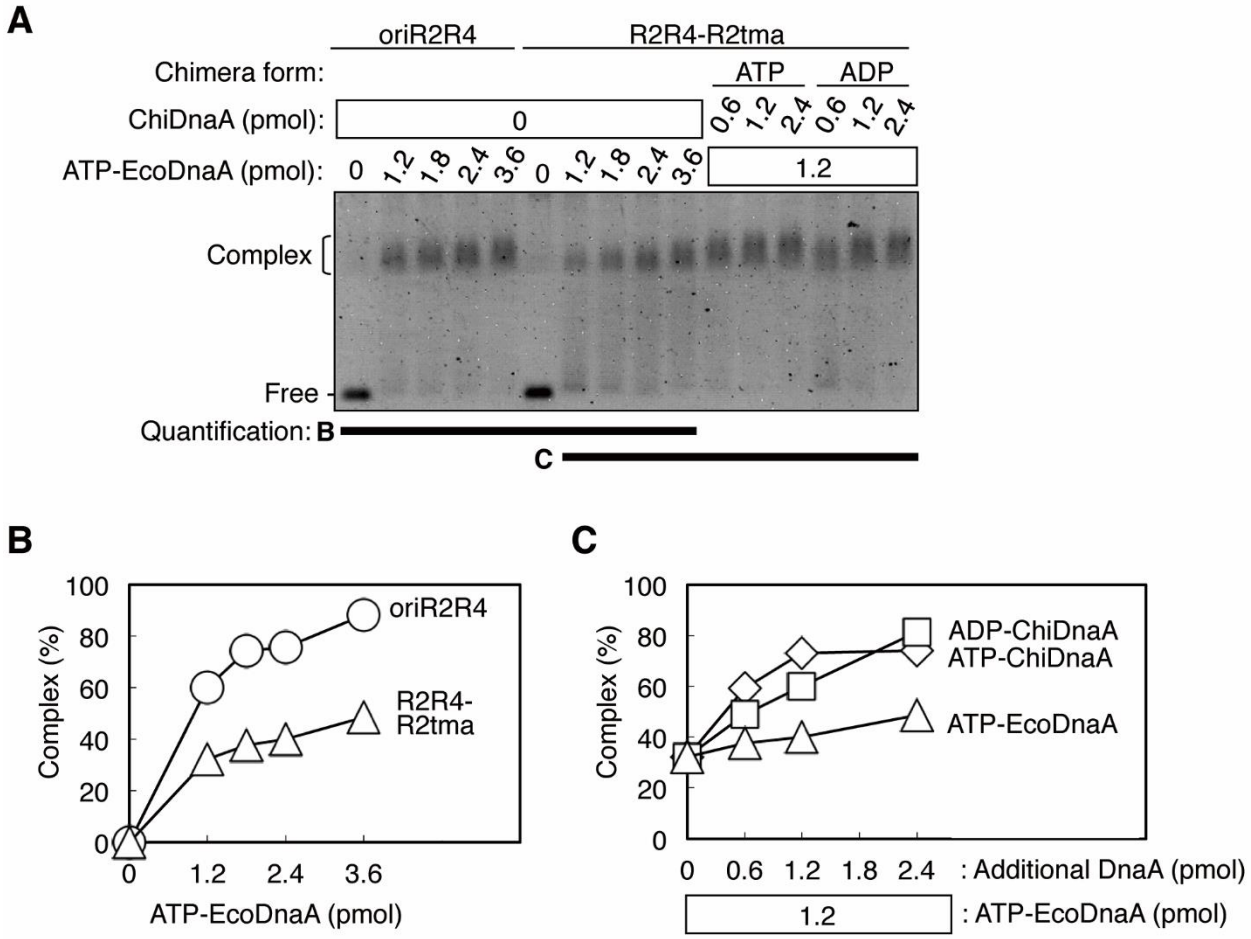


Figure S12. The indicated amounts of ATP-EcoDnaA and ATP/ADP-ChiDnaA were incubated at 30ºC for 10 min in buffer containing 2 mM ATP and the middle and right-half DOR region (oriR2R4) or that bearing the R2 substituted with TmaDnaA box (R2R4-R2tma). DNA fragments used were prepared by PCR amplification using pBRoriC and its derivatives as a template and pairs of primers, ori-2 and oriR2R4 or R2R4tmaR2 f (Table S4). DNA complexes were analyzed by 2% agarose gel electrophoresis and Gelstar staining (A). Percentages of DNA complexed with DnaA to the input DNA were quantified and are shown as 'Complex (%)' (B and C). Data used in each graph are indicated by bold lines in *panel A*.

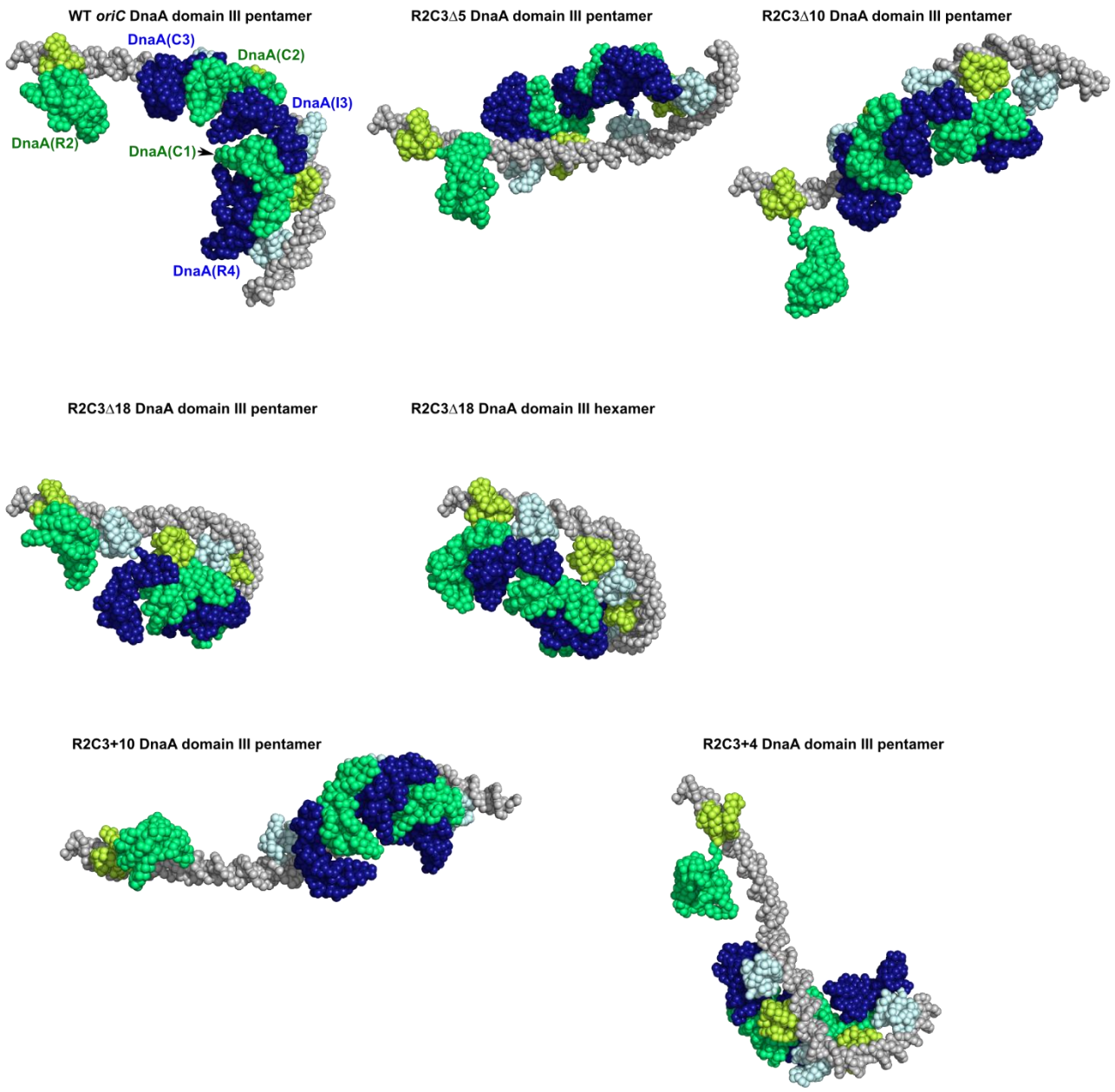


Figure S13. Results of *oriC* mutants simulations. Representative complex models obtained for each of *oriC* mutants bearing altered R2-C3 spacing are drawn. The models are spatially aligned by the left end. The color scheme is identical to Fig. S1. In Fig. 8, we showed the same snapshots as this figure, from a different viewpoint.

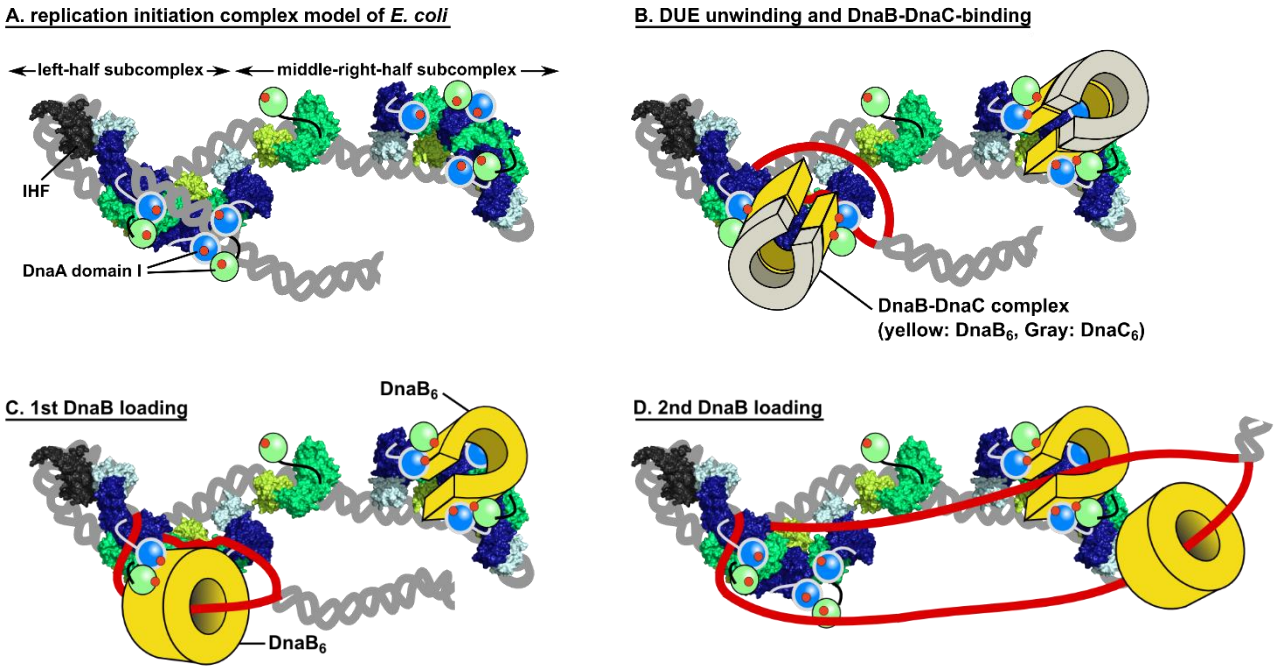


Fig. S14. Model mechanisms of DnaB helicase loading. A) A model of replication initiation complex of *E. coli* is illustrated. Complexes of DnaA domain III-IV are shown based on the present simulations. DnaA domains I and II are simply drawn. DnaA domain I can form homo-dimers (50). B) After DUE unwinding, helical-shaped $DnaB_6$-$DnaC_6$ complexes (51) bind to the left-half and middle–right-half subcomplexes (1). DnaA domain I has a primary DnaB-binding site and multiple DnaA domain I's bind to $DnaB_6$ complex (3, 52). C) $DnaB_6$ bound to the left-half subcomplex is loaded on the unwound DNA region. In this process, N-terminal part of DnaA domain III also may interact with DnaB (53). D) The loaded $DnaB_6$ expands ssDNA region, which allows the second $DnaB_6$ to be loaded on the resultant ssDNA. For DnaA domain III-IV and IHF, the color scheme is identical to Fig. S1. DnaA domain I-II is shown by blue sphere (domain I) and light gray curved line (domain II), or by green sphere (domain I) and black curved line (domain II). DnaB-binding site of DnaA domain I is shown by orange dot. For DNA double-stranded regions are shown by double halical lines colored by gray, and single-stranded regions are shown by red lines, DnaB helicase in yellow and DnaC helicase-loader in gray.