Intrinsic Dynamics of Restriction Endonuclease *Eco*O109I Studied by Molecular Dynamics Simulations and X-Ray Scattering Data Analysis

Tomotaka Oroguchi, Hiroshi Hashimoto, Toshiyuki Shimizu, Mamoru Sato, and Mitsunori Ikeguchi*
International Graduate School of Arts and Sciences, Yokohama City University, Yokohama 230-0045, Japan

ABSTRACT *Eco*O109I is a type II restriction endonuclease that functions as a dimer in solution. Upon DNA binding to the enzyme, the two subunits rotate counterclockwise relative to each other, as the two catalytic domains undergo structural changes to capture the cognate DNA. Using a 150-ns molecular dynamics simulation, we investigated the intrinsic dynamics of the DNA-free enzyme in solution to elucidate the relationship between enzyme dynamics and structural changes. The simulation revealed that the enzyme is considerably flexible, and thus exhibits large fluctuations in the radius of gyration. The small-angle x-ray scattering profile calculated from the simulation, including scattering from explicit hydration water, was in agreement with the experimentally observed profile. Principal component analysis revealed that the major dynamics were represented by the open-close and counterclockwise motions: the former is required for the enzyme to access DNA, whereas the latter corresponds to structural changes upon DNA binding. Furthermore, the intrinsic dynamics in the catalytic domains were consistent with motions capturing the cognate DNA. These results indicate that the structure of *Eco*O109I is intrinsically flexible in the direction of its functional movement, to facilitate effective structural changes for sequence-specific DNA recognition and processing.

INTRODUCTION

Restriction endonucleases (REases) provide antiviral protection for bacteria by degrading the foreign DNA of invading bacteriophages (1). These enzymes recognize specific nucleotide sequences, and cleave both strands of DNA. *Eco*O109I is a type II REase isolated from *Escherichia coli*, and recognizes seven basepairs of double-stranded DNA (ds-DNA). The sequence is RG↓GNCCY, where R = A or G, Y = T or C, and ↓ indicates cleavage position (2). Type II REases are classified into two families, *Eco*RI and *Eco*RV, based on their DNA cleavage pattern. *Eco*O109I is classified into the *Eco*RI family, characterized by the production of 5′-overhang DNA.

The crystal structures of both DNA-free and DNA-bound forms of *Eco*O109I were determined (Fig. 1), and small-angle x-ray solution scattering (SAXS) measurements of the DNA-free form revealed that the homodimer (subunits A and B in Fig. 1 *A*) is the functional unit of the enzyme in solution (2). The subunit of *Eco*O109I consists of two domains: dimerization and catalytic (Fig. 1 *D*). The dimerization domain consists of only α-helices (α1–3 and α9–11), and the catalytic domain, in which the active site for catalysis is located, folds into an α/β structure (α4, β1–2, α6, β3, α7, β4, and α8).

A comparison between the crystal structures of the DNA-free and DNA-bound forms of *Eco*O109I shows that the enzyme undergoes large structural changes upon DNA binding, to allow catalytic residues to interact with DNA. These structural changes can be explained mainly by two movements. The first is the movement of the two catalytic domains in subunits A and B closer together, allowing interactions with DNA (Fig. 1, *A–C*). The second is the twisting motion between the two subunits in a counterclockwise direction, making the two catalytic domains fit together on the major groove of DNA from both sides (Fig. 1, *D–F*). However, in the crystal structure of the DNA-free form, the intervening space between the two catalytic domains is not large enough to be accessed by DNA.

Generally, mechanisms of structural changes in proteins upon substrate binding are explained using the induced-fit model (3), in which substrates induce structural changes in proteins to make the active sites of proteins fit to the substrates. An alternative model to explain structural changes is the preexisting equilibrium model (4), in which protein conformations significantly fluctuate in the substrate-free state, and the equilibrium ensemble before substrate binding includes specific conformations complementary to the substrates as minor components. The substrates then bind selectively to the binding conformations, shifting the equilibrium toward these conformations. This model was supported by experimental studies, e.g., nuclear magnetic resonance relaxation measurements (5–7). Theoretical studies also revealed that protein motions in the substrate-free state correlate well with structural changes upon substrate binding (8–18). In a unified view of the two models, the relationship among protein fluctuations in the substrate-free state, substrate-protein interactions, and structural changes upon substrate binding was formulated using a framework of linear response theory (19). Those studies indicated that intrinsic protein flexibility, evident even in the substrate-free state, is crucial for elucidating the physical mechanisms of functionally important structural changes in proteins. Therefore, we investigated the intrinsic dynamics of *Eco*O109I in the DNA-free state, and examined the

Submitted July 2, 2008, and accepted for publication December 8, 2008.

*Correspondence: ike@tsurumi.yokohama-cu.ac.jp

Editor: Gerhard Hummer.

0006-3495/09/04/2808/15 $2.00

doi: 10.1016/j.bpj.2008.12.3914

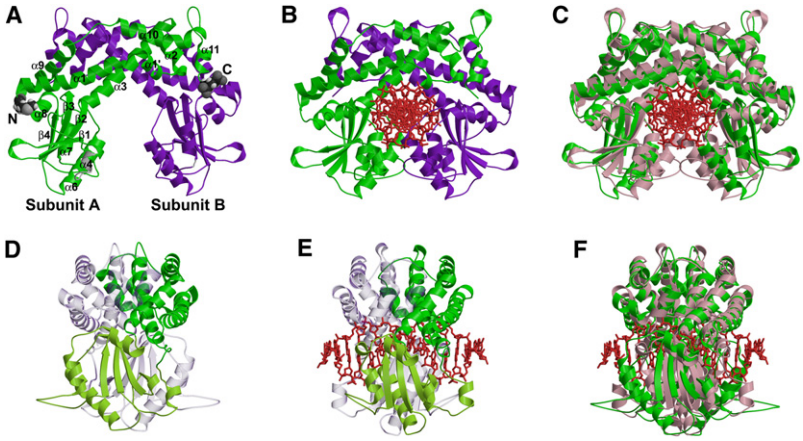

FIGURE 1 Crystal structures of *Eco*O109I. (*A* and *B*) DNA-free and DNA-bound forms, respectively, colored according to subunits A (*green*) and B (*purple*). (*A*) N-terminal and C-terminal residues are rendered in space-filling. Missing residues in crystal structure were modeled (*gray*). Secondary structures $\alpha1$ (residues 2–23), $\alpha1'$ (residues 26–32), $\alpha2$ (residues 36–46), $\alpha3$ (residues 51–75), $\alpha4$ (residues 91–95), $\beta1$ (residues 110–113), $\beta2$ (residues 120–128), $\alpha6$ (residues 133–148), $\beta3$ (residues 157–164), $\alpha7$ (residues 171–178), $\beta4$ (residues 194–198), $\alpha8$ (residues 200–207), $\alpha9$ (residues 211–227), $\alpha10$ (residues 236–252), and $\alpha11$ (residues 262–267) along amino-acid sequence are labeled. (*D* and *E*) Side view of DNA-free and DNA-bound forms, respectively, colored according to dimerization (*green*) and catalytic (*pink*) domains in one subunit, and counterpart subunit (*transparent representation*). (*C* and *F*) Front and side views of superimposition of DNA-free form (*green*) and DNA-bound form (*pink*), respectively.

relationship between the intrinsic dynamics and the structural changes of *Eco*O109I upon DNA binding.

To probe intrinsic dynamics in proteins, all-atom molecular dynamics (MD) simulations are widely used (20,21), and can provide detailed protein motions in solution at an atomic resolution. Structural ensembles obtained by MD simulations should, however, be assessed by means of a comparison with experimental data. Small-angle x-ray solution scattering data are suitable for this purpose (22–28). In fact, whereas x-ray crystallography yields a high-resolution protein structure that is likely to differ from the solution structure, SAXS offers information on the low-resolution structure of proteins in solution (27,28). The structural information of SAXS data involves an average not only over all orientations, but also over all protein conformations. The structural data provided by SAXS are therefore complementary to those of x-ray crystallography. These characteristics of SAXS enabled us to test the structural ensemble obtained by MD simulations by comparing the calculated SAXS profile with the experimentally observed profile.

In this study, we performed a 150-ns MD simulation of the DNA-free form of the enzyme with the explicit water molecules, to elucidate the relationship between the intrinsic dynamics of *Eco*O109I in solution and its structural changes upon DNA binding. From the trajectory of the MD simulation, the SAXS profile, taking scattering from explicit hydration water molecules into account, was calculated and compared with the experimentally observed SAXS profile. The time-resolved analysis of the radius of gyration estimated from the Guinier plot (29) demonstrates that the structure of *Eco*O109I is considerably flexible. Analyses of atomic fluctuations within the subunit and in the dimer identified two rigid cores that undergo rigid-body motions relative to each other. The motions of the two rigid cores were characterized using principal component analysis, and were compared with the DNA-induced structural changes observed in a previous crystallographic study (2). Motions in the catalytic domains were also examined, using principal component analysis (30–32). Determinants of the intrinsic flexibility in *Eco*O109I were pursued using interresidue contact networks in the MD simulation.

METHODS

MD simulations

All simulations were performed at 298 K, using MARBLE (33). The CHARMM22 potential (34) and the TIP3P model (35) were used for the proteins and water, respectively. The periodic boundary condition was applied using the particle mesh Ewald method (36). The symplectic integrator for rigid bodies was used with a time step of 2 fs, in which water and the CH_x, NH_x ($x = 1, 2, 3$), SH, and OH groups were treated as rigid bodies.

The initial structure for the simulation of *Eco*O109I was generated from the crystal structure (Protein Data Bank [PDB] accession code 1WTD) (2). The coordinates for residues 93–100 of subunit A and residues 178–182 of subunit B (Fig. 1, *gray*), which are missing in the PDB coordinates, were modeled after corresponding residues of the other subunit in the crystal structure. The protein and crystal water molecules were immersed in a water box, and 14 Na^+ ions were added in bulk solution, followed by an energy minimization of 5000 steps of the whole system. The total number of atoms was 108,249. To equilibrate the size of the simulation box, a 100-ps run was performed under the NPT condition (33,37), at $T = 298.15$ K and $P = 1.0$ atm, where the positions of all atoms except solvent atoms were constrained with a harmonic force constant of 1 kcal/mol/Å^2. The cube edge length of the simulation box became ~102.1 Å.

After equilibration of the simulation box, two series of product runs were performed under the NPT condition, at $T = 298.15$ K and $P = 1.0$ atm: 1), a 150-ns run without any restraints; and 2), a 20-ns run with restraints on the positions of all atoms of the protein. The procedure for product run 1 was as follows. First, for the equilibration of the structure of the modeled residues, constraints on the positions of only these residues were gradually removed over a period of 100 ps, followed by further equilibration with constraints on the positions of atoms, except those of the modeled residues, for 5 ns. The remaining constraints were then gradually removed over a period of 100 ps, and the 150-ns run was performed without any constraints. Structures were saved every 1 ps in the product runs.

Calculation of SAXS profile

The experimental data of the SAXS profile were taken from a previous study (2). In those experiments, the SAXS profile of a protein was obtained by the

subtraction of the scattering data of buffer solution from those of protein solution. In our calculations, as suggested by Seki et al. (23), we used a similar procedure, i.e., MD simulations for pure solvent were performed in addition to simulations for the protein solution, and the simulated scattering data for pure solvent were subtracted from those of the protein solution. The procedure of the SAXS-profile calculation from MD simulations is summarized as follows: 1), the average density of the bulk water region "bo" (Fig. 2 *A*), ρ_0, is calculated from the simulation for the protein solution; 2), the MD simulation of a pure solvent system, in which some ions are added to make the average density identical to ρ_0, is performed; and 3) the instantaneous scattering intensity $\tilde{I}(Q)$, where Q is defined as $Q = 4\pi \sin\theta/\lambda$ (2θ is the scattering angle, and λ is the wavelength of x-rays), is calculated for each snapshot in the trajectories of the protein solution and pure solvent systems, using the following equation:

$$\tilde{I}(Q) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| i^l \sum_{k}^{N} f_k(Q) j_l(Qr_k) Y_{lm}^*(\Omega_{\mathbf{r}_k}) \right|^2 - \rho_0 \sum_{k}^{N} f_k(Q) \int_S dS(\mathbf{s}_k \cdot \mathbf{n}) \frac{j_1(Qs_k)}{Qs_k} + \rho_0^2 \int_S dS' \int_S dS \left[\frac{(\mathbf{n}' \cdot \mathbf{n})}{Q^2} \cdot \frac{j_1(Qs')}{Qs'} - \{(\mathbf{s}' \cdot \mathbf{n}')(\mathbf{s}' \cdot \mathbf{n})\} \frac{j_2(Qs')}{(Qs')^2} \right], \quad (1)$$

where N is the number of atoms within in region "v" (Fig. 2 *A*), $\mathbf{r}_k$ and $\Omega_{\mathbf{r}_k}$ are the position and the solid angle in real space of the k-th atom, respectively, and $\mathbf{s}_k$ and $\mathbf{s}'$ denote $\mathbf{r} - \mathbf{r}_k$ and $\mathbf{r} - \mathbf{r}'$, respectively. We use $f_k(Q)$ to denote the atomic form factor of the k-th atom, and j_l and Y_{lm} denote a spherical Bessel function and a spherical harmonic function, respectively. We use S to represent the closed surface of region v, and $\mathbf{n}$ is the unit normal vector outward from the infinitesimal surface element dS of the surface S. The details of the derivation of Eq. 1 are described in the Appendix. Continuing the procedure of the SAXS-profile calculation, 4), $\tilde{I}(Q)$ is averaged over all the snapshots for each system, and 5), the final SAXS profile is obtained by the subtraction of the scattering data of the pure solvent from those of the protein solution.

In our method, the calculation of scattering intensity was performed only for region v, which corresponds to the area within a sphere centered at the center of mass of the protein molecule with radius R (Fig. 2 *A*). In the original method, Seki et al. (23) divided region v into three regions: the protein area, its hydration shell, and the bulk water region, to estimate the contribution of scattering from the hydration water to the total scattering intensity. However, as proven theoretically in the Appendix, the final subtracted scattering intensity, calculated only from region v, exactly equals the experimentally obtained intensity if radius R is sufficiently large so that density fluctuations outside the region v (i.e., region "bo") have no correlation with those in the protein area and its hydration shell. This fact simplifies the calculation procedure. In our calculations, we used the sphere $R = 50$ Å as region v, which is sufficiently large for the correct evaluation of the scattering intensity of *Eco*O109I (see Appendix for details).

In addition to the approach described above, we also used a multipole expansion (38) for fast computation of the SAXS profile of a snapshot in the MD trajectory. The multipole expansion is used in the program CRYSOL, developed by Svergun et al. (39), and was applied to a trajectory of MD simulation by Merzel and Smith (24). As shown in these studies, the appropriate truncation of the expansion gives the correct evaluation of a SAXS profile (see Appendix for details). The use of multipole expansion and parallelization of the program (see Appendix) enabled us to calculate the SAXS profile for every snapshot in the trajectory of the 150-ns simulation (the total number of snapshots is 150,000), and analyze the dynamics of *Eco*O109I in terms of the time-resolved SAXS profile.

In both the unrestrained 150-ns and restrained 20-ns MD simulations, the average electron density in the bulk water region bo, ρ_0, was 0.3367 Å^{-3}. We also performed a 2-ns MD simulation of a pure-solvent system, including 3 Na^+ and 3 Cl^- ions, in which the average electron density was identical to ρ_0. The scattering intensities for the pure-solvent system were used in subsequent calculations.

RESULTS AND DISCUSSION

A 150-ns MD simulation of the DNA-free form of *Eco*O109I was performed with explicit water molecules at 298 K. The root mean-square displacement of C^α atoms (C^α-RMSD) from the crystal structure of the DNA-free form reached ~5.0 Å within the first 20 ns of the simulation. However, as described below, the C^α-RMSD of individual rigid cores, i.e., the elements of rigid-body motions, was ~1.0 Å during the simulation, indicating that the rigid cores were stable during the simulation, and that there were certain rigid-body motions between them.

Comparison of SAXS profile calculated from MD simulation with experimentally observed profile

To test the validity of the structural ensemble sampled by the MD simulation, we calculated the SAXS profile from all snapshots in the trajectory, taking scattering from the explicit hydration water molecules into account (for details of the calculation, see Methods and Appendix), and compared it with the experimentally observed profile. As shown in Fig. 3 *A*, the simulation-derived profile was in agreement with the experimentally observed SAXS profile.

In the range of $R_gQ < 1.3$, the scattering intensity $I(Q)$ is approximated to the Guinier approximation,

$$I(Q) \simeq I(0) \exp\left(-Q^2 R_g^2/3 \right), \quad (2)$$

where $I(0)$ and R_g are the scattering intensity at $Q = 0$ (zero-angle) and the radius of gyration, respectively, as estimated from the $\ln[I(Q)]$ versus Q^2 plot (Guinier plot) (29). Fig. 3 *B* compares Guinier plots of the simulation-derived and experimentally observed SAXS profiles, showing that the R_g (referred to as "protein-water R_g" or simply "R_g") estimated from the SAXS profile calculated from the structural ensemble sampled by the MD simulation ($R_g = 28.2$ Å) was in agreement with that obtained from the experimentally observed SAXS profile ($R_g = 28.1 \pm 0.3$ Å).

We then calculated the protein-water R_g for the time window of 500 ps, shifting from $t = 0$–150 ns, to investigate

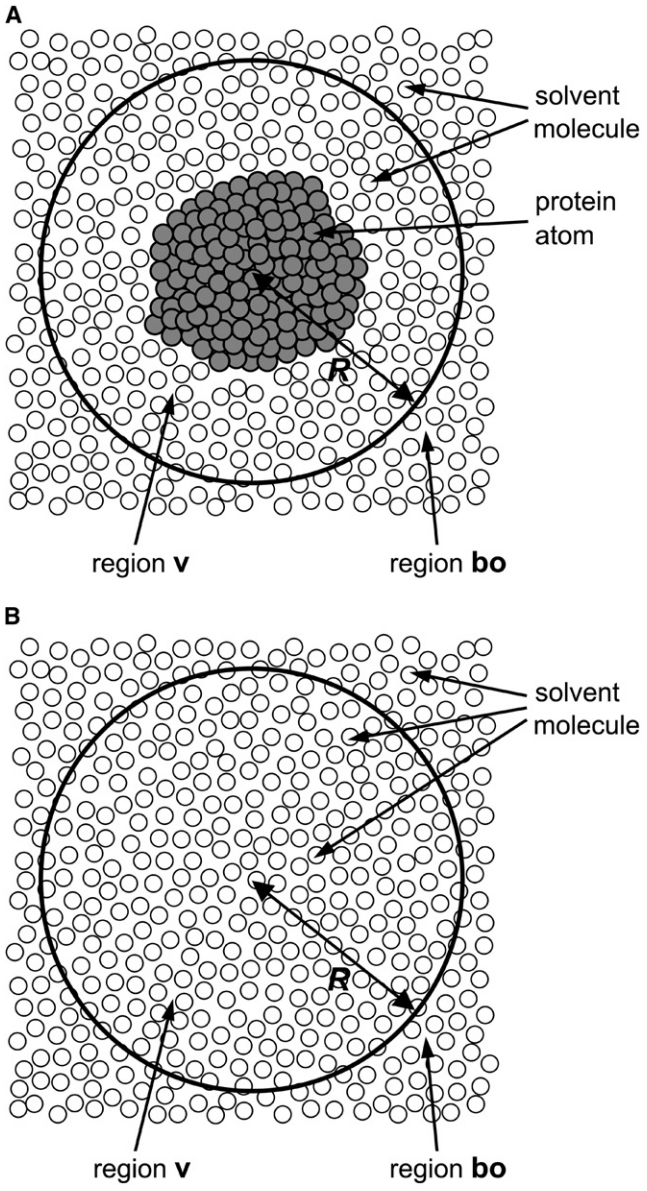

FIGURE 2 Dividing total volume of a system into two regions ''v'' and ''bo'' for a dilute solution of a protein (*A*) and pure solvent (*B*).

the fluctuation of the overall structure of *Eco*O109I. R_g (Fig. 3 *C*, *cyan curve*) exhibited large fluctuation, with a period of ~50–100 ns. There was an ~1.8-Å difference in R_g between the most expanded and closed conformations. However, R_g averaged over all time windows (Fig. 3 *C*, *blue horizontal line*) is in agreement with the R_g obtained from the experimentally observed SAXS profile (Fig. 3 *C*, *pink horizontal line*). In addition, the R_g and the SAXS profile in the range of $Q \leq 0.18$ Å^{-1} almost converged from $t = 100$ ns (data not shown), indicating that at least a 100-ns time length of a simulation is needed to obtain the structural ensemble consistent with the experimentally obtained SAXS data.

We calculated another R_g (referred to as ''protein R_g'') obtained from the mass-weighted atomic structure of the protein without water molecules, $R_g^2 = 1/M \sum_i^N m_i r_i^2$, where $M = \sum_i^N m_i$ (N is the number of protein atoms, and m_i is atomic weight of the ith protein atom) for the time windows described above (Fig. 3 *C*, *black curve*). The time course of protein R_g clearly shows almost the same time dependency as that of the protein-water R_g. This indicates that the fluctuation of the protein-water R_g reflects the fluctuation of the protein moiety, and a 2-Å difference between the protein-water R_g and the protein R_g is attributable to the hydration shell on the surface of the protein molecule (40,41).

To analyze the effects of structural changes on the protein-water R_g during the 150-ns MD simulation, we performed another 20-ns MD simulation in which positions of the protein atoms were restrained to the original crystal positions, and only the solvent atoms were sampled (referred to as ''restraint-MD simulation''). As shown in Fig. 3 *C* (*yellow-green curve*), we calculated the protein-water R_g for the time window of 500 ps, shifting from $t = 0$–20 ns, for the restraint-MD simulation. In this simulation, the protein-water R_g was ~27.6 Å, and the average of the protein-water R_g (Fig. 3 *C*, *green horizontal line*) was still smaller than the experimental value. The difference in protein-water R_g between the restraint-MD simulation and the experimental value indicates that *Eco*O109I adopts a more expanded conformation in solution, as observed in the 150-ns MD simulation, than in a crystalline state. All these results show the validity of the structural ensemble sampled by the 150-ns MD simulation. Next, we investigated the detailed dynamics of *Eco*O109I during a 150-ns MD simulation.

Dynamics within the subunit

To elucidate the dynamics of *Eco*O109I during the 150-ns MD simulation, we first analyzed the fluctuations occurring within the subunit in detail. Based on a superposition on the C^α best fits for each subunit structure, we calculated the root mean-square fluctuation of the C^α atom (C^α-RMSF, Fig. 4 *A*) and the average C^α-RMSD of each residue from the crystal structure of the DNA-free form (data not shown), using all snapshots in the trajectory. As a result, the ''rigid'' segments (Fig. 4 *A*, *blue horizontal lines*), defined as the segments of residues with C^α-RMSF and average C^α-RMSD values <1.2 Å and 2.0 Å, respectively, were localized and constituted one rigid part (Fig. 4 *B*, *blue*) in the center of the subunit. Interestingly, this rigid part of the subunit comprised parts of both the dimerization and catalytic domains, i.e., it comprised α1, the last half of α3 and α9 in the dimerization domain, and β1–3 and α8 in the catalytic domain.

The C^α-RMSD of this rigid part of the subunit from the crystal structure of the DNA-free form was ~0.9 Å in the simulation. In contrast, the C^α-RMSD of the dimerization domain (α1–3 and α9–11) and the catalytic domain (β1–4 and α6–8, excluding the missing residues 93–100 in α4, and 178–182 in the loop between α7 and β4) became ~2.5 Å and ~2.9 Å, respectively. These results indicate that the rigid part of the

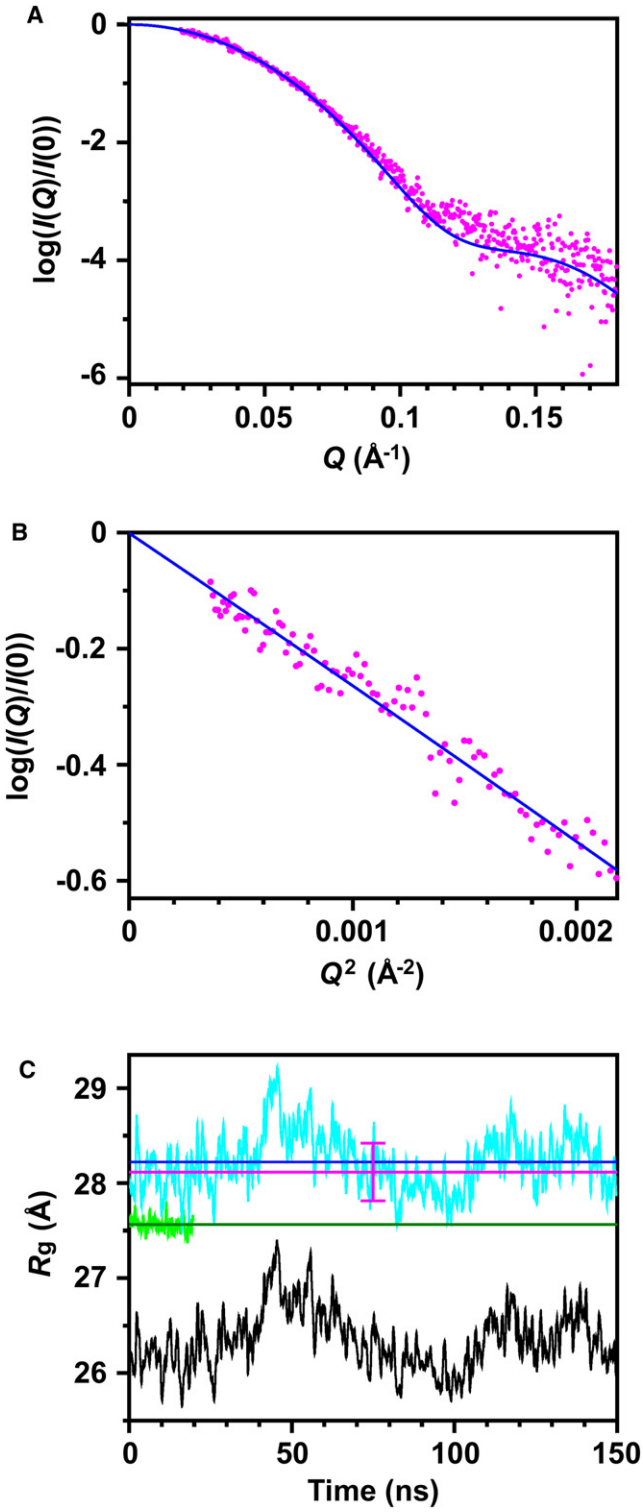

FIGURE 3 Comparison of experimental and simulation-derived SAXS profiles of DNA-free form of *Eco*O109I in solution. The SAXS profiles from the experiments (*pink dots*) and the 150-ns MD simulation (*blue curves*) are shown in logarithmic plot (*A*) and Guinier plot (range of R_gQ, <1.3) (*B*). (*C*) Time evolution of protein-water R_g estimated from Guinier plot for time-window of 500 ps shifting through 150-ns MD trajectory (*cyan curve*) and 20-ns restraint-MD trajectory (*yellow-green curve*). Average R_g is also shown for 150-ns trajectory (*blue horizontal line*) and 20-ns trajectory (*green horizontal line*). Pink horizontal line with error bar shows protein-water Rg estimated from Guinier plot of experimental SAXS profile. Protein R_g was calculated from protein structures without water molecules in 150-ns trajectory (*black*).

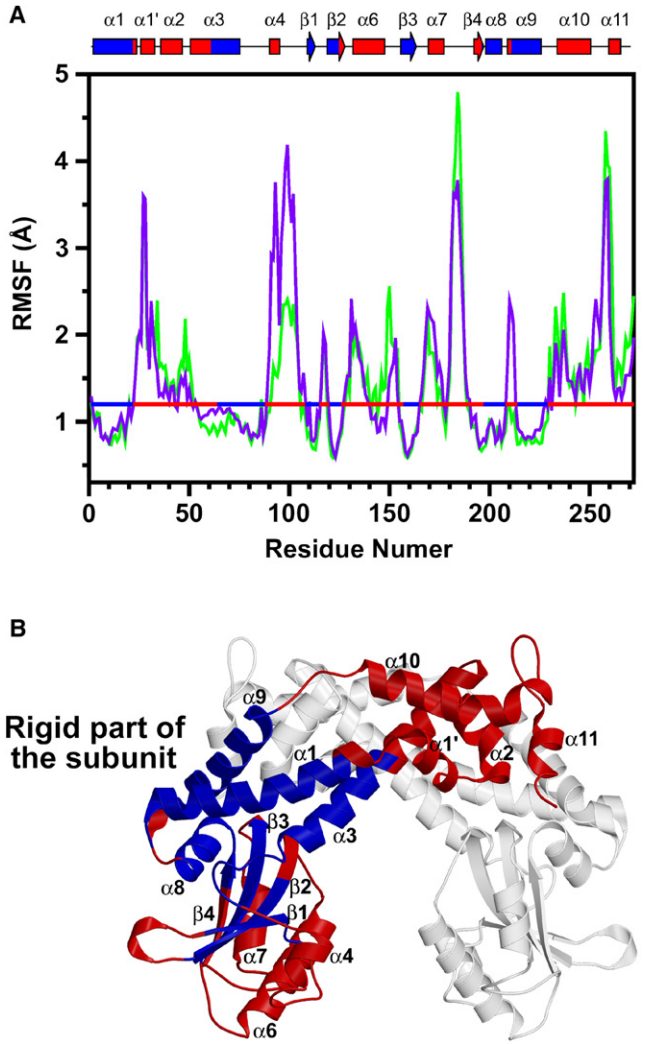

FIGURE 4 Dynamics within subunit during 150-ns simulation. (*A*) C^{α}-RMSF of subunits A (*green*) and B (*purple*). Calculation of C^{α}-RMSF for each subunit was performed, based on superposition using C^{α} best fits for its own subunit. Horizontal line indicates threshold value (1.2 Å) of C^{α}-RMSF, which is used in classification of segments into rigid segments (*blue*; residues 1–21, 64-87, 108–113, 120–125, 157–164, 198–207, and 213–229) and the others (*red*). Assignments of secondary structures are also shown in schematic diagram along top of *A*. (*B*) Rigid segments (*blue*) and others (*red*) of a subunit. All rigid segments are located at center of subunit, and constitute one rigid part of the subunit.

subunit was stable in the simulation, but the dimerization and catalytic domains exhibited certain flexibilities. Furthermore, a comparison of crystal structures between the DNA-free and DNA-bound forms showed that the C^{α}-RMSD of the rigid part of the subunit between the DNA-free and DNA-bound forms was ~0.8 Å, whereas the C^{α}-RMSD of the dimerization and catalytic domains was ~1.5 Å and ~2.1 Å, respectively.

These observations of structural changes in the crystal structures confirmed the results of the MD simulation, i.e., demonstrating the stability of the rigid part of the subunit, and flexibility in the dimerization and catalytic domains.

Dynamics in the dimer: identification of two rigid cores

To investigate dynamics in the dimer, we focused on the dynamics of one subunit relative to the other subunit. We compared the C^{α}-RMSF of the dimeric structure (Fig. 5, *A* and *B*, corresponding to subunits A and B of the dimer, respectively) between two schemes of superposition for the MD trajectory: in one scheme, all snapshots of the dimer were superimposed, using the C^{α} best fits for subunit A (Fig. 5, *A* and *B*, *green curve*); in the other scheme, we used the C^{α} best fits for subunit B (Fig. 5, *A* and *B*, *purple curve*). Interestingly, the C^{α}-RMSF of the two helices α2 and α11 (Fig. 5 *C*, *cyan*; also, Fig. 5, *A* and *B*, *arrows*) in the dimerization domain of one subunit became smaller when dimer structures were best fitted when using the other subunit rather than their own subunit. This indicates that the two helices α2 and α11 in the dimerization domain of one subunit moved collectively with the rigid part (Figs. 4 *B* and 5 *C*, *blue*) of the other subunit rather than with that of the same subunit. Thus, the rigid part of one subunit and the two helices α2 and α11 in the other subunit formed a rigid core (RCA and RCB for subunits A and B, respectively, in Fig. 5 *C*).

The C^{α}-RMSDs of individual rigid cores RCA and RCB from the crystal structure were both ~1.0 Å during the simulation, whereas the C^{α}-RMSD of a combined residue group of two rigid cores became larger than 2.0 Å within the first 10 ns (Fig. 5 *D*). Hence the individual rigid cores were stable in the simulation, and in contrast, there were large rigid-body motions between RCA and RCB. These individual stabilities of RCA and RCB are also supported by a comparison of the crystal structures between the DNA-free and DNA-bound forms of the enzyme. The C^{α}-RMSDs of RCA and RCB between the DNA-free and DNA-bound forms are both 0.8 Å, whereas the C^{α}-RMSD of a combined residue group of the two rigid cores is 2.4 Å, indicating that (in agreement with the results of the MD simulations) individual crystal structures of RCA and RCB remain unchanged upon DNA binding to the enzyme, whereas certain rigid-body motions occur between RCA and RCB.

Rigid-body motions between two rigid cores

For further analysis of rigid-body motions between the rigid cores RCA and RCB in the simulation, principal component analysis was applied for the trajectory of C^{α} atoms of the rigid cores. The first two principal modes accounted for the dominant portion of motions in RCA and RCB, i.e., 25.6% and 22.5% contributions to the total C^{α}-RMSF for the first and second principal modes, respectively (Fig. 6 *A*).

The eigenvector of the first principal mode represents the open-close motion between RCA and RCB (Fig. 6 *D*, *red arrows*). The time course of projections onto the first principal mode is shown in Fig. 6 *B*, in which positive and negative projection values correspond to open and closed conformations, respectively. The oscillation of projections between positive and negative values indicates frequent interchanges between open and closed conformations. Because the R_g of open conformations were larger than those of closed conformations, there was a significant correlation of time courses between projections onto the first principal mode and the time-dependent R_g shown in Fig. 3 *C* (correlation coefficient,

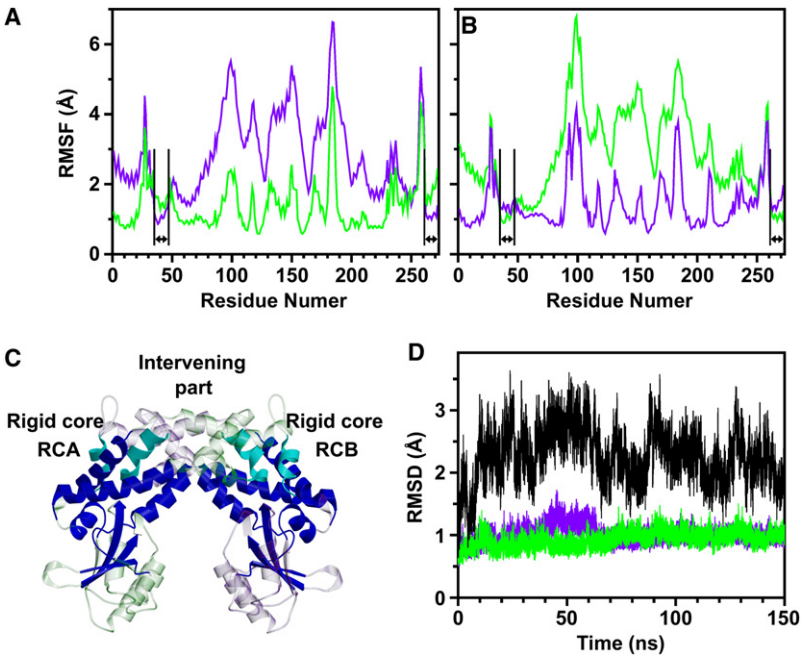

FIGURE 5 Dynamics in dimer during 150-ns simulation. (*A* and *B*) C^{α}-RMSF of dimer is shown for subunits A and B, respectively, calculated using two different superposition schemes. In one scheme, all dimer structures in the simulation were superimposed using C^{α}-atom best fits for subunit A (*green*), and in the other scheme, for subunit B (*purple*). In segments indicated by arrows corresponding to helices α2 (residues 36–46) and α11 (residues 262–272), C^{α}-RMSF was smaller in the case of superimposition on the other subunit than on its own subunit. (*C*) Rigid part of subunit (*blue*) and two helices α2 and α11 of other subunit (*cyan*) form a rigid core (rigid core RCA for subunit A, and RCB for subunit B). (*D*) C^{α}-RMSD from DNA-free crystal structure in 150-ns simulation is shown for RCA (*green*), RCB (*purple*), and a combined residue group of RCA and RCB (*black*).

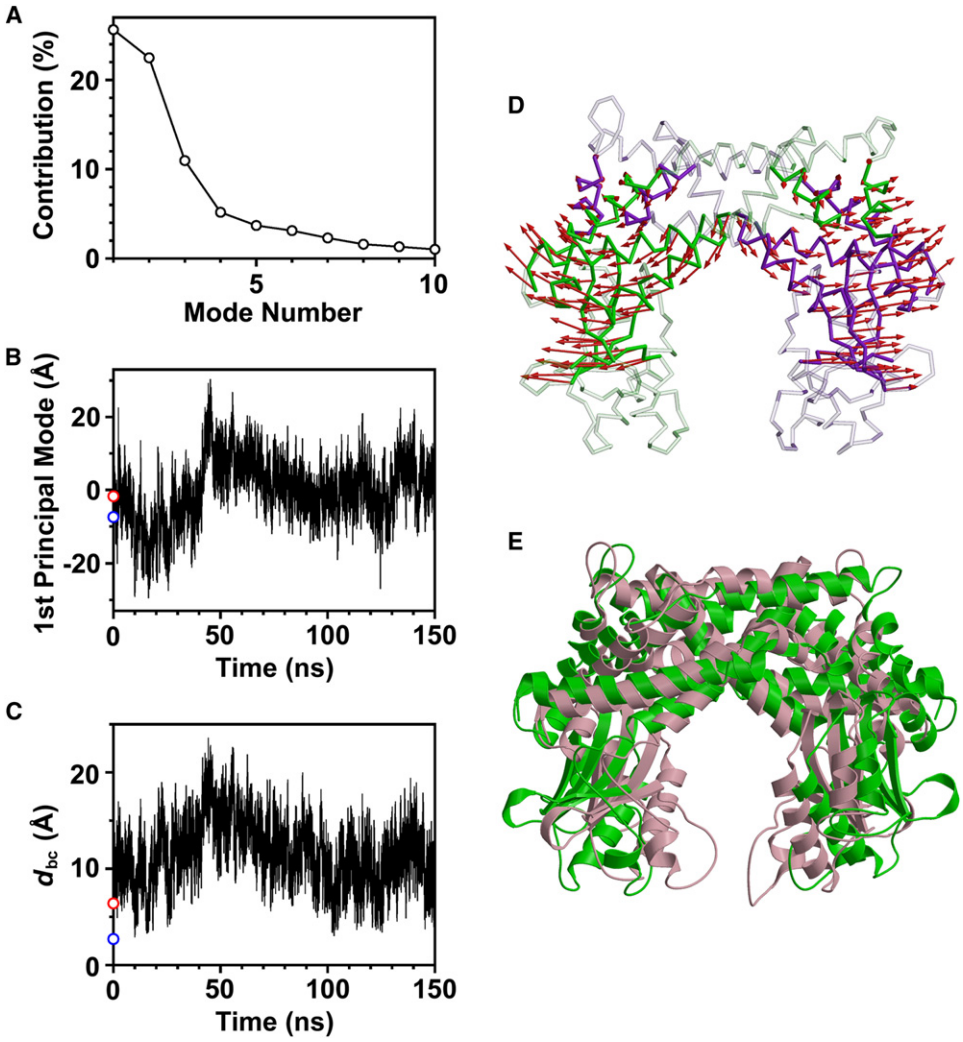

FIGURE 6 Principal component analysis for motions of RCA and RCB. (*A*) Individual contributions of first 10 principal components to total C^{α}-RMSF of rigid cores RCA and RCB. (*B*) Snapshots in trajectory were projected onto first principal mode. The DNA-free (*red open circle*) and DNA-bound (*blue open circle*) crystal structures are also projected onto the first principal mode. (*C*) Distance between two catalytic domains, d_{bc}, in the 150-ns simulation. (*D*) Amplitude (8σ) and direction of C^{α} displacements along first principal mode are represented by red arrows on C^{α} trace of the average structure of RCA and RCB in the 150-ns simulation. Other parts, except for RCA and RCB, are shown in transparent representation. Subunits A and B are shown in green and purple, respectively. (*E*) Superposition of structures with largest projection (*green*) and smallest projection (*pink*) onto first principal mode.

0.74). Moreover, we found that the open-close motion represented by the first principal mode changed the width of the intervening space between the two catalytic domains. The distance between the two catalytic domains, d_{bc} (Fig. 6 *C*), correlated well in the time course with the projections onto the first principal mode (Fig. 6 *B*) (correlation coefficient, 0.65). The domain distance d_{bc} was defined as the closest atomic distance between the two catalytic domains.

In Fig. 6 *E*, the superposition of the structures with the largest and smallest projections onto the first principal mode is shown. The structure with the largest projection value (Fig. 6 *E*, *green*) has a distance between the two catalytic domains, d_{bc}, of 20.9 Å, which is slightly larger than the diameter of DNA (20 Å). In contrast, the structure with the smallest projection value (Fig. 6 *E*, *pink*) has a d_{bc} of 5.2 Å, which is close to the d_{bc} of the crystal structure in the DNA-bound form (2.7 Å). Because the d_{bc} of the crystal structure in the DNA-free form is 6.4 Å, which is much smaller than the diameter of DNA, the open-close motion represented by the first principal mode is necessary for the enzyme to access DNA. Moreover, the small difference between the projection values of crystal structures in the DNA-free and DNA-bound forms (Fig. 6 *B*) indicates that the open-close motion was unexpected in terms of the previous crystallographic study (2).

The eigenvector of the second principal mode represents the twisting motion between RCA and RCB in a counterclockwise direction (Fig. 7 *B*, *red arrows*). The superposition of structures with the largest and smallest projections is shown in Fig. 7 *C*. As seen when comparing Fig. 1 *F* and Fig. 7 *C*, the directions of the second principal mode were significantly similar to the structural changes occurring upon DNA binding. Indeed, the correlation coefficient for the directions between the second eigenvector and the structural change in the DNA-free and DNA-bound crystal structures is 0.90 (Fig. 7 *B*), whereas the correlation coefficient for the first principal mode is 0.11. This indicates that structural changes upon DNA binding clearly correlate with the intrinsic dynamics of *Eco*O109I. In other words, the relative arrangement between RCA and RCB is specifically flexible in a counterclockwise direction in the DNA-free state, and this flexibility effectively facilitates structural changes upon DNA-binding. Furthermore, the crystal structure of

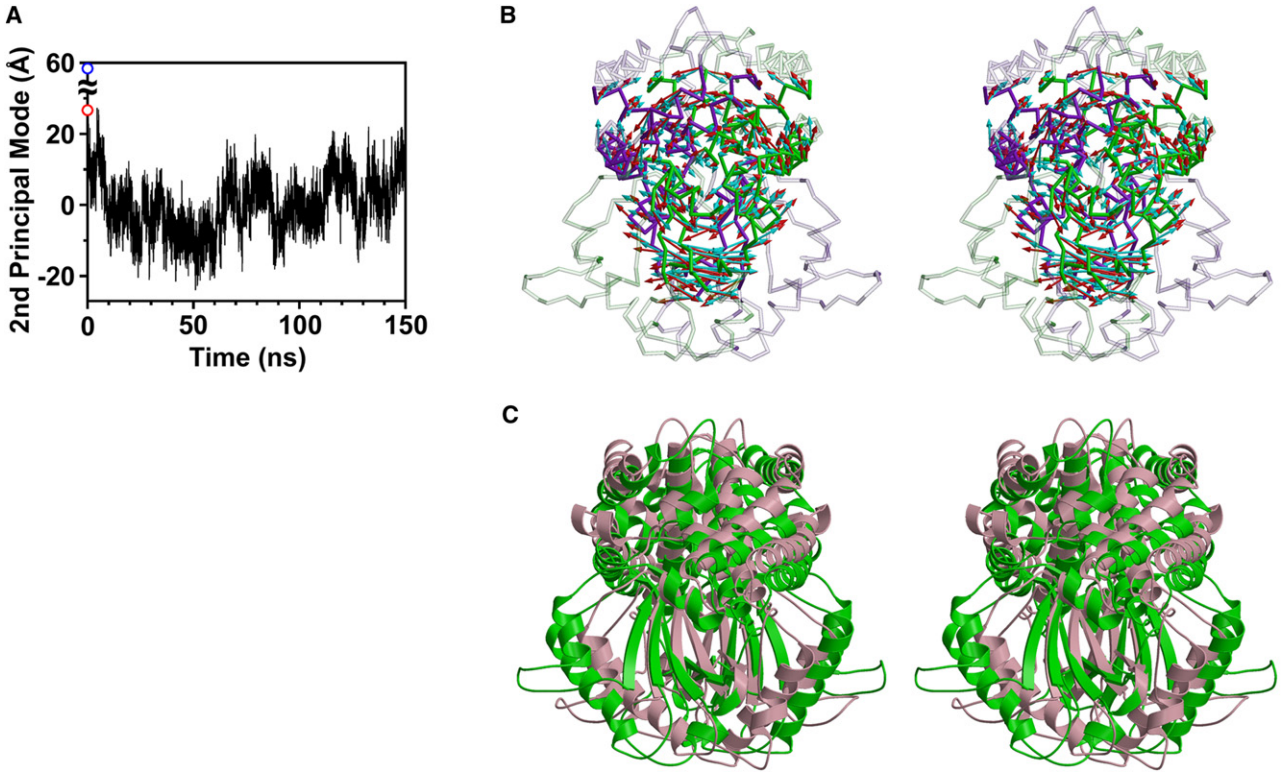

FIGURE 7 Second principal mode of principal component analysis for motions of RCA and RCB. (*A*) Snapshots in trajectory were projected onto second principal mode. The DNA-free (*red open circle*) and DNA-bound (*blue open circle*) crystal structures are also projected onto the second principal mode. (*B*) Amplitude (8σ) and direction of C^{α} displacements along second principal mode are represented by red arrows on C^{α} trace of average structure of RCA and RCB in a side view. Other parts, except for RCA and RCB, are shown in transparent representation. Subunits A and B are shown in green and purple, respectively. The difference in coordinates between the average structure and crystal structure in the DNA-bound form is represented by cyan arrows. (*C*) Superposition of structures with largest projection (*green*) and smallest projection (*pink*) onto second principal mode.

the DNA-free form corresponds to the upper limit of the sampled range of the projections onto the second principal mode (Fig. 7 *A*), indicating that the crystal structure is slightly different in the counterclockwise direction from the average structure, although the crystal structure is within the range of equilibrium. This may be attributable to crystal packing, which was relaxed during the first few nanoseconds of the simulation.

*Eco*RI is another REase in the *Eco*RI family, which functions as a homodimer, and shares a common structural motif in the catalytic domain with *Eco*O109I (42) (Fig. 8, *A* and *B*, *blue*). A normal mode analysis showed that *Eco*RI also has two low-frequency modes: an open-close motion, and a twisting motion, between the two subunits (43). However, the closing direction of the twisting mode in *Eco*RI is opposite to that of *Eco*O109I, i.e., a counterclockwise rotation in *Eco*O109I and a clockwise rotation in *Eco*RI (Fig. 8, *A* and *B*, *yellow arrows*). This difference is consistent with the DNA binding modes and locations of catalytic domains in *Eco*O109I and *Eco*RI. Whereas *Eco*O109I binds DNA from the minor groove side, *Eco*RI binds DNA from the major groove side (Figs. 1 and 8) (2,44). The catalytic domains of *Eco*O109I are located on the side of the DNA opposite the dimer interface (Fig. 8 *A*), and thus interact with the major groove of DNA. Therefore, counterclockwise rotation is necessary for *Eco*O109I to make the catalytic domains fit the major groove of DNA (Fig. 8 *A*, *yellow arrow*). In contrast, the catalytic domains of *Eco*RI are located on the same side of the DNA as the dimer interface (Fig. 8 *B*), and thus interact with the major groove of DNA. Therefore, the clockwise rotation is necessary for *Eco*RI to make the catalytic domains fit the major groove of DNA (Fig. 8 *B*, *yellow arrow*). As a consequence, the intrinsic twisting flexibility, which is different between *Eco*O109I and *Eco*RI, is suitable for facilitating their specific DNA binding and interaction.

Motions of α6, α7, and β4 in the catalytic domain

Next, we focused on the dynamics within the catalytic domain in the simulation. In terms of fluctuations in the MD simulation, the catalytic domain can be divided into three parts: 1), secondary structures belonging to the rigid part of the subunit (β1–3 and α8); 2), unstructured regions (α4 and the loops between secondary structures); and 3), the rest (α6, α7, and β4), located at the bottom of the catalytic domain. Helix α4 and the loop between α7 and β4, which are missing in the crystal structure of the DNA-free form, were also unstructured in the MD simulation (Fig. 4 *A*).

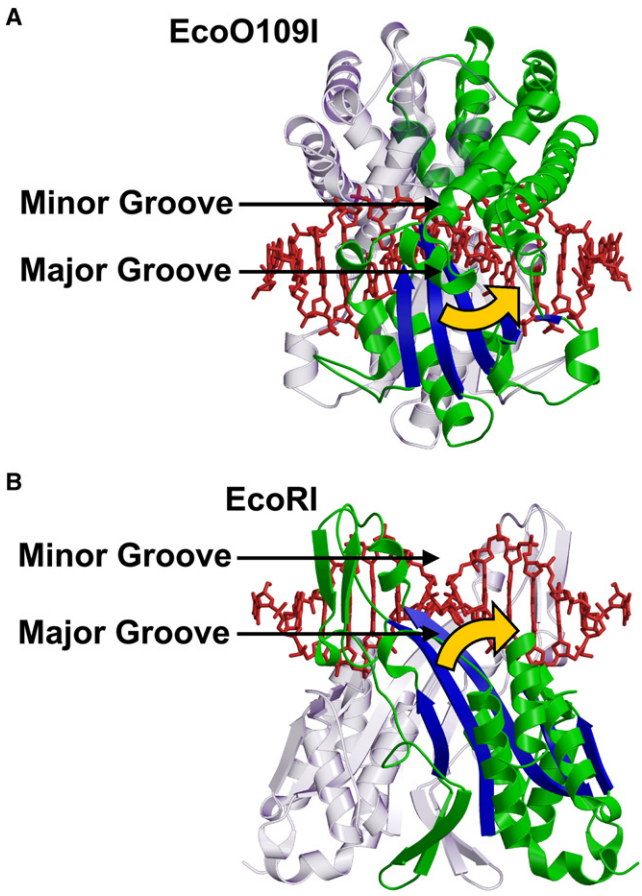

FIGURE 8 Closing direction of twisting mode in *Eco*O109I and *Eco*RI. (*A* and *B*) Side view of DNA-bound form of *Eco*O109I and *Eco*RI, respectively, colored according to common structural motif in the catalytic domain (*blue*) and other parts (*green*) in one subunit. Counterpart subunit is shown in transparent representation. DNA molecules (*red*) are shown, as minor groove faces upward. Yellow arrows represent closing direction of the twisting mode in *Eco*O109I and *Eco*RI, i.e., counterclockwise rotation in *Eco*O109I and clockwise rotation in *Eco*RI.

To investigate the collective motions in the catalytic domains, principal component analysis was performed for the trajectory of the C^{α} atoms of α6, α7, and β4, excluding unstructured regions based on a superposition using the C^{α} best fits for the rigid part of the subunit, where the trajectories of both subunits A and B were used in the calculation.

As a result, the first principal mode accounted for 60.4% of the total fluctuations (Fig. 9 *A*). The motion represented by the eigenvector of the first principal mode (Fig. 9 *C*, *red arrows*) was the closing motion, the direction of which correlated well with that of the structural change (Fig. 9 *C*, *cyan arrows*) upon DNA binding (correlation coefficient, 0.83). As seen in Fig. 9 *C*, α6, α7, and β4 move collectively, relative to the rigid part of the subunit in this principal mode, and this was also evident in the structural change upon DNA binding (Fig. 9 *C*, *cyan arrows*). The small discrepancy in the direction between the first principal mode and the structural change upon DNA binding is attributable to the fact that the movement involves a hinge-bending motion, and the eigenvector points to the tangential direction of this motion.

Time courses of projections onto the first principal mode show no significant correlation between subunits A and B (Fig. 9 *B*). In addition, no significant correlation was found between the projections of α6, α7, and β4 on any principal mode and those of the rigid cores RCA and RCB (Figs. 6 *B*, 7 *A*, and 9 *B*). These results indicate that the collective motions of α6, α7, and β4 were independent of motions in the other parts of the structure during the simulation.

Based on all these results, we conclude that α6, α7, and β4 are intrinsically mobile in the closing direction, to capture DNA. In addition to the rigid-body motions between RCA and RCB, the intrinsic dynamics of α6, α7, and β4 also play an essential role in the function of *Eco*O109I. Furthermore, the DNA-bound crystal structure is located within the sampled range of projections of subunit B (Fig. 9 *B*), indicating that the closed conformation of the catalytic domain observed in the DNA-bound form is accessible within the range of the equilibrium fluctuations in the DNA-free state.

Relationship between intrinsic dynamics and residue interactions of *Eco*O109I

To investigate the relationship between the observed intrinsic dynamics of *Eco*O109I and the residue interactions, we calculated fractions of contacts during the simulation (Fig. 10 *A*). Two residues were seen to be in contact when two nonhydrogen atoms, each belonging to a different residue, were within 4.5 Å of each other. The calculated fraction of contacts was averaged over the two subunits. To identify stable interactions in the structure, a contact maintained at more than 98% during the simulation that was formed between two residues more than 15 residues apart was defined as a ''permanent'' contact, and was mapped onto the three-dimensional structure (Fig. 10 *B*).

As a result, most of the permanent intrasubunit contacts formed within the rigid part of the subunit (Fig. 10 *B*, *blue*), and most of the permanent intersubunit contacts formed between the two helices α2 and α11 (Fig. 10 *B*, *cyan*) in the dimerization domain of one subunit and the rigid part (Fig. 10 *B*, *blue*) of the other subunit. In other words, permanent contacts were localized within each of the two rigid cores (RCA and RCB), forming a ''contact network''. This result is consistent with the fact that RCA and RCB were rigid during the simulation.

In Fig. 10 *B*, the C^{α} atoms of residues that have more than four permanent contacts are shown in a sphere. Such residues work as ''hubs'' that are necessary for the maintenance of a contact network within each of the rigid cores. Interestingly, three residues (Phe-38, Leu-267, and Asn-271), located in helices α2 and α11 of subunit A, worked as

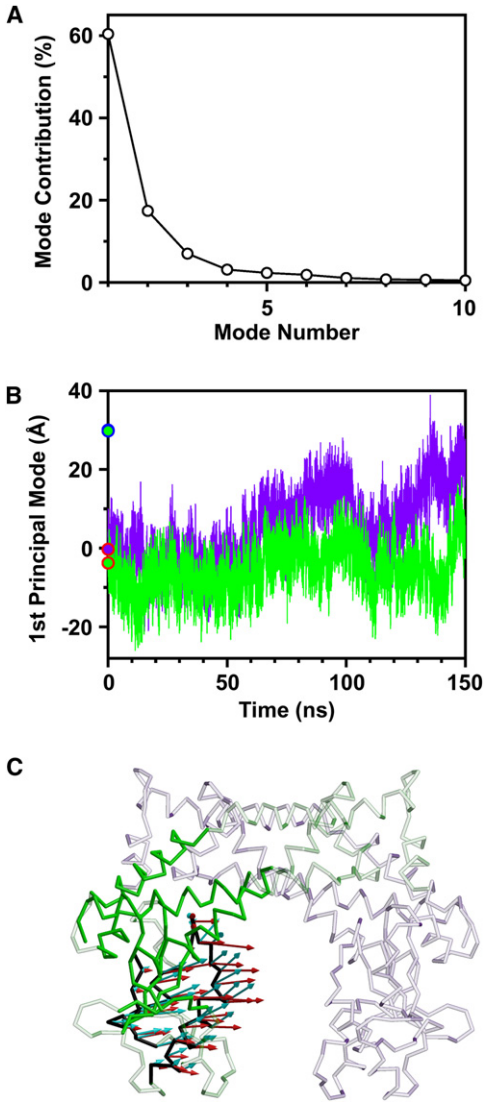

FIGURE 9 Principal component analysis for motions of secondary structures α6, α7, and β4 in catalytic domain, relative to rigid part of subunit. (*A*) Individual contributions of first 10 principal components to total C^{α}-RMSF of secondary structures α6, α7, and β4. Snapshots in trajectory of secondary structures α6, α7, and β4 in subunits A (*green*) and B (*purple*) were projected onto first principal mode (*B*). Blue circle shows projection value for crystal structures in the DNA-bound form. Red circles filled with green and purple show projection values for subunits A and B of the DNA-free crystal structure, respectively. (*C*) Amplitude (3σ) and direction of C^{α} displacements along first principal mode are represented by red arrows on C^{α} trace of average structure, in which are shown the rigid part of the subunit (*green*), the secondary structures α6, α7, and β4 in the catalytic domain (black), and the others (*white*). The difference in coordinates between the average structure and the crystal structure in the DNA-bound form is represented by cyan arrows.

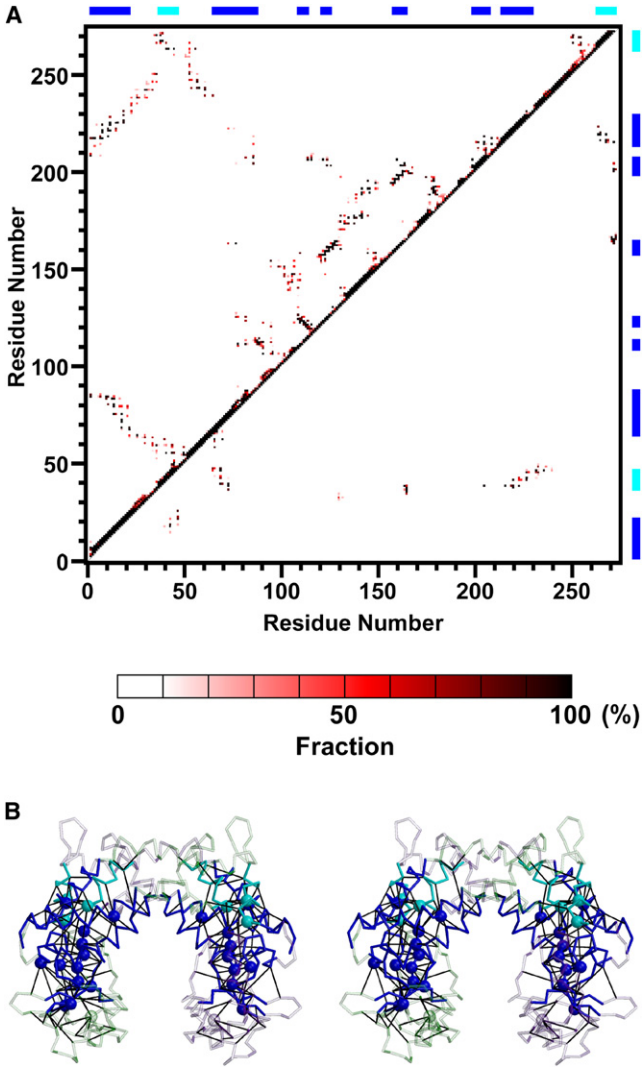

FIGURE 10 Interresidue contact analysis. (*A*) Fractions of contacts present in the simulation. Upper and lower triangles of contact map indicate intrasubunit contacts and intersubunit contacts, respectively. Color bar for fractions of contacts is shown along bottom. Bars along top and right sides indicate rigid part of subunit (*blue*) and helices α2 and α11 (*cyan*) in the dimerization domain. (*B*) Permanent contacts are mapped onto average structure in C^{α} trace. Rigid cores RCA and RCB, composed of the rigid part of the subunit (*blue*) and helices α2 and α11 (*cyan*), are shown in non-transparent representation. Black lines represent permanent contact formations between pairs of residues. The C^{α} atoms of residues with more than four permanent contacts are rendered in space-filling.

hubs of the contact network within RCB (those three residues in subunit B were also hubs within RCA). This indicates that intersubunit interactions within RCA and RCB significantly contributed to the stability of the rigid cores.

In contrast to the contacts within the rigid cores, the intervening part between rigid cores RCA and RCB had few permanent contacts (Fig. 10 *B*), i.e., most of the contact formations in the intervening part were unstable. However, the total number of contacts within this intervening part did not significantly change during the simulation, and was similar to that of the DNA-free crystal structure, i.e., 74 ± 7 and 86 for the average of the simulation and the crystal structure, respectively. This indicates that frequent interchanges of contacts occurred in the intervening part. Most

of the contacts in the intervening part were hydrophobic interactions between the side chains of residues. Few hydrogen bonds and no salt bridges were formed. The characteristics of these interactions allowed for frequent interchanges of contacts without a large change in energy, resulting in flexibility of the intervening part and enabling rigid-body motions between RCA and RCB.

In the catalytic domains, several permanent contacts were formed among α6, α7, and β4, and no significant permanent contact between the rigid part of the subunit and each of α6, α7, and β4 was evident. This is why α6, α7, and β4 moved collectively, without significant correlation with movements of the other parts of the structure in the simulation.

CONCLUSIONS

To elucidate the relationship between the intrinsic dynamics of *Eco*O109I and its function, we performed a 150-ns MD simulation of the DNA-free form of the enzyme, and examined the resultant trajectory in detail.

We calculated the SAXS profile from all snapshots in the trajectory, taking scattering from the explicit hydration water molecules into account. The calculated SAXS profile and R_g estimated from the Guinier plot of the profile are in agreement with the experimentally observed SAXS profile and R_g, supporting the validity of the structural ensemble sampled by the simulation presented here. In addition, a large fluctuation in R_g was observed in the simulation, which indicates considerable flexibility in the overall structure of the enzyme. There was a difference of ~1.8 Å in R_g between its most expanded and pinched forms.

Based on further analyses of the dynamics in the trajectory, we identified two rigid cores, RCA and RCB, which are elements of rigid-body motions in the dimeric structure of *Eco*O109I. Interestingly, the two rigid cores were composed of the rigid part of one subunit and two helices, α2 and α11, of the other subunit. Principal component analysis, as applied to the trajectory of RCA and RCB, revealed that the rigid-body motions between the two rigid cores play an important role in the function of *Eco*O109I. The motion represented by the first principal mode is required for the intervening space between the catalytic domains to be accessed by DNA. The motion represented by the second principal mode has a significant correlation with the structural changes upon DNA binding, and is therefore necessary for *Eco*O109I to achieve effective functionality. These two functional modes accounted for a dominant portion of the fluctuations of RCA and RCB in the DNA-free state.

We also observed functional motions of α6, α7, and β4, which are located at the bottom of the catalytic domain. Principal component analysis applied to the trajectory of α6, α7, and β4 showed that the major dynamic of the catalytic domain is the closing motion relative to the rigid core, which correlates well with the structural change of this domain capturing DNA.

Analysis of contact formations revealed a relationship between intrinsic dynamics and residue interactions. A strong contact network formed within each of the rigid cores, and stable contacts between the rigid part of one subunit and the two helices α2 and α11 of the other subunit worked as hubs in the contact network. In contrast, frequent interchanges of contacts occurred in the intervening part between RCA and RCB, enabling rigid-body motions between two rigid cores. These motions are crucial for the function of the enzyme. Moreover, a small contact network formed among α6, α7, and β4 in the catalytic domain.

Our 150-ns MD simulation of *Eco*O109I revealed that the intrinsic dynamics of this enzyme are important for its function. To achieve functionally important structural changes upon DNA binding, the structure of *Eco*O109I is significantly flexible in the specific direction of functional structural changes, and this flexibility originates in its three-dimensional dimeric structure.

APPENDIX: CALCULATION OF SAXS PROFILE FROM MD TRAJECTORY

Illustration of method for calculation of SAXS profile

Here, we present a method for the calculation of scattering intensities of a solvated protein molecule in MD simulations. Let us consider a simulation box as part of a model system of dilute solution, where one protein molecule is dissolved in a large amount of solvent molecules (Fig. 2 *A*). The origin of the system is set to the center of mass of the protein molecule. The total volume of the model system is divided into two regions, "v" and "bo" (23,24), as illustrated in Fig. 2 *A*. The area corresponding to region v is within a sphere centered at the origin, and that corresponding to region bo is outside the sphere. The radius R of the sphere is chosen to be sufficiently large so that region v includes not only all of the hydration shell around the protein, but also some bulk region. The hydration shell is the area where the solvent density is perturbed from the bulk by the protein molecule.

In a dilute solution, the average electron density of the solution can be assumed to be equal to that of pure solvent, denoted as ρ_0. In our model system, the time-averaged density of the solvent in region bo was used as the ρ_0. Next, denoting electron density at position $\mathbf{r}$ as $\rho(\mathbf{r})$, the instantaneous scattering amplitude of the solution is expressed as

$$\begin{aligned}\tilde{F}(\mathbf{Q}) &= \int_{\mathrm{v}}(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3\mathbf{r} + \int_{\mathrm{bo}}(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3\mathbf{r} \\ &= \tilde{F}_{\mathrm{v}}(\mathbf{Q}) + \tilde{F}_{\mathrm{bo}}(\mathbf{Q}),\end{aligned} \tag{3}$$

where $\mathbf{Q}$ is the scattering vector related to the scattering angle 2θ and the wavelength of the incident radiation λ by the equation $Q = |\mathbf{Q}| = 4\pi \sin\theta/\lambda$. As protein molecules in solution scatter isotropically and undergo configurational changes, the instantaneous scattering intensity $\tilde{I}(\mathbf{Q})$ must be averaged orientationally and configurationally:

$$\begin{aligned}I(Q) &= \left\langle \langle \tilde{I}(\mathbf{Q})\rangle_{\Omega_{\mathbf{Q}}} \right\rangle_{\mathrm{MD}} \\ &= I_{\mathrm{v,v}}(Q) + I_{\mathrm{v,bo}}(Q) + I_{\mathrm{bo,bo}}(Q),\end{aligned} \tag{4}$$

where

$$I_{\lambda\lambda}(Q) = \left\langle \left\langle F^*_\lambda(\mathbf{Q}) \cdot F_\lambda(\mathbf{Q}) \right\rangle_{\Omega_\mathbf{Q}} \right\rangle_{\mathrm{MD}}$$
$$I_{\lambda\mu}(Q) = \left\langle \left\langle F^*_\lambda(\mathbf{Q}) \cdot F_\mu(\mathbf{Q}) + F^*_\mu(\mathbf{Q}) \cdot F_\lambda(\mathbf{Q}) \right\rangle_{\Omega_\mathbf{Q}} \right\rangle_{\mathrm{MD}}$$
$$(\lambda, \mu = \mathrm{v}, \mathrm{bo}), \tag{5}$$

where $F^*(\mathbf{Q})$ is a complex conjugate of $F(\mathbf{Q})$, and $\langle\,\rangle_{\Omega_\mathbf{Q}}$ and $\langle\,\rangle_{\mathrm{MD}}$ denote the orientational average and the average over all snapshots in the trajectory of the MD simulation, respectively ($\Omega_\mathbf{Q}$ is the solid angle in reciprocal space, $\mathbf{Q} = (Q, \Omega_\mathbf{Q})$).

In experiments of solution scattering, the SAXS profile of a protein is obtained by the subtraction of the scattering data of buffer solution from that of the protein solution. In our calculation method, we use the same data treatment as in the experiments, i.e., MD simulations for pure solvent are performed in addition to simulations for the protein solution, and the simulated scattering data for pure solvent are subtracted from those of protein solution.

Let us consider a pure solvent system whose average density equals ρ_0. As well as the dilute solution of the protein, this pure solvent system is divided into two regions, v and bo (Fig. 2 *B*). There is no difference in solvent properties between these two regions in the pure solvent system. Based on the MD simulation of the pure solvent system, the scattering intensity, $I^\mathrm{W}(Q)$, is calculated. Subsequently, $I^\mathrm{W}(Q)$ is subtracted from $I(Q)$, calculated from the MD simulation for protein solution:

$$\begin{aligned}\Delta I(Q) &= I(Q) - I^\mathrm{W}(Q) \\ &= \Delta I_{\mathrm{v,v}}(Q) + \Delta I_{\mathrm{v,bo}}(Q) + \Delta I_{\mathrm{bo,bo}}(Q),\end{aligned} \tag{6}$$

where $\Delta I(Q)$ corresponds to the experimentally obtained scattering intensity of the protein molecule. In our study, to ensure that the average electron density of the pure solvent system is identical to ρ_0, we attempted 15 series of MD simulations for various pure-solvent systems, including different numbers of Na^+ and Cl^- ions, and selected a pure-solvent system in which the average electron density was closest to ρ_0. The procedure for the *i*-th series is as follows: 1), *i* Na^+ and *i* Cl^- ions were added into a water box with a cube edge length of 115.0 Å, followed by an energy minimization of 2000 steps; 2), a 1-ns run was performed under NPT conditions, at $T = 298.15$ K and $P = 1.0$ atm for equilibration; and 3), a 2-ns product run was performed under the same NPT conditions. The edge length of the simulation box became ~113.5 Å in the simulation for every series. Consequently, we selected a pure-solvent system including 3 Na^+ and 3 Cl^- ions, in which the average electron density was identical to ρ_0 calculated from the simulations of *Eco*O109I ($\rho_0 \simeq 0.3367$ Å^{-3}). Numerically, the differences in scattering intensities between the buffers, which were different in terms of average density, exert negligible effects on the calculation of the final SAXS profile, $\Delta I(Q)$ (data not shown). Theoretically, however, the average density of a buffer should be equal to the ρ_0 of the protein solution (see theoretical details below).

As proved below, $\Delta I_{\mathrm{v,bo}}(Q)$ and $\Delta I_{\mathrm{bo,bo}}(Q)$ equal 0 if the radius R of the sphere is sufficiently large so that the density fluctuations, $\rho(\mathbf{r}) - \rho_0$, in the region bo have no correlation with those in the protein area and its hydration shell. Eq. 6 then becomes

$$\begin{aligned}\Delta I(Q) &= \Delta I_{\mathrm{v,v}}(Q) \\ &= \left\langle \left\langle \int_\mathrm{v} d^3\mathbf{r}' \int_\mathrm{v} d^3\mathbf{r} (\rho(\mathbf{r}') - \rho_0)(\rho(\mathbf{r}) - \rho_0) e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle_{\Omega_\mathbf{Q}} \right\rangle_{\mathrm{MD}} \\ &\quad - \left\langle \left\langle \int_\mathrm{v} d^3\mathbf{r}' \int_\mathrm{v} d^3\mathbf{r} (\rho^\mathrm{W}(\mathbf{r}') - \rho_0)(\rho^\mathrm{W}(\mathbf{r}) - \rho_0) e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle_{\Omega_\mathbf{Q}} \right\rangle_{\text{pure solvent MD}},\end{aligned} \tag{7}$$

where $\rho^\mathrm{W}(\mathbf{r})$ is the electron density at position $\mathbf{r}$ in the pure solvent system, and $\langle\,\rangle_{\text{pure solvent MD}}$ denotes the average over all snapshots in the trajectory of the MD simulation of the pure solvent.

Denoting the atomic form factor of the *k*-th atom at position $\mathbf{r}_k$ as $f_k(Q)$ (45), the instantaneous scattering intensity in the bracket $\langle\,\rangle_{\mathrm{MD}}$ of Eq. 7 is rewritten as

$$\begin{aligned}\tilde{I}(Q) &= \left\langle \int_\mathrm{v} d^3\mathbf{r}' \int_\mathrm{v} d^3\mathbf{r} (\rho(\mathbf{r}') - \rho_0)(\rho(\mathbf{r}) - \rho_0) e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle_{\Omega_\mathbf{Q}} \\ &= \left\langle \left| \sum_k^N f_k(Q) e^{-i\mathbf{Q}\cdot\mathbf{r}_k} \right|^2 \right\rangle_{\Omega_\mathbf{Q}} \\ &\quad - 2\rho_0 \sum_k^N f_k(Q) \int_\mathrm{v} d^3\mathbf{r} \left\langle e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}_k)} \right\rangle_{\Omega_\mathbf{Q}} \\ &\quad + \rho_0^2 \int_\mathrm{v} d^3\mathbf{r}' \int_\mathrm{v} d^3\mathbf{r} \left\langle e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')} \right\rangle_{\Omega_\mathbf{Q}},\end{aligned} \tag{8}$$

where N is the number of atoms within region v. For fast computation of the SAXS profile of a snapshot, the first term on the right-hand side of Eq. 8 is calculated, based on a multipole expansion (24,38,39) that uses the following relation: (46):

$$\exp(-i\mathbf{Q}\cdot\mathbf{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(Qr) Y^*_{lm}(\Omega_\mathbf{r}) Y_{lm}(\Omega_\mathbf{Q}), \tag{9}$$

where j_l and Y_{lm} denote a spherical Bessel function and a spherical harmonic function, respectively, and $\Omega_\mathbf{r}$ is the solid angle in real space, $\mathbf{r} = (r, \Omega_\mathbf{r})$. The volume integrals of the second and third terms on the right-hand side of Eq. 8 can be replaced by the surface integral, and this simplifies the numerical procedure of these integrals (47). Eventually, Eq. 8 is rewritten as Eq. 1. Furthermore, if region v is sufficiently large, the shape of surface S is not limited to a sphere.

In summary, the SAXS profile of a protein is obtained from MD simulations as follows: 1), the MD simulation of a protein solution system is performed; 2), the average density of the region bo, ρ_0, is calculated; 3), the MD simulation of a pure solvent system is performed, in which some ions are added to make the average density identical to ρ_0; 4), the instantaneous scattering intensity $\tilde{I}(Q)$ in Eq. 1 is calculated for each snapshot in the trajectories of the protein solution and pure solvent systems; 5), $\tilde{I}(Q)$ is averaged over all snapshots for each system; and 6), the final SAXS profile is obtained by the subtraction of the scattering data of the pure solvent from those of the protein solution.

To examine the size of region v for the correct evaluation of the SAXS profile of *Eco*O109I from the MD simulation, we calculated the scattering intensities $\Delta I(Q)$ by using spheres different in radius by 1 Å in a range of R = 45–51 Å. For the test calculation here, we used 500 snapshots from t = 44,001–44,500 ps, which adopted the most expanded conformations during the 150-ns simulation of *Eco*O109I. The maximum distance between a protein atom and the origin, $\max[|\mathbf{r}_k|]$, was also calculated for each snapshot, and the averages for those 500 snapshots and all snapshots were 46.9 Å and 43.7 Å, respectively. The SAXS profile calculated from the 500 snapshots turned out to converge fully at R = 50 Å, which is ~3 Å larger than $\langle \max[|\mathbf{r}_i|] \rangle_{44001-44501\ \mathrm{ps}}$. These results ensure that the sphere, the radius of which is larger than 50 Å, can be used as region v for the correct evaluation of the SAXS profile from the simulation. Therefore, we used the sphere R = 50 Å as region v for subsequent calculations and analyses.

The multipole expansion of the first term on the right-hand side of Eq. 1 fully converges for $l \le 14$, when evaluating the SAXS profile in a range of $0 \le Q \le 0.18$ Å^{-1}. This is because the spherical Bessel function $j_l(Qr)$ in the expansion becomes almost 0 for $l \ge 15$ in a range of $0 \le Qr \le 9$ ($0 \le r \le 50$ Å). The use of the multipole expansion of $l \le 14$ makes the speed of calculation of scattering intensities ~100 times faster than that of the Debye formula

$$\left\langle \left| \sum_k^N f_k(Q) e^{-i\mathbf{Q}\cdot\mathbf{r}_k} \right|^2 \right\rangle_{\Omega_\mathbf{Q}} = \sum_{i,j}^N f_i(Q) f_j(Q) \frac{\sin(Qr_{ij})}{Qr_{ij}}, \tag{10}$$

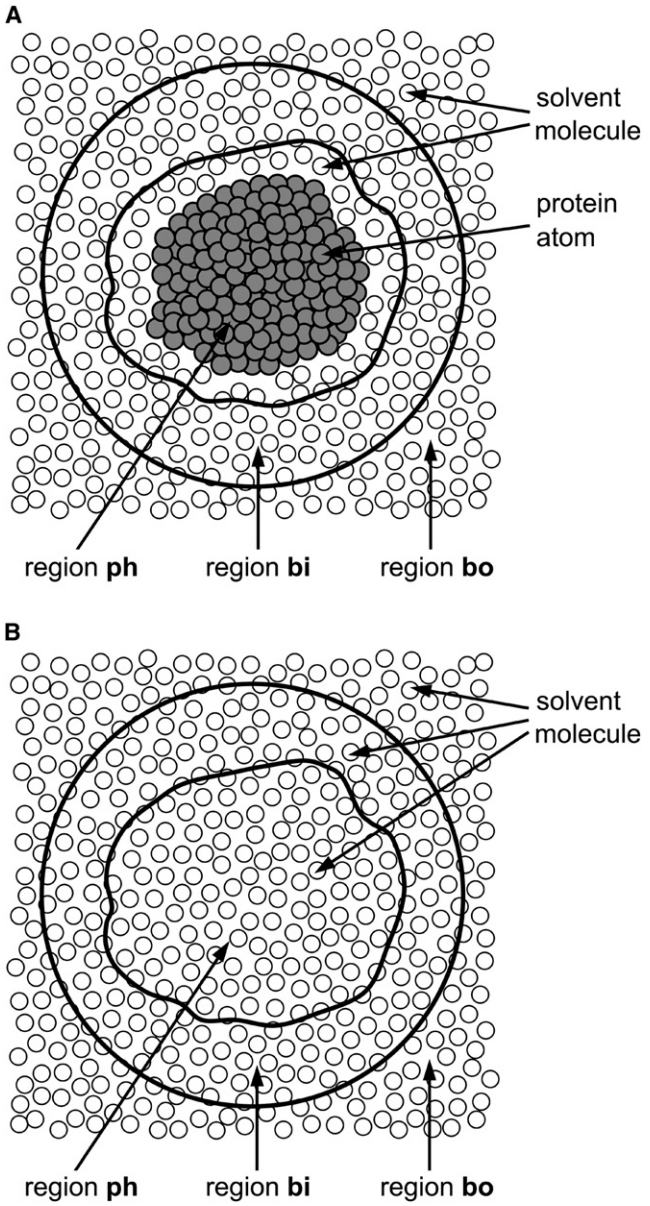

FIGURE 11 Dividing total volume of a system into three regions "ph", "bi", and "bo" produces dilute solution of a protein (*A*) and pure solvent (*B*).

where r_{ij} is the distance between the *i*-th and *j*-th atoms, and N equals ~53,800 for the sphere $R = 50$ Å. When calculating the scattering intensities for 138 data points of Q and using a single CPU of 3.2 GHz, the speed of calculation of the multipole expansion and the Debye formula is ~60 and 6500 s per snapshot, respectively. To make the speed of calculation much faster, we parallelized the program, such that atoms were partitioned by order. In another strategy of parallelization, it is possible to calculate the SAXS profile independently for each snapshot, followed by averaging all profiles. The speed of calculation then becomes 2.5 s per snapshot, when using 128 CPUs. Our program enabled us to calculate the SAXS profile for every snapshot in the trajectory of the 150-ns simulation (the total number of snapshots was 150,000), and to analyze the dynamics of *Eco*O109I in terms of the time-resolved SAXS profile, as shown in Fig. 3 *C*.

Proof of $\Delta I(Q) = \Delta I_{\mathrm{v,v}}(Q)$

Here, we prove that if radius R of the sphere of region v is sufficiently large so that the density fluctuations in region bo have no correlation with those in the protein area and its hydration shell, $\Delta I(Q)$ equals $\Delta I_{\mathrm{v,v}}(Q)$.

First, let us further divide region v of both the protein solution and pure solvent systems into two regions "ph" and "bi", as illustrated in Fig. 11. Region ph consists of the protein and its hydration shell, and the rest of region v is defined as region bi. There is no difference in solvent properties between regions bi and bo in the protein solution system, nor is there any difference among regions ph, bi, and bo in the pure solvent system. The instantaneous scattering amplitude of the protein solution of Eq. 3 is then rewritten as

$$\begin{aligned}\tilde{F}(\mathbf{Q}) &= \int_{\mathrm{ph}}(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3\mathbf{r} + \int_{\mathrm{bi}}(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3\mathbf{r} \\ &\quad + \int_{\mathrm{bo}}(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot\mathbf{r}}d^3\mathbf{r} \\ &= F_{\mathrm{ph}}(\mathbf{Q}) + F_{\mathrm{bi}}(\mathbf{Q}) + F_{\mathrm{bo}}(\mathbf{Q}),\end{aligned} \tag{11}$$

and the scattering intensity of Eq. 4 is rewritten as

$$\begin{aligned}I(Q) &= \left\langle \langle \tilde{I}(\mathbf{Q})\rangle_{\Omega_{\mathbf{Q}}} \right\rangle_{\mathrm{ens}} \\ &= I_{\mathrm{ph,ph}}(Q) + I_{\mathrm{ph,bi}}(Q) + I_{\mathrm{ph,bo}}(Q) \\ &\quad + I_{\mathrm{bi,bi}}(Q) + I_{\mathrm{bi,bo}}(Q) + I_{\mathrm{bo,bo}}(Q),\end{aligned} \tag{12}$$

where $\langle\,\rangle_{\mathrm{ens}}$ represents the configurational average over the ensemble. The subtraction of the scattering data of the pure solvent from those of the protein solution is expressed as

$$\begin{aligned}\Delta I(Q) &= I(Q) - I^{\mathrm{W}}(Q) \\ &= \Delta I_{\mathrm{ph,ph}}(Q) + \Delta I_{\mathrm{ph,bi}}(Q) + \Delta I_{\mathrm{ph,bo}}(Q) \\ &\quad + \Delta I_{\mathrm{bi,bi}}(Q) + \Delta I_{\mathrm{bi,bo}}(Q) + \Delta I_{\mathrm{bo,bo}}(Q) \\ &\quad \left(\Delta I_{\lambda,\mu}(Q) = I_{\lambda,\mu}(Q) - I^{\mathrm{W}}_{\lambda,\mu}(Q)\right).\end{aligned} \tag{13}$$

where $\Delta I(Q)$ corresponds to the experimentally obtained scattering intensity of the protein molecule. The third term of the right-hand side of Eq. 13 is given by

$$\begin{aligned}I_{\mathrm{p,bo}}(Q) &= \left\langle \left\langle F^*_{\mathrm{ph}}(\mathbf{Q})\cdot F_{\mathrm{bo}}(\mathbf{Q}) + F^*_{\mathrm{bo}}(\mathbf{Q})\cdot F_{\mathrm{ph}}(\mathbf{Q})\right\rangle_{\Omega_{\mathbf{Q}}}\right\rangle_{\mathrm{ens}} \\ &= 2\left\langle \left\langle \int_{\mathrm{ph}} d^3\mathbf{r}' \int_{\mathrm{bo}} d^3\mathbf{r}(\rho(\mathbf{r}')-\rho_0)(\rho(\mathbf{r})-\rho_0)e^{-i\mathbf{Q}\cdot(\mathbf{r}-\mathbf{r}')}\right\rangle_{\Omega_{\mathbf{Q}}}\right\rangle_{\mathrm{ens}} \\ &= 2\left\langle \int_{\mathrm{ph}} d^3\mathbf{r}' \int_{\mathrm{bo}} d^3\mathbf{r}\langle(\rho(\mathbf{r}')-\rho_0)(\rho(\mathbf{r})-\rho_0)\rangle_{\mathrm{ens}} e^{-i\mathbf{Q}\cdot\mathbf{r}}\right\rangle_{\Omega_{\mathbf{Q}}}.\end{aligned} \tag{14}$$

If region v is sufficiently large that regions ph and bo are separated enough by region bo, there is no correlation in density fluctuations between regions ph and bo. This relationship can be expressed as

$$\begin{aligned}&\langle(\rho(\mathbf{r}')-\rho_0)(\rho(\mathbf{r})-\rho_0)\rangle_{\mathrm{ens}} = 0 \\ &(\mathbf{r}'\in \mathrm{ph},\ \mathbf{r}\in \mathrm{bo}).\end{aligned} \tag{15}$$

On substituting Eq. 15 in Eq. 14, we find that

$$I_{\mathrm{ph,bo}}(Q) = 0. \tag{16}$$

and in the pure solvent system,

$$I^{\mathrm{W}}_{\mathrm{ph,bo}}(Q) = 0. \tag{17}$$

Considering that there is no difference in solvent properties of regions bi and bo between the protein solution and pure solvent systems, it follows that

$$I_{\mathrm{bi,bi}}(Q) = I^{\mathrm{W}}_{\mathrm{bi,bi}}(Q),\ I_{\mathrm{bi,bo}}(Q) = I^{\mathrm{W}}_{\mathrm{bi,bo}}(Q),$$
$$I_{\mathrm{bo,bo}}(Q) = I^{\mathrm{W}}_{\mathrm{bo,bo}}(Q). \quad (18)$$

On substituting Eqs. 16–18 in Eq. 13, we find that

$$\Delta I(Q) = \Delta I_{\mathrm{ph,ph}}(Q) + \Delta I_{\mathrm{ph,bi}}(Q) + \Delta I_{\mathrm{bi,bi}}(Q) = \Delta I_{\mathrm{v,v}}(Q), \quad (19)$$

where $\Delta I_{\mathrm{bi,bi}}(Q)$ equals 0.

All computations were performed on the Tsurumi Campus of the International Graduate School of Arts and Sciences, Yokohama City University, Yokohama, Japan.

This study was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by grants from the Japanese Science and Technology Agency, and by Research and Development of the Next-Generation Integrated Simulation of Living Matter, a part of the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1. Roberts, R. J., M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, et al. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31:1805–1812.
2. Hashimoto, H., T. Shimizu, T. Imasaki, M. Kato, N. Shichijo, et al. 2005. Crystal structures of type II restriction endonuclease *Eco*O109I and its complex with cognate DNA. *J. Biol. Chem.* 280:5605–5610.
3. Koshland, D. E. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44:98–104.
4. Tsai, C. J., S. Kumar, B. Ma, and R. Nussinov. 1999. Folding funnels, binding funnels, and protein function. *Protein Sci.* 8:1181–1190.
5. Boehr, D. D., D. McElheny, H. J. Dyson, and P. E. Wright. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science.* 313:1638–1642.
6. Ådén, J., and M. Wolf-Watz. 2007. NMR identification of transient complexes critical to adenylate kinase catalysis. *J. Am. Chem. Soc.* 129:14003–14012.
7. Henzler-Wildman, K. A., V. Thai, M. Lei, M. Ott, M. Wolf-Watz, et al. 2007. Intrinsic motions along an enzymatic reaction trajectory. *Nature.* 450:838–844.
8. Kitao, A., and N. Gō. 1999. Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* 9:164–169.
9. Van Aalten, D. M. F., A. Amadei, A. B. M. Linssen, V. G. H. Eijsink, G. Vriend, et al. 1995. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins.* 22:45–54.
10. Berendsen, H. J. C., and S. Hayward. 2000. Collective protein dynamics in relation to function. *Curr. Opin. Struct. Biol.* 10:165–169.
11. Hayward, S., and A. Kitao. 2006. Molecular dynamics simulations of NAD^+-induced domain closure in horse liver alcohol dehydrogenase. *Biophys. J.* 91:1823–1831.
12. Snow, C., G. Qi, and S. Hayward. 2007. Essential dynamics sampling study of adenylate kinase: comparison to citrate synthase and implication for the hinge and shear mechanisms of domain motions. *Proteins.* 67:325–337.
13. Krebs, W. G., V. Alexandrov, C. A. Wilson, N. Echols, H. Yu, et al. 2002. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as useful classifying statistic. *Proteins.* 48:682–695.
14. Tama, F., and C. L. Brooks III. 2002. The mechanism and pathway of pH induced swelling in cowpea chlorotic motte virus. *J. Mol. Biol.* 318:733–747.
15. Valadié, H., J. J. Lacapčre, Y. -H. Sanejouand, and C. Etchebest. 2003. Dynamical properties of the MscL of *Escherichia coli*: a normal mode analysis. *J. Mol. Biol.* 332:657–674.
16. Chunyan, X., D. Tobi, and I. Bahar. 2003. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T ↔ R2 transition. *J. Mol. Biol.* 333:153–168.
17. Tobi, D., and I. Bahar. 2005. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. USA.* 102:18908–18913.
18. Cui, Q., G. Li, J. Ma, and M. Karplus. 2004. A normal mode analysis of structural plasticity in the biomolecular motor F_1-ATPase. *J. Mol. Biol.* 340:345–372.
19. Ikeguchi, M., J. Ueno, M. Sato, and A. Kidera. 2005. Protein structural change upon ligand binding: linear response theory. *Phys. Rev. Lett.* 94:078102.
20. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9:646–652.
21. Karplus, M., and J. Kuriyan. 2005. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA.* 102:6679–6685.
22. Bernadó, P., E. Mylonas, M. V. Petoukhov, M. Blackledge, and D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.
23. Seki, Y., T. Tomizawa, N. N. Khechinashvili, and K. Soda. 2002. Contribution of solvent water to the solution x-ray scattering profile of proteins. *Biophys. Chem.* 95:235–252.
24. Merzel, F., and J. C. Smith. 2002. SASSIM: a method for calculating small-angle x-ray and neutron scattering and the associated molecular envelope from explicit-atom models of solvated proteins. *Acta Crystallogr. D Biol. Crystallogr.* 58:242–249.
25. Shih, A. Y., L. G. Denisov, J. C. Phillips, S. G. Sligar, and K. Schulten. 2005. Molecular dynamics simulations of discoidal bilayers assembled from truncated human lipoproteins. *Biophys. J.* 88:548–556.
26. Lau, A. Y., and B. Roux. 2007. The free energy landscapes governing conformational changes in a glutamate receptor ligand-binding domain. *Structure.* 15:1203–1214.
27. Putnam, C. D., M. Hammel, G. L. Hura, and J. A. Tainer. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40:191–285.
28. Vachette, P., and D. I. Svergun. 2000. Small-angle x-ray scattering by solutions of biological macromolecules. *In* Structure and Dynamics of Biomolecules. E. Fanchon, G. Geissler, J. -L. Hodeau, J. -R. Regnard, and P. A. Timmins, editors. Oxford University Press, New York. 199–237.
29. Guinier, A., and G. Fournet. 1955. Small-angle scattering of X-rays. John Wiley & Sons, New York.
30. Kitao, A., F. Hirata, and N. Gō. 1991. The effect of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melitin in water and in vacuum. *Chem. Phys.* 158:447–472.
31. García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
32. Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins.* 17:412–425.
33. Ikeguchi, M. 2004. Partial rigid-body dynamics in NPT, NPAT and NPγT ensembles for proteins and membranes. *J. Comput. Chem.* 25:529–541.
34. MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

35. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

36. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, et al. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.

37. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.

38. Stuhrmann, H. B. 1970. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle-scattering function. *Acta Crystallogr. A.* 26:297–306.

39. Svergun, D. I., C. Barberato, and M. H. J. Koch. 1995. CRYSOL: a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.

40. Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, et al. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95:2267–2272.

41. Merzel, F., and J. C. Smith. 2002. Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl. Acad. Sci. USA.* 99:5378–5383.

42. Rosenberg, J. M. 1991. Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.* 1:104–113.

43. Doruker, P., L. Nilsson, and O. Kurkcuoglu. 2006. Collective dynamics of *Eco*RI-DNA complex by elastic network model and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 24:1–15.

44. Pingoud, A., and A. Jeltsch. 2001. Structure and function type II restriction endonucleases. *Nucleic Acids Res.* 29:3705–3727.

45. Maslen, E. N., A. G. Fox, and M. A. O'Keefe. 1999. X-ray scattering. *In* International Tables for Crystallography. A. J. C. Wilson and E. Prince, editors. Kluwer Academic Publishers, Dordrecht, The Netherlands. 548–575.

46. Abramowitz, M., and I. A. Stegun. 1970. Handbook of Mathematical Functions. Dover, New York.

47. Soda, K., Y. Miki, T. Nishizawa, and Y. Seki. 1997. New method for incorporating solvent influence into the evaluation of x-ray scattering intensity of proteins in solution. *Biophys. Chem.* 65:45–53.