



Internal dynamics of multidomain protein as revealed by an optimized neutron spin echo measurement and all-atom molecular dynamics simulation

Rintaro Inoue ^{1,*}, Tomotaka Oroguchi,^{2,*} Takashi Oda,^{3,*} Bela Farago,⁴ Anne Martel,⁴ Lionel Porcar,⁴ Mamoru Sato,^{5,†} and Masaaki Sugiyama ^{1,‡}

¹*Institute for Integrated Radiation and Nuclear Science, Kyoto University, Kumatori, Sennan-gun, Osaka 590-0494, Japan*

²*Department of Physics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*

³*Department of Life Science, Rikkyo University, Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*

⁴*Institut Laue Langevin, F-38402 Grenoble, France*

⁵*Graduate School of Medical Life Science, Yokohama City University, Yokohama 230-0045, Japan*



(Received 9 June 2023; accepted 11 October 2023; published 15 November 2023)

Identification of the internal dynamics of multidomain proteins is crucial for clarifying the mechanism of their functions. The neutron spin echo (NSE) technique is well suited for studying internal dynamics. However, the requirement for relatively high protein concentrations and the lack of appropriate analytical methods have impeded the investigation of the internal dynamics with NSE. To overcome these difficulties, we employed a unique approach to study the internal dynamics of a multidomain protein, EcoO109I, whose dynamics was anticipated to be pertinent to DNA degradation. We anticipated a synergetic effect between the NSE measurement at interference-free protein concentration and all-atom molecular dynamics simulation. Through this approach, the internal dynamics of EcoO109I was successfully observed within temporal and spatial scales. Additionally, principal component analysis (PCA) was applied to the internal dynamics trajectory to identify the dominant motion of the internal dynamics. The first PCA mode, which was the most cooperative among all PCA modes, mainly explained the internal dynamics. This dominant mode of EcoO109I exhibited the motion which facilitated both the access of DNA to the recognition site and the cleavage of DNA. Therefore, our approach can identify the functionally relevant internal dynamics of multidomain proteins.

DOI: [10.1103/PhysRevResearch.5.043154](https://doi.org/10.1103/PhysRevResearch.5.043154)

I. INTRODUCTION

Many studies have suggested that the internal dynamics of multidomain proteins, such as locking, cleft opening, and lid opening motions, play significant roles in the development of their biological functions [1,2]. Therefore, experimental techniques that can directly detect these internal dynamics should be established to elucidate the mechanisms of protein functions. Computational approaches predict that internal dynamics are located at the temporal scale ranging from nanoseconds (ns) to microseconds and the spatial scale ranging from ns to several tens of ns, respectively. The neutron spin echo (NSE) [3–11] is the only available experimental technique that can simultaneously access both scales.

NSE mainly observes the combination of translational diffusion, rotational diffusion, and internal dynamics of a multidomain protein in solution [Fig. 1(a)] as the Q dependence of the diffusion coefficient [$D(Q)$] [blue line in

Fig. 1(b)]. Two major problems need to be resolved in the NSE study of multidomain proteins. First, NSE measurement requires a relatively high protein concentration (approximately 50 mg/ml) to secure sufficient statistics [7,8,10–12]. However, high protein concentrations cause interparticle interference and hydrodynamic interactions, making the precise analysis of internal dynamics difficult. Furthermore, high protein concentrations occasionally promote the formation of oligomers and aggregates in the sample solution, which can affect the interpretation of the internal dynamics of the target monomeric protein [12]. Second, an analytical method has not yet been established. For example, the NSE spectrum is interpreted as a combination of several low-frequency normal modes [6–8,10]. However, neither the time information nor the time dependence of displacement of the analyzed modes is considered in this interpretation.

Here, we adopted the following strategies to overcome these problems. First, the NSE measurement was performed at a high neutron flux, owing to the recent development and upgrade of NSE instruments. As a result, the NSE measurement at the lower protein concentration with less interparticle interference and hydrodynamics (interference-free concentration) has been successfully achieved. Second, an all-atom molecular dynamics (MD) simulation [13] was employed to analyze the internal dynamics instead of normal mode analysis (NMA). Notably, the all-atom MD simulation considers the time information and time dependence of displacement

*These authors contributed equally to this work.

†msato@yokohama-cu.ac.jp

‡sugiyama.masaaki.5n@kyoto-u.ac.jp

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

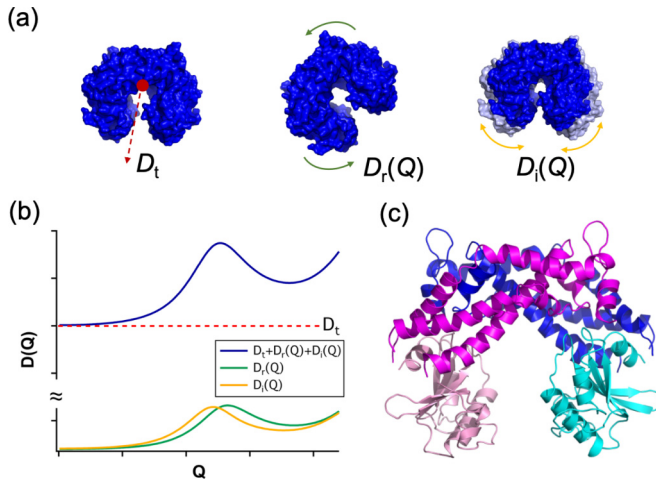


FIG. 1. (a) Schematic view of translational diffusion (D_t), rotational diffusion [$D_r(Q)$], and internal dynamics [$D_i(Q)$]. (b) Q dependences of $D_t + D_r(Q) + D_i(Q)$ (blue line), $D_r(Q)$ (green line), and $D_i(Q)$ (yellow line). Dashed red line corresponds to the translational diffusion constant D_t . (c) Crystallographic structure of DNA-free form of EcoO109I of which the dimeric domain of subunit A, the catalytic domain of subunit A, the dimeric domain of subunit B, and the catalytic domain of subunit B are represented by magenta, pink, blue, and cyan, respectively.

of the analyzed modes. The combination of NSE measurement at interference-free concentrations and all-atom MD simulation could be more advantageous for studying the internal dynamics of multidomain proteins than previous NSE studies.

These combined methods were applied to elucidate the biological function of a multidomain protein EcoO109I [14], which degrades foreign DNA in bacteria. EcoO109I is a homodimer, with subunits comprising dimerized and catalytic domains [Fig. 1(c)] [14]. The foreign DNA is captured at the intervening space between the two catalytic domains of the respective subunits of the homodimer; however, the intervening space in the crystal structure of the DNA-free form is narrower than the diameter of double-strand DNA (DNA) (ca. 20 Å) [14]. Then, a simple question arises: “How can EcoO109I capture DNA?” We expect that internal dynamics exists for capturing DNA. Our preliminary small-angle x-ray scattering (SAXS) study revealed that the solution structure of EcoO109I deviates from its crystal structure [15,16], supporting the existence of internal dynamics in EcoO109I. Although the preliminary studies suggested that the internal dynamics of EcoO109I might be relevant to capturing DNA, more work needs to be done regarding the direct experimental observation of the internal dynamics. Therefore, the direct observation (with an optimized NSE) and analysis of internal dynamics (with an all-atom MD) are mandatory to elucidate its internal dynamics in solution.

In this report, we showed that the internal dynamics of EcoO109I can be successfully observed at interference-free concentration using an NSE experiment and discuss the dynamics from the perspective of biological function.

II. MATERIALS AND METHODS

A. Overexpression and purification of EcoO109I

Escherichia coli HB101 cells were transformed using the plasmid vector pUC118, which encodes EcoO109I. Next, the colony of *E. coli* cells was precultured at 37 °C for 16 h in LB medium (1 g of bactotryptone, 0.5 g of yeast extract, and 1 g of NaCl per 100 ml) containing 100 µg/ml ampicillin. Afterward, the precultured medium was inoculated into LB medium containing 100 µg/ml ampicillin and 1 mM isopropyl β-D-thiogalactopyranoside. After cultivation at 37 °C for 16 h, the cells were harvested by centrifugation and stored at −80 °C. Next, the cells were thawed, resuspended in buffer I [10 mM Tris-HCl pH8.0, 150 mM KCl, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, 5% v/v glycerol, and 0.5 mg/ml hen egg white lysozyme (Sigma-Aldrich)], and disrupted by sonication on ice. Subsequently, the cell lysate was centrifuged for 30 min at 20 400 × g , and streptomycin sulfate was added to the supernatant at a final concentration of 4% w/v. Afterward, the precipitate was removed by centrifuging for 30 min at 20 400 × g , and the supernatant was dialyzed against buffer II (10 mM potassium phosphate buffer pH7.5, 20 mM KCl, 0.1 mM EDTA, 10 mM 2-mercaptoethanol, and 5% v/v glycerol). After dialysis, the precipitate was removed by centrifugation for 30 min at 20 400 × g and filtered. Next, the sample solution was loaded onto a DEAE Sepharose column (Cytiva, Marlborough, MA, USA) equilibrated with Buffer II. EcoO109I was eluted using a linear gradient of 20–1000 mM KCl. Afterward, the eluted fraction was dialyzed against buffer II. Next, the dialyzed samples were filtered and loaded onto a HiTrap heparin column (Cytiva) equilibrated with buffer II. EcoO109I was eluted using a linear gradient of 20–450 mM KCl; the eluted fraction was diluted twice with buffer II. Subsequently, the sample solution was loaded onto a HiTrap Q column (Cytiva) equilibrated with buffer II. EcoO109I was eluted using a linear gradient of 200–400 mM KCl. Next, the eluted fractions were concentrated using an Amicon Ultra-15, 30-kDa MwCo (an ultrafiltration device), flash cooled by liquid nitrogen, and stored at −80 °C. After that, the purity was confirmed using sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS PAGE). Next, EcoO109I was loaded onto a Superdex200 10/300 GL column (Cytiva) equilibrated with buffer III (10 mM potassium phosphate buffer pH7.5, 300 mM NaCl, 0.1 mM EDTA, and 10 mM 2-mercaptoethanol), and the eluted fraction was concentrated using an Amicon Ultra-15, 30-kDa MwCo. For the NSE and small-angle neutron scattering (SANS) experiments, the sample solution was dialyzed against a D₂O buffer with the same composition as buffer III. Next, the protein concentration was determined by measuring the optical density at 280 nm. Notably, the D₂O ratio used for SANS and NSE measurements was calculated as 99.0% from Fourier transform infrared spectroscopy measurement.

B. DNA digestion assay

Plasmid DNA pACYC Duet-1 was used as the substrate for EcoO109I. pACYC Duet-1 was amplified in *E. coli* JM109 cells and extracted using a miniprep kit (Wizard Plus SV Minipreps DNA Purification Systems, Promega). Next, the

extracted pACYC Duet-1 was linearized using NheI digestion. For further purification, linearized pACYC Duet-1 was loaded onto 1% agarose gel with SDS containing loading buffer (0.9% SDS, 50% v/v glycerol, and 0.05% w/v bromophenol blue) and subjected to electrophoresis with conventional Tris-acetate-EDTA buffer (TAE: 40 mM Tris, 40 mM acetic acid, and 1 mM EDTA). Afterward, the pACYC Duet-1 was extracted from the gel using a gel extraction kit (Wizard SV Gel and PCR Clean-Up System; Promega). After that, the extracted pACYC Duet-1 was dialyzed against 0.1 mM EDTA and used in the DNA digestion assay.

A DNA digestion assay was performed at ten different temperatures (4 °C, 10 °C, 15 °C, 20 °C, 25 °C, 30 °C, 37 °C, 45 °C, 50 °C, and 55 °C). The reaction mixture (10 μ l) containing 6.25 ng/ μ l pACYC Duet-1, 0.625 fmol/ μ l EcoO109I, 2.5 mM potassium phosphate buffer pH7.5, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 0.05 mM EDTA, and 5 μ g/ml BSA was incubated at each temperature for 2 h. The reaction was terminated by adding SDS to the loading buffer. Next, the reaction mixture was subjected to electrophoresis using 1% agarose gel in TAE buffer. Products of EcoO109I digestion (1.3 and 2.7 kbp) were detected using ethidium bromide stain and quantified by comparing the band intensity with known amounts of the products. EcoO109I degraded DNA at temperatures >10 °C (Fig. S1 in the Supplemental Material [17]). The activity of EcoO109I remains relatively constant within the temperature range from 15 °C to 37 °C. It should be noted that the long-term stability of EcoO109I decreases at temperatures higher than 37 °C. Considering both the stability of the sample during measurements and its activity, we have decided to perform all the measurements at 15 °C.

C. Small-angle neutron scattering

The SANS experiments were performed using a D22 installed at Institut Laue Langevin (ILL), Grenoble, France. Wavelength and sample-to-detector distances were 5.6 Å and 4.0 m, respectively, and the measurements were performed at 15 °C. The SANS intensity was corrected for background, empty cell, buffer scattering, and transmittance. Additionally, the SANS profiles were converted to the absolute intensities (cm⁻¹) using incident beam flux. All the data reduction procedures were performed with the software program GRASP [18].

D. Dynamic light scattering

Dynamic light scattering measurements were performed using a system equipped with a 22-mW He-Ne laser, an avalanche photodiode (APD, ALV, Germany) mounted on a static/dynamic compact goniometer, ALV/LSE-5003 electronics, and an ALV-5000 correlator (ALV-Laser Vertrieb-Gesellschaft GmbH, Langen, Germany). The measurements were performed at 15 °C. CONTIN analysis [19] was used to calculate the decay rate distribution.

E. Neutron spin echo

NSE measurements were performed using IN15 [20] at ILL. Quartz cells (30 \times 30 mm²) with a 4-mm path length were used. Additionally, three wavelengths (6.3,

8.0, and 10.0 Å) were used, and the measurement was performed at 15 °C.

F. MD simulation

All-atom MD simulations of EcoO109I with explicit water molecules were conducted independently using the AMBER16 [21] software package with the AMBER ff15FB [22] force field and TIP3PFB [23] water model. An atomic model of the crystal structure (PDB accession code 1WTD) [15] was used for the simulation, and the missing regions were modeled as described in a previous study [16]. Moreover, the structure with identified crystal water molecules was solvated in a truncated octahedron box of a water model with a clearance of 20.5 Å between EcoO109I and the face of the box. Next, Na⁺ and Cl⁻ ions were added to adjust the ionic strength of the system to that in the NSE measurement, 350 mM. Electrostatic interactions were treated using the particle-mesh Ewald method [24] with a real-space cutoff of 10 Å. Furthermore, the Lennard-Jones interactions were truncated beyond 10 Å using continuum model correction. Bonds involving hydrogen atoms were constrained using the SHAKE method [25], and the time step was set to 2 fs. An Andersen temperature-coupling scheme [26] was used to control the temperature. The system was first subjected to energy minimization of 1000 steps. Next, the system's temperature was gradually increased from 10 to 293 K using an *NPT* run of 300 ps under 1 atm. For the equilibration of the modeled regions, a 30-ns *NPT* run was conducted with a harmonic restraint for heavy atoms of EcoO109I excluding those in the modeled regions with a force constant of 1.0 kcal mol⁻¹ Å⁻². Lastly, a production run of a 2- μ s *NVT* run was conducted without restraints. The simulations were conducted on an NVIDIA Tesla P100 GPU.

III. RESULTS AND DISCUSSION

A. Small-angle neutron scattering measurement

To determine the optimal condition for NSE measurement, the interference-free concentration, we performed small-angle neutron scattering (SANS) measurements of EcoO109I solutions with protein concentrations *c*, 2.2, 5.3, and 10.0 mg/ml. Figure S2 [17] shows the concentration dependence of the concentration-normalized SANS profiles $[I(Q)/c][Q = (4\pi/\lambda)\sin(\theta)]$, where λ and 2θ are the wavelength of the incident neutron beam and scattering angle, respectively; $I(Q)$: scattering intensity; $I(0)/c$: the concentration-normalized forward scattering intensity; R_g : radii of gyration. The interparticle interference and aggregation were negligible, up to 10.0 mg/ml (refer to the Supplemental Material [17]). Considering the balance between scattering intensity and particle interference, we adopted the EcoO109I solution with 10.0 mg/ml as the concentration with an interference-free and maximized intensity for the present NSE measurement.

B. NSE measurement and analyses

Figure 2 shows the intermediate scattering function $[I(Q, t)/I(Q, 0)]$ of the 10.0 mg/ml EcoO109I solution. $I(Q, t)/I(Q, 0)$

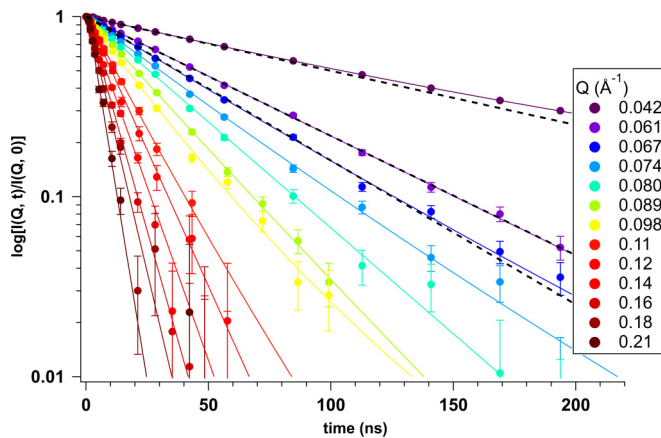


FIG. 2. Semilogarithmic plot of intermediate scattering function $[I(Q, t)/I(Q, 0)]$ from EcoO109I at 10.0 mg/ml. Purple to brown circles correspond to $Q = 0.042$ and 0.21 \AA^{-1} , respectively, and the results of fit with cumulant analysis with Eq. (1) were given by the solid lines with the same color. Dashed black lines at $Q = 0.042$, 0.061 , and 0.067 \AA^{-1} correspond to the results of calculation with $\exp(-D_t Q^2 t)$.

was fitted using the following equation [7],

$$I(Q, t)/I(Q, 0) = \exp(k_1 t + 0.5k_2 t^2), \quad (1)$$

where k_1 and k_2 are the first- and second-order cumulant values, respectively. Since the dynamics other than translational diffusion is observable within the temporal scale accessed by the present NSE measurement, we determined to adopt cumulant expansion for the proper evaluation of the initial decay rate. Additionally, the Q dependence of the effective diffusion coefficient $[D_{\text{eff}}(Q)]$ can be computed from the first-order cumulant (k_1) [7] as follows:

$$k_1 = \lim_{t \rightarrow 0} \frac{\partial}{\partial t} \frac{I(Q, t)}{I(Q, 0)} = -Q^2 D_{\text{eff}}(Q). \quad (2)$$

$D_{\text{eff}}(Q)$ is related to a single protein diffusion coefficient at the infinite dilution limit $[D_0(Q)]$ through the equation below:

$$D_{\text{eff}}(Q) = D_0(Q)H(Q)/S(Q), \quad (3)$$

where $H(Q)$ and $S(Q)$ correspond to Q dependence of the hydrodynamic interaction [8] and structure factor, respectively. For further analysis, we have to compute $D_0(Q)$ from $D_{\text{eff}}(Q)$. As described above, $S(Q)$ is equal to 1 at 10.0 mg/ml due to the interference-free concentration. Then, Eq. (3) reduces to Eq. (4):

$$D_{\text{eff}}(Q) = D_0(Q)H(Q). \quad (4)$$

With Q value approaching to 0, the protein can be treated as a pointlike object, meaning that only the translational diffusion coefficient (D_t) is observable, and consequently, $D_{\text{eff}}(Q)$ is equal to D_t . Therefore, in the lower- Q region the relationship of $D_t = D_0(Q)H(Q)$ holds. Since the present MD simulation box contained only one EcoO109I molecule, it is comparable to the situation of EcoO109I at infinite dilution limit. It is therefore possible to calculate the translational diffusion constant at the infinite diffusion limit ($D_{t, \text{MD}}$). In addition, this $D_{t, \text{MD}}$ is equal to $D_0(Q)$ at the lower- Q region. $D_{t, \text{MD}}$ was calculated to $4.20 \pm 0.06 \text{ \AA}^2/\text{ns}$ utilizing

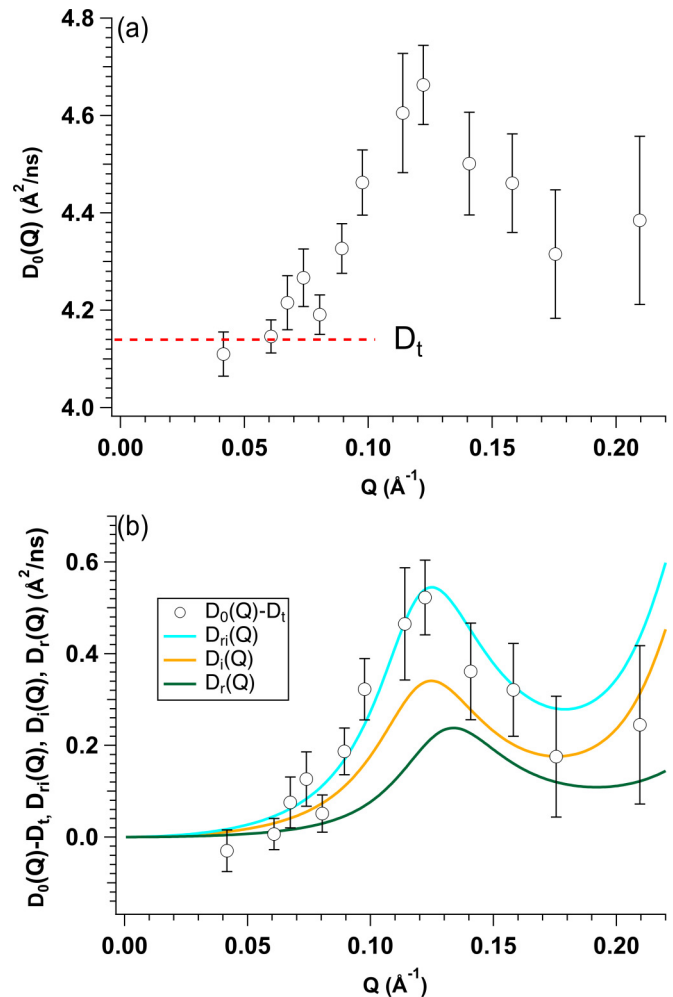


FIG. 3. $D_0(Q)$, $D_0(Q) - D_t$, $D_{ri}(Q)$, $D_i(Q)$, and $D_r(Q)$ are plotted. (a) $D_0(Q)$ (black circle) from EcoO109I at 10.0 mg/ml. Dashed red line corresponds to the translational diffusion constant (D_t). (b) $D_0(Q) - D_t$ (black circle), $D_{ri}(Q)$ (light blue line), $D_i(Q)$ (yellow line), and $D_r(Q)$ (green line), respectively.

HYDROPRO [27]. Through the Q^2 dependence of relaxation rate is measured by dynamic light scattering (DLS) (Fig. S3 [17]), D_t from DLS was calculated to $4.14 \pm 0.02 \text{ \AA}^2/\text{ns}$. It means that D_t is equal to $D_0(Q)$ in the lower- Q region (or $D_{t, \text{MD}}$) within experimental error. Additionally, $H(Q)$ can be approximated as 1 (calculation of $D_0(Q)$ from $D_{\text{eff}}(Q)$ in the Supplemental Material [17]). Therefore, $D_{\text{eff}}(Q)$ can be regarded as $D_0(Q)$ under the present condition without any additional correction. It should be noted that the interference-free concentration is dependent upon protein size and ionic strength and so on. In other words, an interference-free concentration does not universally hold for all the proteins. By choosing the optimum experimental condition in the present work, EcoO109I solution with 10.0 mg/ml finally satisfied the criteria of the interference-free concentration. The interference-free concentration was beneficial for obtaining a reliable $D_0(Q)$ experimentally because the uncertainty of $H(Q)$ could lead to incorrect $D_0(Q)$.

$D_0(Q)$ is approximately expressed as a linear combination of translational diffusion constant (D_t), rotational diffusion

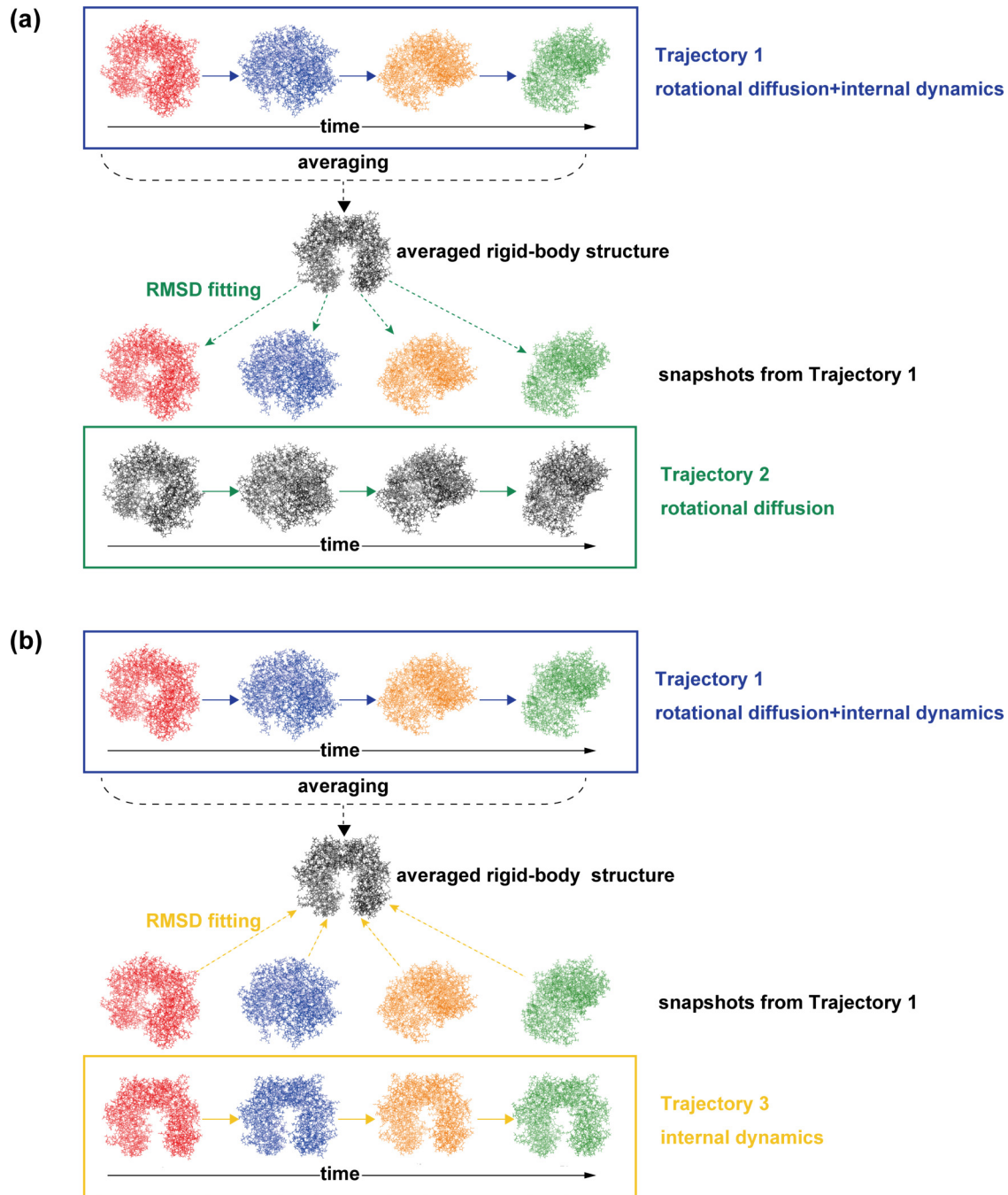


FIG. 4. Schematic views of procedures for reconstructing trajectory 2 originating from the rotational diffusion (a) and trajectory 3 originating from the internal dynamics (b).

$[D_r(Q)]$, and internal dynamics $[D_i(Q)]$ (Fig. 1) as follows:

$$D_0(Q) = D_t + D_r(Q) + D_i(Q) = D_t + D_{ri}(Q), \quad (5)$$

where $D_{ri}(Q)$ is the Q dependence of the linear combination of the rotational diffusion and internal dynamics. At the lower- Q region, $D_0(Q)$ coincided with D_t within experimental error. However, $D_0(Q)$ was higher than D_t in the higher- Q range [$Q > 0.06 \text{ \AA}^{-1}$, Fig. 3(a)]. This indicates that dynamics other than translational diffusion was observed in the Q region.

C. Calculation of $D_{ri}(Q)$, $D_r(Q)$, and $D_i(Q)$ through the trajectories derived from all-atom MD simulation

It is necessary to calculate $D_{ri}(Q)$, $D_r(Q)$, and $D_i(Q)$ for further analysis of the internal dynamics. For this purpose, we focused on the trajectory of an all-atom MD simulation for 2000 ns, which covered the time window of the present NSE measurement. The time course of 20 000 structures extracted from the trajectory every 0.1 ns was termed trajectory 0. To eliminate the contribution of translational diffusion from trajectory 0, the center of mass of each structure in trajectory 0 was translated to the origin of the MD simulation box. The

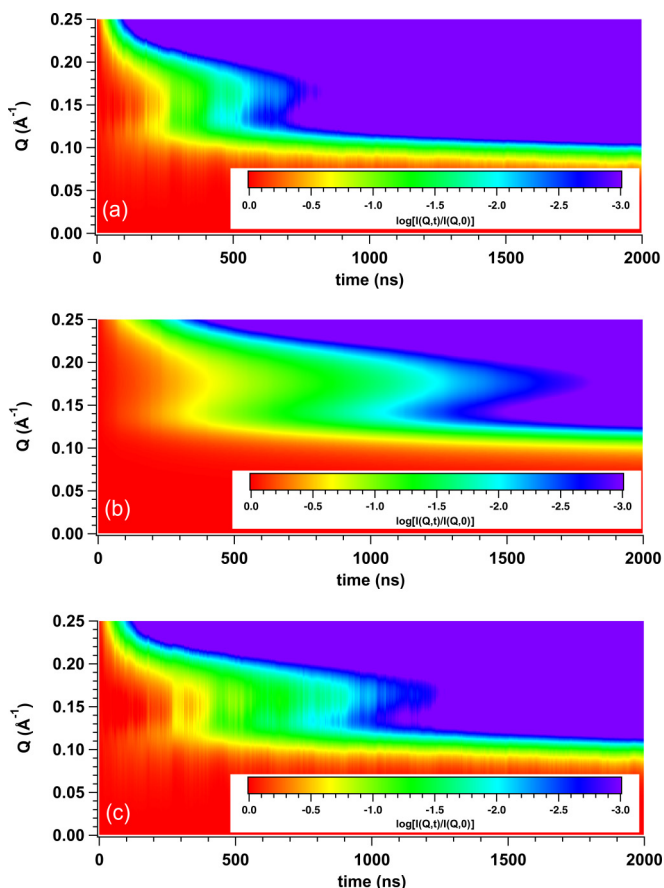


FIG. 5. (a) Two-dimensional $I(Q, t)/I(Q, 0)$ calculated from trajectory 1 (rotational diffusion + internal dynamics). (b) Two-dimensional $I(Q, t)/I(Q, 0)$ calculated from trajectory 2 (rotational diffusion). (c) Two-dimensional $I(Q, t)/I(Q, 0)$ calculated from trajectory 3 (internal dynamics).

time course of these translated structures is termed trajectory 1 (Fig. 4). Notably, trajectory 1 comprises rotational diffusion and internal dynamics. $I(Q, t)/I(Q, 0)$ was calculated using trajectory 1 [Fig. 5(a)], and $D_{ri}(Q)$ was computed from $I(Q, t)/I(Q, 0)$ using Eq. (1). As indicated by the blue line in Fig. 3(b), $D_0(Q) - D_i$ was well reproduced by $D_{ri}(Q)$. Additionally, this means that the set of parameters used for the all-atom MD simulation was experimentally validated by the present NSE result.

To calculate $D_r(Q)$ and $D_i(Q)$, trajectory 1 was decomposed into the contribution of rotational diffusion and internal dynamics using the following procedures. First, the structures in trajectory 1 were aligned with the crystal structure. By averaging these aligned structures, an averaged rigid-body structure was obtained as the template. Next, the averaged rigid-body structure was aligned to each structure in trajectory 1. The time course of the aligned averaged rigid-body structures is termed trajectory 2 [Fig. 4(a)]. Notably, trajectory 2 only exhibits rigid-body rotational diffusion without internal dynamics. Figure 5(b) shows $I(Q, t)/I(Q, 0)$ calculated from trajectory 2, and $D_r(Q)$ is calculated from $I(Q, t)/I(Q, 0)$ using Eq. (1) [refer to the green line in Fig. 3(b)].

Next, the trajectory comprising the internal dynamics was extracted from trajectory 1 using the following proce-

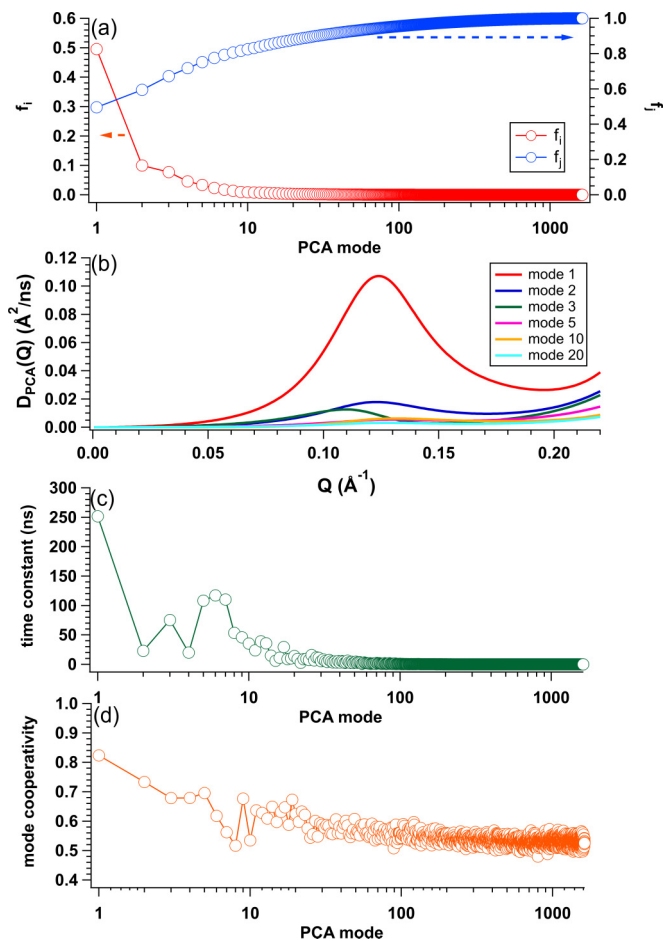


FIG. 6. (a) Fraction of C^α -mean square fluctuation (red circle) and sum fraction (blue circle) as a function of PCA mode number. (b) $D_{PCA}(Q)$ of PCA modes 1 (red line), 2 (blue line), 3 (green line), 5 (pink line), 10 (orange line), and 20 (light blue line). (c) Time constant as a function of PCA mode number. (d) Mode cooperativity as a function of PCA mode number.

dures. First, each structure in trajectory 1 was aligned with an averaged rigid-body structure. The time course of the aligned structures is termed trajectory 3 [Fig. 4(b)]. Trajectory 3 has only internal dynamics without rotational diffusion. Figure 5(c) shows $I(Q, t)/I(Q, 0)$ and $D_i(Q)$ [yellow line in Fig. 3(b)] computed using trajectory 3. Notably, the coupling between the internal dynamics and rotational diffusion is negligible within the spatial and temporal scales covered by the present NSE measurement (comparison of $D_{ri}(Q)$ to $D_r(Q) + D_i(Q)$ in the Supplemental Material and Fig. S4 [17]), thereby validating the application of Eq. (4) for this analysis.

$D_i(Q)$ is obvious in the Q region, indicating that the internal dynamics of EcoO109I was experimentally detected using the NSE measurement.

D. Analysis of internal dynamics contributing to NSE results

Trajectory 3 was analyzed as a linear combination of independent modes using principal component analysis (PCA) [28] to identify the dominant motions in the internal dynamics. In PCA, the structural motion $R(t)$ is represented

by

$$\mathbf{R}(t) = \langle \mathbf{R}(t) \rangle + \sum_{i=1}^M \psi^i(t) \cdot \mathbf{e}^i, \quad (6)$$

where $\langle \mathbf{R}(t) \rangle$ is the averaged rigid-body structure, \mathbf{e}^i and $\psi^i(t)$ are the eigenvector and time dependence of amplitude of the i th mode, respectively. M is the number of modes (544 amino acid residues $\times 3 = 1632$). Figure 6(a) shows the fraction of C^α -mean square fluctuation (C^α -MSF = $\langle \{\psi^i(t)\}^2 \rangle$) of PCA mode i ($1 \leq i \leq 1632$), f_i , to the total C^α -MSF. Additionally, we paid attention to the partial sum fraction (f_j), which is given by $f_j = \sum_{i=1}^j f_i$ ($1 \leq j \leq 1632$). The fraction of the first PCA mode accounted for approximately 50% of total C^α -MSF.

Moreover, the partial sum fraction up to the first three and 20 PCA modes accounted for as much as 67% and 90%, respectively, of the total. This implies that most internal dynamics can be approximated by a linear combination of several PCA modes with large amplitudes. Next, it should be confirmed whether the temporal and spatial scales of the PCA modes with large amplitudes are within the temporal and spatial windows accessed using the NSE measurement. For this purpose, we calculated the $D(Q)$ and time constants of the PCA modes. Notably, our approach provides the time information and time dependence of displacement of the analyzed modes, addressing a limitation of the present NMA.

First, the PCA trajectory of the i th mode $\mathbf{R}_{\text{PCA}}^i(t)$ is constructed as follows:

$$\mathbf{R}_{\text{PCA}}^i(t) = \langle \mathbf{R}(t) \rangle + \psi^i(t) \cdot \mathbf{e}^i. \quad (7)$$

Next the corresponding $I(Q, t)/I(Q, 0)$ and $D(Q)$ of each PCA mode [$D_{\text{PCA}}^i(Q)$] were computed using $\mathbf{R}_{\text{PCA}}^i(t)$.

$D_{\text{PCA}}^i(Q)$'s from several PCA modes are shown in Fig. 6(b). Additionally, peaks were observed for the first, second, and third PCA modes in the Q range, 0.10–0.15 \AA^{-1} . Notably, the first PCA mode, whose peak position coincided with that of $D_i(Q)$, dominated $D_i(Q)$ the most compared with other PCA modes.

Next, we analyzed the time constants of each PCA mode. As given in Eq. (7), the normalized autocorrelation function (ACF) of $\psi^i(t)$ was calculated from the time dependence of displacement $\psi^i(t)$ [Fig. S5(a) [17]].

$$\text{ACF}_i(\tau) = \frac{\langle \psi^i(t + \tau)\psi^i(t) \rangle}{\langle \psi^i(t)^2 \rangle}. \quad (8)$$

The ACFs for the first 20 modes are shown in Fig. S5(b) [17]. It was revealed that all calculated ACFs were not described by simple functions, such as single or stretched exponential functions. Therefore, we regarded the time the ACF decayed to $1/e$ as the time constant. Figure 6(c) shows the time constants as functions of the PCA mode number. Moreover, the calculated time constants were almost within the timeframe of the NSE measurements, particularly the time constant from the first PCA mode (approximately 250 ns).

In a previous analysis utilizing NMA, the amplitude of the analyzed normal mode was calculated and compared to NSE results. It should be claimed that time information and time dependence of displacement of the analyzed modes are not included in NMA; these aspects are explicitly included in all-atom MD simulation. Taking advantage of this information from all-atom MD simulation, both the temporal and spatial information of each PCA mode could be independently calculated without referring to the NSE results in the present analysis. Therefore, PCA modes with large amplitudes are

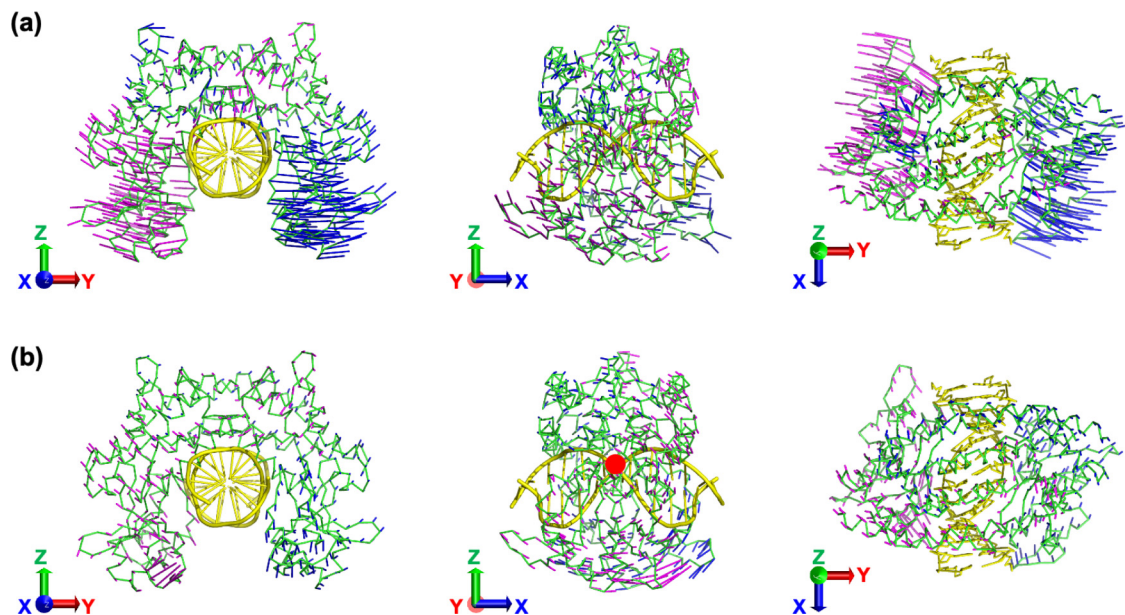


FIG. 7. (a) Domain motion of first PCA mode with the bound DNA in the front (left panel), side (center panel), and top views (right panel). The amplitude (3σ) and direction of C_α displacements along the first PCA mode are represented by arrows, which are colored according to the two homologous subunits. (b) Domain motion of first PCA mode with the bound DNA in the front (left panel), side (center panel), and top views (right panel). The amplitude (3σ) and direction of C_α displacements along the second PCA mode are represented by arrows, which are colored according to the two homologous subunits.

detectable within the temporal and spatial scales accessed by the NSE measurement.

The previous studies have suggested that cooperative internal dynamics is primarily responsible for biologically significant functions [1,2]. Therefore, we analyzed the cooperativity of the PCA modes to identify the biologically relevant ones (see calculation of the cooperativity score for each PCA mode in the Supplemental Material [17]). Figure 6(d) shows the dependence of the cooperativity score on the mode number. The cooperativity score decreased with increasing mode number, suggesting that the analyzed modes became incoherent and local motions in higher PCA modes. The cooperativity scores in the first five modes were higher than the other PCA modes. Additionally, the first and second PCA modes were the most cooperative.

E. Cooperative large-amplitude PCA modes and their functional relevance

To determine the relevance of the first and second PCA modes to function, we focused on the contribution of these modes to motions that capture or cleave the substrate DNA (Fig. 7). First, the averaged rigid-body structure was aligned with the crystal structure of the DNA-bound form [15]. Second, the DNA molecule in the crystal structure of the DNA-bound form was placed in the average rigid-body structure. Third, the trajectories of the PCA modes were constructed for the DNA and averaged rigid-body structure complex. Figure 6 shows the first and second PCA modes in terms of motion against the bound DNA from the front (left panel), side (center panel), and top views (right panel). The front view shows that the first mode is mainly described by an open-close motion expected to capture DNA [the left panel in Fig. 7(a)]. Additionally, the first mode includes motion along the helical axis of the DNA (the side view). Combining this motion with the open-close motion enables the two catalytic domains to fit into the major clefts of the DNA (the top view). The first mode is necessary for catalytic residues to attack the cleavage sites of DNA. Additionally, the second PCA mode is the twisting motion of the two subunits, which minorly contributes to an open-close motion [Fig. 7(b)].

For the quantitative analysis of functional relevance, we calculated the correlation between the PCA modes and the structural change upon DNA binding [Fig. 8(a)] (refer to the Supplemental Material for the calculation details [17]). The highest correlation was observed in the first mode, with a correlation coefficient of 0.78, followed by the second mode (0.47); the other modes exhibited lower correlations. This correlation analysis supports the visual inspection of the functional significance of the first PCA mode.

In the trajectory of the first PCA mode, the distance between the catalytic residue D110 and the DNA cleavage site ($d_{D110, DNA}$) occasionally approached that of the crystal structure of the DNA-bound form [Figs. 8(b) and 8(c)]. However, the variations in $d_{D110, DNA}$ were small in the trajectory of the second PCA mode [Fig. S6(a) [17]]. These results show that the motion of the first PCA mode [left and middle panels in Fig. 7(a)] enabled the two catalytic domains to fit into the major grooves of the DNA (right panel). It contributed to the cleavage of the specific DNA site. Notably, the small motion

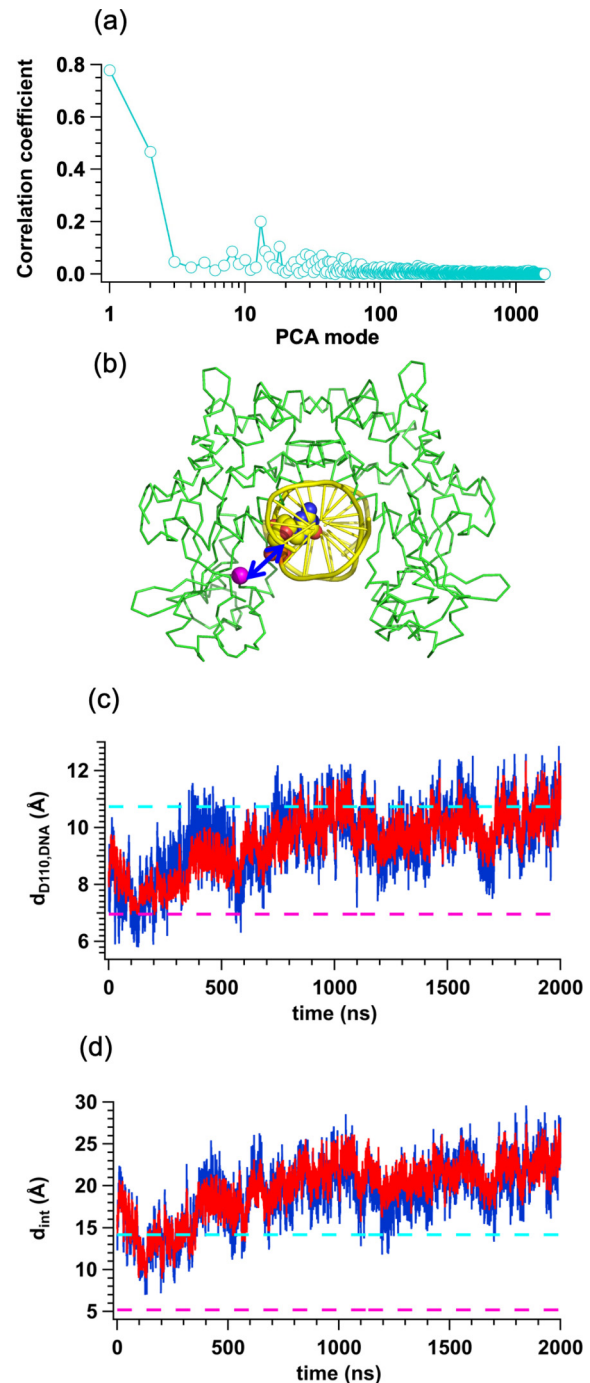


FIG. 8. (a) Correlation coefficient between the PCA modes and the structural change upon DNA binding as a function of PCA mode number. (b) Schematic view of distance of catalytic residue(D110) and cleavage site of DNA; blue arrow corresponds to the distance between D110 and DNA cleavage site. (c) Time dependence of distance between D100 and DNA cleavage site ($d_{D110, DNA}$) from total PCA modes (blue solid line) and that from PCA mode 1 (red solid line). Light blue and pink dotted lines correspond to the $d_{D110, DNA}$ of DNA-free, and DNA-bound form of the crystallographic structure of EcoO109I, respectively. (d) Time dependence of the narrowest intervening space of two domains (d_{int}) from total PCA modes (blue solid line) and that from PCA mode 1 (red solid line). Light blue and pink dotted lines correspond to d_{int} of DNA-free and DNA-bound form of the crystallographic structure of EcoO109I, respectively.

of the catalytic domain along the second PCA mode may be used for the fine arrangement of the DNA fit.

Additionally, we expected that the open-close motion observed in the first PCA mode would affect the narrowest distance between the two catalytic domains (d_{int}). Consistent with our expectation, d_{int} occasionally became larger enough to be assessed using DNA during the 2000-ns MD simulation [Fig. 8(d)]. Conversely, the contributions from the other modes, including the second mode (Fig. S6(b) [17]), were trivial. Moreover, the first PCA mode supported DNA access to the recognition site of EcoO109I.

These results identify the first PCA mode as the only functionally relevant cooperative mode. Notably, this functional mode is the most cooperative motion and has the longest timescale, comparable to that accessed using the present NSE measurement.

IV. SUMMARY

NSE is well suited for studying the internal dynamics of a multidomain protein. However, requirement of relatively high protein concentration and lack of an appropriate analyzing method limited its application to the investigation of internal dynamics. Because of recent development of the NSE instrument and the computational approach, the combination of NSE measurement at interference-free concentration (optimized NSE measurement) and all-atom MD simulation could be a promising technique for investigating the internal dynamics. This technique is applicable to a multidomain protein such as EcoO109I. We demonstrated that $D_{\text{eff}}(Q)$ derived from NSE is directly regarded as a single protein diffusion coefficient $D_0(Q)$ without any correction. In addition, $D(Q)$ of internal dynamics [$D_i(Q)$] was successfully extracted from $D_0(Q)$ with the aid of 2000-ns MD simulation.

To identify the internal dynamics which can dominantly account for the NSE result, PCA was applied to the trajectory

of the internal dynamics. $D(Q)$'s and time constants calculated from the PCA trajectories showed that the modes with large amplitudes were detectable by the present NSE measurement. Especially, the first PCA mode, which was the motion to fit the two catalytic domains into the major clefts of the DNA, dominated the $D_i(Q)$ most among the analyzed modes. The detailed analyses of the trajectories of the PCA modes revealed two interesting aspects of the first PCA one. Firstly, the motion of the catalytic residues to attack the cleavage site of DNA was predominantly described by this mode. Secondly, the open-close motion with this mode occasionally made the space of the recognition site larger than the diameter of the DNA.

From these results, we conclude that the current combination of NSE measurements at interference-free concentrations and all-atom MD simulations has successfully revealed biologically important motions of EcoO109I on both spatial and temporal scales. This is, therefore, an applicable technique for identifying functionally relevant internal dynamics of the other multidomain proteins.

ACKNOWLEDGMENTS

NSE and SANS experiments at ILL were performed under Proposals No. 8-04-715 and No. 8-03-875, respectively. This work was supported by MEXT/JSPS KAKENHI grants (Grants No. JP19KK0071 and No. JP20K06579 to R.I., and Grants No. JP18H05229, No. JP18H05534, and No. JP18H03681 to M.S.). Additionally, this work was partially supported by a project for the construction of a basis for advanced materials science and analytical studies through the innovative use of quantum beams and nuclear sciences at the Institute for Integrated Radiation and Nuclear Science, Kyoto University. This study was also partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research [Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)] from AMED (JP22ama121001j0001) to M.S..

-
- [1] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).
 - [2] D. A. Case, Normal-mode analysis of protein dynamics, *Curr. Opin. Struct. Biol.* **4**, 285 (1994).
 - [3] F. Mezei, *Neutron Spin Echo* (Springer, Berlin, 1979).
 - [4] T. Imae, T. Kanaya, M. Furusaka, and N. Torikai, *Neutrons in Soft Matter* (Wiley, New York, 2011).
 - [5] D. Richter, M. Monkenbusch, A. Arbe, and J. Colmenero, Neutron spin echo in polymer systems, *Adv. Polym. Sci.* **174**, 1 (2005).
 - [6] Z. Bu, R. Biehl, M. Monkenbusch, D. Richter, and D. J. E. Callaway, Coupled protein domain motion in *Taq* polymerase revealed by neutron spin-echo spectroscopy, *Proc. Natl. Acad. Sci. USA* **102**, 17646 (2005).
 - [7] R. Biehl, B. Hoffmann, M. Monkenbusch, P. Falus, S. Preost, R. Merkel, and D. Richter, Direct observation of correlated interdomain motion in alcohol dehydrogenase, *Phys. Rev. Lett.* **101**, 138102 (2008).
 - [8] R. Inoue, R. Biehl, T. Rosenkranz, J. Fitter, M. Monkenbusch, A. Radulescu, B. Farago, and D. Richter, Large domain fluctuations on 50-ns timescale enable catalytic activity in phosphoglycerate kinase, *Biophys. J.* **99**, 2309 (2010).
 - [9] B. Farago, J. Li, G. Cornilescu, D. J. E. Callaway, and Z. Bu, Activation of nanoscale allosteric protein domain motion revealed by neutron spin echo spectroscopy, *Biophys. J.* **99**, 3473 (2010).
 - [10] A. M. Stadler, L. Stingaciu, A. Radulescu, O. Holderer, M. Monkenbusch, R. Biehl, and D. Richter, Internal nanosecond dynamics in the intrinsically disordered myelin basic protein, *J. Am. Chem. Soc.* **136**, 6987 (2014).
 - [11] H. Nakagawa, T. Saio, M. Nagao, R. Inoue, M. Sugiyama, S. Ajito, T. Tominaga, and Y. Kawakita, Conformational dynamics of a multidomain protein by neutron scattering and computational analysis, *Biophys. J.* **120**, 3341 (2021).

- [12] L. R. Stingaciu, O. Ivanova, M. Ohl, R. Biehl, and D. Richter, Fast antibody fragment motion: Flexible linkers act as entropic spring, *Sci. Rep.* **6**, 22148 (2016).
- [13] N. Smolin, R. Biehl, G. R. Kneller, D. Richter, and J. C. Smith, Functional domain motions in proteins on the similar to 1–100 ns timescale: Comparison of neutron spin-echo spectroscopy of phosphoglycerate kinase with molecular-dynamics simulation, *Biophys. J.* **102**, 1108 (2012).
- [14] R. J. Roberts, M. Belfort, T. Bestor, A. S. Bhagwat, T. A. Bickle, J. Bitinaite, R. M. Blumenthal, S. K. Degtyarev, D. T. Dryden, K. Dybvig *et al.*, A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes, *Nucleic Acids Res.* **31**, 1805 (2003).
- [15] H. Hashimoto, T. Shimizu, T. Imasaki, M. Kato, N. Shichijo, K. Kita, and M. Sato, Crystal structures of type II restriction endonuclease EcoO109I and its complex with cognate DNA, *J. Biol. Chem.* **280**, 5605 (2005).
- [16] T. Oroguchi, H. Hashimoto, T. Shimizu, M. Sato, and M. Ikeguchi, Intrinsic dynamics of restriction endonuclease *ecoO109I* studied by molecular dynamics simulations and x-ray scattering data analysis, *Biophys. J.* **96**, 2808 (2009).
- [17] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.043154> for SANS profiles at different concentrations, calculation of $D_0(Q)$ from $D_{\text{eff}}(Q)$, comparison of $D_{\text{ri}}(Q)$ with $D_r(Q) + D_i(Q)$, calculation of cooperativity score for PCA modes, correlation between the PCA modes, and the structural change upon DNA binding.
- [18] <https://www.ill.eu/en/users/support-labs-infrastructure/software-scientific-tools/grasp/>.
- [19] S. W. Provencher, CONTIN: A general purpose constrained regularisation program for inverting noisy linear algebraic and integral equations, *Comput. Phys. Commun.* **27**, 229 (1982).
- [20] B. Farago, Recent neutron spin-echo developments at the ILL (IN11 and IN15), *Phys. B (Amsterdam)* **267**, 270 (1999).
- [21] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, An overview of the AMBER biomolecular simulation package, *WIREs Comput. Mol. Sci.* **3**, 198 (2013).
- [22] L. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martínez, and V. S. Pande, Building a more predictive protein force field: A systematic and reproducible route to AMBER-FB15, *J. Phys. Chem. B* **121**, 4023 (2017).
- [23] L. P. Wang, T. J. Martinez, and V. S. Pande, Building force fields: An automatic, systematic, and reproducible approach, *J. Phys. Chem. Lett.* **5**, 1885 (2014).
- [24] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, Smooth particle mesh ewald method, *J. Chem. Phys.* **103**, 8577 (1995).
- [25] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes, *J. Comput. Phys.* **23**, 327 (1977).
- [26] T. A. Andrea, W. C. Swope, and H. C. Andersen, The role of long ranged forces in determining the structure and properties of liquid water, *J. Chem. Phys.* **79**, 4576 (1983).
- [27] A. Ortega, D. Amorós, and J. García de la Torre, Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models, *Biophys. J.* **101**, 892 (2011).
- [28] A. Kitao, F. Hirata, and N. Go, The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum, *Chem. Phys.* **158**, 447 (1991).