Reaction Pathway Sampling and Free Energy Analyses for Multimeric Protein Complex Disassembly with Employing Hybrid Configuration Bias Monte Carlo/Molecular Dynamics Simulation

Ikuo Kurisaki[*1], Shigenori Tanaka[*1]

[1]Department of Computational Science, Graduate School of System Informatics, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

[*]Ikuo Kurisaki

E-mail: kurisaki@bear.kobe-u.ac.jp, Tel: +81-78-803-6472

[*]Shigenori Tanaka

E-mail: tanaka2@kobe-u.ac.jp, Tel: +81-78-803-6620

Abstract

Physicochemical characterization of multimeric biomacromolecule assembly and disassembly processes is a milestone to understand the mechanisms for biological phenomena at molecular level. Mass spectroscopy (MS) and structural bioinformatics (SB) approaches have become feasible to identify subcomplexes involved in assembly and disassembly, while they cannot provide atomic information sufficient for free energy calculation to characterize transition mechanism between two different sets of subcomplexes. To combine observations derived from MS and SB approaches with conventional free energy calculation protocols, we here designed a new reaction pathway sampling method with employing hybrid configuration bias Monte Carlo/Molecular Dynamics (hcbMC/MD) scheme and applied it to simulate disassembly process of serum amyloid P component (SAP) pentamer. The results we obtained are consistent with those of the earlier MS and SB studies with respect to SAP subcomplex species and the initial stage of SAP disassembly processes. Furthermore, we observed a novel dissociation event, ring-opening reaction of SAP pentamer. Employing free energy calculation combined with the hcbMC/MD reaction pathway trajectories, we moreover obtained experimentally testable observations on (1) reaction time of the ring-opening reaction and (2) importance of Asp42 and Lys117 for stable formation of SAP oligomer.

Introduction

The physicochemical entity of various biological phenomena is multimeric biomolecule complexes formed transiently or stably in the cell. The molecular components of such complexes have been extensively identified and individually assigned to corresponding biological functions (*e.g.*, signal transduction[1-3], DNA repairs[4,5], protein systhesis[6,7], and RNA splicing[8,9]). Since these functions are expressed and regulated through biomolecule assembly and disassembly in response to the cellular environment[10-12], physicochemical characterization of these processes is a milestone to understand the mechanisms for biological phenomena at molecular level.

Multiple biomacromolecule assembly and disassembly processes can be physicochemically characterized by solving the two essential problems coupled with each other: (1) identifying a set of intermediate states characterized by subcomplexes appearing in these processes; (2) elucidating physicochemical mechanisms of transition between such intermediate states.

Mass spectrometry (MS) analyses have become a powerful tool to systematically analyze a possible set of subcomplexes undergoing assembly and disassembly processes at molecular level[13,14]. Meanwhile assembly and disassembly orders have been extensively studied in the field of structural bioinformatics (SB) during the last fifteen

years[15] and can now be predicted with high accuracy based on atomic structures of individual subunits.[16] Thus identification of subcomplexes involved in assembly and disassembly processes has become technically feasible, then providing important clues to characterize intermediate states consisting of assembly and disassembly processes.

As for physicochemical analyses of mechanisms for transition between such intermediate states, atomistic simulations are practically useful at present; we could promptly find a feasible approach for the problem in a class of conventional free energy profile (FEP) calculation protocols, *e.g.*, umbrella sampling combined with reaction pathway sampling techniques[17,18]. FEP such as potential of mean force (PMF) can provide quantitative insights into physicochemical mechanisms for biomacromolecule complex formation in terms of evaluated rate constants of configurational transition between two different intermediate states. Then it could seem practically feasible to examine physicochemical mechanisms for multimeric biomacromolecule assembly and disassembly processes.

Nevertheless, there still remains one technical problem upon combining the subcomplex identification methods with FEP calculation protocols. They need *a priori* knowledge of a complete set of atomic coordinates for all subunits and subcomplexes at intermediate states in assembly or disassembly processes.

In fact, it is still challenging for the above mentioned state-of-the-art MS and SB approaches to provide atomic information sufficient for FEP calculation protocols. MS approaches identify subunits and subcomplex species as particles with molecular mass and net charge, so that their spatial resolution is bound to macromolecular level. SB approaches directly predict a final docking pose and a pair of dissociated subcomplexes. It is therefore out of their methodological design to identify a configuration of a complete set of subunits, at each intermediate state involved in a pre-binding and post-unbound conditions.

To solve the technical problem in molecular simulation, we here propose a new reaction pathway sampling method. The theoretical framework we here employ is a hybrid Monte Carlo (MC)/Molecular Dynamics (MD) scheme, since configuration sampling method in the MD phase is arbitrarily selected for specific research purposes, then resulting in flexible applications of this simulation framework to a variety of molecular phenomena in condensed phases [19-22].

We can find research purpose-oriented flexibility within magnificent examples of hMC/MD simulation studies. Chemical bond formation and scission are empirically simulated by switching potential functions and using effective potential function in the frameworks of REDMOON[20] and KIMMDY[22], respectively. Meanwhile TRS[19] and

hybrid neMD-MC[21] employ randomly assigned force bias and perturbed potential energy to accelerate transformation of biopolymer conformation, respectively.

In the present method, we employed steered molecular dynamics (SMD) method to enhance non-equilibrium subprocesses composing assembly and disassembly processes, namely, intersubunit association and dissociation reactions. This method have been widely used to simulate these reactions involved in dimeric biomacromolecule complexes.[18,23,24] Aiming additional enhancement of configuration sampling, we applied the configuration bias Monte Carlo (cbMC) scheme[25-27] to hybrid MC/MD simulation procedure, recalling the earlier hybrid MC/MD study resulting in several time enhancement of lipid bilayer configuration sampling[28].

The simulation method we propose, referred to as hybrid configuration bias MC/MD (hcbMC/MD) method hereafter, was applied to simulate disassembly processes of Serum Amyloid P component (SAP) homo pentamer (Figure 1), whose subcomplex species appearing in disassembly processes have been extensively studied in the previous MS studies.[14,29,30] We evaluated the performance of our method through comparison with the MS- and SB-derived observations for SAP subcomplex species and earlier stages of SAP pentamer disassembly process.

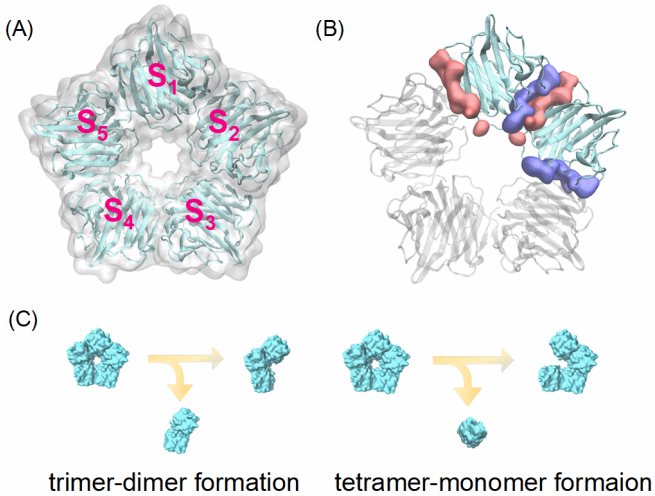


Figure 1. Molecular structures of serum amyloid P component (SAP) homo pentamer. (A) SAP pentamer with annotation for each subunit. (B) Two individual intersubunit interaction surfaces of SAP subunit, highlighted by blue and red colors; S_1 and S_2 are depicted as representative of subunit pair. (C) Initial stages of SAP pentamer disassembly processes predicted by structural bioinformatics approach[14]. Reactions shown in panel C are supposed for any possible monomer and dimer.

Materials and Methods

Setup of serum amyloid P component system

We used atomic coordinates of serum amyloid P component (SAP) homo-pentamer (PDB entry: 4AVS)[31]. Two calcium ions bound to SAP subunits were retained in the atomic coordinates of this SAP system. Nε protonation state was employed for each of

histidine residues, and all carboxyl groups in aspartate and glutamate residues were set to the deprotonated state, where we considered physiological pH condition in the cell. The SAP structure was solvated in the rectangular box with 49305 water molecules and was electrically neutralized by adding 10 Cl^- ion molecules.

To calculate the forces acting among atoms, AMBER force field 14SB[32], TIP3P water model[33,34], and JC ion parameters adjusted for the TIP3P water model[35,36] were employed for amino acid residues, water molecules, and ions, respectively. In addition, the force field parameter sets developed by Bradbrook was applied for divalent calcium ion[37]. Molecular modeling of the SAP system was performed using the LEaP modules in AmberTools 17 package[38]. Molecular mechanics (MM) and molecular dynamics (MD) simulations were performed under the periodic boundary condition with GPU-version PMEMD module in AMBER 17 package[38] based on SPFP algorithm[39] with NVIDIA GeForce GTX1080 Ti. The solvated structure was relaxed through 10-ns NPT MD simulations, then being employed as the initial atomic coordinates for the following hybrid configuration bias MC/MD simulations. Further computational details are described in **S-1** and **S-2** in Supporting Information.

Design concept of hcbMC/MD-based path sampling

Our path sampling simulation method is designed while considering the three essential factors which possibly have prevented MD simulations from their application to the assembly and disassembly processes associated with multiple biomacromolecules: (1) physicochemical events rarely occurring within available computational time; (2) technical complexity to prespecify a sequence of intermediate states, whose possible candidates increase with factorial of the number of intersubunit interface; (3) no *a priori* knowledge for a full set of atomic coordinates of macrobiomolecules at each intermediate state in assembly and disassembly processes.

We cope with the first factor by using two theoretical techniques: (1) steered MD (SMD) method to enhance diffusion of subunit molecules in each MD phase; (2) a configuration bias Monte Carlo scheme in each MC phase (*vide infra*).

Assembly and disassembly processes are sequences of intersubunit association and dissociation reactions, respectively. Even such an elementary subprocess generally takes several micro seconds or longer[40-42], thus rarely occurring within brute-force MD simulations available of today. Enhanced sampling such as SMD method is indispensable to simulate such rare events within realistic computation time[18].

Besides aiming to further accelerate development of molecular systems toward disassembly, we employed biased configuration selection[25-28] for a set of generated

configurations by using a purpose-oriented biased potential energy function, as shown below.

Meanwhile a configuration generated by enhanced sampling is rejected with certain probability by Metropolis algorithm in MC phase. Acceptance judgement of generated configurations via Metropolis algorithm is implemented to circumvent selection of subcomplex configurations anomalously deformed by applying external forces.

The second factor is treated by using a set of the predefined reaction coordinates, where we recall successful usage of predefined set of chemical reactions in REDMOON method [43]. At every hcbMC/MD simulation cycle, a reaction coordinate for intersubunit association and dissociation is randomly chosen from a set of the predefined reaction coordinates and the system is propagated in MD phases (*see* Materials and Methods section for predefined reaction coordinates specific for the SAP system below). Random choice of reaction coordinates is implemented to circumvent technical difficulty to prespecify a complete sequence of intersubunit association and dissociation reactions. As for the case of SAP system, we can simulate the disassembly process without selecting probable sequences of dissociation reactions among the 5! (= 120, derived five intersubunit interface of SAP pentamer) candidates.

Employing a predefined set of reaction coordinates in enhanced sampling simulations

coincidently provides a solution for the third factor; it enables us automatically search reaction pathways without *a priori* knowledge of a complete set of atomic coordinates of all subunits and subcomplexes in molecular system at each intermediate state in the assembly and disassembly processes. In application to disassembly processes, our pathway sampling method needs only initial atomic coordinates of molecular system, which can be obtained from structural biological techniques (*e.g.* X-ray crystallography, electron microscopy and so on). This point differs from the conventional ones, such as Nudged Elastic Bond[44] and targeted MD[45] methods, which at least require a set of atomic coordinates of molecular configurations with respect to both the initial and final states.

Hybrid configuration bias MC/MD simulation scheme

The simulation scheme is designed based on a general configuration bias MC method[25] as follows. We here employed unbiased and SMD simulations to generate old and new sets of atomic configurations of molecular systems, respectively. Intersubunit dissociation reactions are mainly accelerated with SMD simulations. Detailed simulation conditions of MD and SMD-based configuration generation are given in the next subsection.

1. Generate M trial configurations $\{\mathbf{a}_1, \mathbf{a}_2, \cdots \mathbf{a}_M\}$ by an unbiased MD simulation and calculate bias energy u^{bias}, which will be specified later, for each configuration.

2. Assign the last snapshot obtained from the simulation ($\mathbf{a}_M$) as an old configuration, denoted by $\mathbf{x}_o$ and define the Rosenbluth factor[25]

$$W(o) = \exp\left(-\beta u^{bias}\left(\mathbf{x}_o\right)\right) + \sum_{i=1}^{M-1} \exp\left(-\beta u^{bias}\left(\mathbf{a}_i\right)\right) \quad (3)$$

β denotes inverse of k_BT, where k_B and T are Boltzmann constant and system temperature, respectively.

3. Generate M trial configurations $\left\{\mathbf{b}_1, \mathbf{b}_2, \cdots \mathbf{b}_M\right\}$ by a steered MD simulation starting from the old configuration $\mathbf{x}_o$ and calculate bias energy u^{bias} for each configuration.

4. Define the Rosenbluth factor

$$W(n) = \sum_{i=1}^{M} \exp\left(-\beta u^{bias}\left(\mathbf{b}_i\right)\right) \quad (4)$$

and select one configuration, denoted by $\mathbf{x}_n$, with a probability

$$p(\mathbf{x}_n) = \frac{\exp\left(-\beta u^{bias}\left(\mathbf{x}_n\right)\right)}{\sum_{i=1}^{M} \exp\left(-\beta u^{bias}\left(\mathbf{b}_i\right)\right)} \quad (5)$$

5. The configuration change is accepted with a probability

$$acc\left(o \to n\right) = \min\left(1, \frac{W\left(n\right)}{W\left(o\right)} \exp\left\{-\beta\left[U'\left(\mathbf{x}_n\right) - U'\left(\mathbf{x}_o\right)\right]\right\}\right) \quad (6)$$

$$U'\left(\mathbf{x}\right) = U\left(\mathbf{x}\right) - u^{bias}\left(\mathbf{x}\right) \quad (7)$$

where $\mathbf{x}$ and $U\left(\mathbf{x}\right)$ denote a configuration and the potential energy function of the system with configuration of $\mathbf{x}$, respectively.

Considering that configurations of the system are generated by moving all atoms simultaneously with molecular dynamics method, we used the Rosenbluth factor in the forms given in Eqs. (3) and (4). Monte Carlo trial with this type of configuration sampling is particularly referred to as orientation bias Monte Carlo[25].

As for the right-hand side of Eq (4), we assume that probability of generating forward move ($\mathbf{x}_o \rightarrow \mathbf{x}_n$) is equal to that of generating backward move ($\mathbf{x}_o \leftarrow \mathbf{x}_n$). This assumption is employed here for simplification of simulation implementation but not trivial in general. Then with the aim of validating physicochemical observations obtained from an hcbMC/MD trajectory, it is a practical option to perform additional umbrella sampling MD simulations with a set of snapshot structures involved in an intersubunit dissociation event during the trajectory.

A multimeric protein complex disassembly process should be accompanied by gradual breakage of intersubunit native contacts (NC). Then we defined the bias energy as a function of the total number of intersubunit NC formed in a multimeric protein complex (n_{NC}), where the initial atomic coordinates are used as the reference structure:

$$u^{bias}(\mathbf{x}) = k \cdot n_{NC}(\mathbf{x}) \quad (8)$$

The coefficient k is set to β^{-1} at 300 K, that is 0.6 [kcal/mol] by assuming energetic stabilization obtained from hydrophobic atomic contacts[46]. A set of native contacts at

subunit binding interface was defined by using the initial atomic coordinates of SAP pentamer for hcbMC/MD simulations.

Configuration generation at each hcbMC/MD cycle by using unbiased and steered MD simulations

Configurations of SAP system, denoted by $\{\mathbf{a}_1, \mathbf{a}_2, \cdots \mathbf{a}_M\}$, were sampled by using an unbiased NPT MD simulation (300 K, 1 bar). Meanwhile, configurations of the system undergoing a partial intersubunit dissociation, denoted by $\{\mathbf{b}_1, \mathbf{b}_2, \cdots \mathbf{b}_M\}$, were sampled by using a steered molecular dynamics (SMD) simulation under the NPT condition (300 K, 1 bar), starting from the snapshot structure obtained from the unbiased MD simulation. The simulation length is 100 ps for each of MD and SMD simulations. A set of atomic coordinates was recorded by every 10 ps. The system temperature and pressure were regulated by Langevin thermostat with 1-ps^{-1} collision coefficient, and Monte Carlo barostat with attempt of system volume change by every 100 steps, respectively.

In every hcbMC/MD cycle, we randomly chose a pair of monomers in the 10 (= ${}_5C_2$, derived from combination of two subunits among the five) candidates. We employed the distance between centers of gravity of the chosen subunit pair as the reaction coordinate of SMD simulation. A center of gravity for a SAP subunit was calculated by using all of

the 204 C_α atoms.

In an SMD simulation, a target value of the distance was set to $d_0 + \Delta d$, where d_0 and Δd are an initial value of the distance at the cycle and a random integer in the range of 8 to 12, respectively. The harmonic potential with the force constant of 10 kcal/mol/Å^2 was imposed on the reaction coordinate. In each cycle, a reaction coordinate for inter-SAP subunit dissociation reaction and a random seed for Langevin thermostat were changed randomly. Following the above simulation conditions, an unbiased MD simulation was also designed to evaluate the sampling performance of hcbMC/MD simulation, where the 100-ps MD simulation cycle was repeated 2000 times. Additional computational details and a corresponding unbiased MD simulation procedure are described in **S-1** and **S-3** in Supporting Information, respectively.

Umbrella sampling molecular dynamics simulations

Certain numbers of snapshot structures were extracted from a hcbMC/MD trajectory and employed for USMD windows. Following temperature relaxation simulations, 5-ns NVT USMD simulations (300 K) were performed for each of the USMD windows (*see* Table S1 in Supporting Information for details). The system temperature was regulated using Langevin thermostat with 1-ps^{-1} collision coefficient. Each of the last 3-ns USMD

trajectories was used to construct a PMF (*see* Figure S1 for convergence of PMF curves). A reaction coordinate is the distance between centers of gravity of the subunit pair undergoing dissociation reaction. A set of USMD simulations was repeated by 8 times. The further technical details are given in Supporting Information (*see* **SI 4**).

Trajectory analyses

Root mean square deviation (RMSd), and native and non-native atomic contacts were analyzed with the cpptraj module in AmberTools 17 package[38]. We calculated RMSd and native contacts by using the C_α atoms and the non-hydrogen atoms in the initial atomic coordinates for the hcbMC/MD simulations, respectively. The distance criterion for atomic contacts was set to 3.5 Å. Intersubunit dissociation and reassociation conditions are defined as follows: $N_{native} = 0$ and $N_{non\text{-}native} = 0$; $N_{non\text{-}native} \geq 0$ in $N_{native} > 0$, respectively. N_{native} and $N_{non\text{-}native}$ denote the number of native atomic contacts and that of non-native atomic contacts made between a pair of SAP subunit, respectively. Collision cross section (CCS) of SAP protein was analyzed by using IMPACT program[47].

PMF was calculated with Weighed Histogram Analysis Method (WHAM)[48,49] by using each set of USMD trajectories. Statistical errors of PMF values, $\sigma_{PMF}(\xi)$, were estimated by employing bootstrapped sampling[50]:

$$\sigma_{PMF}(\xi)=\left[(N_b-1)^{-1}\sum_{k=1}^{200}\left(W_{b,k}(\xi)-\langle W_b(\xi)\rangle\right)^2\right]^{\frac{1}{2}} \quad (1)$$

Here, N_b, ξ, and $W_{b,k}(\xi)$ denote the number of bootstrapped sampling, the reaction coordinate and the value of kth bootstrapped PMF at each point of ξ, respectively. $\langle W_b(\xi)\rangle$ is the average over all $W_{b,k}(\xi)$, where k ranges from 1 to 200.

Reaction rate, k_{TST}, is estimated by using Eyring's transition state theory:

$$k_{TST}=\frac{k_B T}{h}\exp\left(\frac{-\Delta F^{\dagger}}{k_B T}\right) \quad (2)$$

where k_B and T denote an activation barrier height, Planck constant, Boltzmann constant and a temperature of system, respectively. Reaction time scale, τ_{TST}, is defined as the inverse of k_{TST}. $\Delta F^{\dagger}$ is defined as $F(\xi_0')-F(\xi_0)$, where the value of PMF has a local minimum at ξ_0, and a maximum at ξ_0', which is greater than ξ_0.

Molecular structures were illustrated by using Visual Molecular Dynamics (VMD)[51]. Error bars were calculated from standard error that indicate 95% confidence interval.

Results and Discussion

Subcomplex species and disassembly pathway obtained from the hcbMC/MD simulations

We performed five independent hcbMC/MD simulations for the SAP system. A

hcbMC/MD cycle is repeated 500 times, so that total simulation time length in the MD phase is 100 ns. The hcbMC/MD simulations show total MC trial acceptance ratio of c.a. 40% in average over the five simulations. The sampling efficiency of our method is similar to that of TRS method[19] and seems sufficient to promote disassembly processes.

A ring shape SAP pentamer is converted into a set of subcomplexes and subunits by the 500th cycle of hcbMC/MD simulations (Figure 2). We find a subunit pair with non-specific contacts, which were formed irrespective of intersubunit interaction surfaces of SAP (*see* Figure 1B for comparison). Such non-specific intersubunit reassociation results from a periodic simulation box whose size is insufficient for apparent spatial separation among five SAP subunit. Then we identified subcomplexes by excluding such a reassociated subunit pair that solely makes non-specific intersubunit contacts.

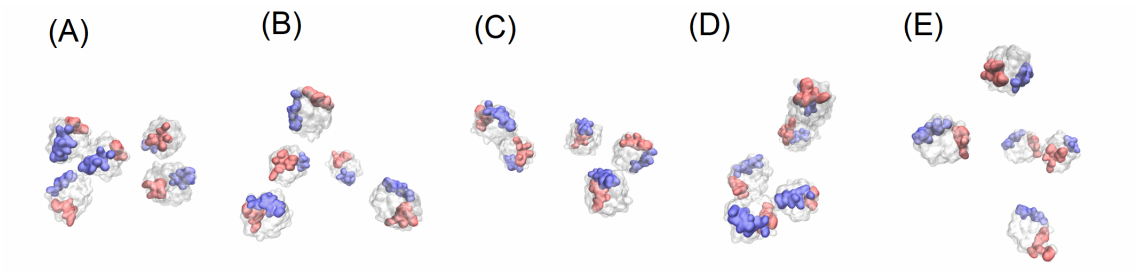


Figure 2. Configurations of SAP subunits at 500th hcbMC/MD cycle. Two individual intersubunit interaction surfaces of each subunit are highlighted by blue and red colors as in the case of Figure 1. Alphabetic annotations for panels, (A)-(E), denote those for

individual simulations hereafter.

Figure 3 shows subcomplex species at each hcbMC/MD cycle for the homo-pentameric SAP system. In our hcbMC/MD simulations, any of subcomplex species (tetramer, trimer and dimer) observed in MS experiments[14] appears. Besides, each of the five simulations results in decomposition into the five isolated SAP subunits. Meanwhile the initial stage of SAP disassembly is classified into two distinct events: trimer-dimer formation and tetramer-monomer formation from the pentamer (Figure 3A, D and E, and Figure 3B and C, respectively; *see* Figure 1C). In Figure 3A, the trimer-dimer pair found at 29th cycle returns to pentamer form, so that we suppose tetramer-monomer formation as the initial subcomplex disassembly. These two different disassembly events were similarly predicted by the structural bioinformatics approach in the earlier MS study.[14]

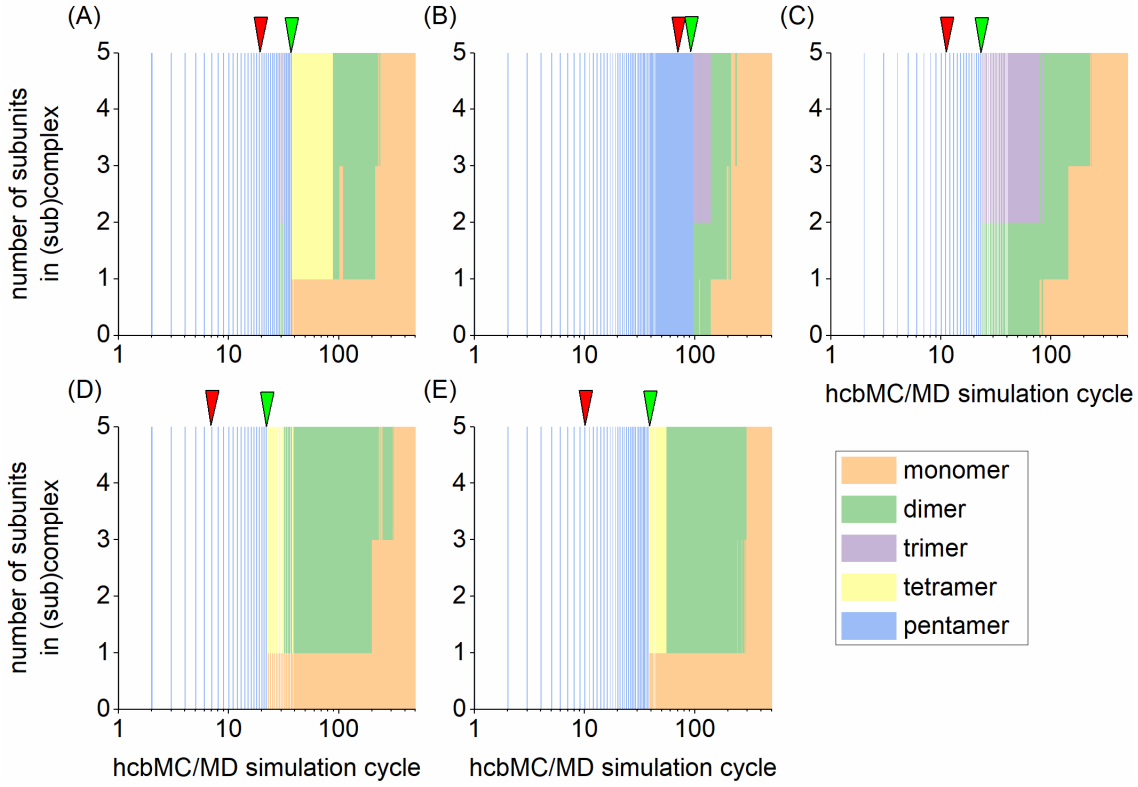


Figure 3. Disassembly of serum amyloid P component (SAP) homo-pentamer through hcbMC/MD cycles. Each panel is for an individual simulation. A distribution of SAP (sub)complex species is illustrated by color. Green and red arrows on panel indicate the initial stage of SAP disassembly (generation of tetramer or trimer) and the first cycle where SAP pentamer undergoes a ring-opening event.

It is thus possible to say that the simulation results are consistent with those of the earlier study[29] with respect to SAP subcomplex species and the initial stage of SAP disassembly processes. Our reaction pathway sampling method could play a role complementary to MS and SB approaches in arrangement of complete sets of atomic

coordinates of all subunits involved in a disassembly process.

Computation performance of hybrid configuration bias MC/MD simulations

The disassembly of pentameric form of SAP can be efficiently simulated due to methodological advantage of our hcbMC/MD approach. We can show this point by using independent five unbiased NPT MD simulations (300 K, 1 bar) with simulation length of 200 ns in total, twice longer than those of hcbMC/MD simulations. As shown in Figure 4, the five SAP molecules retain each intersubunit contacts during the 200-ns simulations, thus keeping to take a pentameric form in each of the unbiased MD simulations.

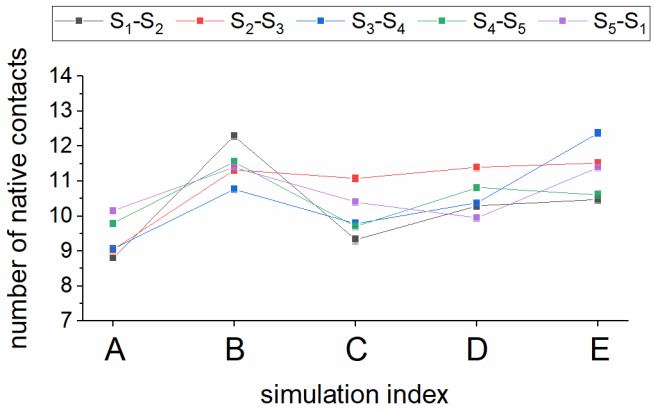


Figure 4. Averaged number of native contacts between SAP subunit pair. Average and statistical error values are calculated using 2000 snapshot structures obtained from each of the five unbiased 200-ns MD simulations (an error bar is buried in the square, then not being apparent in this panel). Each intersubunit native contacts are distinguished by

different color. Annotation of subunit pair is given as in the case of Figure 1. Statistical errors were estimated from standard error by considering 95 % confidence interval.

Furthermore, each of the subunits retains folded configurations though hcbMC/MD cycles. Although we accelerated dissociation reactions using external forces, the values of RMSd for SAP subunits are similar between hcbMC/MD simulations and unbiased 200-ns NPT MD simulations (Table 1). This observation clearly shows that external forces applied to subunits in SMD simulations hardly affect their structures, thus additionally supporting applicability of our hcbMC/MD method to analyze physicochemical properties of proteins under physiological conditions.

Table 1. Root mean square deviation (RMSd)[a] with error value[b] for SAP subunit[c].

	S_1	S_2	S_3	S_4	S_5
hcbMC/MD	0.9 ± 0.2	0.9 ± 0.1	0.8 ± 0.1	0.8 ± 0.1	0.8 ± 0.1
unbiased MD	0.9 ± 0.2	0.8 ± 0.2	0.8 ± 0.2	0.8 ± 0.2	0.8 ± 0.3

[a]RMSd for each SAP subunit was averaged over each 500 hcbMC/MD cycles (100 ns in total) or 2000 unbiased MD cycle (200 ns in total), and averaged over the 5 simulations.

[b]Error values denote CI 95, derived from standard deviation over the 5 simulations.

[c]SAP subunit is distinguished by combination of alphabetic letter and digit as in the case

of Figure 1.

Free energy profile of ring-opening reaction of SAP pentamer

Observing the hcbMC/MD trajectories, we can find a preliminary dissociation event in each of the 5 simulations. One of the five subunit interaction interfaces is broken, then resulting in ring-opened pentameric form of SAP (Figure 5). A ring opening of the SAP pentamer occurs in a relatively early stage of hcbMC/MD cycles (Table 2) and is followd by the initial steps of SAP pentamer disassembly process, that is, tetramer-monomer and trimer-dimer dissociation reactions (*see* Figure 3).

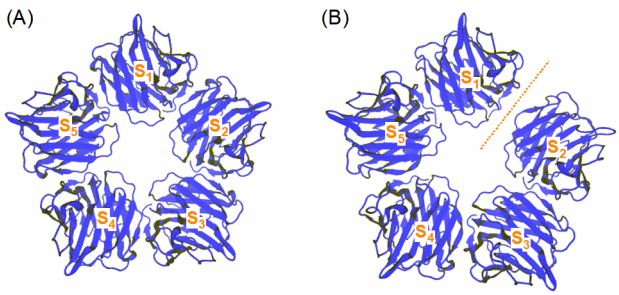


Figure 5. Ring-opened SAP pentamer configuration. (A) X-ray crystallographic structure (PDB entry: 4AVS). (B) Snapshot structure obtained from hcbMC/MD simulation, whose simulation index is A in Table 2. Orange dotted line is depicted in the vicinity of broken subunit interaction interface. SAP subunit is distinguished by combination of alphabetic letter and digit as in the case of Figure 1.

Table 2. Geometrical characterization of SAP pentamer at the cycles where the initial ring-opening reaction occurs.

simulation index	pair of SAP subunit[a]	cycle	CCS[b] [nm^2]
A	S_3-S_4	20th	59.9
B	S_1-S_2	61st	59.8
C	S_3-S_4	11th	59.1
D	S_2-S_3	7th	59.3
E	S_3-S_4	10th	59.8

[a]SAP subunit is distinguished by combination of alphabetic letter and digit as in the case of Figure 1.

[b]CCS denotes collision cross section.

This ring-opening reaction has not been reported in the earlier mass spectroscopy studies[14,29,30], although there are rooms to discuss occurence of the intermediate states under realistic experimental conditions. It is possible that MS observations have not detected such intermediate states due to their limitation for spatial resolution. Values of collision cross section (CCS) for these ring-opened SAP pentamer structures are around 59 nm^2, being larger by c.a. 3 nm^2 than that calculated from the X-ray crystallographically resolved ring-forming pentamer[31]. A ring-opened SAP pentamer appearing in

experimental conditions might be assigned into partially unfolded SAP pentamer in the context of mass spectroscopy.[29]

We then consider possibility of experimental observation of this ring-opening reaction. High-speed timelapse atomic force microscopy (AFM) is now available to observe molecular dynamics of multimeric protein complexes with nanometer spatial resolution and hundreds of milliseconds temporal resolution.[52] This method could be employed to validate the presence of this reaction, if the reaction time scale is found within observation period of AFM. Aiming to estimate the reaction time, we carried out free energy profile calculations with a conventional protocol, umbrella sampling molecular dynamics (USMD) simulations combined with Weighted Histogram Analyses Method (WHAM) [48,49].

We here examined the initial ring-opening events appearing in the hcbMC/MD simulations due to the following observation for the trajectories we obtained. Ring-opened SAP pentamer does not necessarily lead to immediate progress to the initial dissociation steps but may transiently come back to ring-forming intact SAP pentamer (*see* Figure 3A and Figure S2 in Supporting Information). A SAP structure fluctuating between the two pentameric forms might gradually lose inter-SAP subunit contacts; such partially broken contacts would give lower activation barrier of the ring-opening reaction

so that additional free energy calculations are probably needed to estimate a precedent free energy loss. We then arranged a set of initial atomic coordinates for USMD simulations by using snapshot structures appearing at the cycle of initial ring-opening reaction and also those around the neighboring ones if needed. Each of these snapshot structures takes either ring-formed pentamer or ring-opened pentamer. A reaction coordinate is set to the distance between centers of gravity of subunits undergoing breakage of intersubunit contacts.

Each of PMFs has one activation barrier and appears to be up-hill. A steep change of the slope of PMF is found between 37 and 40 Å on the reaction coordinates. The highest activation barrier is 23.7 kcal/mol (Figure 6A) and the corresponding reaction time is 8.2 hour according to eq. (2). This estimated reaction time may fall within observation periods of timelapse AFM, hundreds of milliseconds or longer, so that occurrence of this ring-opening reaction could be validated with employing AFM experiments.

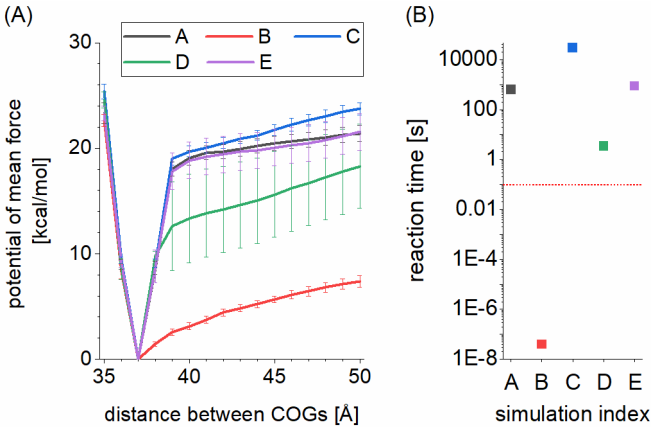


Figure 6. Potential of mean force of SAP pentamer ring-opening reaction (A) and reaction time evaluated with Eyring transition state theory (B). Each simulation is indexed by alphabetic character, that is, (A)-(E). An error bar was calculated from standard deviation and denotes CI95. In panel B, points above the red dotted boaderline have the estimated reaction time falling within observation period with AFM experiments.

Meanwhile, large deviations among the values of activation barrier might denote variations of ring-opening reaction pathways and also following disassembly processes perhaps. Recalling observation period of timelapse AFM experiments, we then classify the five ring-opening events into two groups; that for the simulation annotated by B (sim-B) and those for the other four. The reaction time obtained from sim-B is 39 ns, significantly smaller than those obtained from the other four simulations (Figure 6B). The reaction time of 39 ns means that a ring-opening reaction can spontaneously proceed under thermal fluctuation along this reaction pathway.

The significant decrease of activation barrier could be explained by difference in the number of atomic contacts between a subunit pair which retains contacts of intersubunit interaction surfaces. Figure 7 shows the number of atomic contacts between each subunit pair for the five hcbMC/MD simulations, where the snapshot structures are obtained from the simulation cycle at which we observed the initial ring-opening reaction. With regard to intersubunit atomic contacts retained in ring-opened SAP pentamer, we can find smaller number of atomic contacts in S_5-S_1 interface in sim-B, that is, 8. Meanwhile the remaining three are 15 or larger. It is of note that S_5 is neighboring to S_1, which undergoes ring-opening reaction in sim-B (Table 2). According to such a preliminary deformation at the vicinal interface, the reaction pathway obtained from the sim-B and those obtained from the other four might be separately located in the conformational space of this SAP system.

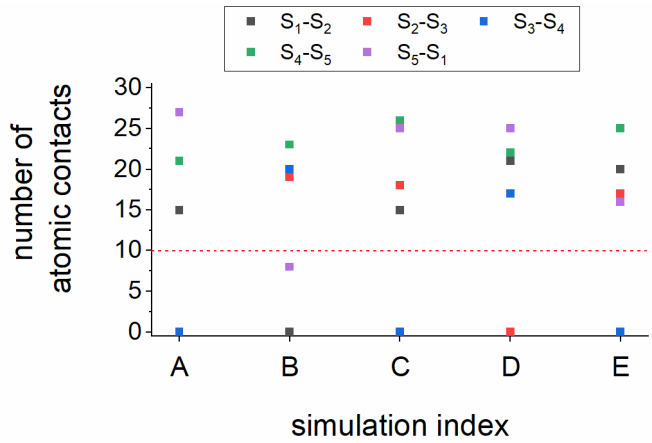


Figure 7. Atomic contacts between subunit pair for snapshot structure at hcbMC/MD cycle for the initial ring-opening reaction. SAP subunit is distinguished by combination of alphabetic letter and digit as in the case of Figure 1.

From the viewpoint of reaction energetics, stepwise breakage of subunit interface seems to more easily occur rather than simultaneous break of multiple subunit interfaces. Recalling the earlier simulation study on stepwise dissociation reaction of protein dimer [53], we can assume that this ring-opening reaction proceeds in a similar manner; allosteric losing of intersubunit contacts is followed by dissociation between a subunit pair.

We suppose that breakage of salt bridge formed between Lys117 and Asp42 in S_5-S_1 interface (Figure 8A, B and C) is particularly important to weaken configurational restraints between S_1 and S_5 in the ring-opened SAP pentamer in sim-B. This salt bridge breakage seems to allow libration motion of S_1 on S_5 and enhances dissociation of S_1 from S_2. The effects of Lys117 and Asp42 on SAP pentamer formation have not been reported to our best knowledge. While considering significant decrease of the activation barrier for sim-B, mutation on either or both of these two charged residues, or strongly acidic and basic conditions, might repress SAP oligomerization. SAP was identified as pathological deposite in 1967 and has been widely studied in the field of amyloid- and

immune-related diseases.[54,55] These two residues might make a theraphetic target in aim of regulation of SAP pemtamer formation.

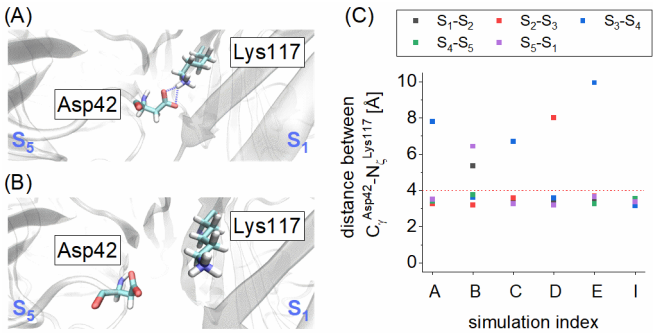


Figure 8. Positional relation between Asp42 and Lys117 at subunit interface. (A) The initial structure of hcbMC/MD simulations. (B) Structure at the timing of initial ring-opening in the simulation indexed with B. (C) Distances between Cγ in Asp42 and Nζ in Lys117 for each interface, where simulation index I denotes the initial structure of hcbMC/MD simulations. SAP subunit is distinguished by combination of alphabetic letter and digit as in the case of Figure 1. In Panel A, blue dotted lines denote hydrogen bonding between Lys117 and Asp42. In Panel C, red dotted line shows level of 4 Å .

According to the above discussion, we can speculate that preliminary deformation at vicinal SAP subunit interface has influence on activation barrier of the ring-opening reaction. We here have discussed the reaction mechanism in terms of one reaction coordinate, distance between centers of gravity of a subunit pair involved in interface

breakage. Besides, we could refine physicochemical characterization of the reaction mechanism by employning an additional reaction coordinate, *e.g.*, the number of allositeric intersubunit contacts at the vicinal interface. Time scale of preliminary deformation could also be elucidated quantitatively. Nontheless, further efforts to address these problems are beyond research scope of this study, thus being left for future studies.

Concluding Remarks

We have here proposed a new reaction pathway sampling method, hcbMC/MD, aiming to simulate complicated assembly and disassembly processes involving multimeric biomacromolecules within realistic computation time.

Our method was applied to disassembly simulations of Serum Amyloid P component (SAP) homo pentamer. Subcomplex species and disassembly pathway we obtained from hcbMC/MD simulations are consistent with the earlier observations. Our simulation approach could play a role complementary to mass spectrometry and structural bioinformatics approaches with regard to arrangement of complete sets of atomic coordinates of a molecular system undergoing disassembly processes.

Regarding observations obtained from mass spectroscopy and structural bioinformatics approaches as molecular and structural landmarks, our atomistic simulation approaches

could provide deeper insights into physicochemical mechanism for biomacromolecular assembly and disassembly processes. Actually we observed a novel dissociation event, ring-opening reaction of SAP pentamer. Employing free energy calculation combined with our hcbMC/MD trajectories, we moreover obtained experimentally testable observations on (1) reaction time of the ring-opening reaction and (2) importance of Asp42 and Lys117 for stable formation of SAP oligomer.

We then suppose that our hcbMC/MD approach opens a new avenue to study physicochemical mechanisms of complicated biological processes associated with multimeric biomacromolecules in the cell.

Supporting Information

See supporting information for detailed procedures for unbiased MD, SMD and USMD simulations.

Acknowledgements

This work was supported by a Grant-Aid for Scientific Research on Innovative Areas "Chemistry for Multimolecular Crowding Biosystems" (JSPS KAKENHI Grand No. JP17H06351).

References

1. Laplante, M.; Sabatini, D. M., mTOR Signaling in Growth Control and Disease. *Cell* **2012,** *149*, 274-293.

2. Saxton, R. A.; Sabatini, D. M., mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **2017,** *168*, 960-976.

3. Shen, K.; Rogala, K. B.; Chou, H. T.; Huang, R. K.; Yu, Z. H.; Sabatini, D. M., Cryo-EM Structure of the Human FLCN-FNIP2-Rag-Ragulator Complex. *Cell* **2019,** *179*, 1319-1329.

4. Lans, H.; Hoeijmakers, J. H. J.; Vermeulen, W.; Marteijn, J. A., The DNA damage response to transcription stress. *Nat. Rev. Mol. Cell Bio.* **2019,** *20*, 766-784.

5. Sancar, A.; Lindsey-Boltz, L. A.; Unsal-Kacmaz, K.; Linn, S., Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **2004,** *73*, 39-85.

6. Bohnsack, K. E.; Bohnsack, M. T., Uncovering the assembly pathway of human ribosomes and its emerging links to disease. *Embo. J.* **2019,** *38*, e100278.

7. Kaczanowska, M.; Ryden-Aulin, M., Ribosome biogenesis and the translation process in Eschetichia coli. *Microbiol. Mol. Biol. R.* **2007,** *71*, 477-494.

8. Shi, Y. G., Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell. Bio.* **2017,** *18*, 655-670.

9. Fica, S. M.; Nagai, K., Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* **2017,** *24*, 791-799.

10. Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A., ATP as a biological hydrotrope. *Science* **2017,** *356*, 753-756.

11. Sumino, A.; Yamamoto, D.; Iwamoto, M.; Dewa, T.; Oiki, S., Gating-Associated Clustering-Dispersion Dynamics of the KcsA Potassium Channel in a Lipid Membrane. *J. Phys. Chem. Lett.* **2014,** *5*, 578-584.

12. Nakajima, K.; Yamazaki, T.; Kimura, Y.; So, M.; Goto, Y.; Ogi, H., Time-Resolved Observation of Evolution of Amyloid-β Oligomer with Temporary Salt Crystals. *J. Phys. Chem. Lett.* **2020,** *11*, 6176-6184.

13. Heck, A. J. R., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **2008,** *5*, 927-933.

14. Hall, Z.; Hernandez, H.; Marsh, J. A.; Teichmann, S. A.; Robinson, C. V., The Role of Salt Bridges, Charge Density, and Subunit Flexibility in Determining Disassembly Routes of Protein Complexes. *Structure* **2013,** *21*, 1325-1337.

15. Soni, N.; Madhusudhan, M. S., Computational modeling of protein assemblies. *Curr. Opin. Struc. Biol.* **2017,** *44*, 179-189.

16. Peterson, L. X.; Togawa, Y.; Esquivel-Rodriguez, J.; Terashi, G.; Christoffer, C.; Roy, A.; Shin, W. H.; Kihara, D., Modeling the assembly order of multimeric heteroprotein complexes. *Plos Comput. Biol.* **2018,** *14*, e1005937.

17. Matsunaga, Y.; Fujisaki, H.; Terada, T.; Furuta, T.; Moritsugu, K.; Kidera, A., Minimum Free Energy Path of Ligand-Induced Transition in Adenylate Kinase. *Plos Comput. Biol.* **2012,** *8*, e1002555.

18. Yamashita, T.; Fujitani, H., On accurate calculation of the potential of mean force between antigen. and antibody: A case of the HyHEL-10-hen egg white lysozyme system. *Chem. Phys. Lett.* **2014,** *609*, 50-53.

19. Kitao, A., Transform and relax sampling for highly anisotropic systems: Application to protein domain motion and folding. *J. Chem. Phys.* **2011,** *135*, 045101.

20. Nagaoka, M.; Suzuki, Y.; Okamoto, T.; Takenaka, N., A hybrid MC/MD reaction method with rare event-driving mechanism: Atomistic realization of 2-chlorobutane racemization process in DMF solution. *Chem. Phys. Lett.* **2013,** *583*, 80-86.

21. Suh, D.; Radak, B. K.; Chipot, C.; Roux, B., Enhanced configurational sampling with hybrid non-equilibrium molecular dynamics-Monte Carlo propagator. *J. Chem. Phys.*

2018, *148*, 014101.

22. Rennekamp, B.; Kutzki, F.; Obarska-Kosinska, A.; Zapp, C.; Grater, F., Hybrid Kinetic Monte Carlo/Molecular Dynamics Simulations of Bond Scissions in Proteins. *J. Chem. Theory. Comput.* **2020,** *16*, 553-563.

23. Kurisaki, I.; Takayanagi, M.; Nagaoka, M., Bound Na^+ is a Negative Effecter for Thrombin-Substrate Stereospecific Complex Formation. *J. Phys. Chem. B* **2016,** *120*, 4540-4547.

24. Kurisaki, I.; Tanaka, S., ATP Converts $A\beta_{42}$ Oligomer into Off-Pathway Species by Making Contact with Its Backbone Atoms Using Hydrophobic Adenosine. *J. Phys. Chem. B* **2019,** *123*, 9922-9933.

25. Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applicaltions*, Computational Science Series. Academic Press: 2002.

26. Smit, B.; Karaborni, S.; Siepmann, J. I., Computer-Simulations of Vapor-Liquid Phase-Equilibria of N-Alkanes. *J. Chem. Phys.* **1995,** *102*, 2126-2140.

27. Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I.; Krishna, R., Improving the efficiency of the configurational-bias Monte Carlo algorithm. *Mol. Phys.* **1998,** *94*, 727-733.

28. Chiu, S. W.; Clark, M. M.; Jakobsson, E.; Subramaniam, S.; Scott, H. L.,

Application of combined Monte Carlo and molecular dynamics method to simulation of dipalmitoyl phosphatidylcholine lipid bilayer. *J. Comput. Chem.* **1999,** *20*, 1153-1164.

29. Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V., Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* **2012,** *134*, 3429-3438.

30. Aquilina, J. A.; Robinson, C. V., Investigating interactions of the pentraxins serum amyloid P component and C-reactive protein by mass spectrometry. *Biochem. J.* **2003,** *375*, 323-328.

31. Kolstoe, S. E.; Jenvey, M. C.; Purvis, A.; Light, M. E.; Thompson, D.; Hughes, P.; Pepys, M. B.; Wood, S. P., Interaction of serum amyloid P component with hexanoyl bis(D-proline) (CPHPC). *Acta Crystallogr. D* **2014,** *70*, 2232-2240.

32. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.

33. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79*, 926-935.

34. Kusalik, P. G.; Svishchev, I. M., The Spatial Structure in Liquid Water. *Science*

1994, *265* (5176), 1219-1221.

35. Joung, I. S.; Cheatham, T. E., III, Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009,** *113*, 13279-13290.

36. Joung, I. S.; Cheatham, T. E., Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008,** *112*, 9020-9041.

37. Bradbrook, G. M.; Gleichmann, T.; Harrop, S. J.; Habash, J.; Raftery, J.; Kalb, J.; Yariv, J.; Hillier, I. H.; Helliwell, J. R., X-ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes - Relating structure to thermodynamics of binding. *J. Chem. Soc. Faraday T.* **1998,** *94*, 1603-1611.

38. Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III ; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; *et al. Amber 17*, University of California, San Francisco, 2017.

39. Le Grand, S.; Gotz, A. W.; Walker, R. C., SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013,** *184*, 374-380.

40. Tran, D. P.; Kitao, A., Dissociation Process of a MDM2/p53 Complex

Investigated by Parallel Cascade Selection Molecular Dynamics and the Markov State Model. *J. Phys. Chem. B* **2019,** *123*, 2469-2478.

41. Cossio-Perez, R.; Pierdominici-Sottile, G.; Sobrado, P.; Palma, J., Molecular Dynamics Simulations of Substrate Release from Trypanosoma cruzi UDP-Galactopyranose Mutase. *J. Chem. Inf. Model.* **2019,** *59*, 809-817.

42. Barducci, A.; Bonomi, M.; Prakash, M. K.; Parrinello, M., Free-energy landscape of protein oligomerization from atomistic simulations. *P. Natl. Acad. Sci. USA* **2013,** *110*, E4708-E4713.

43. Takenaka, N.; Suzuki, Y.; Sakai, H.; Nagaoka, M., On Electrolyte-Dependent Formation of Solid Electrolyte Interphase Film in Lithium-Ion Batteries: Strong Sensitivity to Small Structural Difference of Electrolyte Molecules. *J. Phys. Chem. C* **2014,** *118*, 10874-10882.

44. Henkelman, G.; Uberuaga, B. P.; Jonsson, H., A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000,** *113*, 9901-9904.

45. Mashl, R. J.; Jakobsson, E., End-point targeted molecular dynamics: large-scale conformational changes in potassium channels. *Biophys J.* **2008,** *94*, 4307-19.

46. Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T., Molecular

interactions between estrogen receptor and its ligand studied by the ab initio fragment molecular orbital method. *J. Phys. Chem. B* **2006,** *110*, 16102-10.

47. Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P., Collision Cross Sections for Structural Proteomics. *Structure* **2015,** *23*, 791-799.

48. Grossfield, A. *WHAM: The Weighted Histogram Analysis Method, version 2.0.9*, University of Rochester Medical Center: Rochester, NY, 2013.

49. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992,** *13*, 1011-1021.

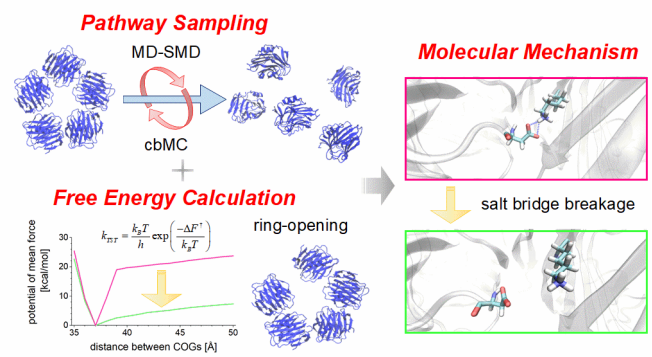
50. Hub, J. S.; de Groot, B. L.; van der Spoel, D., g_wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010,** *6* , 3713-3720.

51. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graphics.* **1996,** *14*, 33-38.

52. Ando, T.; Uchihashi, T.; Scheuring, S., Filming Biomolecular Processes by High-Speed Atomic Force Microscopy. *Chem. Rev.* **2014,** *114*, 3120-3188.

53. Zhang, L. Q.; Borthakur, S.; Buck, M., Dissociation of a Dynamic Protein Complex Studied by All-Atom Molecular Simulations. *Biophys J.* **2016,** *110*, 877-886.

54. Xi, D.; Luo, T. T.; Xiong, H. W.; Liu, J. C.; Lu, H.; Li, M. H.; Hou, Y. Q.; Guo, Z. G., SAP: structure, function, and its roles in immune-related diseases. *Int. J. Cardiol.* **2015,** *187*, 20-26.

55. Cathcart, E. S.; Shirahama, T.; Cohen, A. S., Isolation and Identification of a Plasma Component of Amyloid. *Biochim. Biophys. Acta* **1967,** *147*, 392-393.

TOC graphics