Ph.D. Thesis

Development and Application Studies on

the Generalized-Ensemble Algorithms:

Stability and Aggregation of Proteins

Department of Structural Molecular Science

School of Physical Sciences

The Graduate University for Advanced Studies, SOKENDAI

Masataka Yamauchi

March 8, 2021

Contents

Chapter 1

General Introduction

Proteins perform biological functions by folding into the native structures in physiological conditions. According to Anfinsen's dogma [1], the native structure is the minimum free-energy state. Therefore, the amino-acid sequence of proteins determines the native structure under given environments. The proteins are denatured by changes in environmental conditions such as pH, temperature, and pressure. The denaturation of proteins is a reversible process under a low concentration condition because the proteins recover the native structures when environments return to physiological conditions. On the other hand, the proteins aggregate and form fibril structures when they lack folding ability [2]. The aggregation and fibril formation of the proteins tend to occur under high concentration conditions. They are usually an irreversible process. These aggregates and fibrils are supposed to associate with amyloid diseases such as Alzheimer's and Parkinson's diseases. Therefore, understanding the stability, folding process, and mechanisms of aggregation and fibril formation of the proteins are fundamental issues in protein science. Molecular dynamics (MD) simulation is a powerful tool to investigate the characteristics of the proteins at atomic-resolution, which complements experimental and theoretical studies.

Aggregation of Amyloidogenic Polypeptides

Proteins and polypeptides inherently convert to the amyloid aggregates and fibrils that are distinct from the native structures. The amyloid aggregates and fibrils contain characteristic β-rich structures, which is referred to as cross-β structure. X-ray diffraction from the amyloid aggregated and fibrils shows a characteristic pattern (cross-β diffraction) that is observed at about 4.7 Å and 10 Å on perpendicular axes [3]. The diffraction at 4.7 Å derives from the intervals between inter β-strands within a β-sheet. The diffraction at 10 Å derives from the intervals between inter β-sheets that are stacked one another. To date, X-ray diffraction, solid-state NMR, and cryo-electron microscopy have reported a variety of amyloid fibril structures in which they typically form parallel β-sheets [4–14]. Whereas they share the common features of the amyloids, they have different diameters, twists, packing patterns between protofilaments, and salt-bridges. Namely, the amyloid fibrils show heterogeneous amyloid folds referred to as polymorphism.

Amyloid fibril formation appears to be an intrinsic property that proteins possess in general, rather than a unique property of the amyloidogenic peptides [15–17]. Dobson *et al.* reported that apomyoglobin, which folds into an α-helical rich structure in the native state, converts to the fibril structures under the destabilized conditions of the native structure [15]. They also found that even in polyamino acids peptides such as poly-L-lysine, -glutamic acid, and -threonine can adapt to the fibril structures under carefully selected conditions [16]. These observations suggest an inverse side chain effect in the amyloid formation; the amyloid formation is primarily defined by main chain interaction and a distinct process from protein folding in which side chain contacts play an important role. In other words, the absence of folding propensity results in the amyloid formation.

These aggregates and fibrils are thought to be a cause of neurodegenerative diseases. Amyloid-β and α-synuclein are famous as the diseases-related amyloidogenic polypeptides because deposits and abnormal accumulations of these peptides were recognized as components in amyloid from Alzheimer's and Parkinson's disease patients. Alzheimer's disease does not cure even though the amyloid deposits are removed from the brains of Alzheimer's patients; no cure has been established

yet for these diseases. Furthermore, differences in the aggregates and fibrils structures result in different cytotoxicity and disease symptoms [9, 18–21]. Therefore, understanding the mechanisms underlying the aggregation and fibril formation is necessary to establish treatment for the diseases.

Molecular Dynamics Simulation on the Biomolecule Systems

In 1977, the first application of the MD simulation to a protein was reported by Karplus *et al.* [22]. They performed an MD simulation of the bovine pancreatic trypsin inhibitor for 9.2 ps. Since then, timescale [23, 24], size of systems [25, 26], and accuracy [27–29] have been central issues in the MD studies of biomolecule systems. The MD simulations in the canonical [30–33] and isothermal–isobaric [33–35] ensembles have been widely applied to the biomolecule systems to gain molecular insights into the structures, conformational changes, functions, and thermodynamics. However, the conformational ensemble obtained from the conventional MD simulations is limited only from a few local-minimum states in many cases. This is because proteins have many local-minimum free-energy states and tend to get trapped in one of them, especially under a low-temperature condition. A specially designed supercomputer system called Anton has succeeded in the observations of the folding and unfolding process of the proteins by extending the upper limit of the time scale up to milliseconds [23, 24]. However, except for this special example, the typical time-scale that we can access by the all-atom MD simulations is several hundred nanoseconds to microseconds. This time-scale is too short for the proteins to overcome the free energy barrier. Hence, this situation makes it difficult to perform an accurate estimation of physical quantities such as free energy differences.

Generalized-Ensemble Algorithms

The generalized-ensemble algorithms [36] are one of the powerful frameworks to overcome the sampling problems, which is classified into three methods: multicanonical

algorithm (MUCA) [37–40], simulated tempering (ST) [41, 42], and replica-exchange method (REM) [43, 44].

The MUCA simulation is based on a flat probability distribution of potential energy. The flat probability distribution, P_{MUCA}, is expressed by the product of the density of states, $n(U)$, and a non-Boltzmann weight factor, $W_{\mathrm{MUCA}}(U)$, as follows:

$$
\begin{aligned}
P_{\mathrm{MUCA}} &= n(U)W_{\mathrm{MUCA}}(U) \\
&= n(U)\exp[-\beta_0 U_{\mathrm{MUCA}}(U;\ T_0)] = \mathrm{constant},
\end{aligned}
\tag{1.1}
$$

where, T_0 is a reference temperatures, β_0 is a inverse temperature at T_0, and U_{MUCA} is the multicanonical potential energy. Because this artificial non-Boltzmann statistical ensemble realizes a random walk in the potential energy space, the target system does not get trapped in local-minimum free-energy states.

The ST simulation regards temperature as a dynamic variable; that is, temperature varies during the simulation. A non-Boltzmann weight factor

$$
W_{\mathrm{ST}}(U;\ T) \equiv \exp\{-\beta U + f(T)\}
\tag{1.2}
$$

is introduced to realize the uniform distribution of the temperature as follows:

$$
\begin{aligned}
P_{\mathrm{ST}}(T) &= \int dU\ n(U)W_{\mathrm{ST}}(U;\ T) \\
&= \int dU\ n(U)\exp\{-\beta U + f(T)\} = \mathrm{constant}.
\end{aligned}
\tag{1.3}
$$

In the ST, therefore, the random walk in the temperature space induces the random walk in the potential space, and thus, the target system can escape from a local-minimum free-energy state. The non-Boltzmann weight factors introduced in the MUCA and ST simulations are not known *a priori*; therefore, we have to determine it in advance by iterative procedures.

The REM utilizes multiple non-interaction copies (or replicas) of the target system at different temperatures. The REM is performed through the following procedures: (i) MD or Monte Carlo (MC) simulations at different temperatures are performed in parallel and (ii) temperature exchange trials between two replicas are

attempted at every certain time step using the Metropolis algorithm. Typically, two replicas that have neighboring temperatures are chosen as a pair of exchange trials. Each replica realizes a random walk in the temperature space during the simulation, which induces the random walk in the potential energy space. Thus, the target system can escape from a local-minimum free-energy state. The REM is widely used nowadays because it is not necessary to determine the non-Boltzmann weight factor in advance. We can perform the replica-exchange simulation only by choosing the temperatures for the replicas. Importantly, the generalized-ensemble simulations can reproduce the statistical average of physical quantities by the reweighting techniques after the simulations.

Thanks to the recent improvement of computational hardware and software, we can apply the MD simulation to the large-scale molecular systems. On the other hand, the generalized-ensemblte algorithms have drawbacks for applying to the large molecular systems. In the case of REM, the number of replicas needed increases as $\mathcal{O}(\sqrt{N})$ where N is the degrees of freedom of the system. In other words, the number of required replicas increases as the system size becomes large. Therefore, the application of the REM to the large-scale molecular systems is still difficult due to computational costs. Many algorithms have been proposed to solve the drawback, which can be classified into two strategies.

The first strategy is an extension of the REM to the multidimensional spaces. The multidimensional REM [45] (also known as Hamiltonian REM [46]) and its variants have been proposed; for instance, replica-exchange umbrella sampling [45], replica-exchange solute tempering [47], and coulomb replica-exchange [48] methods. The multidimensional REM introduces not only temperature but also any parameters as the exchange parameters. This allows us to decrease the degrees of freedom regarding replica-exchange by only focusing on the degrees of freedom of interest. The multidimensional REM can also induce conformational changes of interest by carefully selecting the exchange parameters.

The second strategy is to improve the replica-exchange procedure itself, for example, increasing the transition ratio (acceptance ratio) of the temperature values. To achieve this, we have to improve algorithms to calculate transition probabilities

and select candidates for the next state. A replica-permutation method (RPM) [49] is a good example of this strategy. The RPM performs temperature permutation among more than two replicas during the simulation, whereas the REM performs temperature exchange between two replicas. The transition probability is calculated by the Suwa–Todo algorithm [50] instead of the Metropolis algorithm [51]. The Suwa–Todo algorithm satisfies only the balance condition without imposing the detailed balance condition. This method minimizes the rejection ratio of the MC trials. The RPM succeeds in improve sampling efficiency both in temperature and conformational spaces.

Importance of the Isothermal–Isobaric Ensemble

The generalized-ensemble algorithms were originally developed in the canonical ensemble. The canonical ensemble simulation does not control the pressure because the volume of the simulation cell is constant. That is, the canonical ensemble simulation cannot treat the phenomena involving volume change. In contrast, temperature and pressure are controlled in experimental conditions. The physical quantities reported by experiments follow the isothermal–isobaric ensemble. In particular, the temperature and pressure are significant to characterize the thermodynamics properties of the proteins. Once we obtain temperature and pressure dependence of the difference in the Gibbs free energy $\Delta G(T, P)$, we can calculate the following thermodynamics quantities from the partial derivatives or fitting parameter of $\Delta G(T, P)$ [52]: the differences in the partial molar enthalpy ΔH, the partial molar entropy ΔS, partial molar heat capacity ΔC_p, partial molar volume ΔV, thermal expansivity factor $\Delta \alpha$, and compressibility factor $\Delta \beta$. Therefore, the isothermal–isobaric MD simulation is significant to reflect experimental conditions.

According to Le Châtelie's principle, smaller volume states of a substance realize under high-pressure conditions; namely, the probabilities of the states with smaller volume increase as pressure increases. This principle tells us that the pressure perturbation induces the population changes of the conformational states. Let us suppose that there are various metastable states between the native and denatured

states. Here, the metastable states are supposed to have a lower population than the native state in the physiological condition. The populations of the metastable states that have lower volume than the native state increase as pressure increases. Finally, the denatured state appears because the denatured states usually has the smallest partial molar volume. High-pressure experiments have been succeeded in detecting the metastable states of proteins, conformational changes, and estimation of thermodynamics quantities among the native, denatured, and metastable states [53–62]. High-pressure nuclear magnetic resonance experiments further succeeded in determining the metastable structure of the protein [60]. Therefore, pressure is a robust tool to detect these metastable states and provide thermodynamics information.

It is also significant to understand the biological meanings of the metastable state; that is, the relation among structures of the metastable states, functions, and folding pathways. The MD simulation complements experimental results and provides atomic-resolution pictures of the volume changes by cavities, water penetration, and solute–solvent interactions [63–69]. As the author mentioned above, however, the conformation of proteins tend to get trapped in local-minimum free-energy states in the MD simulation, which prevents us from accurate evaluation of thermodynamics quantities. The consistency of the thermodynamics quantities between the simulations and experiments should be ensured before gaining molecular insight from the simulation. Thus, we need an efficient simulation method in the isothermal–isobaric ensemble that can cover a wide range of temperature and pressure space. The REM, MUCA, and ST have extended to the isothermal–isobaric ensemble [63, 70–75]; however, the RPM has not been extended to the isothermal–isobaric ensemble yet.

The Purpose of This Thesis

In this doctoral thesis, the author addresses the following four research subjects:

1. Development of an isothermal–isobaric replica-permutation method to realize an efficient molecular simulation in a wide range of temperature and pressure.

2. Development of a replica sub-permutation method beyond the replica-exchange and replica-permutation methods.

3. Application of the isothermal–isobaric replica-permutation method to reveal temperature and pressure dependence of a small protein, chignolin.

4. Investigation of the initial process of the disease-related protein aggregation by the isothermal–isobaric replica-permutation method.

The former half of this thesis is devoted to the development and improvement of the replica-permutation method. In Chapter 2, the author introduces the Markov chain Monte Carlo algorithms and the formulation of the REM and RPM. In Chapter 3, the author describes the development of the isothermal–isobaric replica-permutation methods. The Suwa–Todo [50], Metropolis [51], and heat bath [76] algorithms are compared to investigate the effects on the efficiency of the parameter permutation. In Chapter 4, the author describes the development of the replica sub-permutation method. In this method, a new permutation algorithm termed "sub-permutation" is introduced to improve the sampling efficiency in the parameter permutation trials.

The latter half of this thesis is devoted to the application studies of the isothermal–isobaric replica-permutation method developed by the author. In Chapter 5, the author investigates the temperature and pressure dependence of the chignolin. The chignolin consists of ten amino-acid residues and has folded, misfolded, and unfolded states. Both folded and misfolded conformations have unique β-hairpin structures. The author evaluates thermodynamics quantities among the three states. The author finds that the pressure dependence of folded and misfolded chignolin is completely different and reveals its mechanisms. In Chapter 6, the author investigates the dimerization of α-synuclein fragments as the initial stage of the aggregation process. The α-synuclein is an intrinsically disordered protein consisting of 140 amino acid residues. The aggregates and fibrils of the α-synuclein are believed to be associated with Parkinson's disease. Here, the author focuses on an 11-residue segment that is an essential region for both fibril formation and cytotoxicity [9]. The author investigates the stabilities and structural properties of the fragments along the

dimerization process. The dimerization mechanism is also presented. Chapter 7 is devoted to conclusions.

References

[1] C B Anfinsen, E Haber, M. Sela, and F H White. The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Popypeptide Chain. *Proceedings of the National Academy of Sciences*, 47(9):1309–1314, 1961.

[2] Fabrizio Chiti and Christopher M. Dobson. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry*, 86(1):27–68, 2017.

[3] E. D. Eanes and G. G. Glenner. X-ray Diffraction Studies on Amyloid Filaments. *Journal of Histochemistry & Cytochemistry*, 16(11):673–677, 1968.

[4] Aneta T. Petkova, Yoshitaka Ishii, John J. Balbach, Oleg N. Antzutkin, Richard D. Leapman, Frank Delaglio, and Robert Tycko. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16742–16747, 2002.

[5] T. Luhrs, Christiane Ritter, Marc Adrian, D. Riek-Loher, Bernd Bohrmann, H. Dobeli, David Schubert, and Roland Riek. 3D structure of Alzheimer's amyloid-β(1-42) fibrils. *Proceedings of the National Academy of Sciences*, 102(48):17342–17347, 2005.

[6] Jun-Xia Lu, Wei Qiang, Wai-Ming Yau, Charles D. Schwieters, Stephen C. Meredith, and Robert Tycko. Molecular Structure of β-Amyloid Fibrils in Alzheimer's Disease Brain Tissue. *Cell*, 154(6):1257–1268, 2013.

[7] Yiling Xiao, Buyong Ma, Dan McElheny, Sudhakar Parthasarathy, Fei Long, Minako Hoshi, Ruth Nussinov, and Yoshitaka Ishii. Aβ(1-42) fibril structure

illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural and Molecular Biology*, 22(6):499–505, 2015.

[8] Michael T. Colvin, Robert Silvers, Qing Zhe Ni, Thach V. Can, Ivan Sergeyev, Melanie Rosay, Kevin J. Donovan, Brian Michael, Joseph Wall, Sara Linse, and Robert G. Griffin. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *Journal of the American Chemical Society*, 138(30):9663–9674, 2016.

[9] Jose A Rodriguez, Magdalena I Ivanova, Michael R Sawaya, Duilio Cascio, Francis E Reyes, Dan Shi, Smriti Sangwan, Elizabeth L Guenther, Lisa M Johnson, Meng Zhang, Lin Jiang, Mark A Arbing, Brent L Nannenga, Johan Hattne, Julian Whitelegge, Aaron S Brewster, Marc Messerschmidt, Sébastien Boutet, Nicholas K Sauter, Tamir Gonen, and David S. Eisenberg. Structure of the toxic core of α-synuclein from invisible crystals. *Nature*, 525(7570):486–490, 2015.

[10] Marcus D. Tuttle, Gemma Comellas, Andrew J. Nieuwkoop, Dustin J. Covell, Deborah A. Berthold, Kathryn D. Kloepper, Joseph M. Courtney, Jae K. Kim, Alexander M. Barclay, Amy Kendall, William Wan, Gerald Stubbs, Charles D. Schwieters, Virginia M Y Lee, Julia M. George, and Chad M. Rienstra. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nature Structural & Molecular Biology*, 23(5):409–415, 2016.

[11] Ricardo Guerrero-Ferreira, Nicholas M.I. Taylor, Daniel Mona, Philippe Ringler, Matthias E. Lauer, Roland Riek, Markus Britschgi, and Henning Stahlberg. Cryo-EM structure of alpha-synuclein fibrils. *eLife*, 7(i):1–18, 2018.

[12] Binsen Li, Peng Ge, Kevin A. Murray, Phorum Sheth, Meng Zhang, Gayatri Nair, Michael R. Sawaya, Woo Shik Shin, David R. Boyer, Shulin Ye, David S. Eisenberg, Z. Hong Zhou, and Lin Jiang. Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications*, 9(1):1–10, 2018.

[13] Ricardo Guerrero-Ferreira, Nicholas M.I. Taylor, Ana-Andreea Arteni, Pratibha Kumari, Daniel Mona, Philippe Ringler, Markus Britschgi, Matthias E. Lauer,

Ali Makky, Joeri Verasdonck, Roland Riek, Ronald Melki, Beat H. Meier, Anja Böckmann, Luc Bousset, and Henning Stahlberg. Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife*, 8:1–24, 2019.

[14] T Strohaker, B C Jung, S H Liou, C O Fernandez, D Riedel, S Becker, G M Halliday, M Bennati, W S Kim, S J Lee, and M Zweckstetter. Structural heterogeneity of alpha-synuclein fibrils amplified from patient brain extracts. *Nat Commun*, 10(1):5535, 2019.

[15] Marcus Fändrich, Matthew A. Fletcher, and Christopher M. Dobson. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825):165–166, 2001.

[16] M. Fandrich. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal*, 21(21):5682–5690, 2002.

[17] Yoshikazu Aso, Kentaro Shiraki, and Masahiro Takagi. Systematic analysis of aggregates from 38 kinds of non disease-related proteins: Identifying the intrinsic propensity of polypeptides to form amyloid fibrils. *Bioscience, Biotechnology and Biochemistry*, 71(5):1313–1321, 2007.

[18] Benedetta Mannini, Estefania Mulvihill, Caterina Sgromo, Roberta Cascella, Reza Khodarahmi, Matteo Ramazzotti, Christopher M. Dobson, Cristina Cecchi, and Fabrizio Chiti. Toxicity of protein oligomers is rationalized by a function combining size and surface hydrophobicity. *ACS Chemical Biology*, 9(10):2309–2317, 2014.

[19] Giuliana Fusco, Serene W. Chen, Philip T. F. Williamson, Roberta Cascella, Michele Perni, James A. Jarvis, Cristina Cecchi, Michele Vendruscolo, Fabrizio Chiti, Nunilo Cremades, Liming Ying, Christopher M. Dobson, and Alfonso De Simone. Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science*, 358(6369):1440–1443, 2017.

[20] Mohammad Shahnawaz, Abhisek Mukherjee, Sandra Pritzkow, Nicolas Mendez, Prakruti Rabadia, Xiangan Liu, Bo Hu, Ann Schmeichel, Wolfgang

Singer, Gang Wu, Ah-Lim Tsai, Hamid Shirani, K. Peter R. Nilsson, Phillip A. Low, and Claudio Soto. Discriminating α-synuclein strains in Parkinson's disease and multiple system atrophy. *Nature*, 578(7794):273–277, 2020.

[21] David R. Boyer, Binsen Li, Chuanqi Sun, Weijia Fan, Kang Zhou, Michael P. Hughes, Michael R. Sawaya, Lin Jiang, and David S. Eisenberg. The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7):3592–3602, 2020.

[22] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, 1977.

[23] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[24] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How Fast-Folding Proteins Fold. *Science*, 334(6055):517–520, 2011.

[25] Jaewoon Jung, Takaharu Mori, Chigusa Kobayashi, Yasuhiro Matsunaga, Takao Yoda, Michael Feig, and Yuji Sugita. GENESIS: a hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(4):310–323, 2015.

[26] Yoshimichi Andoh, Noriyuki Yoshii, Kazushi Fujimoto, Keisuke Mizutani, Hidekazu Kojima, Atsushi Yamada, Susumu Okazaki, Kazutomo Kawaguchi, Hidemi Nagao, Kensuke Iwahashi, Fumiyasu Mizutani, Kazuo Minami, Shin-ichi Ichikawa, Hidemi Komatsu, Shigeru Ishizuki, Yasuhiro Takeda, and Masao Fukushima. MODYLAS: A Highly Parallelized General-Purpose Molecular Dynamics Simulation Program for Large-Scale Systems with Long-Range Forces Calculated by Fast Multipole Method (FMM) and Highly Scalable Fine-

Grained New Parallel Processing Algorithms. *Journal of Chemical Theory and Computation*, 9(7):3201–3209, 2013.

[27] W L Jorgensen and J Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[28] Christopher I Bayly, Kenneth M Merz, David M Ferguson, Wendy D Cornell, Thomas Fox, James W Caldwell, Peter A Kollman, Piotr Cieplak, Ian R Gould, David C Spellmeyer, Christopher I Bayly, Ian R. Gould, Jr. Kenneth M. Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, ames W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[29] Alexander D. MacKerell, Bernard Brooks, Charles L. Brooks, Lennart Nilsson, Benoit Roux, Youngdo Won, and Martin Karplus. CHARMM: The Energy Function and Its Parameterization. In *Encyclopedia of Computational Chemistry*. John Wiley & Sons, Ltd, Chichester, UK, 2002.

[30] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[31] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[32] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[33] Glenn J. Martyna, Mark E. Tuckerman, Douglas J. Tobias, and Michael L. Klein. Explicit reversible integrators for extended systems dynamics. *Molecular Physics*, 87(5):1117–1157, 1996.

[34] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[35] M. Parrinello and A. Rahman. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters*, 45(14):1196–1199, 1980.

[36] Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, 60(2):96–123, 2001.

[37] Bernd A Berg and Thomas Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267(2):249–253, 1991.

[38] Bernd A. Berg and Thomas Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68(1):9–12, 1992.

[39] Ulrich H.E. Hansmann, Yuko Okamoto, and Frank Eisenmenger. Molecular dynamics, Langevin and hydrid Monte Carlo simulations in a multicanonical ensemble. *Chemical Physics Letters*, 259(3-4):321–330, 1996.

[40] Nobuyuki Nakajima, Haruki Nakamura, and Akinori Kidera. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *The Journal of Physical Chemistry B*, 101(5):817–824, 1997.

[41] Enzo Marinari and Giorgio Parisi. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)*, 19(6):451–458, 1992.

[42] A P Lyubartsev, A A Martsinovski, S V Shevkunov, and P N Vorontsov - Velyaminov. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics*, 96(3):1776–1783, 1992.

[43] Koji Hukushima and Koji Nemoto. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[44] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[45] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *Journal of Chemical Physics*, 113(15):6042–6051, 2000.

[46] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058–9067, 2002.

[47] Pu Liu, Byungchan Kim, Richard A Friesner, and B J Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences*, 102(39):13749–13754, 2005.

[48] Satoru G. Itoh and Hisashi Okumura. Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules. *Journal of Computational Chemistry*, 34(8):622–639, 2013.

[49] Satoru G Itoh and Hisashi Okumura. Replica-Permutation Method with the Suwa–Todo Algorithm beyond the Replica-Exchange Method. *Journal of Chemical Theory and Computation*, 9(1):570–581, 2013.

[50] Hidemaro Suwa and Synge Todo. Markov Chain Monte Carlo Method without Detailed Balance. *Physical Review Letters*, 105(12):120603, 2010.

[51] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[52] S. A. Hawley. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry*, 10(13):2436–2442, 1971.

[53] Kazuyuki Akasaka. Probing Conformational Fluctuation of Proteins by Pressure Perturbation. *Chemical Reviews*, 106(5):1814–1835, 2006.

[54] Jerson L Silva, Andrea C. Oliveira, Tuane C R G Vieira, Guilherme A. P. de Oliveira, Marisa C Suarez, and Debora Foguel. High-Pressure Chemical Biology and Biotechnology. *Chemical Reviews*, 114(14):7239–7267, 2014.

[55] Takahiro Takekiyo, Takashi Imai, Minoru Kato, and Yoshihiro Taniguchi. Temperature and Pressure Effects on Conformational Equilibria of Alanine Dipeptide in Aqueous Solution. *Biopolymers*, 73(2):283–290, 2004.

[56] Hiroshi Imamura and Minoru Kato. Effect of pressure on helix-coil transition of an alanine-based peptide: An FTIR study. *Proteins: Structure, Function and Bioinformatics*, 75(4):911–918, 2009.

[57] Sabine Neumaier, Maren Büttner, Annett Bachmann, and Thomas Kiefhaber. Transition state and ground state properties of the helix-coil transition in peptides deduced from high-pressure studies. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52):20988–20993, 2013.

[58] R. Kitahara, H. Yamada, and K. Akasaka. Two folded conformers of ubiquitin revealed by high-pressure NMR. *Biochemistry*, 40(45):13556–13563, 2001.

[59] Ryo Kitahara and Kazuyuki Akasaka. Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3167–3172, 2003.

[60] Ryo Kitahara, Shigeyuki Yokoyama, and Kazuyuki Akasaka. NMR Snapshots of a Fluctuating Protein Structure: Ubiquitin at 30 bar–3 kbar. *Journal of Molecular Biology*, 347(2):277–285, 2005.

[61] Lydia Nisius and Stephan Grzesiek. Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network. *Nature Chemistry*, 4(9):711–717, 2012.

[62] Layara Akemi Abiko, Anne Grahl, and Stephan Grzesiek. High Pressure Shifts the β 1 -Adrenergic Receptor to the Active Conformation in the Absence of

G Protein. *Journal of the American Chemical Society*, 141(42):16663–16670, 2019.

[63] Masataka Yamauchi, Yoshiharu Mori, and Hisashi Okumura. Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble. *Biophysical Reviews*, 11(3):457–469, 2019.

[64] Hiroaki Hata, Masayoshi Nishiyama, and Akio Kitao. Molecular dynamics simulation of proteins under high pressure: Structure, function and thermodynamics. *Biochimica et Biophysica Acta - General Subjects*, 1864(2):1–10, 2020.

[65] Ryan Day and Angel E. García. Water penetration in the low and high pressure native states of ubiquitin. *Proteins: Structure, Function, and Bioinformatics*, 70(4):1175–1184, 2008.

[66] Takashi Imai and Yuji Sugita. Dynamic correlation between pressure-induced protein structural transition and water penetration. *Journal of Physical Chemistry B*, 114(6):2281–2286, 2010.

[67] Y Mori and H Okumura. Pressure-Induced Helical Structure of a Peptide Studied by Simulated Tempering Molecular Dynamics Simulations. *J Phys Chem Lett*, 4(12):2079–2083, 2013.

[68] Y Mori and H Okumura. Molecular dynamics of the structural changes of helical peptides induced by pressure. *Proteins*, 82(11):2970–2981, 2014.

[69] Y Mori and Y Okamoto. Conformational changes of ubiquitin under high pressure conditions: A pressure simulated tempering molecular dynamics study. *Journal of Computational Chemistry*, 38(15):1167–1173, 2017.

[70] Tsuneyasu Okabe, Masaaki Kawata, Yuko Okamoto, and Masuhiro Mikami. Replica-exchange Monte Carlo method for the isobaric–isothermal ensemble. *Chemical Physics Letters*, 335(5-6):435–439, 2001.

[71] Yoshiharu Mori and Yuko Okamoto. Generalized-Ensemble Algorithms for the Isobaric–Isothermal Ensemble. *Journal of the Physical Society of Japan*, 79(7):074003, 2010.

[72] Hisashi Okumura and Yuko Okamoto. Monte Carlo simulations in multibaric – multithermal ensemble. *Chemical Physics Letters*, 383(3-4):391–396, 2004.

[73] Hisashi Okumura and Yuko Okamoto. Monte Carlo simulations in generalized isobaric–isothermal ensembles. *Physical Review E*, 70(2):026702, 2004.

[74] Hisashi Okumura and Yuko Okamoto. Molecular dynamics simulations in the multibaric–multithermal ensemble. *Chemical Physics Letters*, 391(4-6):248–253, 2004.

[75] Hisashi Okumura and Yuko Okamoto. Multibaric–multithermal ensemble molecular dynamics simulations. *Journal of Computational Chemistry*, 27(3):379–395, 2006.

[76] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

Chapter 2

Simulation Methods

This chapter introduces Markov chain Monte Carlo algorithms [1], and then reviews formulation of the replica-exchange method (REM) [2] and replica-permutation method (RPM) [3].

2.1 Markov Chain Monte Carlo Algorithm

This section introduces the basics of Markov chain Monte Carlo algorithms. Here, the author assumes a system that has n states. Each state i has a weight w_i and transits to another state j with a transition probability $P(i \rightarrow j)$. A stochastic flow $v(i \rightarrow j)$ from the state i to the state j is defined as follows:

$$v(i \rightarrow j) = w_i P(i \rightarrow j). \tag{2.1}$$

To converge the unique distribution, Markov chain should satisfy ergodicity. That is, the Markov chain are *irreducibility* and *aperiodicity*. In addition, the following equations are imposed to obtain the required equilibrium distribution such as isothermal–isobaric ensemble:

$$\sum_{i=1}^{n} P(i \rightarrow j) = 1, \tag{2.2}$$

$$\sum_{i=1}^{n} w_i P(i \rightarrow j) = w_j. \tag{2.3}$$

Equation (2.2) is the normalization condition and Eq. (2.3) is the balance condition. Eqs. (2.2) and (2.3) are summarized by using the stochastic flow as follows:

$$\sum_{i=1}^{n} v(i \rightarrow j) = \sum_{i=1}^{n} v(j \rightarrow i). \tag{2.4}$$

By performing state transitions with $v(i \rightarrow j)$ that satisfy Eq. (2.4), the required statistical ensemble is obtained.

The Metoropolis algorithm [4] (its generalization is called Metropolis–Hastings algorithm [5]) and heat bath algorithms (also referred to as Gibbs sampler) are the two most practical implementations of the MCMC algorithms. These algorithms only satisfy the detailed balance condition, which is the sufficient condition of Eq. (2.4):

$$v(i \rightarrow j) = v(j \rightarrow i). \tag{2.5}$$

In the Metropolis algorithms, the amount of stochastic flow and transition probability from state i to state j is given as follows:

$$v(i \rightarrow j) = \frac{1}{n-1} \min \left[w_i, \ w_j \right], \ i \neq j, \tag{2.6}$$

$$P(i \rightarrow j) = \frac{v(i \rightarrow j)}{w_i} = \frac{1}{n-1} \min \left[1, \ \frac{w_j}{w_i} \right], \ i \neq j, \tag{2.7}$$

where the coefficient $1/(n-1)$ arises from the random selection of state j from $n-1$ candidates except state i. In the Metropolis algorithm, transition probability is calculated by using only the weight of current state w_i and the weight of candidate state w_j. In the heat bath algorithms, the amount of stochastic flow and transition probability from state i to state j are given as

$$v(i \rightarrow j) = \frac{w_i w_j}{\sum_{k=1}^{n} w_k}, \ \forall i, \ j, \tag{2.8}$$

$$P(i \rightarrow j) = \frac{v(i \rightarrow j)}{w_i} = \frac{w_j}{\sum_{k=1}^{n} w_k}, \ \forall i, \ j. \tag{2.9}$$

The schematic figure of the Metropolis and heat bath algorithms is presented in Figs. 2.1(a) and (b).

The author then introduces the Suwa–Todo algorithm. This algorithm satisfies
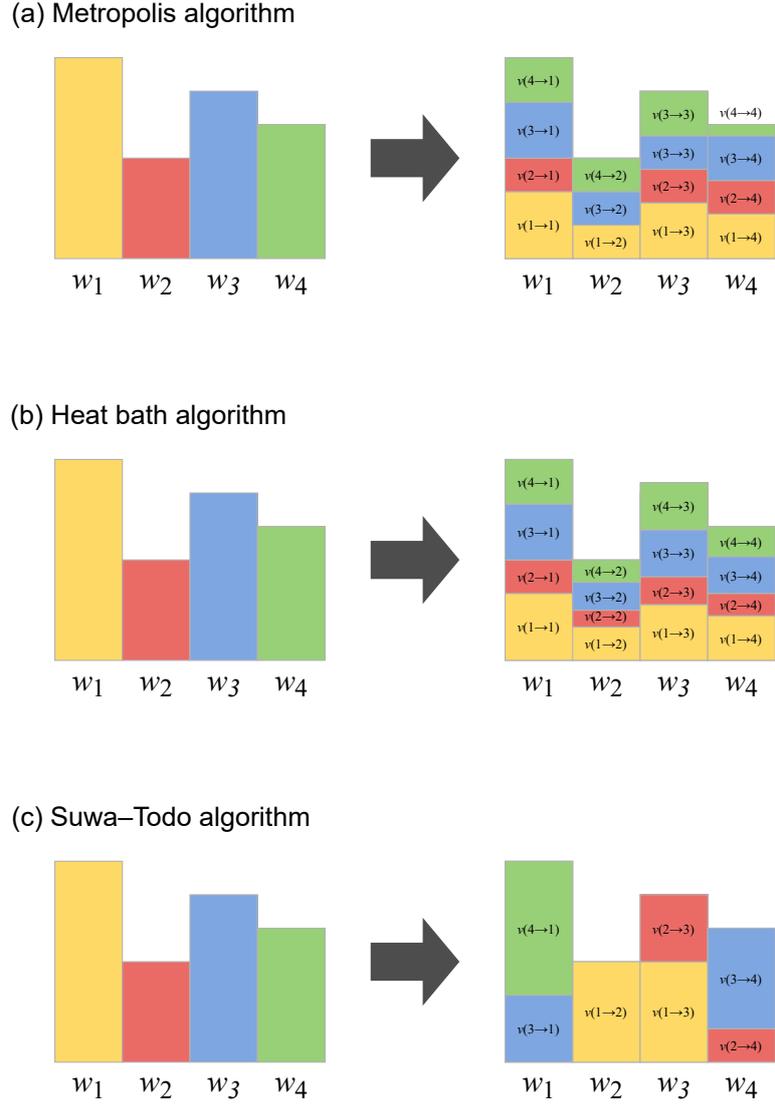
Figure 2.1: Schematic figures of the (a) Metropolis, (b) heat bath, and (c) Suwa–Todo algorithms.

Eq. (2.4) without imposing the detailed balance condition Eq. (2.5). The Suwa–Todo algorithm is explained by the following procedure, also Fig. 2.1(c) shows the schematic of weight allocation. (i) Determine a state that has the maximum weight factor. If two or more states have the maximum weight factors, one of them is chosen. The maximum weight is labeled by w_1 without loss of generality. Other weights are labeled as $w_2, \ldots, w_n$ in any order. The following procedure is like filling a "box" w_j by "liquid" i originally in the box w_i. The stochastic flow $v(i \to j)$ is the amount of liquid i that moves from box w_i to box w_j; this corresponds to the stochastic flow from state i to state j. (ii) Fill the box w_2 as much as possible with

liquid 1 in box w_1. The amount of this liquid is $v(1 \to 2)$. If liquid 1 in box w_1 still remains, then fill the next box w_3 with as much liquid as possible by the rest of liquid 1. This procedure should be continued until all of liquid 1 is moved to other boxes. The last partially filled box is referred to as w_k. The stochastic flows $v(1 \to 2)$, ..., $v(1 \to k)$ are determined via this process. (ⅲ) By using liquid 2, fill the last partially filled box w_k to the maximum extent possible. If box w_k is filled, then fill box w_{k+1}. This procedure should be continued until all of liquid 2 is moved to other boxes w_{k+2}, ..., w_l. The stochastic flows $v(2 \to k)$, ..., $v(2 \to l)$ are obtained. (ⅳ) Once the last box w_n is filled, then fill box w_1 by all the remaining liquids. All liquids are moved to other boxes in this manner.

The amount of stochastic flow of the Suwa–Todo algorithm is formulated as follows:

$$v(i \to j) = \max \left[0, \ \min \left[\Delta_{ij}, w_i + w_j - \Delta_{ij}, \ w_i, w_j\right]\right], \tag{2.10}$$

where

$$\Delta_{ij} \equiv S_i - S_{j-1} + w_1, \tag{2.11}$$

$$S_i \equiv \sum_{k=1}^{i} w_k, \tag{2.12}$$

$$S_0 \equiv S_n, \tag{2.13}$$

and w_1 is assumed to have the maximum weight factor among n states. The transition probability, $P(i \to j) = v(i \to j)/w_i$, is obtained by using Eq. (2.10). The stochastic flow from state i to state i is calculated as follows:

$$v(i \to i) = \begin{cases} \max[0, \ 2w_1 - S_n], & i = 1 \\ 0, & i \geq 2. \end{cases} \tag{2.14}$$

This equation implies that the rejection ratio is zero when $w_1 \leq S_n/2$. Therefore, the Suwa–Todo algorithm minimizes the rejection ratio. Note that the algorithm is consistent with the Metropolis algorithm when $n = 2$.

2.2 Replica-Exchange Method

Let us consider a system of N atoms and M non-interacting copies (replicas). Each replica realizes the canonical ensemble at a different temperature T_m ($m = 1, 2, \ldots, M$). In the REM, temperature and replica correspond one-to-one. Therefore, the replica index i ($i = 1, \ldots, M$) is a permutation of the temperature index m ($m = 1, 2 \ldots, M$) and vice versa:

$$\begin{cases} i = i(m) \equiv f(m) \\ m = m(i) \equiv f^{-1}(i), \end{cases} \tag{2.15}$$

where $f(m)$ is a mapping function from the temperature index to the replica index, and $f^{-1}(i)$ is the inverse. A state in the REM is expressed by a combination of the temperature indices and the replica indices as follows:

$$X_\mu = \left[x_1^{[i(1)]}, \ x_2^{[i(2)]}, \ \ldots, \ x_M^{[i(M)]} \right] = \left[x_{m(1)}^{[1]}, \ x_{m(2)}^{[2]}, \ \ldots, \ x_{m(M)}^{[M]} \right]. \tag{2.16}$$

State X_μ is specified by the M sets of replicas. The state of each replica is expressed by a set of coordinates and momenta:

$$x_m^{[i]} \equiv (q^{[i]}, \ p^{[i]})_m. \tag{2.17}$$

The Hamiltonian H in state $x_m^{[i]}$ is given by the sum of the kinetic energy K and potential energy U:

$$H(x_m^{[i]}) = K(p_m^{[i]}) + U(q_m^{[i]}). \tag{2.18}$$

In the canonical ensemble at temperature T_m, each state $x_m^{[i]}$ is weighted by the Boltzmann factor:

$$w_{\mathrm{B}} = \exp[-\beta_m H(x_m^{[i]})], \tag{2.19}$$

where $\beta_m = 1/k_{\mathrm{B}} T_m$ (k_{B} is the Boltzmann constant). Because each replica does not interact, the weight factor of state X_μ is given by the product of the Boltzmann

factor of each replica i

$$w_{\mathrm{R}}(X_\mu) = \prod_{i=1}^{M} \exp\left[-\beta_{m(i)} H(x_{m(i)}^{[i]})\right] = \prod_{m=1}^{M} \exp\left[-\beta_m H(x_m^{[i(m)]})\right]. \tag{2.20}$$

Every certain step during the simulation, the REM tries to exchange replicas i and j at temperatures T_m and T_n, respectively ($i = f(m)$ and $j = f(n)$). The exchanging a pair of replicas in the REM is expressed as follows:

$$X_\mu = [\ldots, x_m^{[i]}, \ldots, x_n^{[j]}, \ldots] \to X_\nu = [\ldots, x_m^{[j]'}, \ldots, x_n^{[i]'}, \ldots]. \tag{2.21}$$

Figure 2.2(a) shows the schematic figure of the REM. The amount of stochastic flow and transition probability for this replica exchange are obtained by the Metropolis criteria in Eqs. (2.6) and (2.7):

$$v(X_\mu \to X_\nu) = C \min\left[w_{\mathrm{R}}(X_\mu),\ w_{\mathrm{R}}(X_\nu)\right], \tag{2.22}$$

$$P(X_\mu \to X_\nu) = C \min\left[1,\ \frac{w_{\mathrm{R}}(X_\nu)}{w_{\mathrm{R}}(X_\mu)}\right], \tag{2.23}$$

where $C = 1/_M C_2$. When the replica-exchange trial is accepted, the momenta of the replicas are scaled to satisfy the condition $\langle K(p_m^{[i]})\rangle_{T_m} = \frac{3}{2} N k_{\mathrm{B}} T_m$ as follows:

$$p_n^{[i]'} = \sqrt{\frac{T_n}{T_m}} p_m^{[i]}, \tag{2.24}$$

$$p_m^{[j]'} = \sqrt{\frac{T_m}{T_n}} p_n^{[j]}. \tag{2.25}$$

This rescaling of moment cancel out the kinetic energy term, K, in Eq. (2.23). Thus, the weight factor

$$w_{\mathrm{R}}(X_\mu) = \prod_{m=1}^{M} \exp[-\beta_m U(x_m^{[i(m)]})] \tag{2.26}$$

is used instead of Eq. (2.19). The transition probability is calculated as follows:

$$P(X_\mu \to X_\nu) = C \min\left[1,\ \exp(-\Delta)\right], \tag{2.27}$$
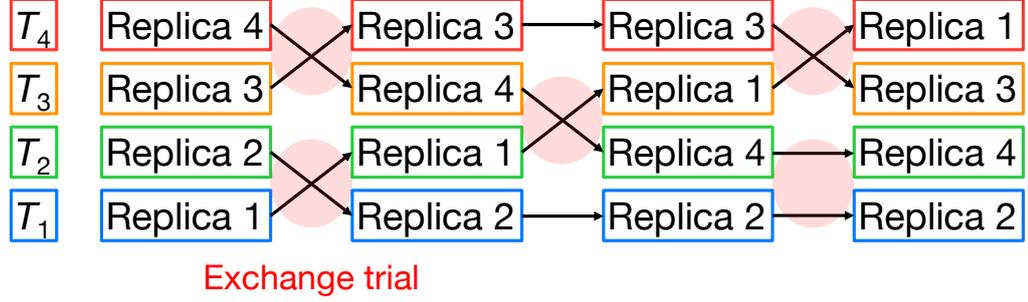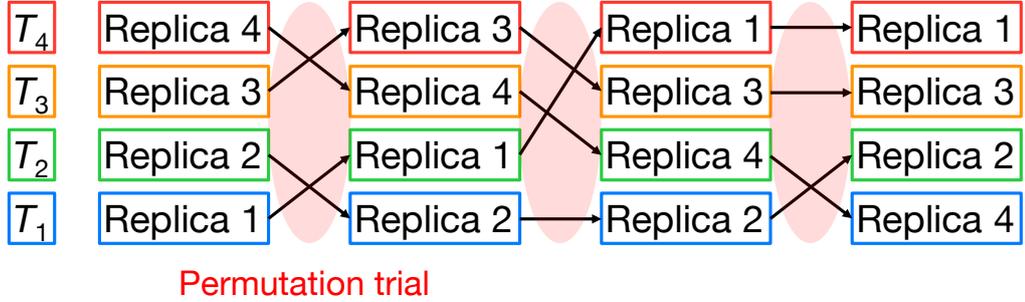
Figure 2.2: Schematic figures of the (a) replica-exchange and (b) replica-permutation methods.

where Δ is given as

$$\Delta = (\beta_m - \beta_n)(U(q^{[j]}) - U(q^{[i]})). \qquad (2.28)$$

2.3 Replica-Permutation Method

Let us consider performing replica permutation among all M parameters

$$X_\mu = \left[x_1^{[i(1)]}, \ldots, x_M^{[i(M)]} \right] \to X_\nu = \left[x_1^{[j(1)]}, \ldots, x_M^{[j(M)]} \right], \qquad (2.29)$$

or parameter permutation among all M replicas

$$X_\mu = \left[x_{m(1)}^{[1]}, \ldots, x_{m(M)}^{[M]} \right] \to X_\nu = \left[x_{n(1)}^{[1]}, \ldots, x_{n(M)}^{[M]} \right], \qquad (2.30)$$

Table 2.1: An example of replica-permutation candidates for three replicas system.

State index μ	Replica 1	Replica 2	Replica 3
1	T_1	T_2	T_3
2	T_2	T_1	T_3
3	T_1	T_3	T_2
4	T_2	T_3	T_1
5	T_3	T_1	T_2
6	T_3	T_2	T_1

where i, j, m, and n are permutation functions. In this method, temperature permutation among more than two replicas are allowed. Note that the number of all possible combinations between the replica and temperatures is $M!$. Therefore, the range of indices μ and ν are between 1 and $M!$. Table 2.1 shows an example of replica-permutation candidates in a three replica system. Figure 2.2(b) shows the schematic figure of the RPM.

The weight factor in Eq. (2.26) is also used in the RPM because momenta are modified when the replica permutation is accepted as follows:

$$p_{n(i)}^{[i]} = \sqrt{\frac{T_{n(i)}}{T_{m(i)}}} p_{m(i)}^{[i]}. \tag{2.31}$$

In the RPM, the Suwa–Todo algorithm, instead of the Metropolis algorithm, is used to calculate the transition probability. If the Metropolis algorithm is applied to replica permutation, the transition probability $P(X_\mu \to X_\nu)$ becomes lower [3]. In other words, the rejection ratio becomes large. It should be noted that the heat bath algorithm is also applicable. This method has been proposed by Plattner *et al.* as an infinite swapping method [6, 7]. The Suwa–Todo algorithm minimizes the rejection ratio. Therefore, it is suitable for replica permutation in comparison with the Metropolis and heat bath algorithms. The amount of stochastic flow $v(X_\mu \to X_\nu)$ is determined by only replacing the weight w_i in Eq. (2.10) with $w_R(X_\mu)$ as follows:

$$v(X_\mu \to X_\nu) = \max[0, \min[\Delta_{\mu\nu}, w_R(X_\mu) + w_R(X_\nu) - \Delta_{\mu\nu}, w_R(X_\mu), w_R(X_\nu)]], \tag{2.32}$$

where

$$\Delta_{\mu\nu} \equiv S_\mu - S_{\nu-1} + w_{\mathrm{R}}(X_1), \tag{2.33}$$

$$S_\mu \equiv \sum_{\nu=1}^{\mu} w_{\mathrm{R}}(X_\nu), \tag{2.34}$$

$$S_0 \equiv S_{M!}. \tag{2.35}$$

If $w_{\mathrm{R}}(X_\gamma)$ is the maximum weight, Eqs. (2.33) and (2.34) are modified as follows:

$$\Delta_{\mu\nu} \equiv S_\mu - S_{\nu-1} + w_{\mathrm{R}}(X_\gamma), \tag{2.36}$$

$$S_\mu \equiv \begin{cases} \displaystyle\sum_{\nu=\gamma}^{\mu} w_{\mathrm{R}}(X_\nu), & \text{for } \mu \geq \gamma \\ \displaystyle\sum_{\nu=\gamma}^{M!} w_{\mathrm{R}}(X_\nu) + \sum_{\nu=1}^{\mu} w_{\mathrm{R}}(X_\nu), & \text{for } \mu < \gamma. \end{cases} \tag{2.37}$$

The transition probability, $P(X_\mu \to X_\nu) = v(X_\mu \to X_\nu)/w(X_\mu)$, for $\nu = 1, \ldots, M!$ is obtained using Eqs. (2.26) and (2.32). Consequently, the replica permutation from X_μ to X_ν is performed following the obtained transition probability.

To perform the RPM, the permutation functions should be prepared in advance. The author defines a permutation $P_i^{(M)}$ for M elements (replicas) is defined as follows:

$$\begin{aligned} P_i^{(M)} &\equiv \left[\ \sigma_i(1),\ \ \sigma_i(2),\ \ \sigma_i(3),\ \ \cdots,\ \ \sigma_i(M)\ \right] \\ &= \begin{pmatrix} 1, & 2, & 3, & \cdots, & M \\ \sigma_i(1), & \sigma_i(2), & \sigma_i(3), & \cdots, & \sigma_i(M) \end{pmatrix}. \end{aligned} \tag{2.38}$$

The author also defines a set of all permutations as follows:

$$
\{P^{(M)}\} \equiv
\left\{
\begin{array}{c}
P_1^{(M)} \\
P_2^{(M)} \\
P_3^{[M]} \\
\vdots \\
P_{M!}^{(M)}
\end{array}
\right\}
=
\left\{
\begin{array}{c}
\left[\begin{array}{ccccc} 1, & 2, & 3, & \cdots, & M \end{array}\right] \\
\left[\begin{array}{ccccc} 2, & 1, & 3, & \cdots, & M \end{array}\right] \\
\left[\begin{array}{ccccc} 1, & 3, & 2, & \cdots, & M \end{array}\right] \\
\vdots \\
\left[\begin{array}{ccccc} M, & M-1, & M-2, & \cdots, & 1 \end{array}\right]
\end{array}
\right\},
\qquad (2.39)
$$

where $\{P^{(M)}\}$ is an $M! \times M$ matrix. The matrix $\{P^{(M)}\}$ can be constructed based on $\{P^{[M-1]}\}$; the first step is to add M at the M-th column of $\{P^{[M-1]}\}$:

$$
\begin{aligned}
\{P^{(M)}\}_M &\equiv \left\{ [P_i^{(M-1)},\ M]\colon \text{for } i = 1, \cdots, (M-1)! \right\} \\[4pt]
&=
\left\{
\begin{array}{c}
\left[\begin{array}{cccccc} 1, & 2, & 3, & \cdots, & M-1, & M \end{array}\right] \\
\left[\begin{array}{cccccc} 2, & 1, & 3, & \cdots, & M-1, & M \end{array}\right] \\
\left[\begin{array}{cccccc} 1, & 3, & 2, & \cdots, & M-1, & M \end{array}\right] \\
\vdots \\
\left[\begin{array}{cccccc} M-1, & M-2, & M-3, & \cdots, & 1, & M \end{array}\right]
\end{array}
\right\},
\qquad (2.40)
\end{aligned}
$$

where the subscript i of the braces indicates that $\{P^{(M)}\}_i$ is a set of permutations that added number M is at i-th column. Here, a transposition $\tau_{j,k}$ that is an operator that exchanges j-th and k-th columns of $P_i^{(M)}$ is introduced as follows:

$$
\tau_{j,k} \left[\begin{array}{ccccccc} 1, & \cdots, & j, & \cdots, & k, & \cdots, & M \end{array}\right] = \left[\begin{array}{ccccccc} 1, & \cdots, & k, & \cdots, & j, & \cdots, & M \end{array}\right].
$$
$$
(2.41)
$$

The next step is to repeat the adjacent transposition operation on $\{P^{(M)}\}_M$ until

the number M in $\{P^{(M)}\}_M$ moves to the first column. That is,

$$\{P^{(M)}\}_{M-1} = \tau_{M-1,M} \{P^{(M)}\}_M,$$

$$\{P^{(M)}\}_{M-2} = \tau_{M-2,M-1} \{P^{(M)}\}_{M-1},$$

$$\vdots$$

$$\{P^{(M)}\}_2 = \tau_{2,3} \{P^{(M)}\}_3,$$

$$\{P^{(M)}\}_1 = \tau_{1,2} \{P^{(M)}\}_2.$$

As a result, the set of all permutations for M elements is obtained as follows:

$$\{P^{(M)}\} = \left\{\{P^{(M)}\}_i : \text{for } i = M, M-1, \cdots, 1\right\}. \tag{2.42}$$

References

[1] B. A. Berg. *Markov chain Monte Carlo simulations and their statistical analysis.* World Scientific, 2004.

[2] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[3] Satoru G Itoh and Hisashi Okumura. Replica-Permutation Method with the Suwa–Todo Algorithm beyond the Replica-Exchange Method. *Journal of Chemical Theory and Computation*, 9(1):570–581, 2013.

[4] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[5] W K Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97, 1970.

[6] Nuria Plattner, J D Doll, Paul Dupuis, Hui Wang, Yufei Liu, and J E Gubernatis. An infinite swapping approach to the rare-event sampling problem. *The Journal of Chemical Physics*, 135(13):134111, 2011.

[7] Nuria Plattner, J D Doll, and Markus Meuwly. Overcoming the Rare Event Sampling Problem in Biological Systems with Infinite Swapping. *Journal of Chemical Theory and Computation*, 9(9):4215–4224, 2013.

Chapter 3

Development of Isothermal–Isobaric Replica-Permutation Method

3.1 Introduction

Molecular dynamics simulation is a powerful tool to study a biomolecule because it provides us with an atomic picture of the biomolecule. However, sometimes its conformation gets trapped in a local-minimum free energy state, which hampers the efficient simulation. To overcome this problem, generalized-ensemble algorithms have been developed [1].

Multicanonical algorithm [2–5] and simulated tempering [6, 7] are the most popular generalized-ensemble algorithms. These algorithms are based on non-Boltzmann weight factors and realize a random work in the energy spaces. However, the non-Boltzmann weight factors are not known *a priori* and have to be determined in advance by iterative procedures. The replica-exchange method (REM) [8, 9] is another successful algorithm. The REM uses some non-interacting copies (replicas) of a target system. Different temperatures are assigned to the replicas. The canonical molecular dynamics simulations are performed independently and simultaneously. By exchanging the temperatures between the replicas, random walks of the replicas in the temperature (or energy) space are realized. Thus, the simulation can escape from the local-minimum free-energy states. In the REM, the product of the Boltzmann factors for all replicas is used as the probability weight factor. Therefore, we can carry out a replica-exchange simulation only by assigning the temperature for each replica without performing the iterative procedures.

A replica-permutation method (RPM) [10] has been proposed as an improved alternative to the REM. In the RPM, temperatures are permuted among more than two replicas. To allow this permutation, the Suwa–Todo algorithm [11], instead of

the Metropolis algorithm [12], is employed. The Suwa–Todo algorithm satisfies only the global balance condition without imposing the detailed balance condition so that the rejection ratio can be minimized. The Hamiltonian RPM [13–15] was also proposed, wherein the parameters introduced in the Hamiltonian were permuted. The Suwa–Todo algorithm was also used in the simulated tempering method [16]. However, these previous RPM and simulated tempering methods are applicable only in the canonical ensembles. That is, the volume cannot be changed, and pressure cannot be specified. Although the REM and simulated tempering method with the Metropolis algorithm were extended to the isothermal–isobaric ensemble [17, 18], the RPM has not been extended yet. Thus, the phenomena involving volume change cannot be treated so far in RPM. On the other hand, in almost all experiments, the pressure as well as temperature is controlled and proteins change their conformation via pressure changes [19]. In this regard, it is important to sample the conformations of biomolecules at various temperatures and pressures more efficiently.

In this study, the author extended the RPM to the isothermal–isobaric ensemble to propose a new generalized-ensemble algorithm for molecular dynamics and Monte Carlo simulations. This algorithm is referred to as an isothermal–isobaric RPM. Not only the temperatures but also the pressures can be permuted so that two-dimensional random walks in temperature and pressure spaces can be realized. This is the first example of a two-dimensional RPM. Through the isothermal–isobaric replica-permutation simulation, one can obtain accurate physical quantities at any temperature and pressure within a range of employed values. Thus, we can compare the simulation results with experimental results obtained under identical conditions.

The author considered the effects of employed MC algorithms on sampling efficiency. This study compares three MC algorithms in the replica-permutation trial: The Suwa–Todo [11], heat bath [20], and metropolis [12] algorithms. The replica-permutation trial with the heat bath algorithm is also known as the infinite swapping method (INS) [21, 22]. Both RPM and INS are known to improve the sampling efficiency in comparison with the REM. The difference between the RPM and INS is that RPM uses the Suwa–Todo algorithm in the replica-permutation trial, whereas INS uses the heat bath algorithm. The INS is originally developed in the canonical

ensemble, but it can be extended to the isothermal–isobaric ensemble in a similar manner with RPM. Suwa and Todo pointed out that the autocorrelation time by the Suwa–Todo algorithm is 2.7 times faster than that by the heat bath algorithm in the four-state Potts model [11]. However, the sampling efficiency of INS and RPM were not compared so far. This study compares the sampling efficiencies of these methods and reveals the most suitable MC algorithm of the replica-permutation trials

3.2 Theory

3.2.1 Isothermal–Isobaric Ensemble

The isothermal–isobaric ensemble is characterized by the number of particles N, pressure P, and temperature T. Here, let us consider N-particle system in a box with volume V. The probability distribution in the isothermal–isobaric ensemble P_{NPT} of potential energy U and volume V is given as follows:

$$P_{NPT}(U,\ V;\ T,\ P) = n(U,\ V)\exp[-\beta(U+PV)], \tag{3.1}$$

where $n(U,V)$ is the density of states and $\beta = (1/k_\mathrm{B}T)$ is the inverse temperature (k_B is the Boltzmann constant). The potential energy of the system is a function of the coordinates and volume: $U(r,\ V)$, where r is a set of coordinates defined as $r \equiv \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N\}$. The isothermal–isobaric molecular simulation uses the scaled coordinates [23, 24]. The scaled coordinates is defined as $s = V^{-1/3}r$ when an isotropic box is used. Accordingly, the weight factor is transformed as

$$\begin{aligned}
\exp[-\beta(U+PV)] &= \exp\{-\beta[U(s,\ V)+PV]\}V^N ds \\
&= \exp\{-\beta[U(s,\ V)+PV-Nk_\mathrm{B}T\log V]\}ds
\end{aligned} \tag{3.2}$$

3.2.2 Isothermal–Isobaric Replica-Permutation Method

Let us consider a system of N atoms and M non-interacting copies (or replicas) of the system. Each replica is supposed to follow the different thermodynamics

ensemble specified by the thermodynamics state, λ_m, where m is a thermodynamics state index. In the isothermal–isobaric ensemble, the thermodynamics state is characterized by temperature, T_m, and pressure, P_m: $\lambda_m = [T_m, \ P_m]$.

The isothermal–isobaric RPM consider M_0 different temperatures, T_{m_0} ($m_0 = 1, \ldots, M_0$), and M_1 different pressures, P_{m_1} ($m_1 = 1, \ldots, M_1$). Hence, the total number of replicas is $M = M_0 \times M_1$. A parameter permutation among all M replicas is expressed as follows:

$$X_\mu = \left[x^{[1]}_{m_{(1)}}, \ \ldots, \ x^{[M]}_{m_{(M)}} \right] \to X_\nu = \left[x^{[1]}_{n_{(1)}}, \ \ldots, \ x^{[M]}_{n_{(M)}} \right], \tag{3.3}$$

where $m(i)$ and $n(i)$ are permutation functions from the replica index to the thermodynamics state index. The parameter permutation can be divided into temperature and pressure permutations as follows:

$$\begin{aligned} X_\mu &= \left[x^{[1]}_{\{m_0(1), \ m_1(1)\}}, \ \ldots, \ x^{[M]}_{\{m_0(M), \ m_1(M)\}} \right] \\ &\to X_\nu = \left[x^{[1]}_{\{n_0(1), \ n_1(1)\}}, \ \ldots, \ x^{[M]}_{\{n_0(M), \ n_1(M)\}} \right], \end{aligned} \tag{3.4}$$

where $m_0(i)$ and $n_0(i)$ are permutation functions from the replica index to the temperature index, and $m_1(i)$ and $n_1(i)$ are permutation functions from the replica index to the pressure index. Index $m_0(i)$ takes a value between 1 and M_0, and $m_1(i)$ takes a value between 1 and M_1. One-to-one correspondence exists between the replicas and sets of temperatures and pressures. In the isothermal–isobaric RPM, a weight factor of the state X_μ is

$$w_{\mathrm{R}}(X_\mu) = \prod_{i=1}^{M} \exp\left[-\beta_{m_0(i)} \left(H(q^{(i)}, \ p^{(i)}) + P_{m_1(i)} V^{(i)} \right) \right]. \tag{3.5}$$

The momenta of the replicas are scaled as

$$p^{[i]}_{\{n_0(i), \ n_1(i)\}} = \sqrt{\frac{T_{n_0(i)}}{T_{m_0(i)}}} \, p^{[i]}_{\{m_0(i), \ m_1(i)\}} \tag{3.6}$$

when the replica-permutation trial is accepted. This rescale of momenta cancels out kinetic terms in Eq. (3.5) when calculating the transition probability. Thus, a

weight factor without the kinetic energy

$$w_{\mathrm{R}}(X_\mu) = \prod_{i=1}^{M} \exp\left[-\beta_{m_0(i)}(U(q^{(i)}) + P_{m_1(i)}V^{(i)})\right] \tag{3.7}$$

is usually used instead of Eq. (3.5).

The isothermal–isobaric replica-permutation simulation is performed by the following procedures:

Step. 1 The index μ ($\mu = 1, \ldots, M!$) is assigned to all combinations between the replicas and sets of temperatures and pressures.

Step. 2 For all replicas, isothermal–isobaric molecular dynamics or Monte Carlo simulations are performed at the assigned temperature and pressure simultaneously and independently over certain steps.

Step. 3 A replica-permutation trial is performed by the Suwa–Todo algorithm. First, calculate the weight of each state by Eq. (3.7), and determine the maximum weight $w_{\mathrm{R}}(X_\gamma)$. Second, calculate the amount of stochastic flow $v(X_\mu \to X_\nu)$ in Eq. (2.32) and the transition probability $P(X_\mu \to X_\nu) = v(X_\mu \to X_\nu)/w_{\mathrm{R}}(X_\mu)$ for $\nu = 1, \ldots, M!$. Finally, transition from state X_μ to state X_ν is accepted stochastically with the probability $P(X_\mu \to X_\nu)$. If another algorithm is employed to calculate the transition probabilities, such as the Metropolis and heat bath algorithms, this step must modify following the procedures of each algorithm.

The replica-permutation molecular dynamics (RPMD) or Monte Carlo simulation in the isothermal–isobaric ensemble is performed by repeating step 2 and step 3.

Because the number of replicas tends to be large in the isothermal–isobaric RPM, considerable computational cost is incurred to calculate the stochastic flow. To reduce the computational cost, the author proposes that temperature permutation and pressure permutation are carried out alternately. Two subsets in which replicas

have the same number of temperatures or pressures are introduced as follows:

$$X_{\mu,\ [T_{m_0}]} = \begin{bmatrix} (T_{m_0},\ P_1) \\ (T_{m_0},\ P_2) \\ \cdots \\ (T_{m_0},\ P_{M_1}) \end{bmatrix},\ X_{\mu,\ [P_{m_1}]} = \begin{bmatrix} (T_1,\ P_{m_1}) \\ (T_2,\ P_{m_1}) \\ \cdots \\ (T_{M_0},\ P_{m_1}) \end{bmatrix}. \tag{3.8}$$

In this case, the weight factor of each subset is modified as follows:

(i) For a temperature permutation at each pressure P_{m_1} $(m_1 = 1,\ \ldots,\ M_1)$, the weight factor for state $X_{\mu,\ [P_{m_1}]}$ is given by

$$w_{\mathrm{R}}(X_{\mu,\ [P_{m_1}]}) = \prod_{i=1}^{M_0} \exp\left[-\beta_{m_0(i)}\left(U(q^{(i)}) + P_{m_1}V^{(i)}\right)\right]. \tag{3.9}$$

(ii) For a pressure permutation at each temperature T_{m_0} $(m_0 = 0,\ \ldots,\ M_0)$, the weight factor for state $X_{\mu,\ [T_{m_0}]}$ is given by

$$w_{\mathrm{R}}(X_{\mu,\ [T_{m_0}]}) = \prod_{i=1}^{M_1} \exp\left[-\beta_{m_0}\left(U(q^{(i)}) + P_{m_1(i)}V^{(i)}\right)\right]. \tag{3.10}$$

The term $\exp[-\beta_{m_0}U(q^{(i)})]$ is canceled out when calculating the transition probability because both $v(X_\mu \to X_\nu)$ and $w_{\mathrm{R}}(X_\mu)$ contain this term. Therefore, the weight factor becomes

$$w_{\mathrm{R}}(X_{\mu,\ [T_{m_0}]}) = \prod_{i=1}^{M_1} \exp\left[-\beta_{m_0}P_{m_1(i)}V^{(i)}\right] \tag{3.11}$$

when calculating the stochastic flow and transition probability.

If M_0 or M_1 is still large, replicas in the subsets can be divided into smaller subsets with the following two rules: (i) the number of replicas in one subset is the same. (ii) the temperature or pressure indices in one subset are sequential and

cyclic. The subset $X_{\mu,\ [P_{m_1}]}$ can be divided into

$$X^1_{\mu,\ [P_{m_1}]} = \begin{bmatrix} (T_1,\ P_{m_1}) \\ (T_2,\ P_{m_1}) \\ (T_3,\ P_{m_1}) \\ (T_4,\ P_{m_1}) \end{bmatrix}, \ X^3_{\mu,\ [P_{m_1}]} = \begin{bmatrix} (T_5,\ P_{m_1}) \\ (T_6,\ P_{m_1}) \\ (T_7,\ P_{m_1}) \\ (T_8,\ P_{m_1}) \end{bmatrix} \tag{3.12}$$

and

$$X^2_{\mu,\ [P_{m_1}]} = \begin{bmatrix} (T_3,\ P_{m_1}) \\ (T_4,\ P_{m_1}) \\ (T_5,\ P_{m_1}) \\ (T_6,\ P_{m_1}) \end{bmatrix}, \ X^4_{\mu,\ [P_{m_1}]} = \begin{bmatrix} (T_1,\ P_{m_1}) \\ (T_2,\ P_{m_1}) \\ (T_7,\ P_{m_1}) \\ (T_8,\ P_{m_1}) \end{bmatrix}. \tag{3.13}$$

Likewise, the subset $X_{\mu,\ [T_{m_0}]}$ can be divided as follows:

$$X^1_{\mu,\ [T_{m_0}]} = \begin{bmatrix} (T_{m_0},\ P_1) \\ (T_{m_0},\ P_2) \\ (T_{m_0},\ P_3) \\ (T_{m_0},\ P_4) \end{bmatrix}, \ X^3_{\mu,\ [T_{m_0}]} = \begin{bmatrix} (T_{m_0},\ P_5) \\ (T_{m_0},\ P_6) \\ (T_{m_0},\ P_7) \\ (T_{m_0},\ P_8) \end{bmatrix} \tag{3.14}$$

and

$$X^2_{\mu,\ [T_{m_0}]} = \begin{bmatrix} (T_{m_0},\ P_3) \\ (T_{m_0},\ P_4) \\ (T_{m_0},\ P_5) \\ (T_{m_0},\ P_6) \end{bmatrix}, \ X^4_{\mu,\ [T_{m_0}]} = \begin{bmatrix} (T_{m_0},\ P_1) \\ (T_{m_0},\ P_2) \\ (T_{m_0},\ P_7) \\ (T_{m_0},\ P_8) \end{bmatrix}. \tag{3.15}$$

If the subset division mentioned above are employed, the procedure of the isothermal–isobaric RPM in step 3 is slightly modified. One of the following procedures is carried out alternately instead of step 3:

Step. 3-1 Temperature permutations for $X^1_{\mu,\ [P_{m_1}]}$ and $X^3_{\mu,\ [P_{m_1}]}$ are performed at each pressure P_{m_1} ($m_1 = 1, \ldots, M_1$).

Step. 3-2 Pressure permutations for $X^1_{\mu,[\ T_{m_0}]}$ and $X^3_{\mu,\ [T_{m_0}]}$ are performed at each temperature T_{m_0} ($m_0 = 1,\ \ldots,\ M_0$).

Step. 3-3 Temperature permutations for $X^2_{\mu,\ [P_{m_1}]}$ and $X^4_{\mu,\ [P_{m_1}]}$ are performed at

39

each pressure P_{m_1} ($m_1 = 1, \ldots, M_1$).

Step. 3-4 Pressure permutations for $X^2_{\mu, [T_{m_0}]}$ and $X^4_{\mu, [T_{m_0}]}$ are performed at each temperature T_{m_0} ($m_0 = 1, \ldots, M_0$).

3.2.3 Reweighting Techniques: Estimation of Physical Quantities

After carrying out the isothermal–isobaric replica-permutation simulation, we can obtain the trajectory data on the multiple equilibrium states (M states, which are a combination of $T = T_1, \ldots, T_{M_0}$ and $P = P_1, \ldots, P_{M_1}$). The arithmetic mean of the trajectory data from the one equilibrium state gives the isothermal–isobaric average of a physical quantity, A, at the specified temperature and pressure values for the simulation, which is as follows:

$$\langle A \rangle_{T_{m_0}, P_{m_1}} = \frac{1}{n_m} \sum_{i=1}^{n_m} A(x_m(i)), \tag{3.16}$$

where m indicates the thermodynamics state at $T = T_{m_0}$ and $P = P_{m_1}$, n_m is the number of trajectory data, and $x_m(i)$ is the state of the system.

A weighted histogram analysis method (WHAM) [25, 26], which is also called multiple-histogram reweighting technique [27], and a multistate Bennett acceptance ratio (MBAR) method [28] are other choices to calculate the isothermal–isobaric averages of the physical quantities. These methods enable us to evaluate the physical quantities more precisely because they considers all trajectory data from multiple equilibrium states. The WHAM gives the density of states, whereas the MBAR cannot directly calculate it. However, the histogram construction of the WHAM makes bias due to the binning. In adition, the binning is a time-consuming process. The MBAR does not require making the histogram; thus, the MBAR provides the weight for each snapshot and the direct estimation of statistical errors.

Weighted Histogram Analysis Method

The statistical average of a physical quantity, A, at temperature T and pressure P is given as follows:

$$\langle A \rangle_{T,P} = \frac{\int dU \int dV \; \omega(U, \, V) A(U, V) \exp[-\beta(U + PV)]}{\int dU \int dV \; \omega(U, \, V) \exp[-\beta(U + PV)]} \tag{3.17}$$

where $\omega(U, \, V)$ is the density of states in the isothermal–isobaric ensemble. The WHAM provides the procedure to calculate the density of states $\omega(U, \, V)$ and dimensionless Gibbs free energy $g_m = g_{ij} \equiv g(T_{m_0}, \, P_{m_1})$ as follows:

$$\omega(U, V) = \frac{\sum_{i=1}^{M_0} \sum_{j=1}^{M_1} (1 + 2\tau_{ij})^{-1} h_{ij}(U, V)}{\sum_{i=1}^{M_0} \sum_{j=1}^{M_1} (1 + 2\tau_{ij})^{-1} n_{ij} \exp[g_{ij} - \beta_i(U + P_j V)]}, \tag{3.18}$$

$$\exp(-g_{ij}) = \sum_U \sum_V \omega(U, V) \exp[-\beta_i(U + P_j V)], \tag{3.19}$$

where h_{ij} is the histogram of U and V, n_{ij} is the number of the data, and τ_{ij} is the integrated autocorrelation time in the simulation at T_i and P_j. In many cases for biomolecule systems, τ_{ij} may be set to a constant value [25]. By solving self-consistent Eqs. (3.18) and (3.19), we can obtain the density of states.

Multistate Bennett Acceptance Ratio

In the MBAR method, the statistical average of a physical quantity A at thermodynamics state α is calculated as follows.

$$\langle A \rangle_\alpha = \frac{\sum_{j=1}^{M} \sum_{k=1}^{N_j} A(x_{jk}) W_{jk}}{\sum_{j=1}^{M} \sum_{k=1}^{N_j} W_{jk}} = \sum_{j=1}^{M} \sum_{k=1}^{N_j} w_{jk} A(x_{jk}) \tag{3.20}$$

where N_j is the number of sampling data obtained from each replica j. Index k indicates the k-th data of replica j. w_{jk} is the weight factor for each data in the required ensemble, and W_{jk} in Eq. (3.20) is calculated as

$$W_{jk} = \frac{\exp[-u_\alpha(x_{jk})]}{\sum_{l=1}^{M} N_l \exp[g_m - u_l(x_{jk})]}. \tag{3.21}$$

In the isothermal–isobaric ensemble, u_α is expressed as

$$u_\alpha(x_{jk}) = \beta_\alpha[U(x_{jk}) + P_\alpha V(x_{jk})]. \tag{3.22}$$

The dimensionless Gibbs free energy g_m at temperature $T_{m_0(l)}$ and pressure $P_{m_1(l)}$ can be calculated by solving the following equation iteratively:

$$g_{m'} = -\ln \sum_{j=1}^{M} \sum_{k=1}^{N_j} \frac{\exp[-u_{m'}(x_{jk})]}{\sum_{m=1}^{M} N_l \exp[g_m - u_l(x_{jk})]}. \tag{3.23}$$

The probability for the appropriate reaction coordinate ξ, that is divided into n bins, is given by

$$P_i(T, \ P; \ \xi) = \sum_{j=1}^{M} \sum_{k=1}^{N_j} w_{jk} \chi_i(x_{jk}; \ \xi), \tag{3.24}$$

where χ_i is an indicator function that assumes the value of one if the system is in bin i and zero otherwise. The potential of mean force can then be computed from $P_i(T, P; \xi)$ up to an irrelevant additive constant as

$$F_i(T, \ P; \ \xi) = -k_B T \ln \left(\frac{P_i(T, \ P; \ \xi)}{\nu_i} \right), \tag{3.25}$$

where ν_i is the relative width of bin i.

3.3   Computational Details

In this study, the author performed three isothermal–isobaric RPMD simulations with (1) the Suwa–Todo algorithm, (2) the heat bath algorithm, and (3) the Metropolis algorithm. Note that the RPMD with the heat bath algorithm is the identical to the INS. Hereafter, the RPMD with the Metropolis algorithm will be referred to as M-RPMD, respectively. The isothermal–isobaric replica-exchange MD (REMD) simulation was also performed for comparison.

A chignolin [29] in explicit water was used as the test system. The amino acid sequence of the chignolin is GYDPETGTWG. The simulation system was consists of one chignolin molecule, 1,621 water molecules, and two sodium ions Na^+ as counter

Table 3.1: Parameter index m to indicate T and P indices of the replicas.

	P_1	P_2	P_3	P_4
T_1	1	2	3	4
T_2	5	6	7	8
T_3	9	10	11	12
T_4	13	14	15	16
T_5	17	18	19	20
T_6	21	22	23	24

ions. A fully extended chignolin structure was used as the initial structure. A cubic unit cell with periodic boundary conditions was used. The simulations were performed using the Generalized-Ensemble Molecular Biophysics (GEMB) program, which was developed by Prof. Hisashi Okumura and the author and has been applied to several biomolecules [30–32]. The AMBER parm14SB [33] force field was used for chignolin, and the TIP3P [34] rigid-body model was used for the water molecules. The temperature was controlled by the Nosé-Hoover thermostat [35–37] and the pressure was controlled by the Andersen barostat [24]. The electrostatic potential energy was calculated by the particle mesh Ewald method [38, 39]. The cutoff distance was $r_c = 12.0$ Å for both electrostatic and Lennard-Jones potential. The multiple time-step method [40] was also applied. The time step was taken as $\Delta t = 0.5$ fs for the bonded interactions of the protein atoms; $\Delta t = 2.0$ fs for the non-bonded interactions of the protein atoms and those between the protein atoms and solvent molecules; and $\Delta t = 4.0$ fs for the interaction between the solvent molecules. Because the symplectic rigid-body algorithm was used for the water molecules, Δt can be taken as 4.0 fs [41–43]. The following six temperature and four pressure values were used: 300.0, 307.6, 315.4, 323.4, 331.6, and 340.0 K for temperature; 0.1, 30, 60, and 100 MPa for pressure. Therefore, the number of replicas is 24. A new parameter index m was introduced to indicate both the temperature and pressure indices, as listed in Table 3.1.

Because the sampling efficiency may depend on the number of subsets, the author tested six different settings for the replicas to divide into subsets. The numbers of temperature and pressure subsets for each pressure and temperature value are as follows: (T subsets, P subsets) = (1,1), (1,4), (4,1), (4,4), (6,1), and (6,1). That is,

each subset has the following numbers of replicas in the temperature and pressure spaces: (T replicas, P replica) = (6,4), (6,2), (3,4), (3,2), (2,4), and (2,2). These simulations will be referred to as RPMD1, RPMD2, RPMD3, RPMD4, RPMD5, and REMD. Note that dividing into three temperature subsets and two pressure subsets, i.e., (T subsets, P subsets) = (6,4), is the identical to the REMD because each subset has only two replicas in the temperature and pressure spaces: (T replicas, P replicas) = (2,2). The number of replicas in the subsets of INS and M-RPMD are same as RPMD1, i.e., (T subset, P subset)=(1,1) and (T replicas, P replicas) = (6,4).

Trials of replica permutation or exchanges were performed every 0.5 ps. Here, temperature permutations/exchanges at each pressure $P_{m_1} = P_1$, ..., P_4 were performed simultaneously at odd-numbered trials, and pressure permutations/exchanges at each temperature $T_{m_0} = T_1$, ..., T_6 were performed simultaneously at even-numbered trials. The trajectory data were stored every 0.5 ps. Above simulations were performed for 30 ns per replica.

3.4 Results and Discussion

3.4.1 Comparison of Sampling Efficiencies among RPM, INS and REM

Figure 3.1 show the time series of temperature and pressure indices of replica 1 of the RPMD1, REMD, INS, and M-RPMD. In the RPMD1, REMD, and INS, each replica took all temperature and pressure values during the simulation. Because of the random walks in the temperature-pressure space, the simulation can escape from local-minimum free-energy state. However, the parameter value does not frequently transit in the M-RPMD simulation.

The number of tunneling events in the temperature and pressure spaces is listed in Table 3.2 to quantitatively compare sampling efficiency among the RPMD, REMD, INS, and M-RPMD. Here, one tunneling event is defined as a round trip between the highest and lowest values in the temperature or pressure space. Results are listed in
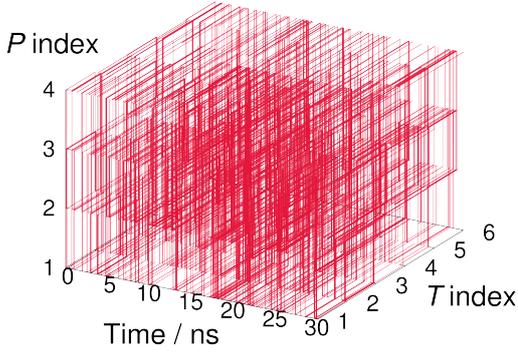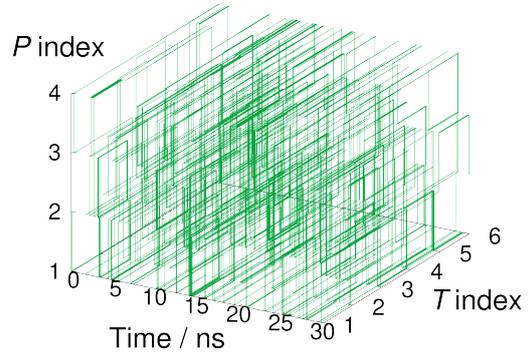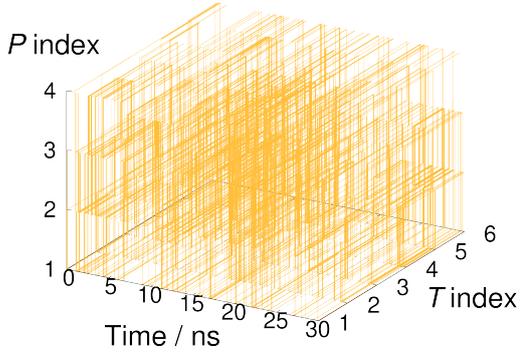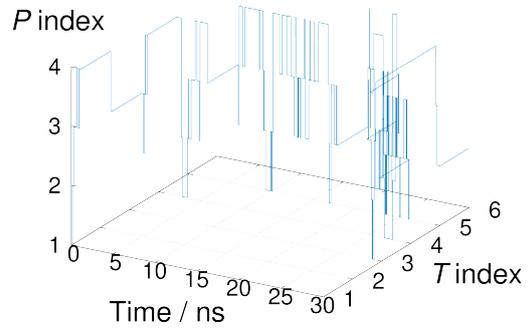
Figure 3.1: Time series of the temperature and pressure indices of replica 1 in the (a) RPMD1, (b) REMD, (c) INS, and (d) M-RPMD simulations.

Table 3.2: The number of tunneling events in the temperature and pressure spaces.

Method	Temperature space	Pressure space
RPMD1	64 ± 1	146 ± 2
RPMD2	65 ± 1	37 ± 1
RPMD3	41 ± 1	151 ± 2
RPMD4	43 ± 1	38 ± 1
RPMD5	33 ± 1	148 ± 1
REMD	35 ± 1	38 ± 1
INS	50 ± 1	113 ± 2
M-RPMD	0 ± 0	8 ± 0.4

Table 3.2. The error bars were calculated by the bootstrap method [44]. Tunneling events in the RPMD1 are larger than those in the INS, REMD, and M-RPMD both in the temperature and pressure spaces.

Figure 3.2 presents the transition ratio of the parameters to see parameter transition. In this figure, the longitudinal axis is the parameter index before transition,
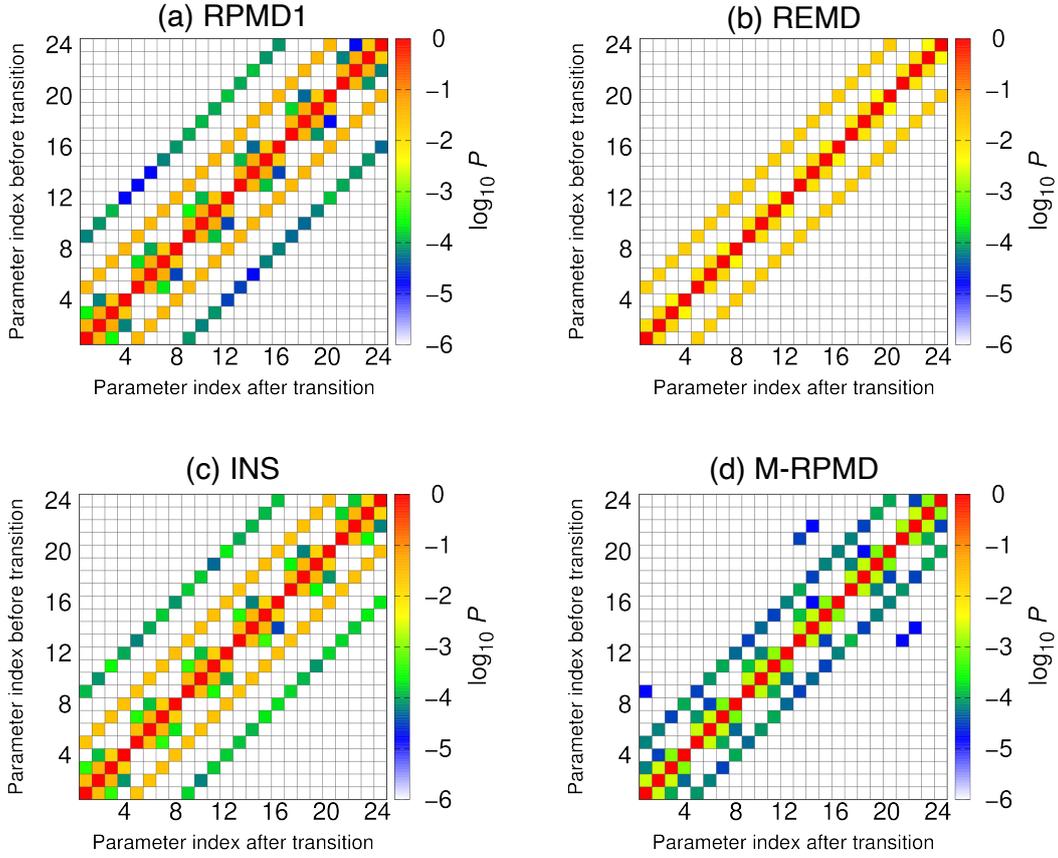
Figure 3.2: Transition of the parameter index of the (a) RPMD1, (b) REMD, (c) INS, and (d) M-RPMD simulations. The longitudinal axis is the parameter index before transition, and the horizontal axis is the parameter index after transition. Each cell shows the transition ratio in a logarithmic scale.

and the horizontal axis is the parameter index after transition. Let us focus on parameter index $m = 10$ as an example. In the REMD, the parameter index transits from 10 to 6, 9, 10, 11, and 14. This means that the temperature and pressure indices changed to only the neighboring values (Table 3.2). On the other hand, in the RPMD1, INS, the parameter index transits from 10 to 2, 6, 8, 9, 10, 11, 12, 14, and 18. This means that the temperature and pressure indices can transit not only to the neighboring values but also the non-neighboring values (Table 3.2). In the M-RPMD, the parameter index is changed from 10 to 6, 9, 10, 11, 12, and 14. The parameter transition to the non-neighboring values occur in the pressure space, but it does not occur in the temperature spaces.

Figure 3.3 shows the transition ratio to other indices in the temperature and

pressure spaces. Here, the transition ratio is defined as follows:

$$\text{Transition ratio of index } i = \frac{\text{Total transitions from index } i \text{ to other indices}}{\text{Total replica-permutation trials}}.$$

The transition ratio of RPMD1 is clearly higher than the ratios of INS, REMD, and M-RPMD. These results clearly indicate that the isothermal–isobaric RPMD realizes more efficient simulation than the isothermal–isobaric INS, REMD, and M-RPMD. The replica-permutation simulation with the Suwa–Todo algorithm realize higher transition ration than that with the heat bath algorithm. One of the reasons is that the Suwa–Todo algorithm minimizes the rejection ratio (in many cases, it is rejection free) as Eq. (2.14) indicates, whereas the heat bath algorithm has a finite rejection ratio. The author also found that the Metropolis algorithm does not work well in the replica-permutation simulation. The transition probability of the Metropolis algorithmis given in Eq. (2.7). In Eq. (2.7), $\min(1, w_i/w_j)$ takes a value from 0 to 1. The coefficient $1/(n-1)$ becomes small as the number of states, n, increases. The replica-permutation trial considers $M!$ states; and consequently, the transition probability from state i to state j is in proportion to $1/(M!-1)$. Therefore the transition probabilities become significantly small as the number of replicas increases.

3.4.2   Effects of Subset Division on the Sampling Efficiency

Let us compared the sampling efficiency of the isothermal–isobaric RPM among different numbers of subsets. The number of tunneling events are listed in Table 3.2. In the temperature space, as the number of replicas in one subset increases, the number of tunneling events in temperature also increases. Similarly, in the pressure space, as the number of replicas in one subset increases, the number of tunneling events in pressure also increases. Fig. 3.4 shows the transition ratio to other indices. Here, the transition ratio at different pressure indices is shown separately. As the number of replicas in one subset increases, the transition ratio also becomes large. These results indicate that the RPM becomes more efficient with a large number of replicas in one subset.
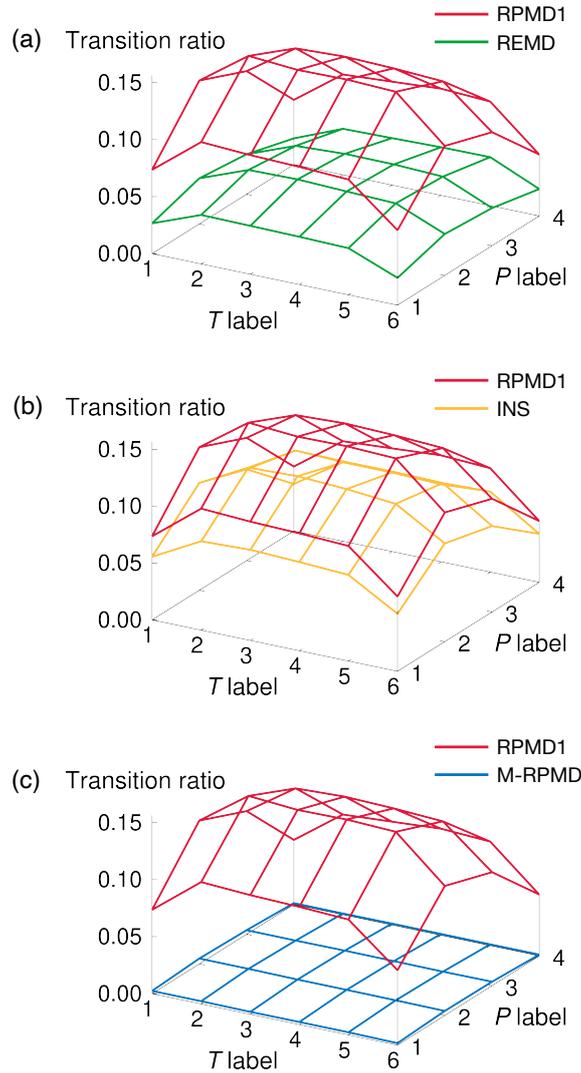
Figure 3.3: Comparison of the transition ratio between (a) RPMD1 and REMD, (b) RPMD1 and INS, and (c) RPMD1 and M-RPMD.

To explain these results, the author considers the permutation of four replicas, as shown in Fig. 3.5 (a). The number of the candidates of the next state is $4! = 24$. Let us focus on whether replica 1 transits to another index or not. The number of candidate states that replica 1 does not transit to another index is six (green round rectangle). Therefore, the proportion of the rejection candidates in all candidates is $6/24 = 1/4$. Likewise, let us consider the permutation (exchange) of two replicas as shown in Fig. 3.5 (b). The number of the candidates of the next state is $2! = 2$ and the number of candidate states that replica 1 does not transit to another index is 1. Therefore, the proportion of the rejected candidates for replica 1 among all
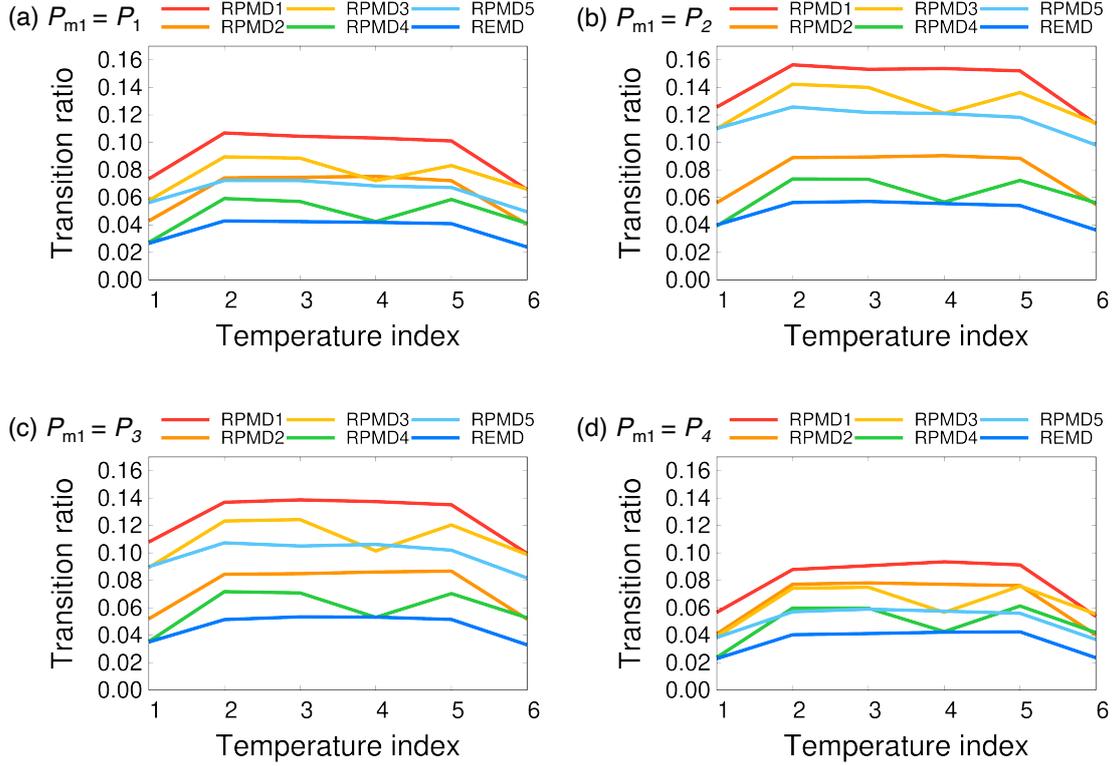
Figure 3.4: Comparison of the transition ratio among the replica-permutation simulations with a different number of subsets.

candidates is $1/2$. This simple example show that if the number of replicas in one subset increases, then the proportion of rejected candidates for some replica 1 in all candidates decreases. More generally, the author considers N replicas and $2n$ subsets. The number of replicas in one subset is defined as $M = N/n$. Here, N, n, and M are natural numbers. Let us focus on whether replica i transits to another index. The number of total candidates of the next state is $M!$. The number of rejected candidates for replica i is $(M-1)!$ because this is the permutation of $(M-1)$ replicas, except for replica i in one subset. Thus, the proportion of the candidates that replica i does not transit to any index in all candidates is $(M-1)!/M! = 1/M$. Therefore, if the number of replicas in one subset increases, the proportion of the rejection candidates for replica i in all candidates decreases in proportion to $1/M$. Therefore, more efficient simulation can be realized when one subset includes more numbers of replicas.

(a) Each subset has 4 replicas (RPMD)

State number μ_1 μ_2 μ_3 μ_4 μ_5 μ_6 μ_7 μ_8 ... μ_{22} μ_{23} μ_{24}

	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8		μ_{22}	μ_{23}	μ_{24}
Replica 1	T_1	T_1	T_1	T_1	T_1	T_1	T_2	T_2		T_4	T_4	T_4
Replica 2	T_2	T_2	T_3	T_3	T_4	T_4	T_1	T_1		T_2	T_3	T_3
Replica 3	T_3	T_4	T_2	T_4	T_2	T_3	T_3	T_4		T_3	T_1	T_2
Replica 4	T_4	T_3	T_4	T_2	T_3	T_2	T_4	T_3		T_1	T_2	T_1

Subset Current state Candidates of the next state
Candidates that replica 1 does not change temperature

(b) Each subset has 2 replicas (REMD)

State number μ_1 μ_2

	μ_1	μ_2
Replica 1	T_1	T_2
Replica 2	T_2	T_1
Replica 3	T_3	T_3
Replica 4	T_4	T_4

(c) Ratio of the number of candidates

$M = 2$ $M = 4$ $M = 8$

All candidates

Candidates that replica i does not change the parameter value
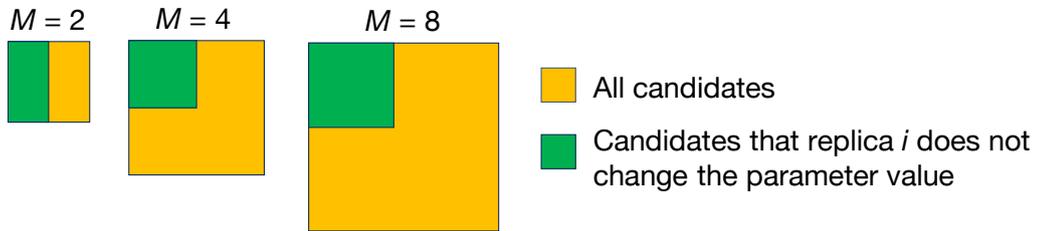
Figure 3.5: Schematic figure of the (a) replica-permutation and (b) replica-exchange candidates

3.5   Conclusions

In this study, the author developed the isothermal–isobaric RPM. The isothermal–isobaric RPMD performs temperature and pressure permutations among more than two replicas, which allows efficient sampling from a wide temperature and pressure range. The author compared the effects of employed MC algorithms in the parameter permutation trials on the sampling efficiency among the Suwa–Todo, heat bath, and Metropolis algorithms. The author revealed that the Suwa–Todo algorithm achieves the most efficient sampling. The author also pointed out that introducing subsets

effects the transition ratio of the parameters. In the RPM, one subset should include as many replicas as possible to realize more efficient sampling; however, it is a trade-off between sampling efficiency and the computational costs of the permutation trials.

References

[1] Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers - Peptide Science Section*, 60(2):96–123, 2001.

[2] Bernd A Berg and Thomas Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267(2):249–253, 1991.

[3] Bernd A. Berg and Thomas Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68(1):9–12, 1992.

[4] Ulrich H.E. Hansmann, Yuko Okamoto, and Frank Eisenmenger. Molecular dynamics, Langevin and hydrid Monte Carlo simulations in a multicanonical ensemble. *Chemical Physics Letters*, 259(3-4):321–330, 1996.

[5] Nobuyuki Nakajima, Haruki Nakamura, and Akinori Kidera. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *The Journal of Physical Chemistry B*, 101(5):817–824, 1997.

[6] Enzo Marinari and Giorgio Parisi. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)*, 19(6):451–458, 1992.

[7] A P Lyubartsev, A A Martsinovski, S V Shevkunov, and P N Vorontsov - Velyaminov. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics*, 96(3):1776–1783, 1992.

[8] Koji Hukushima and Koji Nemoto. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[9] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[10] Satoru G Itoh and Hisashi Okumura. Replica-Permutation Method with the Suwa–Todo Algorithm beyond the Replica-Exchange Method. *Journal of Chemical Theory and Computation*, 9(1):570–581, 2013.

[11] Hidemaro Suwa and Synge Todo. Markov Chain Monte Carlo Method without Detailed Balance. *Physical Review Letters*, 105(12):120603, 2010.

[12] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[13] Satoru G Itoh and Hisashi Okumura. Hamiltonian replica-permutation method and its applications to an alanine dipeptide and amyloid-β(29-42) peptides. *Journal of Computational Chemistry*, 34(29):2493–2497, 2013.

[14] Satoru G. Itoh and Hisashi Okumura. Dimerization process of amyloid-β(29-42) studied by the hamiltonian replica-permutation molecular dynamics simulations. *Journal of Physical Chemistry B*, 118(39):11428–11436, 2014.

[15] Satoru G Itoh and Hisashi Okumura. Oligomer Formation of Amyloid-β(29–42) from Its Monomers Using the Hamiltonian Replica-Permutation Molecular Dynamics Simulation. *The Journal of Physical Chemistry B*, 120(27):6555–6561, 2016.

[16] Y Mori and H Okumura. Simulated tempering based on global balance or detailed balance conditions: Suwa–Todo, heat bath, and Metropolis algorithms. *Journal of Computational Chemistry*, 36(31):2344–2349, 2015.

[17] Tsuneyasu Okabe, Masaaki Kawata, Yuko Okamoto, and Masuhiro Mikami. Replica-exchange Monte Carlo method for the isobaric–isothermal ensemble. *Chemical Physics Letters*, 335(5-6):435–439, 2001.

[18] Yoshiharu Mori and Yuko Okamoto. Generalized-Ensemble Algorithms for the Isobaric–Isothermal Ensemble. *Journal of the Physical Society of Japan*, 79(7):074003, 2010.

[19] Jerson L Silva, Andrea C. Oliveira, Tuane C R G Vieira, Guilherme A. P. de Oliveira, Marisa C Suarez, and Debora Foguel. High-Pressure Chemical Biology and Biotechnology. *Chemical Reviews*, 114(14):7239–7267, 2014.

[20] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

[21] Nuria Plattner, J D Doll, Paul Dupuis, Hui Wang, Yufei Liu, and J E Gubernatis. An infinite swapping approach to the rare-event sampling problem. *The Journal of Chemical Physics*, 135(13):134111, 2011.

[22] Nuria Plattner, J D Doll, and Markus Meuwly. Overcoming the Rare Event Sampling Problem in Biological Systems with Infinite Swapping. *Journal of Chemical Theory and Computation*, 9(9):4215–4224, 2013.

[23] I. R. McDonald. NpT-ensemble monte carlo calculations for binary liquid mixtures. *Molecular Physics*, 23(1):41–58, 1972.

[24] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[25] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[26] John D Chodera, William C Swope, Jed W Pitera, Chaok Seok, and Ken A Dill. Use of the Weighted Histogram Analysis Method for the Analysis of

Simulated and Parallel Tempering Simulations. *Journal of Chemical Theory and Computation*, 3(1):26–41, 2007.

[27] Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Physical Review Letters*, 63(12):1195–1198, 1989.

[28] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12):124105, 2008.

[29] Shinya Honda, Kazuhiko Yamasaki, Yoshito Sawada, and Hisayuki Morii. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure*, 12(8):1507–1518, 2004.

[30] Hisashi Okumura. Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins: Structure, Function, and Bioinformatics*, 80(10):2397–2416, 2012.

[31] Hisashi Okumura and Satoru G Itoh. Amyloid Fibril Disruption by Ultrasonic Cavitation: Nonequilibrium Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 136(30):10549–10552, 2014.

[32] Hisashi Okumura and Satoru G Itoh. Structural and fluctuational difference between two ends of Aβ amyloid fibril: MD simulations predict only one end has open conformations. *Scientific Reports*, 6(1):38422, 2016.

[33] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[34] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[35] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[36] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[37] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[38] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[39] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[40] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[41] Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[42] T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna. Symplectic quaternion scheme for biophysical molecular dynamics. *The Journal of Chemical Physics*, 116(20):8649–8659, 2002.

[43] Hisashi Okumura, Satoru G. Itoh, and Yuko Okamoto. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric–isothermal, and related ensembles. *The Journal of Chemical Physics*, 126(8):084103, 2007.

[44] B Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

Chapter 4

Development of Replica Sub-Permutation Method

4.1   Introduction

Molecular dynamics (MD) simulation is a powerful tool for gaining insight into the molecular properties of biomolecules. To study the properties of biomolecules through MD simulation, an efficient conformational search is required. Through recent advances in computational hardware and software, the time scale that we can observe by an MD simulation has been extended up to several hundred microseconds [1, 2]. However, obtaining sufficient conformations from the phase space remains difficult because the conformation of the biomolecules gets trapped in a local-minimum free-energy state. Generalized-ensemble algorithms [3–10] have been developed to overcome the sampling problem. Among the generalized-ensemble algorithms, the replica-exchange method (REM) [11, 12] is one of the most widely used algorithm nowadays.

In the past decade, a number of ideas and algorithms have been applied to the REM to improve sampling efficiency. For example, velocity rescaling criteria [13] and temperature spacing [14–16] are examined to optimize acceptance probability. The exchange-attempt frequency has also been studied, suggesting that frequent exchange attempts achieve fast convergence of the distribution [17, 18]. The replica-exchange trial scheme and the method for calculating transition probability also affect the sampling efficiency. An all-pair exchange scheme [19] using a kinetic MC algorithm and the variations [20–22] of this scheme have been proposed to enhance parameter mixing. Instead of the Metropolis algorithm, the Gibbs sampler has been applied with the modified replica-exchange scheme that is called independence sampling updates [23].

A replica-permutation method (RPM) [24] has been proposed as an improved alternative to the REM. The RPM performs parameter permutation among more than

two replicas, whereas the REM performs parameter exchange between two replicas only. The Suwa–Todo algorithm [25] is used to calculate the transition probability, as it can minimize the rejection ratio of the MC trials. However, this algorithm is equivalent to the Metropolis algorithm when there are only two candidates. The parameter exchange between two replicas is exactly the case, and the REM cannot make the merit of the Suwa–Todo algorithm. To increase the number of candidates in the MC trials, parameter permutation between more than two replicas are introduced. Previous studies have compared sampling efficiency for the RPM using the Metropolis algorithm [26], heat bath algorithm [27] (known as an infinite swapping method [28, 29]), Suwa–Todo algorithm with the detailed balance condition (DBC) [30], and Suwa–Todo algorithm without the DBC [25]. The Suwa–Todo algorithm without the DBC has been shown to be the most suitable algorithm for replica permutation in terms of the sampling efficiency [24, 31, 32].

As the author has revealed in Chapter 3, it is desirable to consider all of the permutations of the parameters associated with the replicas. When the number of replicas is not particularly large, typically not exceeding ten, the RPM considers all possible combinations between replicas and parameters. However, as the number of replicas increases, a considerable computational cost is incurred upon calculating the transition probability. Therefore, in the majority of cases, it is impractical to consider all possible combinations. To reduce the number of combinations, the replicas are divided into several subsets in the RPM. However, employing subsets is not desirable because it induces a zigzag distribution of the transition ratio, and the sampling efficiency decreases [32].

In this study, the author propose a replica sub-permutation method (RSPM) for MD and MC simulations. To reduce the number of combinations between replicas and parameters without the loss of sampling efficiency, the RSPM introduces a new permutation algorithm referred to as a sub-permutation. In the RSPM, parameter permutation is performed based on the sub-permutation. The author has applied the RSPM to an artificially designed protein, chignolin [33], in explicit water. The chignolin comprises ten amino acid residues (GYDPETGTWG) and folds into a β-hairpin structure. Chignolin has been thoroughly studied from a theoret-

ical standpoint [34–38] and is therefore suitable as a test system. For comparison, the RPM and REM were also used. The results demonstrate that among the three methods, the RSPM is the most efficient.

4.2 Theory

4.2.1 Replica Sub-Permutation Method

Let us consider a system of N atoms and M non-interacting copies (or replicas) of the system. As in the RPM, the RSPM considers performing replica permutation among all M parameters

$$X_\mu = \left[x_1^{[i(1)]}, \ldots, x_M^{[i(M)]} \right] \to X_\nu = \left[x_1^{[j(1)]}, \ldots, x_M^{[j(M)]} \right], \tag{4.1}$$

or parameter permutation among all M replicas

$$X_\mu = \left[x_{m(1)}^{[1]}, \ldots, x_{m(M)}^{[M]} \right] \to X_\nu = \left[x_{n(1)}^{[1]}, \ldots, x_{n(M)}^{[M]} \right], \tag{4.2}$$

where $x_m^{[i]}$ is the state of replicas, i and j are permutation functions from the thermodynamics state to the replica index, and m and n are those from the replica index to the thermodynamics state.

The RSPM introduces a new permutation algorithm to reduce the number of candidates for the next state X_ν. The idea of the RSPM is to use permutations containing transitions only to neighboring parameter values. Here, let us consider temperature permutation among four replicas as an example. In the RPM, all possible candidates between replicas and parameters are considered; thus, there are 24 ($= 4!$) permutation candidates as shown in Fig. 4.1 (a). On the other hand, in the RSPM, only permutations satisfying the following transitions are considered: (i) the replica at temperature T_1 transits to T_1 or T_2; (ii) the replica at temperature T_2 transits to T_1, T_2, or T_3; (iii) the replica at temperature T_3 transits to T_2, T_3, or T_4; (iv) the replica at temperature T_4 transits to T_3 or T_4. Permutations that satisfy the above four conditions from all 24 ($= 4!$) permutations are selected as candidates

of the next state X_ν. The author term this permutation "sub-permutation". The sub-permutation functions $S_i^{(M)}$ for M elements (replicas) system are defined as follows:

$$S_i^{(M)} \equiv \left[\begin{array}{ccccc} \sigma_i(1), & \sigma_i(2), & \sigma_i(3), & \cdots, & \sigma_i(M) \end{array} \right]$$
$$= \left(\begin{array}{ccccc} 1, & 2, & 3, & \cdots, & M \\ \sigma_i(1), & \sigma_i(2), & \sigma_i(3), & \cdots, & \sigma_i(M) \end{array} \right), \qquad (4.3)$$

where

$$\sigma_i(l) = \begin{cases} l-1, \ l, \ \text{or} \ l+1 & \text{when } 1 < l < M, \\ 1 \text{ or } 2 & \text{when } l = 1, \\ M-1 \text{ or } M & \text{when } l = M. \end{cases} \qquad (4.4)$$

The sub-permutation function is similar to permutation function in Eq. (2.38), but limitation on the values of $\sigma_i(l)$ is imposed. A set of sub-permutations is defined as follows:

$$\left\{ S^{(M)} \right\} \equiv \left\{ S_i^{(M)} : \text{for } i = 1, \cdots, N_{\{S^{(M)}\}} \right\}, \qquad (4.5)$$

where $N_{\{S^{(M)}\}}$ is the total number of sub-permutations and $\{S^{(M)}\}$ is an $N_{\{S^{(M)}\}} \times M$ matrix. Figure 4.1 (b) presents the sub-permutations for four replicas.

The sub-permutations $\{S^{(M)}\}$ is constructed based on $\{S^{(M-1)}\}$ and $\{S^{(M-2)}\}$ as follows:

$$\{S^{(M)}\} = \left\{ \begin{array}{ll} \left[S_i^{(M-1)}, \ M \right] & : \text{for } i = 1, \cdots, N_{\{S^{(M-1)}\}} \\ \left[S_i^{(M-2)}, \ M, \ M-1 \right] & : \text{for } j = 1, \cdots, N_{\{S^{(M-2)}\}} \end{array} \right\} \qquad (4.6)$$

where

$$\left\{ S^{(1)} \right\} = [1],$$
$$\left\{ S^{(2)} \right\} = \left\{ \left[\begin{array}{cc} 1, & 2 \\ 2, & 1 \end{array} \right] \right\}.$$

(a) Replica-permutation candidates for four replicas

	$P_1^{(4)}$	$P_2^{(4)}$	$P_3^{(4)}$	$P_4^{(4)}$	$P_5^{(4)}$	$P_6^{(4)}$	$P_7^{(4)}$	$P_8^{(4)}$	$\cdots$	$P_{22}^{(4)}$	$P_{23}^{(4)}$	$P_{24}^{(4)}$
Replica 1 at T_1	T_1	T_1	T_1	T_1	T_1	T_1	T_2	T_2		T_4	T_4	T_4
Replica 2 at T_2	T_2	T_2	T_3	T_3	T_4	T_4	T_1	T_1	$\cdots$	T_2	T_3	T_3
Replica 3 at T_3	T_3	T_4	T_2	T_4	T_2	T_3	T_3	T_4		T_3	T_1	T_2
Replica 4 at T_4	T_4	T_3	T_4	T_2	T_3	T_2	T_4	T_3		T_1	T_2	T_1

(b) Replica sub-permutation candidates for four replicas

	$S_1^{(4)}$	$S_2^{(4)}$	$S_3^{(4)}$	$\times$	$\times$	$\times$	$S_4^{(4)}$	$S_5^{(4)}$	$\cdots$	$\times$	$\times$	$\times$
Replica 1 at T_1	T_1	T_1	T_1	T_1	T_1	T_1	T_2	T_2		T_4	T_4	T_4
Replica 2 at T_2	T_2	T_2	T_3	T_3	T_4	T_4	T_1	T_1	$\cdots$	T_2	T_3	T_3
Replica 3 at T_3	T_3	T_4	T_2	T_4	T_2	T_3	T_3	T_4		T_3	T_1	T_2
Replica 4 at T_4	T_4	T_3	T_4	T_2	T_3	T_2	T_4	T_3		T_1	T_2	T_1

Figure 4.1: Schematic figure of replica-permutation candidates for the (a) replica permutation and (b) replica sub-permutation of a four-replica system.

In the case of $M = 3$, for example, the sub-permutations can be constructed as follows:

$$\left\{ [S_i^{(M-1)}, \ M]: \text{for } i = 1, \cdots, N_{\{S^{(M-1)}\}} \right\} = \left\{ \begin{bmatrix} 1, & 2, & 3 \\ 2, & 1, & 3 \end{bmatrix} \right\},$$

$$\left\{ [S_j^{(M-2)}, \ M, \ M-1]: \text{for } j = 1, \cdots, N_{\{S^{(M-2)}\}} \right\} = \left\{ \begin{bmatrix} 1, & 3, & 2 \end{bmatrix} \right\}.$$

In a similar manner, the sub-permutations in the case of $M = 4$ can be obtained as follows:

$$\left\{ [S_i^{(M-1)}, \ M]: \text{for } i = 1, \cdots, N_{\{S^{(M-1)}\}} \right\} = \left\{ \begin{bmatrix} 1, & 2, & 3, & 4 \\ 2, & 1, & 3, & 4 \\ 1, & 3, & 2, & 4 \end{bmatrix} \right\},$$

$$\left\{ [S_j^{(M-2)}, \ M, \ M-1]: \text{for } j = 1, \cdots, N_{\{S^{(M-2)}\}} \right\} = \left\{ \begin{bmatrix} 1, & 2, & 4, & 3 \\ 2, & 1, & 4, & 3 \end{bmatrix} \right\}.$$

The number of sub-permutations $N_{\{S^{(M)}\}}$ clearly satisfies a following recurrence for-

mula:

$$N_{\{S^{(M)}\}} = N_{\{S^{(M-1)}\}} + N_{\{S^{(M-2)}\}} \tag{4.7}$$

where $N_{\{S^{(1)}\}} = 1$ and $N_{\{S^{(2)}\}} = 2$. By solving the recurrence formula, the number of sub-permutations for a system with M elements is obtained:

$$N_{\{S^{(M)}\}} = \frac{1}{\sqrt{5}} \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^{M+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{M+1} \right\}. \tag{4.8}$$

This is the Fibonacci sequence: $N_{\{S^{(M)}\}} = 1, 2, 3, 5, 8, 13, 21, \cdots$. The number of sub-permutations $N_{\{S^{(M)}\}}$ is smaller than that of all permutations $M!$, as illustrated in Fig. 4.2 and Table 4.1. It should be noted that the number of all permutations for eight and ten replicas in the original RPM ($8! = 40,320$ and $10! = 3,628,800$) are almost the same with those for 23 and 32 replicas in the RSPM ($N_{\{S^{(23)}\}} = 46,368$ and $N_{\{S^{(32)}\}} = 3,524,578$), respectively. While the original RPM can perform parameter permutations among up to eight to ten replicas, the RSPM can perform parameter permutations among up to 23 to 32 replicas. Thus, the replica sub-permutation method succeeds in extending the upper limit for the number of replicas that can be permutated at one permutation trial.

Although, so far, the author has only discussed the sub-permutations containing transitions from the parameter index l to between $l - 1$ and $l + 1$, the sub-permutations can be extended more generally. For example, it is possible to consider sub-permutations that contain transitions not only to the adjacent parameter indices, but also to those two or three parameter indices away. Here, ϵ is defined as a sub-permutation range. The replica at temperature T_m transits to a temperature between $T_{m-\epsilon}$ and $T_{m+\epsilon}$ when $\epsilon + 1 \le m \le M - \epsilon$. The replica at temperature T_m transits to a temperature between T_1 and $T_{m+\epsilon}$ when $m < \epsilon + 1$, and between $T_{m-\epsilon}$ and T_M when $m > M - \epsilon$.

Sub-permutation $\{S^{(M)}\}$ with a range of ϵ can be constructed based on $\{S^{(M-1)}\}$. The first step is to add M at the M-th column of the sub-permutation $\{S^{(M-1)}\}$: $\{S^{(M)}\}_M \equiv \{[S_i^{(M-1)}, M]: \text{for } i = 1, \cdots, N_{\{S^{(M-1)}\}}\}$, where subscript i of the braces indicates that $\{S^{(M)}\}_i$ is a set of sub-permutations that added number M is at
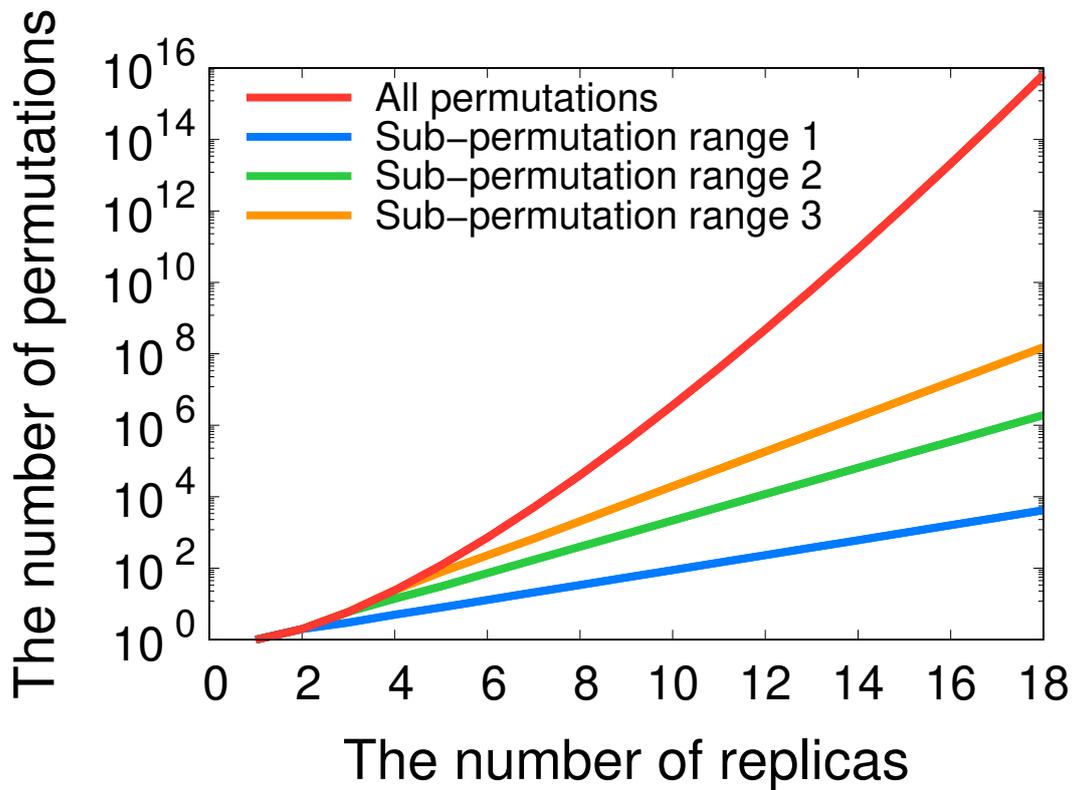
63

Figure 4.2: Comparison of the number of permutations

Table 4.1: The number of sub-permutations and permutations.

Replicas	Range 1	Range 2	Range 3	All permutations
1	1	1	1	1
2	2	2	2	2
3	3	6	6	6
4	5	14	24	24
5	8	31	78	120
6	13	73	230	720
7	21	172	675	5,040
8	34	400	2,069	40,320
9	55	932	6,404	362,880
10	89	2,177	19,708	3,628,800
11	144	5,081	60,216	39,916,800
12	233	11,854	183,988	479,001,600
13	377	27,662	563,172	6,227,020,800
14	610	64,554	1,725,349	87,178,291,200
15	987	150,639	5,284,109	1,307,674,368,000
16	1,597	351,521	16,177,694	20,922,789,888,000
17	2,584	820,296	49,526,506	355,687,428,096,000
18	4,181	1,914,208	151,635,752	6,402,373,705,728,000

i-th column. The next step is to perform adjacent transposition to $\{S^{(M)}\}_M$: $\tau_{M-1,M}\{S^{(M-1)}\}_M$. At this step, $N_{\{S^{(M)}\}_M}$ possible permutations are generated. Sub-permutations $\{S^{(M)}\}_{M-1}$ is defined as a set of permutations such that the number at the M-th column is not less than $M-\epsilon$ in the possible permutations. Similarly, the adjacent transposition operation to $\{S^{(M)}\}_i$ (i.e., $\tau_{i-1,i}\{S^{(M)}\}_i$) is repeated. After every transposition, the permutations in which the number at the M-th column is less than $M-\epsilon$ have to be eliminated. The remaining permutations are defined as $\{S^{(M)}\}_{i-1}$. The adjacent transposition operations are repeated until number M in the permutations moves to the $(M-\epsilon)$-th column. As a result, the set of all sub-permutations with the range of ϵ is obtained:

$$\left\{S^{(M)}\right\} = \left\{\{S^{(M)}\}_i\text{: for } i = M,\ M-1,\ \cdots,\ M-\epsilon\right\}. \qquad (4.9)$$

The number of sub-permutations with ranges of 1, 2, and 3 are presented in Fig. 4.1 and Table 4.1 Note that the sub-permutation with a range of $\epsilon = M$ gives the original RPM.

To finalize this section, let us discuss an application involving more than one permutation parameter (i.e. multidimensional parameter space), such as coupled temperature and Hamiltonian replica sub-permutation. The replica sub-permutation described here limits transitions between similar thermodynamic states, and requires one-dimensional ordering of the states. In the case of involving more than one permutation parameter, the thermodynamic similarity between states can not be determined straightforwardly, which makes difficult to order the thermodynamic states in one dimension; therefore, it is difficult to apply the replica sub-permutation directly. One strategy to employ more than one permutation parameter, i.e. multidimensional RSPM, is to perform replica sub-permutation for each parameter individually. That is, only a sub-permutation of one parameter is performed at one sub-permutation trial, and a sub-permutation of another parameter is performed at another trial. The parameter to be permuted is sequentially changed for every sub-permutation trial. For example, in the temperature and pressure RSPM, temperature and pressure sub-permutations are performed alternately. The temperature sub-permutation at-

tempts are performed among the replicas that have the same pressure value, and the pressure sub-permutation attempts are performed among the replicas that have the same temperature value. Although each sub-permutation trial is one-dimensional for temperature or pressure, the random walk in two-dimensional parameter space are realized during the simulation.

4.3   Computational Details

All simulations in this study used the system comprising one chignolin molecule [33], 1,621 water molecules, and two sodium ions as counter ions. A fully extended structure was used as the initial structure. The MD simulations were performed using the Generalized-Ensemble Molecular Biophysics (GEMB) program developed by the authors. This program has been applied to several proteins [38–41]. A cubic unit cell with periodic boundary conditions was used. The AMBER parm14SB force field [42] was used for the chignolin, and the TIP3P rigid-body model [43] was used for the water molecules. The temperature was controlled by the Nosé-Hoover thermostat [44–46], and the pressure was controlled by the Andersen barostat [47]. Electrostatic interactions were calculated using the particle mesh Ewald method (PME) [48, 49]. The cutoff distance was set to 12.0 Å for the van der Waals interactions and the electrostatic interactions in the direct space sum for PME. The multiple time step method was used [50]. The time step was set to $\Delta t = 0.5$ fs for bonded interactions of protein atoms, $\Delta t = 2.0$ fs for non-bonded interactions of protein atoms and for non-bonded interactions between protein atoms and solvent molecules, and $\Delta t = 4.0$ fs for interactions between solvent molecules. The time step was able to be set to 4.0 fs because the symplectic rigid-body algorithm [51–53] was used for the water molecules. The other simulation details described below are summarized in Table 4.2.

Table 4.2: Details of simulation set up and parameters.

	Method	Permutation/exchange algorithm	Replicas	Temperature (K)	Pressure (MPa)	Time (ns)
Simulation 1	RSPM	range 1	8	300.0 – 360.0	0.1	45
		range 2				
		range 3				
	RPM	without subsets				
		with subsets (4 subsets)				
	REM	neighboring exchange				
Simulation 2	RSPM	range 1	16	300.0 – 450.0	0.1	2000
	RPM	with subsets (4 subsets)				
	REM	neightboring exchange				
Simulation 3	RSPM	range 1	160	300.0 – 450.0	0.1 – 500	180
	RPM	with subsets (8 subsets)		(16 temperatures)	(10 pressures)	
	REM	neightboring exchange				

4.3.1 Simulation 1: Comparison of Sampling Efficiency between the RSPM, RPM with and without Subsets, and REM

To compare how the sub-permutation range and subset division affect the sampling efficiency in the parameter space, the author performed three replica sub-permutation MD (RSPMD) simulation, two replica permutation MD (RPMD) simulations with and without subsets, and one replica exchange MD (REMD) simulation. There were eight replicas used in the simulations. The temperatures, T_1–T_8, were distributed from 300 to 360.0 K: 300.0, 307.9, 316.0, 324.4, 332.9, 341.7, 350.7, and 360.0 K. The pressure was controlled at 0.1 MPa. In the RSPM, three sub-permutations with ranges of 1, 2, and 3 were used to compare the sampling efficiency. Each simulation is referred to as RSPM range 1, RSPM range 2, and RSPM range 3. In the RPMD simulation with subsets, the replicas and temperatures were divided into four subsets. Each subset, therefore, contained four replicas and temperatures:

$$X_\mu^1 = [x_1^{[i(1)]}, \ x_2^{[i(2)]}, \ x_3^{[i(3)]}, \ x_4^{[i(4)]}],$$

$$X_\mu^2 = [x_5^{[i(5)]}, \ x_6^{[i(6)]}, \ x_7^{[i(7)]}, \ x_8^{[i(8)]}],$$

$$X_\mu^3 = [x_1^{[i(1)]}, \ x_2^{[i(2)]}, \ x_7^{[i(7)]}, \ x_8^{[i(8)]}],$$

$$X_\mu^4 = [x_3^{[i(3)]}, \ x_4^{[i(4)]}, \ x_5^{[i(5)]}, \ x_6^{[i(6)]}],$$

where i is the permutation function from the temperature index to the replica index. Replica permutations in subsets X_μ^1 and X_μ^2 were conducted independently at odd-numbered permutation trials, while replica permutations in subsets X_μ^3 and X_μ^4 were conducted independently at even-numbered permutation trials. In the REMD simulation, pairs of replicas corresponding to two adjacent temperatures were exchanged simultaneously. Here, two pairing options, namely, $[T_1, T_2]$, $[T_3, T_4]$, $[T_5, T_6]$, $[T_7, T_8]$, and $[T_2, T_3]$, $[T_4, T_5]$, $[T_6, T_7]$, $[T_1, T_8]$, were used alternately. Replica permutation or exchange trials were performed every 0.5 ps, and the trajectory data were stored every 0.5 ps. The simulations were performed for 50 ns per replica, including 5 ns equilibration per replica.

4.3.2 Simulation 2: Comparison of Sampling Efficiency for Chignolin Folding

The author then performed RSPMD, RPMD, and REMD simulations to evaluate the efficiency of protein folding for chignolin. To enhance protein folding and unfolding, 16 replicas and temperatures were employed. Temperatures T_1–T_{16} were distributed from 300.0 to 450.0 K: 300.0, 308.2, 316.7, 325.3, 334.3, 343.4, 352.8, 362.5 372.4, 382.6, 393.1, 403.9, 414.9, 426.3, 438.0, and 450.0 K. In the RSPM simulation, the sub-permutation that contained transitions only to adjacent temperatures (i.e., range of 1) was employed. In the RPMD simulation, replicas and temperatures were divided into four subsets. Therefore, each subset contained eight replicas and temperatures:

$$X_\mu^1 = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]}\ ],$$

$$X_\mu^2 = [x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}],$$

$$X_\mu^3 = [x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]}],$$

$$X_\mu^4 = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}].$$

Replica permutations in subsets X_μ^1 and X_μ^2 were conducted independently at odd-numbered permutation trials, while replica permutations in subsets X_μ^3 and X_μ^4 were

conducted independently at even-numbered permutation trials. Replica permutation or exchange trials were performed every 0.5 ps, and the trajectory data were stored every 0.5 ps. The simulations were performed for 2,100 ns, including 100 ns equilibration per replica.

4.3.3   Simulation 3: Two-dimensional Replica Sub-Permutation Simulation

To demonstrate a replica sub-permutation simulation for multiple thermodynamic parameters, the author performed two-dimensional isothermal–isobaric RSPMD simulation. For comparison purpose, two-dimensional isothermal–isobaric RPMD [4, 32] and REMD [54, 55] simulations were also performed. Sixteen temperature and ten pressure values were employed and the total number of replicas is 160. Temperatures T_1–T_{16} were distributed from 300.0 to 450.0 K: 300.0, 308.2, 316.7, 325.3, 334.3, 343.4, 352.8, 362.5 372.4, 382.6, 393.1, 403.9, 414.9, 426.3, 438.0, and 450.0 K. Note that the number of temperatures and the temperature values are the same with Simulation 2. The pressures, P_1–P_{10}, were distributed from 0.1 to 500.0 MPa: 0.1, 35.0, 75.0, 120.0, 170.0, 230.0, 290.0, 355.0, 425.0, and 500.0 MPa. Trials of replica sub-permutations/permutations/exchanges were performed every 0.5 ps. Replica sub-permutations/permutations/exchanges for temperatures at each pressure $P_{m_1} = P_1, \ldots, P_{10}$ were performed simultaneously at odd-numbered trials, and those for pressures at each temperature $T_{m_0} = T_1, \ldots, T_{16}$ were performed simultaneously at even-numbered trials. Thus, replica sub-permutation/permutation/exchange trials are performed every 1.0 ps for both temperatures and pressures. In the RPMD simulation, replicas are further divided into four subsets for the temperature and pressure permutations. That is, subsets for the temperature permutation at each

pressure value contained eight replicas:

$$X^1_{\mu,\ [P_{m_1}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]}\ ],$$

$$X^2_{\mu,\ [P_{m_1}]} = [x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}],$$

$$X^3_{\mu,\ [P_{m_1}]} = [x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]}],$$

$$X^4_{\mu,\ [P_{m_1}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}],$$

and subsets for the pressure permutation at each temperature value contained five replicas:

$$X^1_{\mu,\ [T_{m_0}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]}\ ],$$

$$X^2_{\mu,\ [T_{m_0}]} = [x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]}],$$

$$X^3_{\mu,\ [T_{m_0}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]}],$$

$$X^4_{\mu,\ [T_{m_0}]} = [x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]}\ ].$$

Here, i is permutation function from the temperature or pressure index to the replica index. Replica permutations in subsets X^1_μ and X^2_μ were conducted independently in odd-numbered temperature/pressure permutation trial, while replica permutations in subsets X^3_μ and X^4_μ were conducted independently in even-numbered temperature/pressure permutation trial. The trajectory data were stored every 0.5 ps. The simulations were performed for 200 ns per replica, including 20 ns equilibration run per replica.

4.4 Results and Discussion

4.4.1 Simulation 1: Comparison of Sampling Efficiency between the RSPM, RPM with and without Subsets, and REM

Figure 4.3 shows the transition ratios of the replicas from one temperature to other temperatures. Here, the transition ratio at a certain temperature is defined as
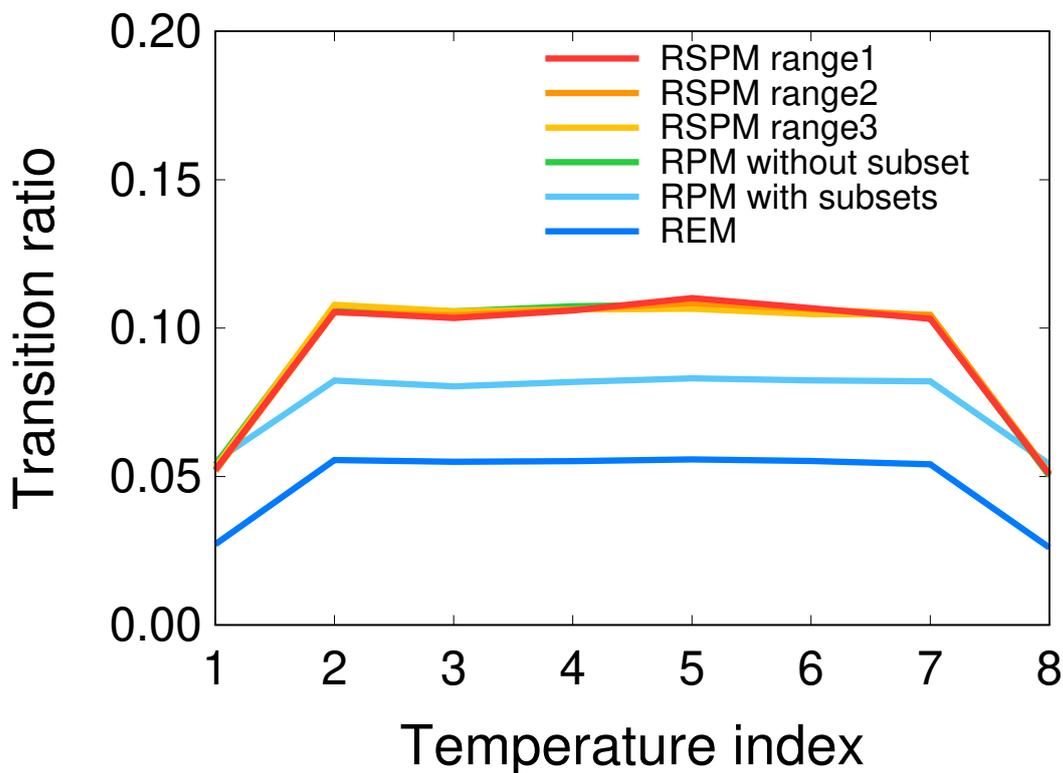
Figure 4.3: Comparison of transition ratios between the RSPMs, RPM with and without subsets, and REM.

a probability with which the replica at that temperature is transferred to other temperatures. The RSPM and RPM achieve higher transition ratios than the REM. The transition ratios of the RSPM and RPM without subsets are nearly identical to one another and higher than those of the RPM with subsets.

Table 4.3 list the number of tunneling events of the replicas in the temperature space. A tunneling event is a widely used measure for evaluating sampling efficiency in the parameter space. Here, a tunneling event is defined as a round trip between the lowest and highest temperature. Here, errors were estimated by calculating standard deviation. As expected from the transition ratio, the number of tunneling events of the RSPM and RPM without subsets is higher than that of the REM and RPM with subsets. It should be noted that the tunneling events of the RSPM and RPM without subsets are in good agreement with one another within the error range. Thus, the RSPM and RPM without subsets achieve the highest number of tunneling events among the six simulations.

These results indicate that introducing subsets to reduce the number of replica-

Table 4.3: The number of tunneling events in the temperature space in Simulation 1.

	RSPM		RPM		REM
range 1	range 2	range 3	without subsets	with subsets	
73 ± 6	70 ± 4	71 ± 3	75 ± 7	56 ± 4	43 ± 4

permutation candidates is unfavorable with respect to sampling efficiency in the parameter space. Conversely, the sampling efficiency of the RSPM is nearly identical to the RPM in which all possible combinations between replicas and parameters are considered. Therefore, the RSPM succeeds in reducing the number of candidates for the next state without the loss of sampling efficiency, especially the replica sub-permutation containing transitions only to the adjacent temperatures (i.e., RSPM range 1).

To provide a rationale for selecting sub-permutation with a range of 1, the author calculated the transition ratio in more detail. The results are presented in Fig. 4.4. In the RSPM range 1 and the REM, the temperature of each replica was transferred only to the adjacent temperatures. The RPM with and without subsets allows replicas to transfer not only to neighboring temperatures, but also to non-neighboring temperatures. The same holds for RSPM with ranges of 2 and 3. However, the transition ratios to non-adjacent temperatures are significantly lower than the transition ratio to the adjacent temperatures. This is because the overlap between the probability distributions at two temperatures in the phase space becomes small as the temperature difference increases. These results indicate that permutations including transitions to non-adjacent temperatures contribute little to the sampling efficiency. Therefore, even if transitions to non-adjacent temperatures are omitted, the sampling efficiency does not decrease; in other words, the RSPM with a range of 1 is sufficient.
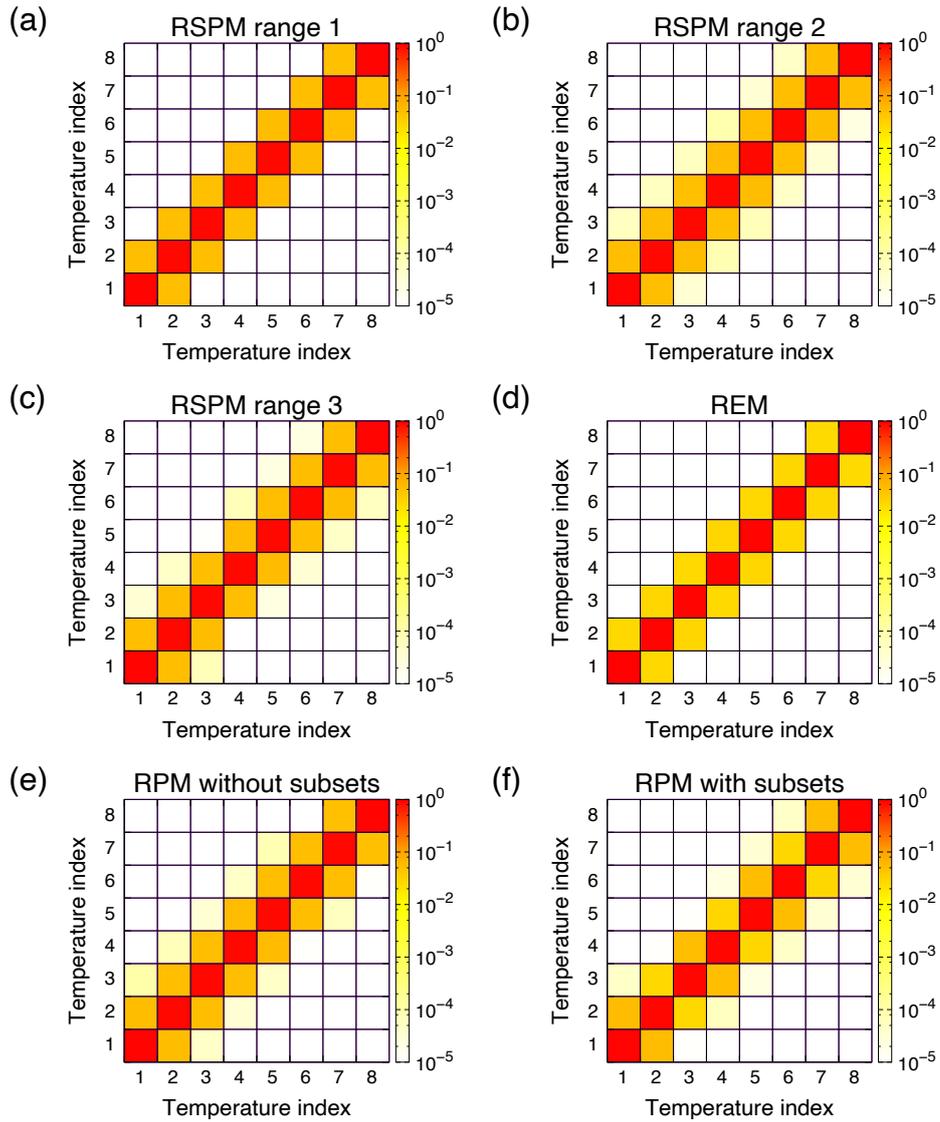
Figure 4.4: Transition probability from one temperature to other temperatures in the (a) RSPM range 1, (b) RSPM range 2, (c) RSPM range 3, (d) REM, (e) RPM without subsets, and (f) RPM with subsets. The vertical axis is the temperature index before transition, and the horizontal axis is the temperature index after transition.

4.4.2 Simulation 2: Comparison of Sampling Efficiency for Chignolin Folding

Transition Ratio and Tunneling Events in the Temperature Space

Figure 4.5 presents the transition ratios from one temperature to other temperatures to evaluate sampling efficiency in the temperature space. The RSPM and RPM achieve higher transition ratios than the REM. However, uneven transition

Table 4.4: Comparison of measures of the sampling efficiency in Simulation 2.

Method	Tunneling events	Folding-unfolding events per replica	Folding-unfolding events at 300 K	Autocorrelation time (ps)
RSPM	412 ± 14	31 ± 5	5079	755.2
RPM	388 ± 17	30 ± 5	4719	926.7
REM	263 ± 13	29 ± 5	2883	1553.9

ratios are seen in the RPM with subsets. Lower transition ratios are observed in temperature indices 1, 4, 5, 8, 9, 12, 13, and 16 which are temperatures at the subset terminals. This is because the subsets restricts the temperature transition of the replicas at the subset terminals. For example, temperature exchanges between temperature indices 4 and 5 and between temperature indices 12 and 13 can occur only at odd-numbered permutation trials, and a temperature exchange between temperature indices 8 and 9 can occur only at even-numbed permutation trials. Therefore, transition ratios at the terminal temperatures in the subsets become lower than those at other temperatures. This unevenness is unfavorable for an efficient simulation. In contrast, the RSPM achieves a smoother transition ratio. The number of tunneling events in the temperature space is displayed in Table 4.4. Of the three methods, the RSPM achieves the highest number of tunneling events. Due to a higher and smoother transition ratio of the replicas in the temperature space, the RSPM improves tunneling events in comparison with other methods. Thus, the author conclude that the RSPM is the most efficient of the three methods in the temperature space.

Validation of Statistical Ensemble and Free-Energy Landscape

Figure 4.6 shows the probability distributions of the potential energy and volume at each simulated temperature and pressure values to ensure the statistical ensemble. The probability distributions, $f(x; T_{m_0}, P_{m_1})$, were obtained by normalizing potential energy or volume histograms for each simulated temperature and pressure so that the probability distributions satisfy $\int_{-\infty}^{\infty} dE f(E; T_{m_0}, P_{m_1}) = 1$ or $\int_{-\infty}^{\infty} dV f(V; T_{m_0}, P_{m_1}) = 1$. The probability distributions are almost the same between the RSPMD, RPMD, and REMD simulations both in the potential energy
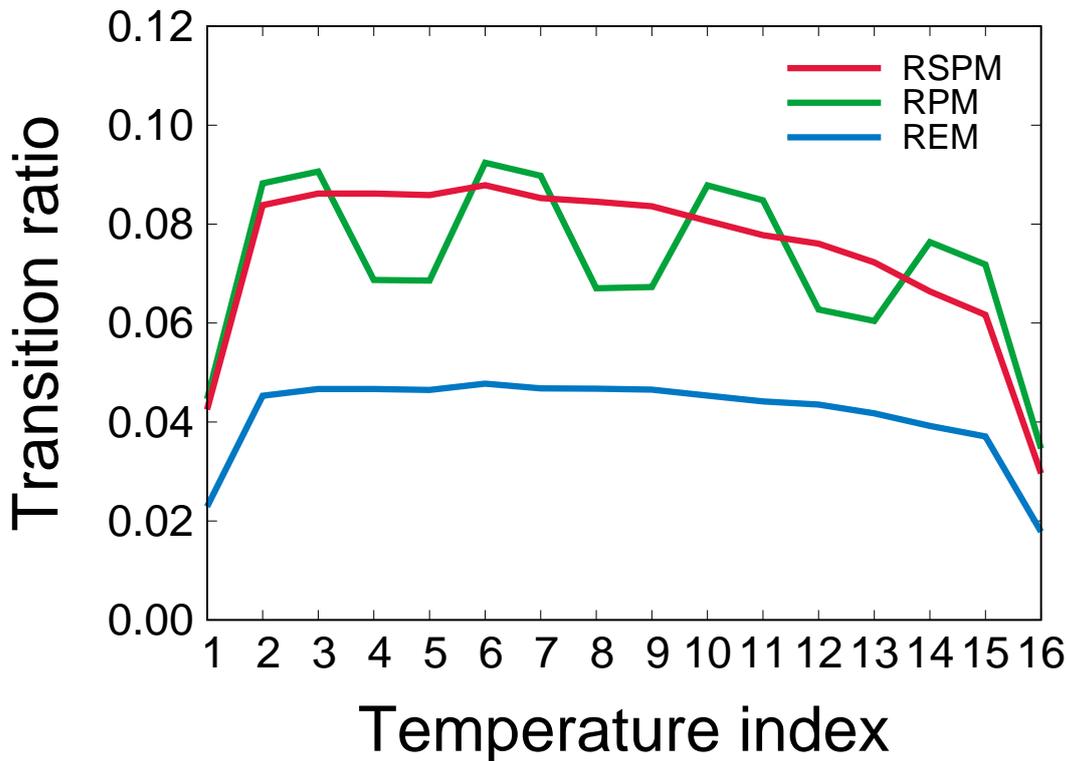
Figure 4.5: Comparison of transition ratios between RSPM, RPM, and REM.

and volume spaces. These results support the convergence of the isothermal–isobaric ensembles; therefore, the RSPM correctly generates the statistical ensembles.

To evaluate the sampling quality of the RSPMD, RPMD, and REMD simulations, the author computed free-energy landscapes (or potential of mean force) from snapshots at $T = 300$ K. Free-energy as a function of the reaction coordinates ξ_1 and ξ_2 is calculated by

$$F(\xi_1, \xi_2) = -k_B T \ln P(\xi_1, \xi_2), \tag{4.10}$$

where $P(\xi_1, \xi_2)$ is the probability distribution. Figures 4.7(a)–(c) present free-energy landscapes as a function of $\xi_1 = \mathrm{d}(\mathrm{Aps3N} - \mathrm{Gly7N})$ and $\xi_2 = \mathrm{d}(\mathrm{Asp3N} - \mathrm{Thr8O})$, which are the distance between Asp3O and Gly7N atoms and the distance between Asp3N and Thr8O atoms, respectively. The two-dimensional free-energy landscapes have a global minimum state at $(\xi_1,\ \xi_2) = (6.5\ \text{Å},\ 3.0\ \text{Å})$, which corresponds to the folded state. There are other minimum states at $(\xi_1,\ \xi_2) = (4.0\ \text{Å},\ 7.0\ \text{Å})$ and $(\xi_1,\ \xi_2) = (8.5\ \text{Å},\ 11.5\ \text{Å})$, which correspond to the misfolded state and the
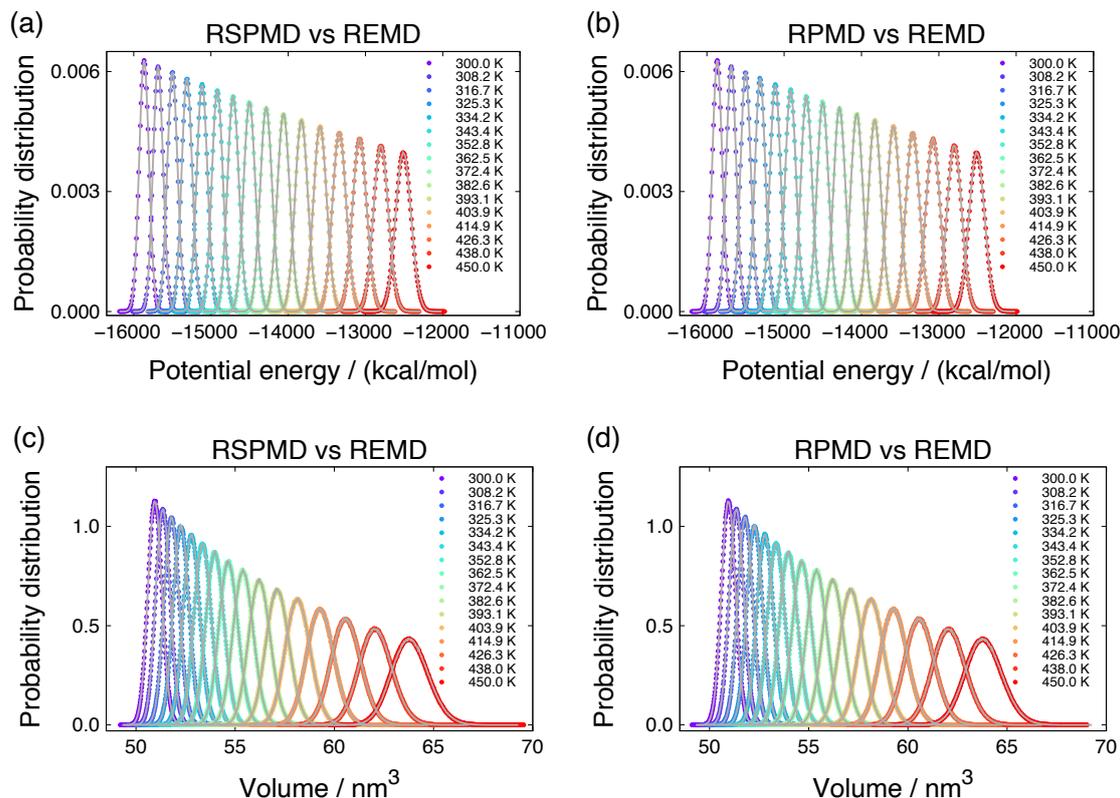
Figure 4.6: Potential-energy probability-distributions for (a) RSPMD and REMD, and (b) RPMD and REMD at all simulated temperatures. Volume-probability distributions for (c) RSPMD and REMD, and (d) RPMD and REMD at all simulated temperatures. RSPMD and RPMD data are shown with points, and REMD data are shown with solid lines.

intermediate state, respectively. Typical conformations in the folded, misfolded, and intermediate states are illustrated in Figure 4.7(d)–(f). The trends of the 2D free-energy landscapes between the three simulations are similar, indicating that each 2D free-energy landscape is well converged. It is also verified that a wider range of conformations was sampled; thus, each simulation achieved a sufficient conformational search.

Comparison of the Sampling Efficiency in the Conformational Space

The root mean square deviation (RMSD) was calculated to evaluate the sampling efficiency in the conformational space. Figure 4.8 presents the time series of the
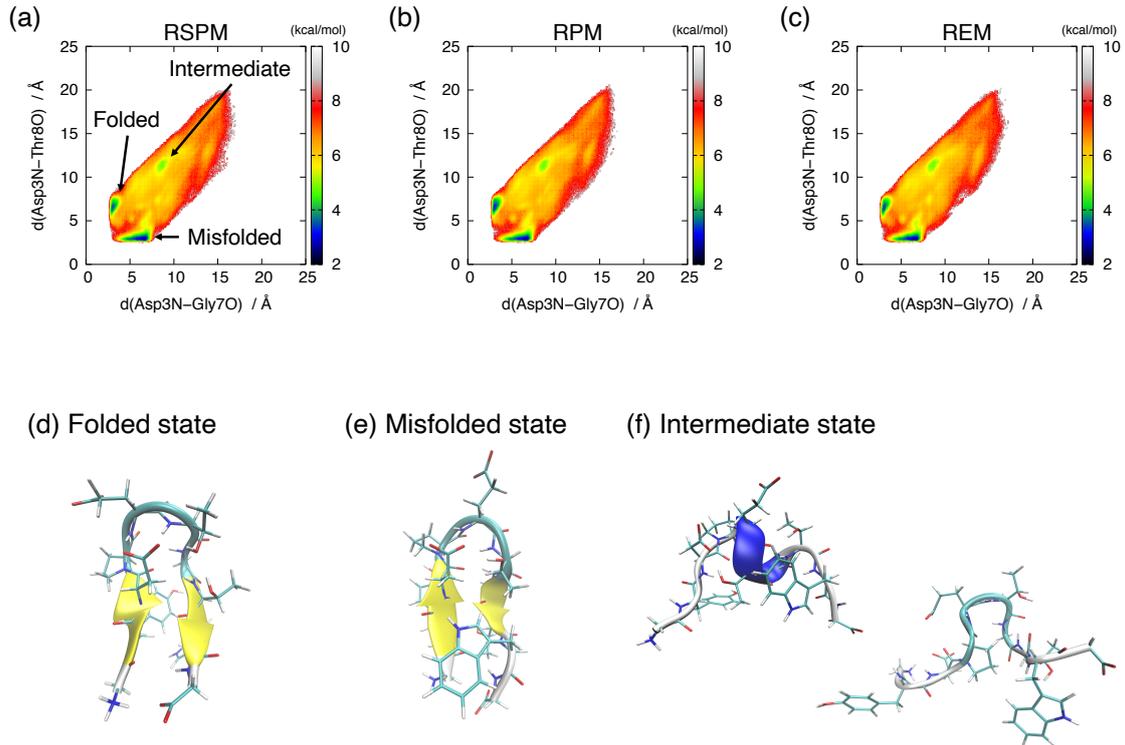
Figure 4.7: Two-dimensional free-energy landscapes of the chignolin calculated from the (a) RSPM, (b) RPM, and (c) REM. Typical structure of the chignolin in the (d) folded, (e) misfolded, and (f) intermediate states.

RMSD of replica 1. The RMSD is defined as

$$\text{RMSD} = \min\left\{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_i^0)^2}\right\}, \tag{4.11}$$

where N is the number of atoms, $\boldsymbol{r}_i$ is the set of coordinates obtained from the simulation, and $\boldsymbol{r}_i^0$ is the set of coordinates of the reference structure. The RMSD was calculated for heavy atoms (C, N, and C_α) at the backbone with respect to the reference conformation. The nuclear magnetic resonance (NMR) structure (PDB ID: 1UAO, Model 1) was used as the reference conformation. Minimization was performed considering the rigid translation and rotation of $\boldsymbol{r}_i$ with respect to the center of geometry [56]. The RMSD increased and decreased repeatedly; this variation reflects the folding and unfolding of chignolin.

Here, folding–unfolding events is defined as a measure of sampling efficiency. One folding-unfolding event is counted when the chignolin molecule completes a cy-
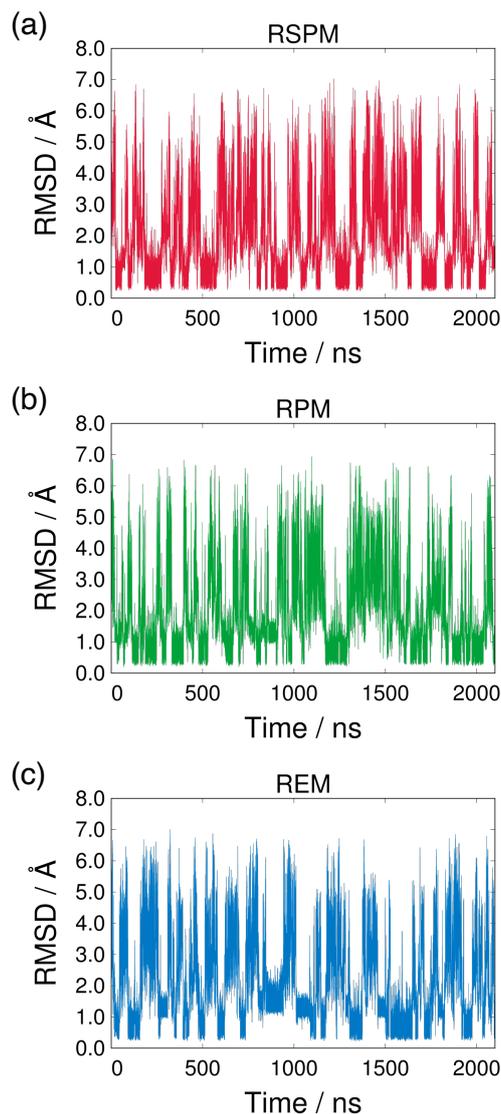
Figure 4.8: Time series of the RMSD from the NMR structure (PDB ID: 1UAO) of replica 1 in the (a) RSPM, (b) RPM, and (c) REM.

cle from the folded (unfolded) state to the (unfolded) folded state and back to the folded (unfolded) state. The structure is considered to be in the folded state when the RMSD is smaller than 0.6 Å, and in the unfolded state when the RMSD is larger than 4.0 Å. The number of folding–unfolding events per replica is listed in Table 4.4. No significant difference was observed among the three methods. However, it is important to compare the folding-unfolding events at a specific temperature, especially at a low temperature, because physical quantities are often calculated at that temperature. The author calculated the RMSD and the number of folding–unfolding events at $T = 300$ K. The time series of the RMSD at $T = 300$ K are shown in Fig-
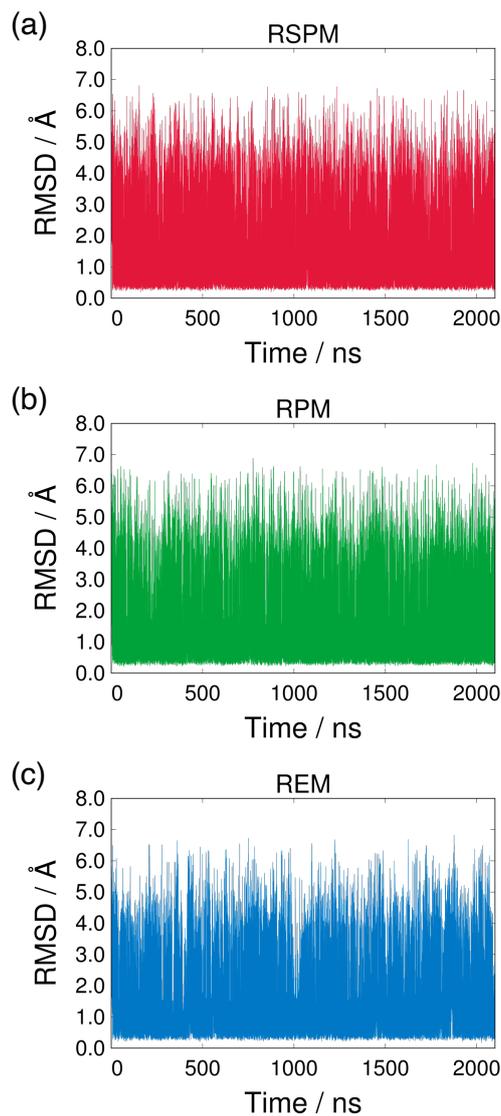
Figure 4.9: Time series of the RMSD from the NMR structure (PDB ID: 1UAO) at 300 K in the (a) RSPM, (b) RPM, and (c) REM.

ure 4.9. The number of folding–unfolding events at $T = 300$ K is listed in Table 4.4. The RSPM achieves the highest number of folding–unfolding events among the three methods. This is because replica replacement occurs more frequently due to the improvement of transition probability, and different conformations are observed one after another.

To evaluate sampling efficiency in the conformation space more quantitatively, the author calculated the autocorrelation function of the RMSD and the autocorre-

lation time at $T = 300$ K. The autocorrelation function was calculated as follows:

$$C_x(t) \equiv \frac{\langle x(0)x(t) \rangle - \langle x(0) \rangle \langle x(t) \rangle}{\langle x(0)^2 \rangle - \langle x(0) \rangle^2}, \tag{4.12}$$

where

$$\langle x(0)x(t) \rangle = \frac{1}{\mathcal{N} - \frac{t}{\Delta t}} \sum_{i=1}^{\mathcal{N} - \frac{t}{\Delta t}} x(t_i)x(t_i + t), \tag{4.13}$$

t_i is time at the i-th snapshot, and $\mathcal{N}$ is the number of snapshots. The autocorrelation time was estimated by

$$\tau = \sum_{i=1}^{t_{\max}/\Delta t} C_x(t_i)\Delta t. \tag{4.14}$$

In this study, Δt and $t_{\max}$ were set to 0.5 ps and 50 ns, respectively. The correlation of samples affect the convergence of physical quantities. The autocorrelation function at a certain temperature is useful for evaluating how fast snapshots become independent at the temperature. In other words, the autocorrelation function and autocorrelation time at a certain temperature are measures about how many uncorrelated samples were obtained. Note that the autocorrelation function and the autocorrelation time calculated here do not mean the real physical quantities of the chignolin in solution. This is because the structure at $T = 300$ K is replaced when replica exchange/permutation trial is accepted. Figure 4.10 presents the autocorrelation functions of the RMSD from the NMR structure at $T = 300$ K. The autocorrelation times are listed in Table 4.4. The autocorrelation functions converge to approximately zero in all simulations. From the inset in Figure 4.10, it can be seen that the decay of the autocorrelation functions of the RSPM and RPM are faster than that of the REM. It is also observed that the autocorrelation function in the RSPM decays slightly faster than that in the RPM. The autocorrelation time of the RSPM is the smallest among the three methods indicating that a larger number of independent samples can be obtained using the RSPM. Therefore, the convergence of other physical quantities is expected to be improved as well. In summary, the
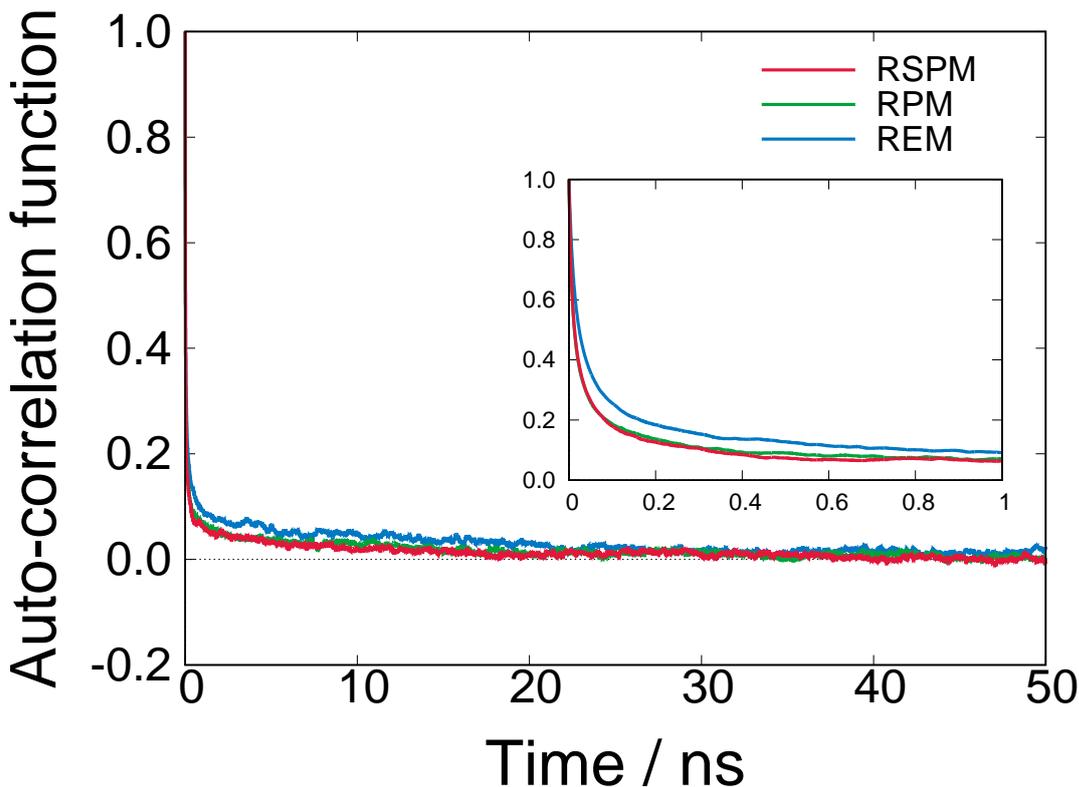
Figure 4.10: Autocorrelation functions of the RMSD from the NMR structure (PDB ID: 1UAO) at 300 K. The inset shows the same autocorrelation functions with a different time scale.

RSPM is the most efficient among the three methods not only in the temperature space but also in the conformational space.

Convergence of the Population of the Folded State

To confirm the convergence of physical quantities, the author calculated the cumulative average for a fraction of the folded state, which is defined as follows:

$$f_{\text{Fold}}(t) = \frac{1}{t} \int_0^t dt' \, \chi(t') \tag{4.15}$$

where $\chi(t)$ is an indicator function that takes 1 when the structure at time t is in the folded state or zero otherwise. Here, a folded chignolin is defined as a structure that has three hydrogen bonds between the backbone N and O: Asp3O–Gly7N, Asp3O–Thr8N, and Asp3N–Thr8O [33]. These hydrogen bonds were determined using the Define Secondary Structure of Proteins (DSSP) algorithm [57]. Figure 4.11
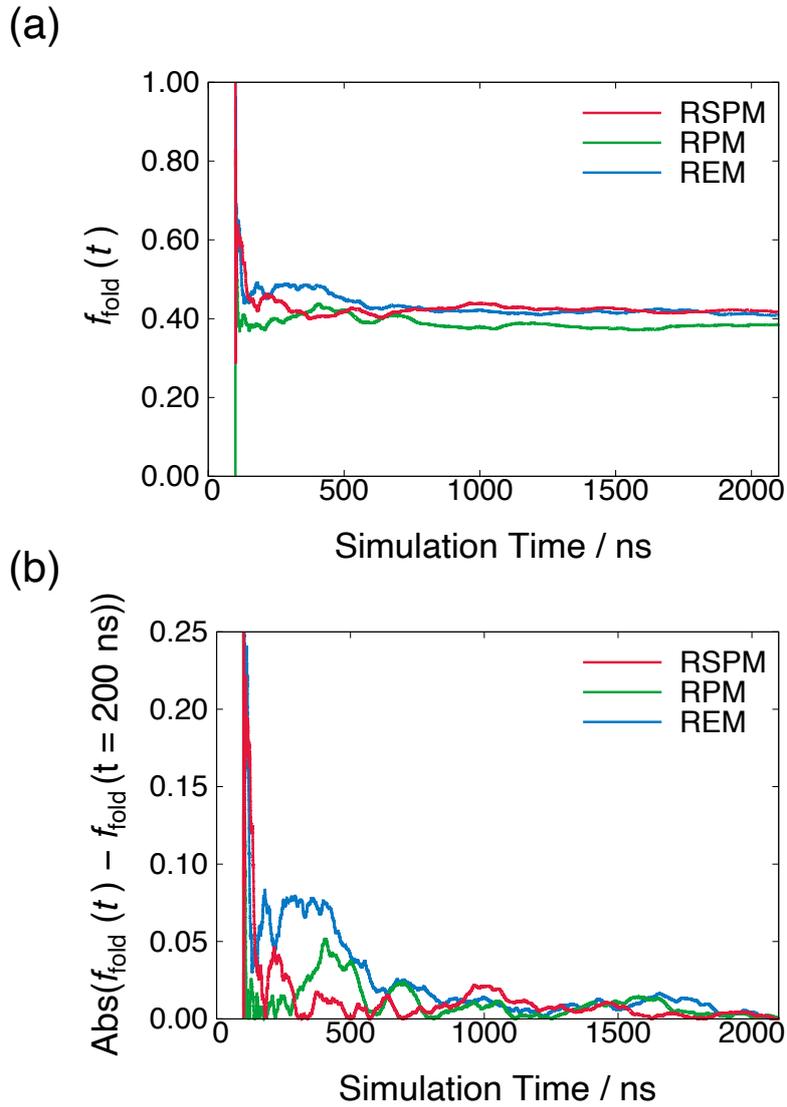
(a)

(b)

Figure 4.11: Convergence of the probability of the folded state. (a) Cumulative average of folded state $f_{\text{fold}}(t)$ until time t. (b) Absolute value of the difference in $f_{\text{fold}}(t)$ between time t and at the final time $t = 2{,}100$ ns.

(a) presents the cumulative average for the fraction of the folded state at $T = 300$ K, while Figure 4.11 (b) presents the absolute value of the difference between the cumulative average at time t and $t = 2{,}100$ ns. Note that this average was taken after the equilibration period of 100 ns. The fraction of the folded state are well converged, although it differs slightly, and the RSPM achieves faster convergence than the RPM and REM.

4.4.3   Simulation 3: Two-dimensional Replica Sub-Permutation Simulation

Let us discuss the sampling efficiency for the two-dimensional isothermal–isobaric RSPMD, RPMD, and REMD simulations using 16 temperature and ten pressure values. Figure 4.12 presents the transition ratios from one parameter value to other parameter values. Among the three simulations, the RSPM achieves higher transition ratios in almost all parameter values. As discussed in the previous subsection, the transition ratios of the parameter decrease in the RPMD with subsets because the transition ratios of the parameter values at the subset terminals tend to be lower than those at other parameters values. The same holds for the two-dimensional isothermal–isobaric RPMD using subsets. In the case of 16 temperatures, the subset boundaries exist between temperature indices 4 and 5, 8 and 9, and 12 and 13. In the case of ten pressures, the boundaries are between pressure indices 2 and 3, 5 and 6, and 7 and 8. The transition ratio decreases more at the intersection points of the temperature boundaries and the pressure boundaries. Generally, in the the multidimensional parameter space, the number of intersection points of the subset terminals increases with increasing the dimensions of the parameter space. Therefore, transition ratios of the RPMD with subsets tend to decrease in higher-dimensional parameter spaces.

In order to examine the details, the author calculated transition ratios only at the temperature or pressure permutation trials. Figure 4.13 (a) presents the transition ratios for the temperature-permutation trials. As expected, smoother transition ratio is observed in the RSPMD. In the RPMD, the transition ratios of the temperature values at the subset terminals such as temperature indices 1, 4, 5, 8, 9, 12, 13, and 16 take lower values than those in the RSPMD while the peak values of the transition ratio in the RPMD are higher than the RSPMD. The transition ratios for the temperature-permutation trials increase as pressure increases in the three methods. In addition, at the peaks of the transition ratios in the RPMD, the differences between the RSPMD and the RPMD become large with the pressure increases, especially at the high temperature values. Figure 4.13
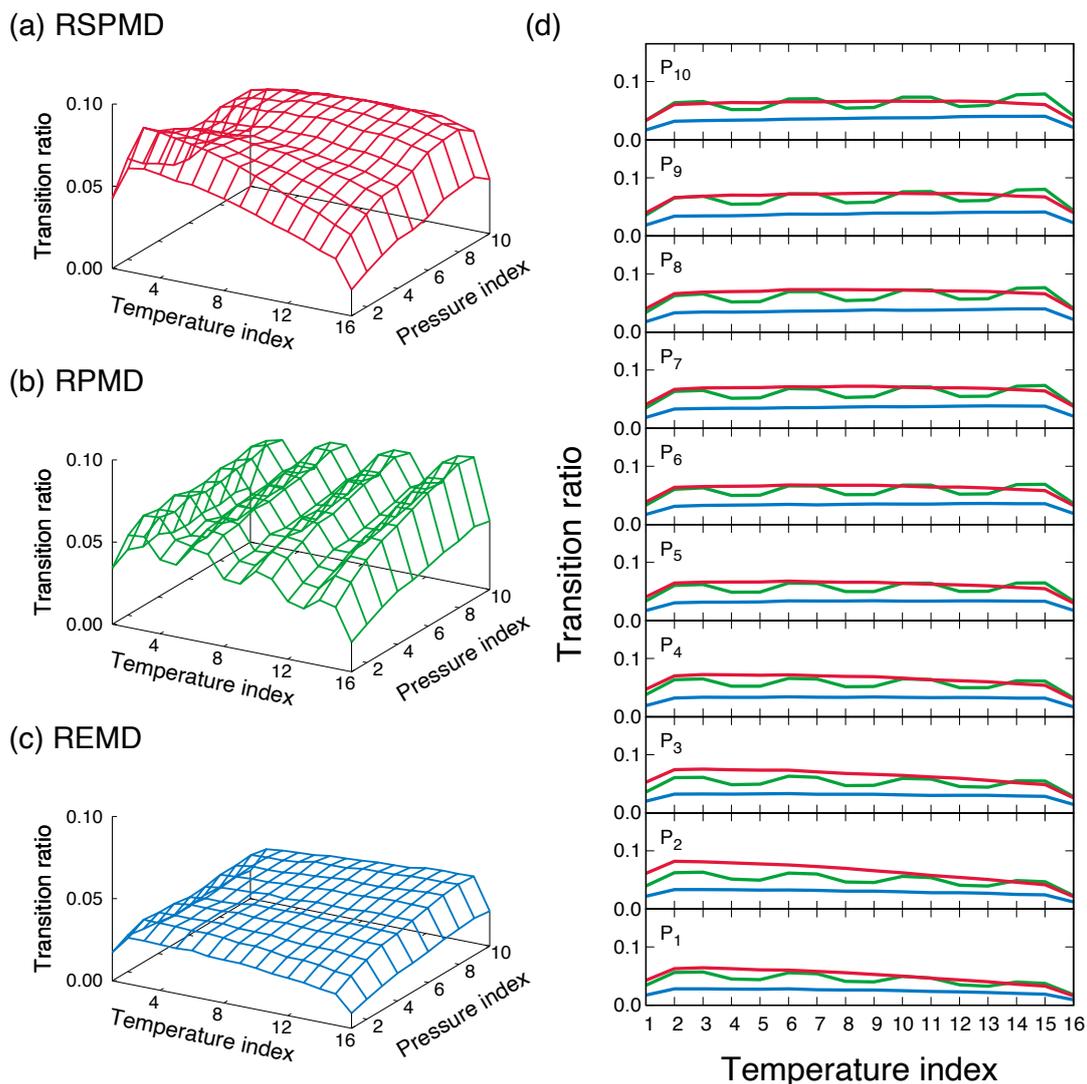
Figure 4.12: Transition ratios from one parameter to other parameters for the isothermal–isobaric (a) RSPMD, (b) RPMD, and (c) REM using 16 temperatures and ten pressures. (d) The transition ratios for the parameter permutation sliced at each pressure. Red, green, and blue lines show the data for RSPMD, RPMD, and REMD, respectively.

(b) presents the transition ratios for the pressure-permutation trials. The RSPMD realizes the highest transition ratio for the pressure permutation at all temperature and pressure values. In the RPMD, small peaks are observed at pressure indices 4 and 9 because pressure values except for indices 4 and 9 are at the subset terminals. The transition ratios for the pressure-permutation trials decrease as temperature increase, especially at the low pressure values.

The temperature and pressure dependence of the transition ratios is understood
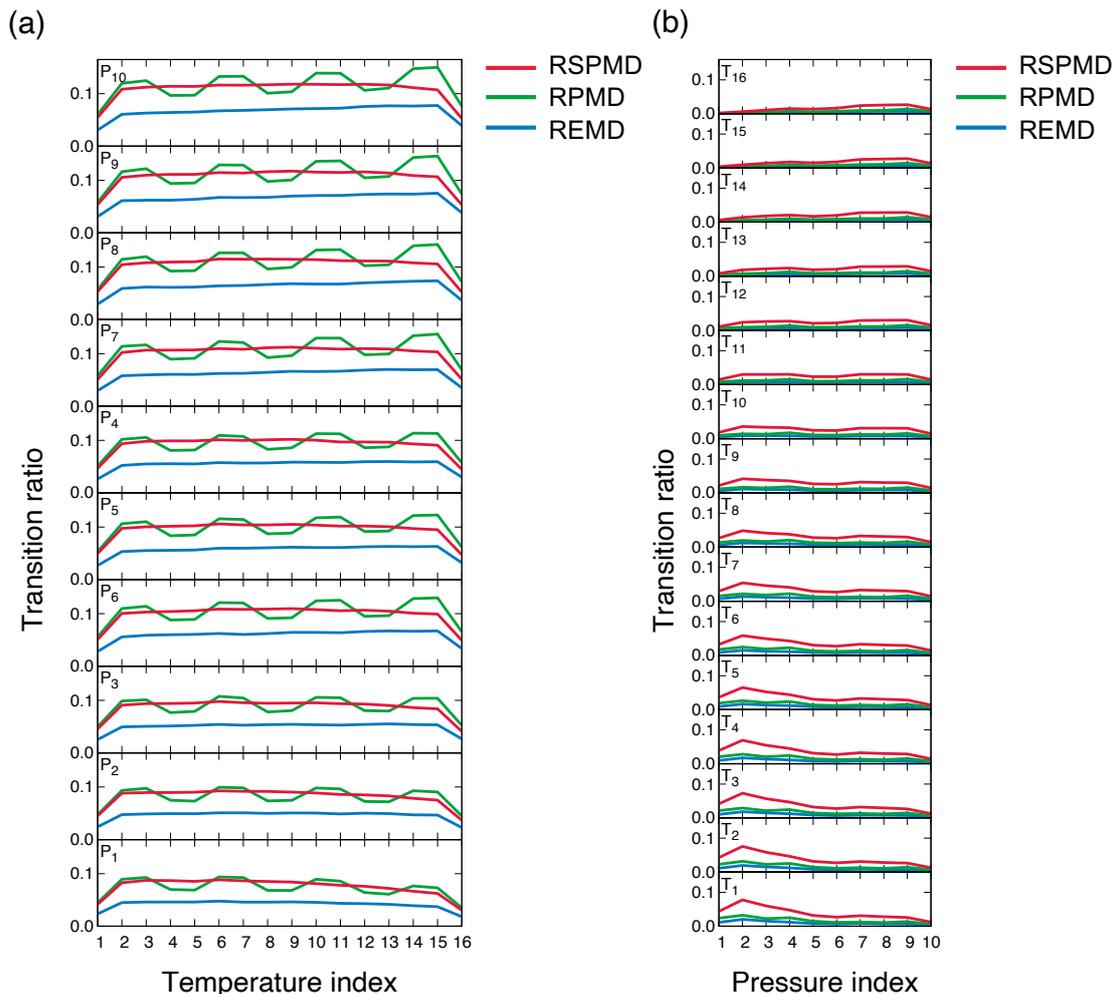
Figure 4.13: Transition ratios from one parameter to other parameters (a) for temperature permutation trials and (b) for pressure permutation trials. Red, green, and blue lines show the data for RSPMD, RPMD, and REMD, respectively.

from the overlap of probability distributions in the potential energy and volume spaces. Figure 4.14 (a) and 4.15 (a) present the probability distributions of the potential energy and volume, respectively at each simulated pressure value in the RSPMD and RPMD simulations. The RSPMD and RPMD have almost the same potential energy and volume distributions. The pressure dependence of the transition ratios for temperature-permutation trial can be explained from these distributions. The peak positions of the potential energy and volume at each temperature shift to lower values with increasing pressure. In particular, larger peak shifts are found at higher temperature values. With the peak shifts, overlap between distributions become large both in the potential energy and volume distributions. To

estimate the overlap between two distributions f_1 and f_2, the author calculated following quantity:

$$\text{overlap}[f_1,\ f_2] = \int_{-\infty}^{\infty} dE \ \min[f_1(E),\ f_2(E)] \qquad (4.16)$$

for the potential energy distributions and

$$\text{overlap}[f_1,\ f_2] = \int_{-\infty}^{\infty} dV \ \min[f_1(V),\ f_2(V)] \qquad (4.17)$$

for the volume distributions. $\text{overlap}[f_1,\ f_2] \simeq 1$ if distributions f_1 and f_2 are similar to each other, and $\text{overlap}[f_1,\ f_2] = 0$ if there is no overlap between f_1 and f_2. The overlaps between two distributions at adjacent temperatures are shown in Figures 4.14 (b) and 4.15 (b), and those with second adjacent temperature are shown in Figures 4.14 (c) and 4.15 (c). The overlaps with the adjacent and the second adjacent temperatures become large as pressure increases, which causes higher transition ratios for temperature-permutation trial at high pressure values. The increase in the overlap with the second adjacent temperatures causes the increase in the transition probabilities for the permutation candidates that contain the parameter transitions to the second adjacent temperatures. The RPMD includes such permutation candidates, whereas the RSPMD does not include. As a result, increase in the transition ratios for the temperature-permutation trials is larger in the RPMD than in the RSPMD. This tendency stands out more at the high temperature than low temperature values corresponding to the peak shifts and overlaps in the energy and volume distributions. Then, the author discuss the temperature dependence of the transition ratios for pressure-permutation trial. Figure 4.16 (a) shows the probability distributions of the volume at each simulated temperature value, and Figure 4.16 (b) and (c) show the overlaps between two distributions at adjacent and second adjacent pressures, respectively. Note that the probability distributions of the volume are enough to discuss the transition ratio because the potential energy term is canceled out when the transition probability for the pressure permutation is calculated [32]. As temperature increases, the peak positions of the volume distributions at each pressure shift to larger volume values, and the overlap between
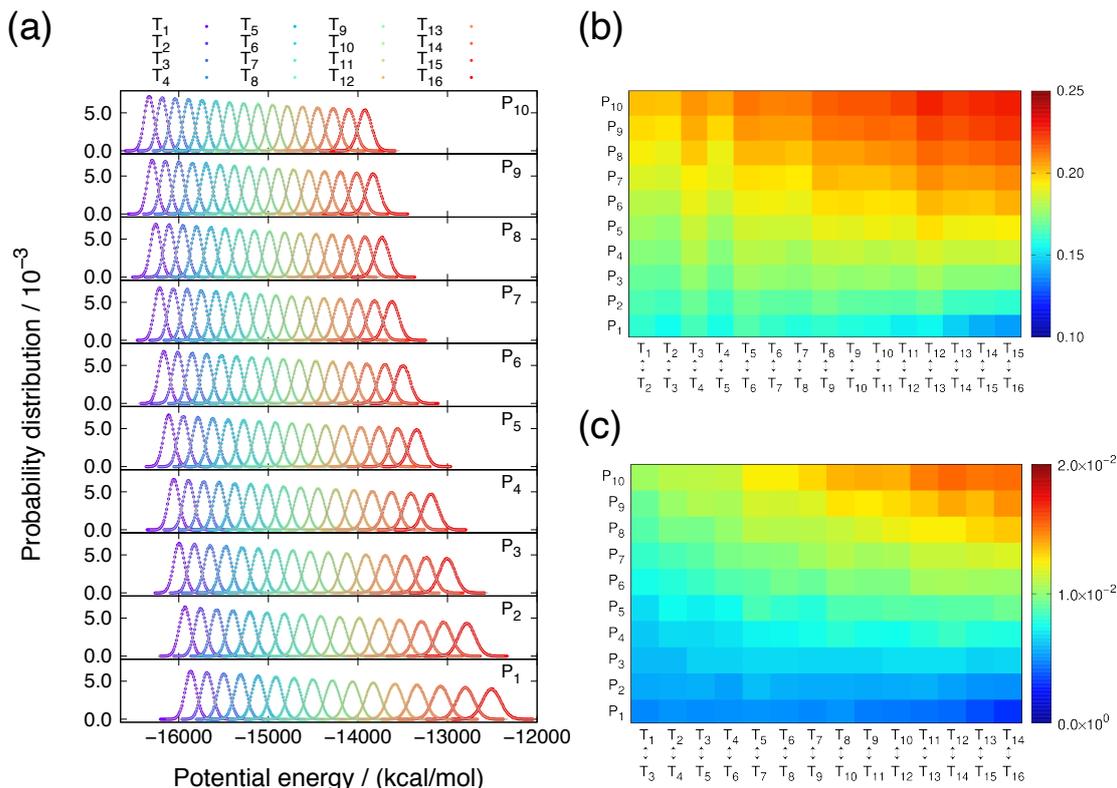
Figure 4.14: (a) Probability distributions of the potential energy at each pressure value. The RSPMD data are shown with colored points and the RPMD data are shown with gray solid lines. The overlap of the potential-energy probability-distributions (b) between adjacent temperatures and (c) between second adjacent temperatures calculated from the RPMD simulation. It should be noted that the overlaps of the probability distributions calculated from the RSPMD and REMD simulations are almost the same with the RPMD.

distributions become small, especially at the low pressure values. As a result, the transition ratios decrease at higher temperature values.

Table 4.5 lists the numbers of tunneling events of the replicas. Here, a tunneling event in the two-dimensional parameter space is calculated as a round trip between the lowest temperature and pressure values, (T_1, P_1), and the highest temperature and pressure values, (T_{16}, P_{10}). In the parameter space, the number of tunneling events in the RSPMD is the largest among the three methods as expected from the transition ratios. The numbers of tunneling events in the temperature space for the RSPMD and RPMD are the same, and higher than that of the REMD. One might think that this result seems to be contradictory from the conclusion of Simulation 2 in the previous subsection because the RSPMD simulation in Simulation 2 shows
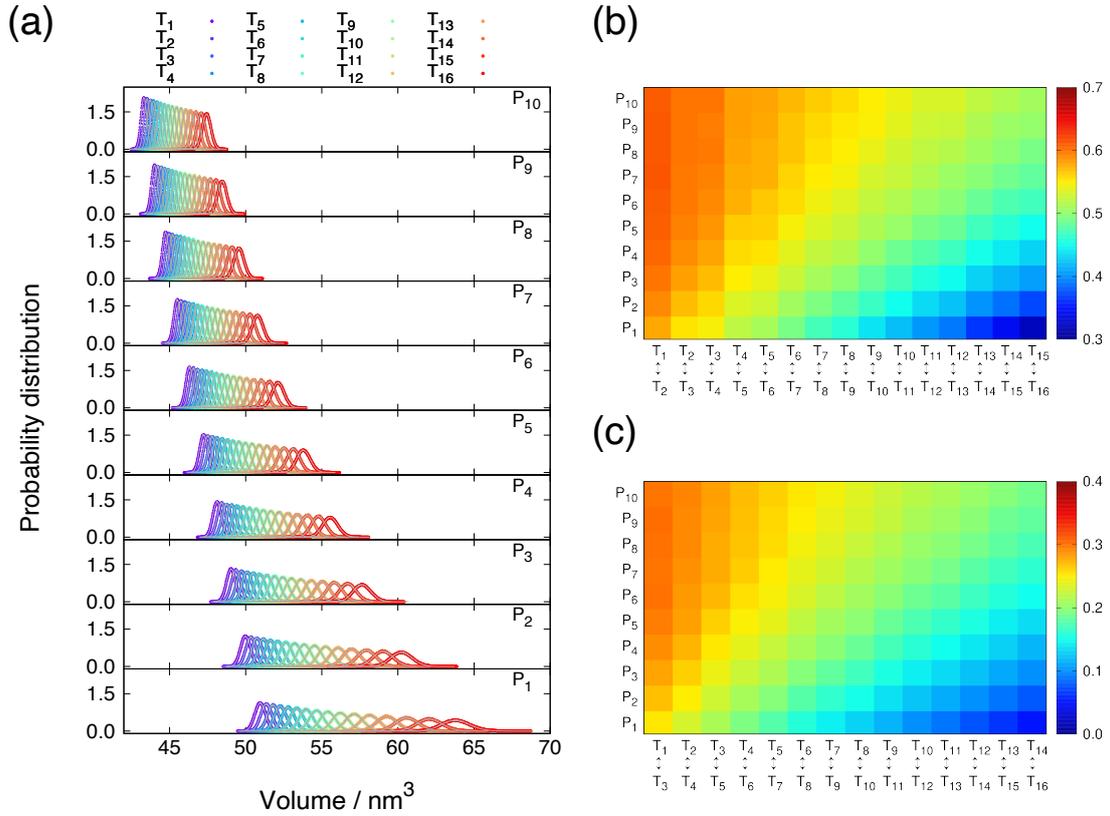
Figure 4.15: (a) Probability distributions of the volume at each pressure value. The RSPMD data are shown with colored points and RPMD data are shown with gray solid lines. The overlaps of the volume-probability distributions (b) between adjacent temperatures and (c) between second adjacent temperatures calculated from the RPMD simulation. It should be noted that the overlaps of the probability distributions calculated from the RSPMD and REMD simulations are almost the same with the RPMD.

the improvement of the tunneling events in the temperature space. These results are explained as follows: The transition ratios for the temperature permutation at the pressure value P_1 are almost identical to Simulation 2 when compared Figure 4.5 to the bottom panel of Figure 4.13 (a). Therefore, it can be reasonable to think that the RSPMD realizes more efficient temperature transition than the RPMD at lower pressure values. However, the peak values of the transition ratios in the RPMD increase, and the difference from the RSPMD become large as pressure increases, especially at high temperature values as discussed. The RPMD possibly achieves the efficient temperature transition in comparison with RSPMD at high pressure values. As a result, the average numbers of tunneling events for the RSPMD and RPMD simulations are the same in the temperature space. In the pressure space, on
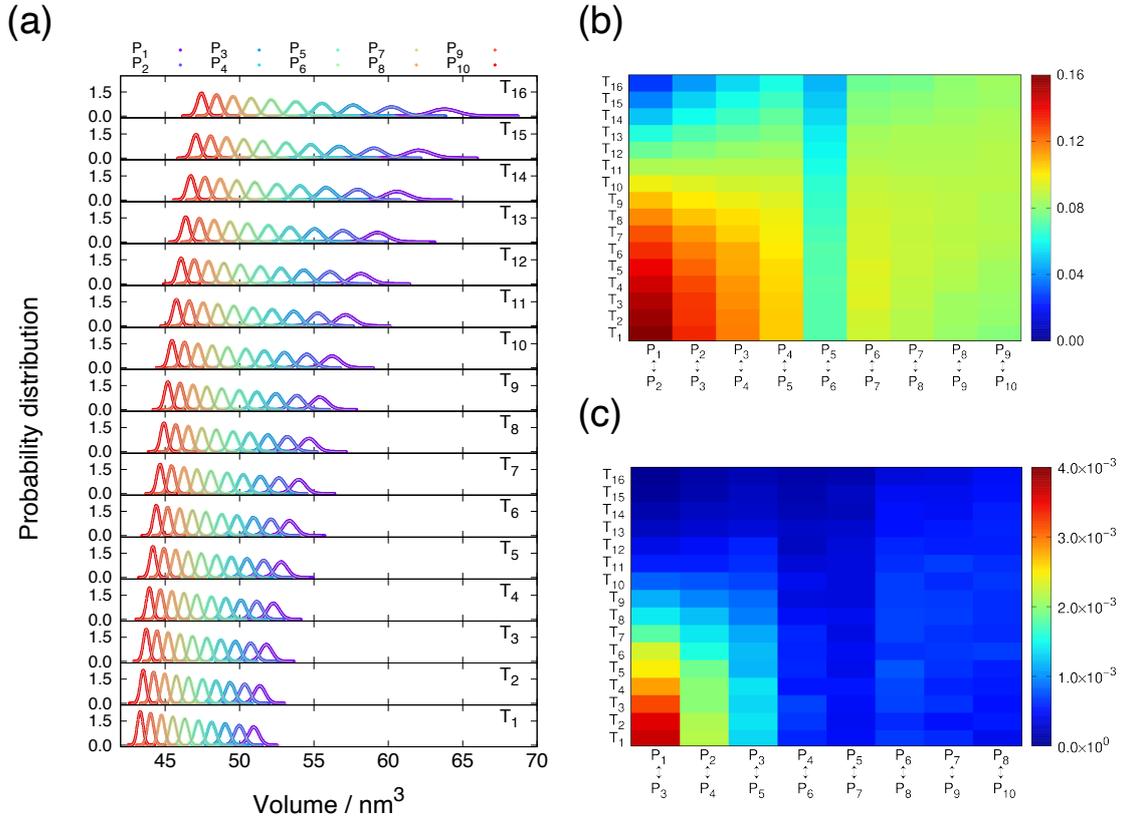
Figure 4.16: (a) Probability distributions of the volume at each temperature value. RSPMD data are shown with colored points and RPMD data are shown with gray solid lines. The overlaps of the volume-probability distributions (b) between adjacent pressures and (c) between second adjacent pressures calculated from RPMD simulation. It should be noted that the overlaps of the probability distributions calculated from the RSPMD and REMD simulations are almost the same with the RPMD.

Table 4.5: The number of tunneling events in the parameter space for two-dimensional RSPMD, RSPMD and REMD simulations in Simulation 3.

Method	Whole parameter space	Temperature space	Pressure space
RSPMD	7 ± 2	28 ± 4	27 ± 4
RPMD	5 ± 1	28 ± 4	10 ± 2
REMD	3 ± 1	18 ± 3	7 ± 2

the other hand, the number tunneling events in the RSPMD is the highest among the three methods as expected from the transition ratios.

4.5　Conclusions

In this study, the author proposed a replica sub-permutation method (RSPM) for MD and MC simulations. This method introduces a new permutation algorithm called sub-permutation to reduce the number of candidates in parameter permutation. The sub-permutation is a set of permutations that include transitions only to neighboring parameters. One might think that the RSPM is similar to the REM using neighboring exchange scheme. However, it should be emphasized that the RSPM employs the Suwa–Todo algorithm, while the REM employs the Metropolis algorithm. RSPM can make the most of the merit of the Suwa–Todo algorithm because the sub-permutation gives the more than two candidates for the next state.

The author examined sampling efficiency for the RSPM, RPM with and without subsets, and REM. The results indicate that among the methods, the RSPM and RPM without subsets are the most efficient in the temperature space. This result indicates that parameter transition by replica sub-permutation is nearly identical to replica permutation that uses all possible combinations of replicas and parameters. Thus, the RSPM succeeds in reducing the number of candidates for the next state without the loss of sampling efficiency.

The author performed folding simulations of chignolin in explicit water to compare the efficiency of the conformational search for the RSPM, RPM, and REM. The RSPM achieves the most efficient simulation in the temperature space. In addition, the author computed the autocorrelation function and autocorrelation time of the RMSD from the NMR structure at $T = 300$ K. The RSPM achieves the fastest decay of the autocorrelation function and, as a result, the shortest autocorrelation time. These results demonstrate that a larger number of uncorrelated samples can be obtained from the RSPM simulation, which improves the convergence of physical quantities. Therefore, the RSPM is the most efficient method not only in the parameter space but also in the conformational space.

References

[1] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[2] Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[3] Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, 60(2):96–123, 2001.

[4] Masataka Yamauchi, Yoshiharu Mori, and Hisashi Okumura. Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble. *Biophysical Reviews*, 11(3):457–469, 2019.

[5] Bernd A Berg and Thomas Neuhaus. Multicanonical algorithms for first order phase transitions. *Physics Letters B*, 267(2):249–253, 1991.

[6] Bernd A Berg and Thomas Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Physical Review Letters*, 68(1):9, 1992.

[7] Ulrich HE Hansmann, Yuko Okamoto, and Frank Eisenmenger. Molecular dynamics, langevin and hydrid monte carlo simulations in a multicanonical ensemble. *Chemical Physics Letters*, 259(3-4):321–330, 1996.

[8] Nobuyuki Nakajima, Haruki Nakamura, and Akinori Kidera. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conforma-

tional sampling of peptides. *The Journal of Physical Chemistry B*, 101(5):817–824, 1997.

[9] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. New approach to monte carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics*, 96(3):1776–1783, 1992.

[10] Enzo Marinari and Giorgio Parisi. Simulated tempering: a new monte carlo scheme. *Europhysics Letters*, 19(6):451, 1992.

[11] Koji Hukushima and Koji Nemoto. Exchange monte carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[12] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[13] Walter Nadler and Ulrich H. E. Hansmann. Optimizing replica exchange moves for molecular dynamics. *Physical Review E*, 76(5):057102, 2007.

[14] Helmut G Katzgraber, Simon Trebst, David A Huse, and Matthias Troyer. Feedback-optimized parallel tempering monte carlo. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(03):P03018, 2006.

[15] Simon Trebst, Matthias Troyer, and Ulrich H. E. Hansmann. Optimized parallel tempering simulations of proteins. *The Journal of Chemical Physics*, 124(17):174903, 2006.

[16] Walter Nadler and Ulrich H. E. Hansmann. Generalized ensemble and tempering simulations: A unified view. *Physical Review E*, 75(2):026109, 2007.

[17] Daniel Sindhikara, Yilin Meng, and Adrian E Roitberg. Exchange frequency in replica exchange molecular dynamics. *The Journal of Chemical Physics*, 128(2):01B609, 2008.

[18] Daniel J Sindhikara, Daniel J Emerson, and Adrian E Roitberg. Exchange often and properly in replica exchange molecular dynamics. *Journal of Chemical Theory and Computation*, 6(9):2804–2808, 2010.

[19] F Calvo. All-exchanges parallel tempering. *The Journal of Chemical Physics*, 123(12):124106, 2005.

[20] Paul Brenner, Christopher R Sweet, Dustin VonHandorf, and Jesús A Izaguirre. Accelerating the replica exchange method through an efficient all-pairs exchange. *The Journal of Chemical Physics*, 126(7):074103, 2007.

[21] Manuel Athènes and Florent Calvo. Multiple-replica exchange with information retrieval. *ChemPhysChem*, 9(16):2332–2339, 2008.

[22] Hiroko X Kondo and Makoto Taiji. Enhanced exchange algorithm without detailed balance condition for replica exchange method. *The Journal of Chemical Physics*, 138(24):244113, 2013.

[23] John D Chodera and Michael R Shirts. Replica exchange and expanded ensemble simulations as gibbs sampling: Simple improvements for enhanced mixing. *The Journal of Chemical Physics*, 135(19):194110, 2011.

[24] Satoru G Itoh and Hisashi Okumura. Replica-permutation method with the suwa–todo algorithm beyond the replica-exchange method. *Journal of Chemical Theory and Computation*, 9(1):570–581, 2012.

[25] Hidemaro Suwa and Synge Todo. Markov chain monte carlo method without detailed balance. *Physical Review Letters*, 105(12):120603, 2010.

[26] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[27] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

[28] Nuria Plattner, J. D. Doll, Paul Dupuis, Hui Wang, Yufei Liu, and J. E. Gubernatis. An infinite swapping approach to the rare-event sampling problem. *The Journal of Chemical Physics*, 135(13):134111, 2011.

[29] Nuria Plattner, J. D. Doll, and Markus Meuwly. Overcoming the rare event sampling problem in biological systems with infinite swapping. *Journal of Chemical Theory and Computation*, 9(9):4215–4224, 2013.

[30] Hidemaro Suwa. Geometric allocation approach in markov chain monte carlo. In *Geometrically Constructed Markov Chain Monte Carlo Study of Quantum Spin-phonon Complex Systems*, pages 11–42. Springer, 2014.

[31] Hiroaki Nishizawa and Hisashi Okumura. Comparison of replica-permutation molecular dynamics simulations with and without detailed balance condition. *Journal of the Physical Society of Japan*, 84(7):074801, 2015.

[32] Masataka Yamauchi and Hisashi Okumura. Development of isothermal-isobaric replica-permutation method for molecular dynamics and monte carlo simulations and its application to reveal temperature and pressure dependence of folded, misfolded, and unfolded states of chignolin. *The Journal of Chemical Physics*, 147(18):184107, 2017.

[33] Shinya Honda, Kazuhiko Yamasaki, Yoshito Sawada, and Hisayuki Morii. 10 residue folded peptide designed by segment statistics. *Structure*, 12(8):1507–1518, 2004.

[34] David van der Spoel and M Marvin Seibert. Protein folding kinetics and thermodynamics from atomistic simulations. *Physical Review Letters*, 96(23):238102, 2006.

[35] Daisuke Satoh, Kentaro Shimizu, Shugo Nakamura, and Tohru Terada. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS letters*, 580(14):3422–3426, 2006.

[36] Atsushi Suenaga, Tetsu Narumi, Noriyuki Futatsugi, Ryoko Yanai, Yousuke Ohno, Noriaki Okimoto, and Makoto Taiji. Folding dynamics of 10-residue β-hairpin peptide chignolin. *Chemistry - An Asian Journal*, 2(5):591–598, 2007.

[37] Ryuhei Harada and Akio Kitao. Exploring the folding free energy landscape of a β-hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method. *The Journal of Physical Chemistry B*, 115(27):8806–8812, 2011.

[38] Hisashi Okumura. Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins*, 80(10):2397–2416, 2012.

[39] Hisashi Okumura and Satoru G Itoh. Amyloid fibril disruption by ultrasonic cavitation: nonequilibrium molecular dynamics simulations. *Journal of the American Chemical Society*, 136(30):10549–10552, 2014.

[40] Hisashi Okumura and Satoru G Itoh. Structural and fluctuational difference between two ends of aβ amyloid fibril: Md simulations predict only one end has open conformations. *Scientific Reports*, 6:38422, 2016.

[41] Rupali Gupta, Shigeru Saito, Yoshiharu Mori, Satoru G Itoh, Hisashi Okumura, and Makoto Tominaga. Structural basis of trpa1 inhibition by hc-030031 utilizing species-specific differences. *Scientific Reports*, 6:37460, 2016.

[42] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[43] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[44] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[45] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[46] William G Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[47] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[48] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[49] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[50] M Tuckerman, Bruce J Berne, and Glenn J Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[51] Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[52] T. F. Miller III, M Eleftheriou, P Pattnaik, A Ndirango, D Newns, and G. J. Martyna. Symplectic quaternion scheme for biophysical molecular dynamics. *The Journal of Chemical Physics*, 116(20):8649–8659, 2002.

[53] Hisashi Okumura, Satoru G Itoh, and Yuko Okamoto. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric-isothermal, and related ensembles. *The Journal of Chemical Physics*, 126(8):084103, 2007.

[54] Tsuneyasu Okabe, Masaaki Kawata, Yuko Okamoto, and Masuhiro Mikami. Replica-exchange monte carlo method for the isobaric–isothermal ensemble. *Chemical Physics Letters*, 335(5-6):435–439, 2001.

[55] Yoshiharu Mori and Yuko Okamoto. Generalized-ensemble algorithms for the isobaric–isothermal ensemble. *Journal of the Physical Society of Japan*, 79(7):074003, 2010.

[56] Evangelos A Coutsias, Chaok Seok, and Ken A Dill. Using quaternions to calculate rmsd. *Journal of Computational Chemistry*, 25(15):1849–1857, 2004.

[57] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

Chapter 5

Unusual Behavior of Chignolin Under High-Pressure Conditions

5.1   Introduction

Several studies have revealed that pressure induces denaturation of proteins [1]. According to Le Châtelie's principle, smaller partial volume states of a substance realize under high-pressure conditions. This means that the pressure-induced conformational changes of the biomolecules are explained by changes in the partial molar volume. The changes in the partial molar volume of the biomolecules is thought to be composed of three contributions as follows:

$$\Delta V = \Delta V_{\text{vdw}} + \Delta V_{\text{cavity}} + \Delta V_{\text{hydration}}, \tag{5.1}$$

where V_{vdw} is the volume of van der Waals, V_{cavity} is the volume of cavity (solvent-inaccessible region) inside the biomolecules, and $V_{\text{hydration}}$ is the volume resulting from solute–solvent interactions [2]. Although it is established that $\Delta V_{\text{vdw}} \simeq 0$, it has been under discussion which term (i.e., ΔV_{cavity} or $\Delta V_{\text{hydration}}$) in Eq. (5.1) contributes to the changes in the partial molar volume.

The pressure-induced denaturation of the proteins typically explained by the smaller partial molar volume of the denatured state of proteins due to the lack of cavities [3]. On the other hand, high pressure studies using model peptides were conducted to investigate the molecular mechanisms of pressure-induced conformational changes when the biomolecules do not have cavities in the native state [4–7]. The model peptides that form a single secondary structure, such as α-helix and β-hairpin, are useful to understand how $\Delta V_{\text{hydration}}$ term contributes to the stability of conformation. This is because they do not have cavity in the native state, and can be regarded as $\Delta V_{\text{cavity}} \simeq 0$. Interestingly, α-helical peptides such as AK16 pep-

tide and C-peptide analogs are stabilized under high-pressure conditions, whereas β-hairpin mini protein chignolin [8] is denatured under high-pressure conditions. These results imply that only the cavity term in the partial molar volume does not always characterize the changes in the partial volume. Further studies are necessary to understand the pressure-induced conformational changes of the biomolecules.

In this study, the author focuses on the high pressure effects on the β-hairpin mini protein, chignolin. The chignolin has ten amino-acid residues (GYDPETGTWG) and unique folded and misfolded structures. Both folded and misfolded structures have similar β-hairpin structures. The chignolin is regarded as a suitable model protein because of its size and unique secondary structures. The chignolin has been studied in detail both experimentally and computationally. The temperature dependence of the folded conformation was measured experimentally by Honda *et al.* [8], and its pressure dependence was measured by Imamura and Kato [6]. A computational study on the temperature and pressure dependence of chignolin was performed by Okumura [9]. However, these studies only focused on the folded and unfolded states. The author performed isothermal–isobaric replica-permutation simulations of a β-hairpin mini protein, chignolin, in explicit water solvent to reveals temperature and pressure dependence of folded, misfolded, and unfolded states of chignolin. From the simulation, the author obtains a new insight into the temperature and pressure dependence of the misfolded chignolin. This is the first theoretical work to investigate the misfolded state under high pressure conditions.

5.2 Computational Details

This study focus on the stability of artificially designed protein, chignolin, under various temperature and pressure conditions. The amino acid sequence of the chignolin is GYDPETGTWG. The author employed the isothermal–isobaric replica-permutation molecular dynamics method [10] to achieve the efficient conformational sampling from wider range of temperature and pressure values. A fully extended structure was used as the initial structure. The system consists of one chignolin molecule [8], 1,621 water molecules, and two sodium ions as counter ions. The MD

simulations were performed using the Generalized-Ensemble Molecular Biophysics (GEMB) program. A cubic unit cell with periodic boundary conditions was used. The AMBER parm14SB force field [11] was used for the chignolin, and the TIP3P rigid-body model [12] was used for the water molecules. The temperature was controlled by the Nosé-Hoover thermostat [13–15], and the pressure was controlled by the Andersen barostat [16]. Electrostatic interactions were calculated using the particle mesh Ewald method (PME) [17, 18]. The cutoff distance was set to 12.0 Å for the van der Waals interactions and the electrostatic interactions in the direct space sum for PME. The multiple time step method was used [19]. The time step was set to $\Delta t = 0.5$ fs for bonded interactions of protein atoms, $\Delta t = 2.0$ fs for non-bonded interactions of protein atoms and for non-bonded interactions between protein atoms and solvent molecules, and $\Delta t = 4.0$ fs for interactions between solvent molecules. The time step was able to be set to 4.0 fs because the symplectic rigid-body algorithm [20–22] was used for the water molecules.

Sixteen temperature and ten pressure values were employed in this simulation. The total number of replicas is 160. Temperatures T_1–T_{16} were distributed from 300.0 to 450.0 K: 300.0, 308.2, 316.7, 325.3, 334.3, 343.4, 352.8, 362.5 372.4, 382.6, 393.1, 403.9, 414.9, 426.3, 438.0, and 450.0 K. Replica permutation trials for temperatures at each pressure $P_{m_1} = P_1, \ldots, P_{10}$ were performed simultaneously at odd-numbered trials, and those for pressures at each temperature $T_{m_0} = T_1, \ldots, T_{16}$ were performed simultaneously at even-numbered trials. Thus, replica permutation trials are performed every 1.0 ps for both temperatures and pressures. In the RPMD simulation, replicas are further divided into four subsets for the temperature and pressure permutations. That is, subsets for the temperature permutation at each pressure value contained eight replicas:

$$X^1_{\mu,\,[P_{m_1}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]}\ ],$$
$$X^2_{\mu,\,[P_{m_1}]} = [x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}],$$
$$X^3_{\mu,\,[P_{m_1}]} = [x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]},\ x_{11}^{[i(11)]},\ x_{12}^{[i(12)]}],$$
$$X^4_{\mu,\,[P_{m_1}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_{13}^{[i(13)]},\ x_{14}^{[i(14)]},\ x_{15}^{[i(15)]},\ x_{16}^{[i(16)]}],$$

and subsets for the pressure permutation at each temperature value contained five replicas:

$$X^1_{\mu,\ [T_{m_0}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]}\ ],$$

$$X^2_{\mu,\ [T_{m_0}]} = [x_6^{[i(6)]},\ x_7^{[i(7)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]}],$$

$$X^3_{\mu,\ [T_{m_0}]} = [x_1^{[i(1)]},\ x_2^{[i(2)]},\ x_8^{[i(8)]},\ x_9^{[i(9)]},\ x_{10}^{[i(10)]}],$$

$$X^4_{\mu,\ [T_{m_0}]} = [x_3^{[i(3)]},\ x_4^{[i(4)]},\ x_5^{[i(5)]},\ x_6^{[i(6)]},\ x_7^{[i(7)]}\ ].$$

Here, i is permutation function from the temperature or pressure index to the replica index. Replica permutations in subsets X^1_μ and X^2_μ were conducted independently in odd-numbered temperature/pressure permutation trial, while replica permutations in subsets X^3_μ and X^4_μ were conducted independently in even-numbered temperature/pressure permutation trial. The trajectory data were stored every 10 ps. The simulations were performed for 200 ns per replica, including 20 ns equilibration run per replica; thus, the total simulation time was 32 μs.

In the analysis, the multistate Bennett acceptance ration method [23] was used to calculate statistical ensemble averages of physical quantities.

5.3   Results and Discussion

5.3.1   Validation of Sampling Quality

Figure 5.1 presents the time series of the temperature and pressure indices of replicas 1 and 40, 80, and 120. To confirm sampling efficiency in the parameter spaces, the author counted the tunneling events in the temperature and pressure spaces. Here, one tunneling event was defined as a round trip between the lowest and highest values in the temperature or pressure space. The average number of tunneling events per one replica was 28.4 ± 0.3 in the temperature space and 10.4 ± 0.1 in the pressure space. Figure 5.1 presents the time series of a root mean square deviation (RMSD) to confirm the sampling quality in the conformational space. The RMSD was calculated for the backbone N, C_α, and C atoms with respect to the reference conformation. The nuclear magnetic resonance (NMR) structure (PDB ID: 1UAO,

Model 1) was used as the reference conformation. The RMSD is defined as

$$\mathrm{RMSD} = \min\left\{ \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_i^0)^2} \right\}, \tag{5.2}$$

where N is the number of backbone N, C_α, and C atoms; $\boldsymbol{r}_i^0$ is the set of the coordinates of the reference conformation; and $\boldsymbol{r}_i$ is the set of the coordinates obtained by the isothermal–isobaric RPMD simulation. The minimization is performed considering the rigid translations and rigid rotation of $\boldsymbol{r}_i$ with respect to the center of geometry [24]. The RMSD becomes larger and smaller repeatedly. This means that conformation of the chignolin changes to the native (RMSD $\approx$ 0.4 Å – 1.0 Å), metastable (RMSD $\approx$ 1.2 Å – 1.8 Å), and unfolded structures repeatedly. To count the folding-unfolding events, the author regarded the folded state as that with as the RMSD less than 0.6 Å and the unfolded state as that with the RMSD more than 4.0 Å. One folding-unfolding event is defined as when the chignolin molecule completes a cycle from the folded (unfolded) state to the unfolded (folded) state and back to the folded (unfolded) state. The average number of folding-unfolding events per one replica was 2.3 ± 0.1. These results show that this simulation realize efficient conformation sampling under various temperature and pressure conditions.

5.3.2   Potential of Mean Force as a Function of RMSD

Figure 5.2 shows the potential of mean force (PMF) or free energy landscape as a function of the RMSD. When $T = 300$ K and $P = 0.1$ MPa, there are two minimum states: the native state at RMSD = 0.6 Å that corresponds to the folded state [8] and the metastable state at RMSD = 1.3 Å that is known as the misfolded state [25, 26]. As the temperature or pressure increases, the unfolded state at RMSD = 2.0 – 4.0 Å become more stable. This means that chignolin is denatured by both temperature and pressure. Interestingly, as pressure increases, the misfolded state gradually becomes stable as compared to the folded state. This is the first time that the pressure dependence of the misfolded state has been reported.

To discuss the temperature and pressure dependence of chignolin in detail, the
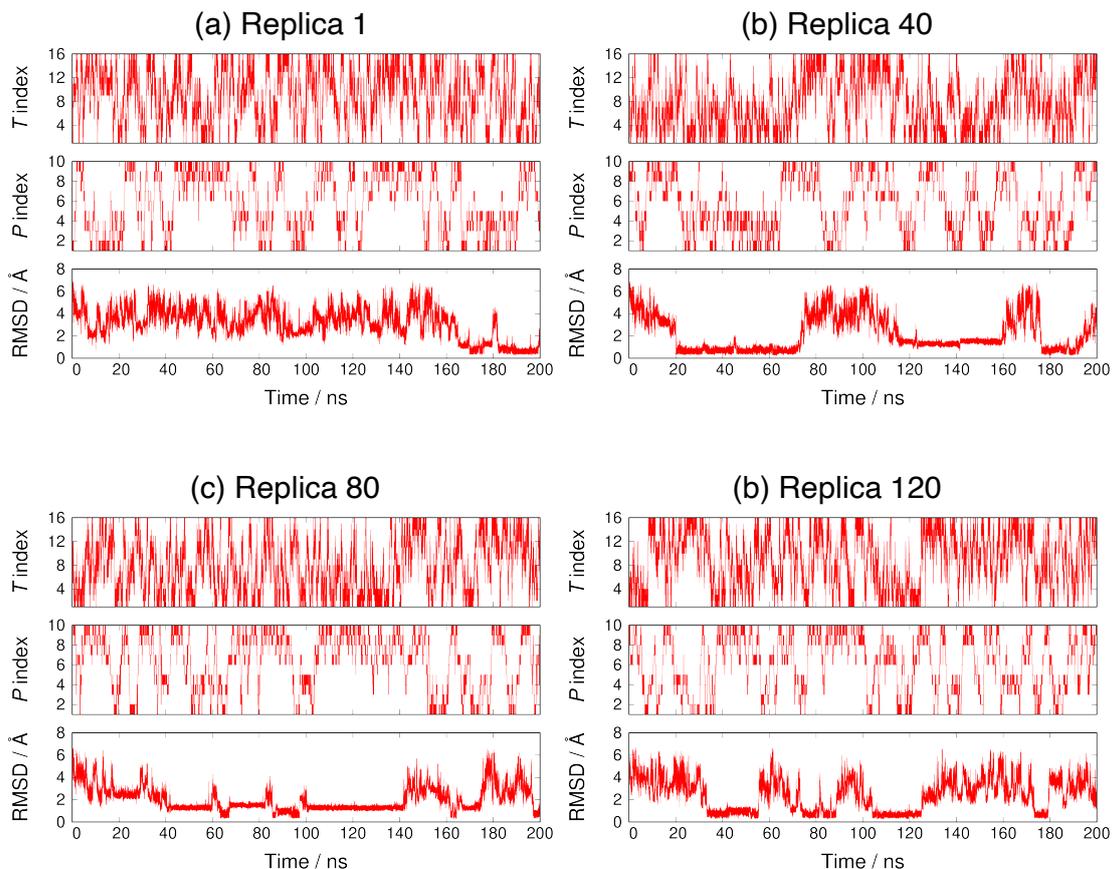
Figure 5.1: Time series of the RMSD, temperature index, and pressure index of (a) replica 1 (b) replica 40, (c) replica 80, and (d) replica 120.

author give an account of the definitions and characteristics of the folded and misfolded states. According to Honda *et al.* [8], the folded structure has β-hairpin conformation and forms three dominant hydrogen bonds between the backbone N and O: Asp3O–Gly7N, Asp3O–Thr8N, and Asp3N–Thr8O. The misfolded structure is another β-hairpin conformation. Previous computational studies revealed that the hydrogen bonds Asp3O–Gly7N and Asp3N–Gly7O are typically formed in the misfolded structure. Hence, the folded and misfolded structure can be distinguished by the hydrogen bonds in the main chain. One can use Define Secondary Structure of Proteins (DSSP) algorithm [27] to determine the states. However, the DSSP is too severe to recognize the folded and misfolded structures because some structures that are quite similar to an ideal folded or misfolded structure are not recognized as the folded or misfolded state. Even some models of the native NMR structure (PDB ID 1UAO) are not recognized as folded.
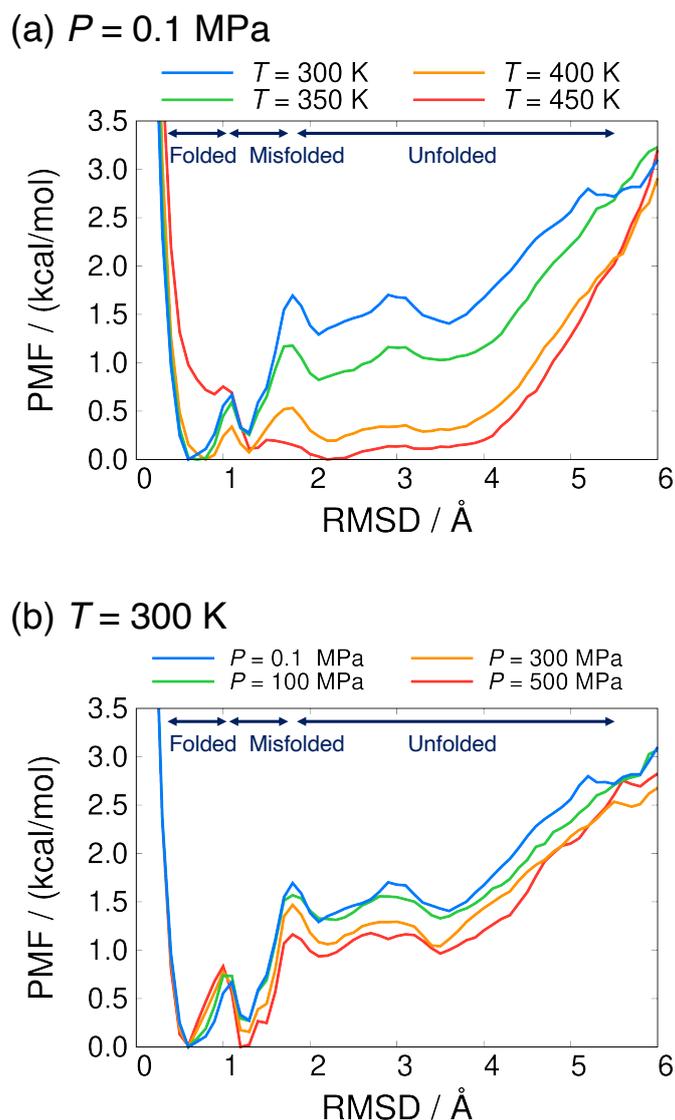
(a) $P = 0.1$ MPa

(b) $T = 300$ K

Figure 5.2: Potential of mean force as a function of RMSD (a) $P = 0.1$ MPa and (b) $T = 300$ K.

To avoid this problem, the author constructed RMSD-based definition to determine the folded, misfolded and unfolded states. The procedure for redefining folded, misfolded, and unfolded states is as follows: First, the author determined whether the structures obtained from the simulation were consistent with folded, misfolded, or unfolded state based on the definition based on the hydrogen bonds. Here, the hydrogen bonds were determined by the DSSP. At this time, the misfolded structure that had the highest hydrogen bond energy was chosen as the representative structure of the misfolded state. Next, the author calculated the RMSD with respect to the NMR structure (RMSD1) and the misfolded structure that are cho-

Figure 5.3: (a) Potential of mean force as a function of RMSD 1 and RMSD2 at $T = 300$ K and $P = 0.1$ MPa. (b) Definition of the folded and misfolded states. The minimum value is set to zero.

sen earlier (RMSD2). Then, the author calculated the two-dimensional PMF as a function of RMSD1 and RMSD2. Here, the PMF was calculated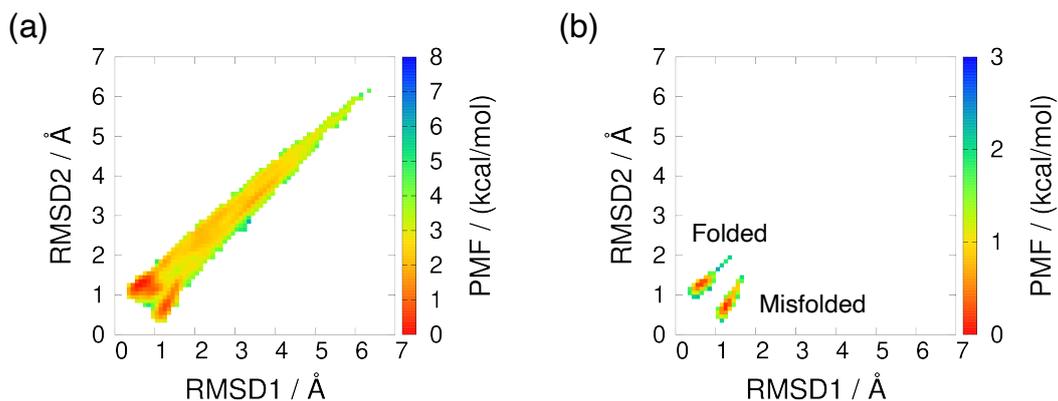 at $T = 300$ K and $P = 0.1$ MPa and the minimum value was set to zero. The result is shown in 5.3 (a). There are two minimum states around (RMSD1, RMSD2) = (0.6, 1.2) and (1.3, 0.6). To divide two minima clearly, the author also calculated PMF using only the folded or misfolded structures that satisfied the definition based on hydrogen bonds. There are two clusters that correspond to the folded and misfolded structures in Fig. 5.3 (b). Both the folded state minimum and the misfolded state minimum shown in Fig. 5.3 (b) were set to zero independently. The region where the PMF was less than 3 kcal/mol was considered as a folded or misfolded state. That is, when the RMSD1 and RMSD2 of the chignolin were in these regions, the structure was regarded as the corresponding state (i.e., folded or misfolded state).

The typical conformations of the folded and misfolded states obtained from our simulation are shown in Fig. 5.4. The folded and misfolded structures have a similar β-hairpin structure. However, some differences exist. One of the two β-strands in the β-hairpin structure is shifted along the other β-strand by one residue. Another difference is that Tyr2 and Trp9 in the folded state form a hydrophobic core, whereas the side chains of these residues are on opposite sides in the misfolded state.

(a) Folded state

Hydrophobic core

Side view Front view Back view

(b) Misfolded state

Opposite sides

Side view Front view Back view
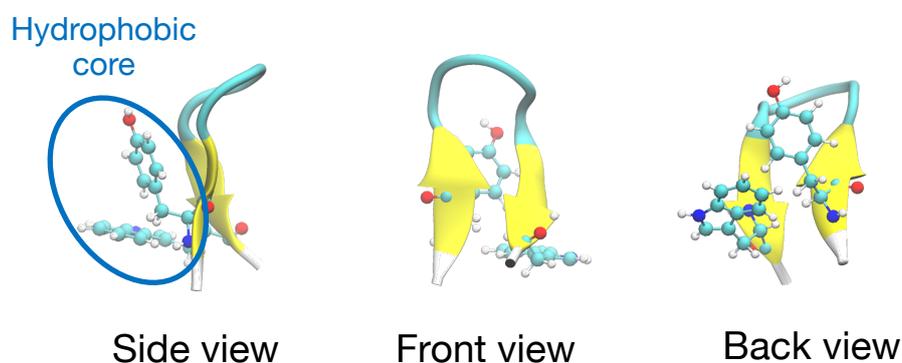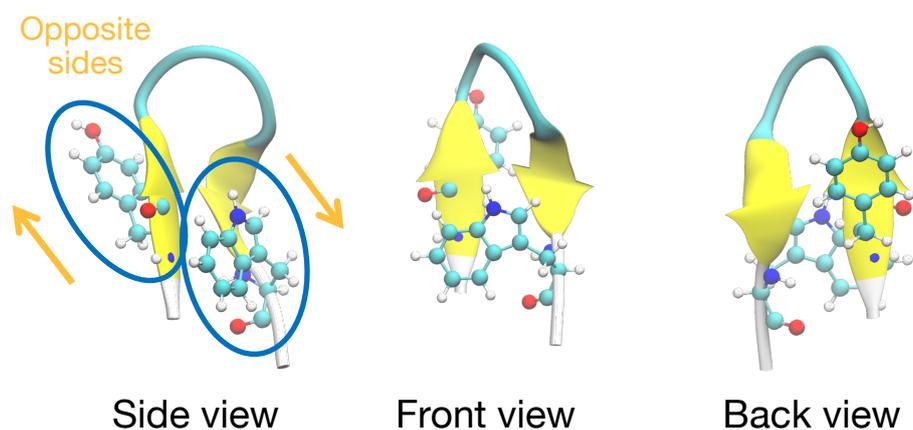
Figure 5.4: Typical conformations in the (a) folded state and (b) misfolded state.

5.3.3 Thermodynamics Quantities between Folded and Unfolded States

The RMSD-based definition was used to calculate the fractions of the folded, misfolded, and unfolded states. Figure 5.5(a) shows temperature dependence of the fraction. As the temperature increases, the fractions of the folded and misfolded states decrease. This indicates that folded or misfolded chignolin is denatured by temperature. Figure 5.5(b) shows their pressure dependence of the fraction. As the pressure increases, the fraction of the folded state decreases, whereas that of the misfolded state increases at $T = 300$ K. This result implies that folded chignolin is denatured by pressure, while misfolded chignolin becomes more stable by pressure.
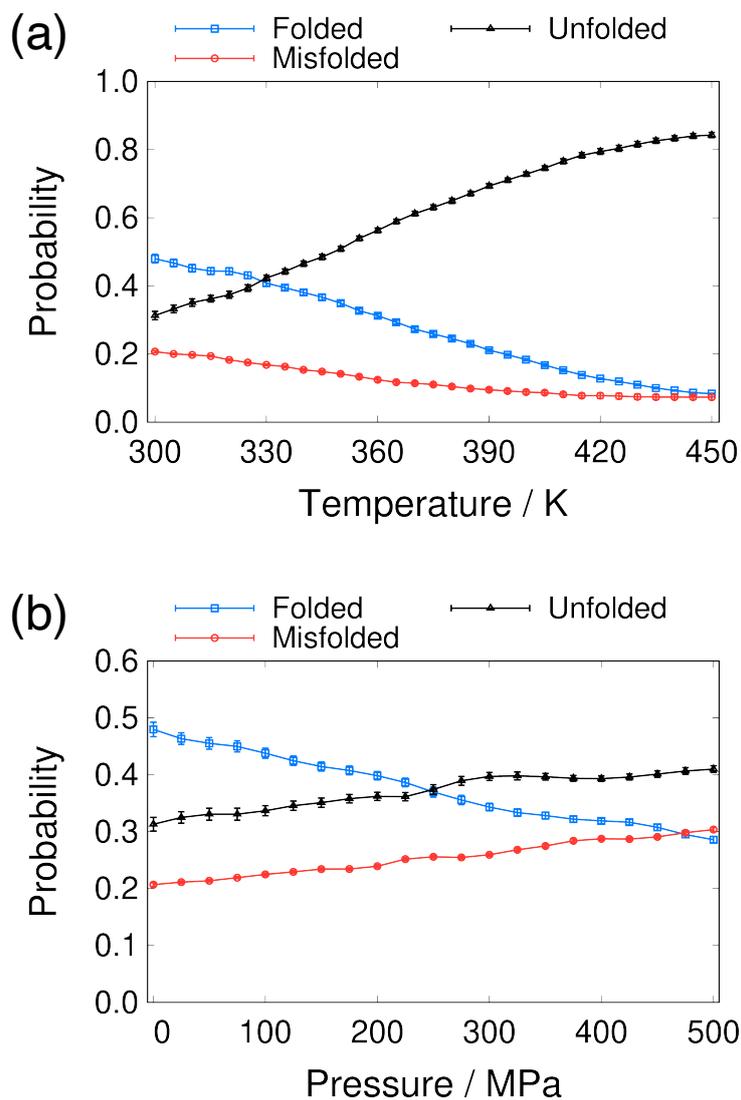
Figure 5.5: (a) Temperature and (b) pressure dependence of the fractions of the folded, misfolded, and unfolded states.

The folded and misfolded chignolin exhibit different behavior against pressure in spite of their similar structures.

Thermodynamics quantities are obtained from the fractions. The difference in the Gibbs free energy ΔG between the unfolded state and the folded state is calculated as follows:

$$\Delta G^{\mathrm{unfold}} = G_{\mathrm{unfold}} - G_{\mathrm{fold}} = RT \log \left(\frac{f_{\mathrm{fold}}}{f_{\mathrm{unfold}}} \right), \tag{5.3}$$

where R is the gas constant, $R = 8.3145$ J/(mol K). The Gibbs free energy difference

ΔG^{unfold} was fitted using the equation proposed by Hawley [28]:

$$\Delta G(T, P) = \Delta G_0 - \Delta S_0(T - T_0) - \Delta C_P \left[T \left\{ \ln\left(\frac{T}{T_0}\right) - 1 \right\} + T_0 \right]$$
$$+ \Delta V_0(P - P_0) + \frac{\Delta\beta}{2}(P - P_0)^2 + \Delta\alpha(P - P_0)(T - T_0), \tag{5.4}$$

where T_0 and P_0 are the reference temperature and pressure, respectively, and ΔG_0, ΔS_0, and ΔV_0 are the differences in the Gibbs free energy, entropy, and partial molar volume, respectively at T_0 and P_0. Here, ΔG_0 was obtained by the RPMD simulation at T_0 and P_0. ΔC_P is the difference in the heat capacity. Further, $\Delta\alpha$ and $\Delta\beta$ are the differences in the thermal expansivity factor and compressibility factor, respectively; these factors are related to, but are not the same as, the thermal expansion coefficient and compressibility in thermodynamics, respectively. Based on Eq. (5.4), temperature and pressure dependence of ΔG were fitted by the following equations:

$$\Delta G(T) = \Delta G_0 - \Delta S_0(T - T_0) - \Delta C_P \left[T \left\{ \ln\left(\frac{T}{T_0}\right) - 1 \right\} + T_0 \right], \tag{5.5}$$

$$\Delta G(P) = \Delta G_0 + \Delta V_0(P - P_0) + \frac{\Delta\beta}{2}(P - P_0)^2. \tag{5.6}$$

The denaturation temperature T_d and denaturation pressure P_d were defined as the values where $\Delta G^{\text{unfold}} = 0$.

Figure 5.6 shows the temperature and pressure dependence of ΔG^{unfold}. Thermodynamic parameters obtained from fitting of $\Delta G^{\text{unfold}}(T)$ and $\Delta G^{\text{unfold}}(P)$ are listed in Tables 5.1 and 5.2. Thermodynamic parameters obtained from fitting of $\Delta G^{\text{unfold}}(T, P)$ are listed in Table 5.3. Here, the error bars of these physical quantities were calculated by the jackknife method [29, 30].

The differences in the partial molar enthalpy of unfolding ΔH^{unfold}, the heat capacity $\Delta C_P^{\text{unfold}}$, and the partial molar volume ΔV^{unfold} of unfolding were calculated as follows:

$$\Delta H^{\text{unfold}} = \left[\frac{\partial \left(\Delta G^{\text{unfold}}/T\right)}{\partial \left(1/T\right)} \right]_P = R \left[\frac{\partial \log(f_{\text{fold}}/f_{\text{unfold}})}{\partial(1/T)} \right]_P, \tag{5.7}$$

$$\Delta C_p^{\text{unfold}} = -T \left(\frac{\partial^2 \Delta G^{\text{unfold}}}{\partial T^2} \right)_P, \tag{5.8}$$

$$\Delta V^{\text{unfold}} = \left[\frac{\partial \Delta G^{\text{unfold}}}{\partial P} \right]_T = RT \left[\frac{\partial \log(f_{\text{fold}}/f_{\text{unfold}})}{\partial P} \right]_T. \tag{5.9}$$

By substituting Eq. (5.4) into Eq. (5.7) and Eq. (5.9), the following equations are obtained:

$$\begin{aligned}
\Delta H^{\text{unfold}}(T, P) =& \Delta G_0 + T_0 \Delta S_0 + (T - T_0) \Delta C_p \\
& + \Delta V_0 (P - P_0) + \frac{\Delta \beta}{2} (P - P_0)^2 \\
& - \Delta \alpha (P - P_0) T_0,
\end{aligned} \tag{5.10}$$

$$\Delta V^{\text{unfold}}(T, P) = \Delta V_0 + \Delta \beta (P - P_0) + \Delta \alpha (T - T_0). \tag{5.11}$$

In this study, the differences in partial molar enthalpy ΔH^{unfold} and partial molar volume ΔV^{unfold} were estimated at $T = T_{\text{d}}$, and the differences in the internal energy ΔU^{unfold} and entropy ΔS^{unfold} were also estimated at T_{d}. The results are listed in Table 5.4. $\Delta H^{\text{unfold}} = 15.1 \pm 1.0$ kJ/mol and $T_{\text{d}} = 327 \pm 3$ K at $P = 0.1$ MPa were obtained. In the experiment by Honda et $al.$ [8], ΔH^{unfold} and T_{d} were estimated as 25.9–32.2 kJ/mol and 311–316 K, respectively. The thermodynamics quantities from the simulation is in reasonably good agreement with the experiment results of Honda et al. Note that Honda et $al.$ set $\Delta C_p^{\text{unfold}} = 0$ to fit their experimental data because their ΔG^{unfold} was almost a straight line. Then, the author also estimated ΔH^{unfold} and T_{d} assuming that $\Delta C_p^{\text{unfold}} = 0$. That is, ΔG^{unfold} for temperature dependence was fitted using a first-order polynomial. In this condition, $\Delta H^{\text{unfold}} = 18.9 \pm 0.4$ kJ/mol and $T_{\text{d}} = 318 \pm 3$ K are obtained. These results are in better agreement with the experimental results. The difference in the partial molar enthalpy ΔH^{unfold} is positive, meaning that the partial molar enthalpy of the folded structures is lower than that of the unfolded structures. Here, ΔU^{unfold} is dominant for ΔH^{unfold} at any pressure. That is, that the folded structures are enthalpically (in particular, energetically) stable. In terms of molecular mechanics, this implies that the folded structures have more intramolecular hydrogen bonds than the unfolded structures. Further, ΔS^{unfold} decreases as pressure increases. The unfolded structures include

Table 5.1: Thermodynamics parameters obtained from fitting of $\Delta G(T)$.

	P (MPa)	T_0 (K)	ΔG_0 (kJ/mol)	ΔS_0 (J/mol/K)	ΔC_p (kJ/mol/K)
$\Delta G = G_{\text{unfold}} - G_{\text{fold}}$	0.1	300	1.06 ± 0.16	32.5 ± 1.6	0.1590 ± 0.0049
	100	300	0.66 ± 0.11	33.6 ± 0.9	0.1184 ± 0.0026
	200	300	0.24 ± 0.10	27.3 ± 0.7	0.1467 ± 0.0021
	300	300	-0.36 ± 0.09	21.9 ± 0.3	0.1662 ± 0.0005
	400	300	-0.52 ± 0.06	24.9 ± 0.2	0.1423 ± 0.0008
	500	300	-0.90 ± 0.08	21.8 ± 0.7	0.1595 ± 0.0028
$\Delta G = G_{\text{misfold}} - G_{\text{fold}}$	0.1	300	2.10 ± 0.09	-26.8 ± 0.5	0.170 ± 0.010
	100	300	1.66 ± 0.06	-17.4 ± 0.7	0.092 ± 0.007
	200	300	1.27 ± 0.06	-18.6 ± 0.4	0.083 ± 0.009
	300	300	0.70 ± 0.06	-18.1 ± 0.4	0.071 ± 0.008
	400	300	0.26 ± 0.05	-28.2 ± 0.7	0.106 ± 0.006
	500	300	-0.15 ± 0.06	-32.9 ± 0.5	0.131 ± 0.008
$\Delta G = G_{\text{unfold}} - G_{\text{misfold}}$	0.1	300	-1.03 ± 0.10	59.3 ± 1.5	-0.011 ± 0.012
	100	300	-1.01 ± 0.07	51.0 ± 1.5	0.026 ± 0.007
	200	300	-1.03 ± 0.06	45.9 ± 0.9	0.063 ± 0.009
	300	300	-1.06 ± 0.06	40.0 ± 0.7	0.095 ± 0.008
	400	300	-0.78 ± 0.04	53.1 ± 0.9	0.036 ± 0.006
	500	300	-0.75 ± 0.05	54.7 ± 0.5	0.028 ± 0.008

Table 5.2: Thermodynamics parameters obtained from fitting of $\Delta G(P)$.

	T (K)	P_0 (MPa)	ΔG_0 (kJ/mol)	ΔV_0 (cm³/mol)	$\Delta \beta$ (cm³/mol/MPa)
$\Delta G = G_{\text{unfold}} - G_{\text{fold}}$	300	0.1	1.06 ± 0.16	-4.90 ± 0.40	0.0040 ± 0.0010
	350	0.1	-1.09 ± 0.10	-2.88 ± 0.18	-0.0002 ± 0.0005
	400	0.1	-4.58 ± 0.08	-0.73 ± 0.06	-0.0043 ± 0.0002
	450	0.1	-8.63 ± 0.09	2.16 ± 0.22	-0.0135 ± 0.0008
$\Delta G = G_{\text{misfold}} - G_{\text{fold}}$	300	0.1	2.10 ± 0.09	-4.6 ± 0.2	0.0005 ± 0.0007
	350	0.1	2.62 ± 0.11	-4.4 ± 0.1	0.0036 ± 0.0004
	400	0.1	2.43 ± 0.25	-3.7 ± 0.6	0.0075 ± 0.0019
	450	0.1	0.50 ± 0.37	7.0 ± 0.6	-0.0278 ± 0.0022
$\Delta G = G_{\text{unfold}} - G_{\text{misfold}}$	300	0.1	-1.0 ± 0.1	-0.3 ± 0.3	0.0036 ± 0.0008
	350	0.1	-3.7 ± 0.1	1.5 ± 0.1	-0.0038 ± 0.0005
	400	0.1	-7.0 ± 0.3	2.9 ± 0.6	-0.0118 ± 0.0020
	450	0.1	-9.1 ± 0.4	-4.9 ± 0.7	0.0143 ± 0.0024

not only coil-like structures, but also helical structures. The decrease in ΔS^{unfold} implies that some structures are suppressed by high pressure. The difference in heat capacities are positive at all pressures. The folded structures are stabilized by the hydrogen bonds, and the fluctuations are constrained as compared to the unfolded state. Therefore, the folded structures have lower heat capacity than the unfolded structures.

Table 5.3: Thermodynamic parameters obtained from fitting of $\Delta G^{\text{unfold}}(T,\ P)$.

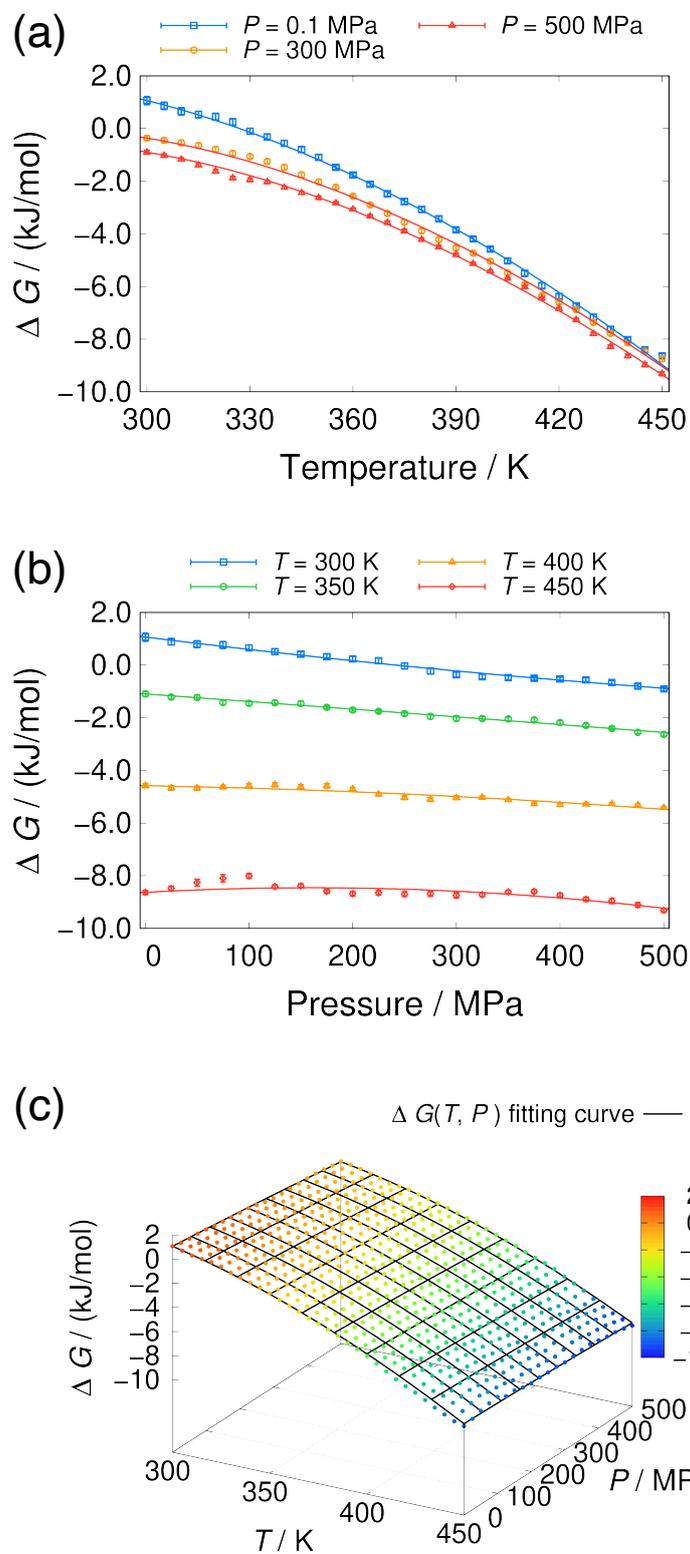| Thermodynamics parameters	value
T_0 (K)	300
P_0 (MPa)	0.1
ΔG_0 (kJ/mol)	1.06
ΔS_0 (kJ/mol/K)	0.034 ± 0.001
ΔC_P (kJ/mol/K)	0.136 ± 0.004
ΔV_0 (cm^3/mol)	-3.5 ± 0.3
$\Delta \alpha$ (cm^3/mol/K)	0.018 ± 0.001
$\Delta \beta$ (cm^3/mol/MPa)	-0.0021 ± 0.0004

Figure 5.6: The differences in Gibbs free energy between folded and unfolded states. (a) Temperature, (b) pressure, (c) both temperature and pressure dependence.

Table 5.4: Differences in thermodynamics quantities between the folded and unfolded states at the denaturation temperature T_d.

P (MPa)	T_d (K)	ΔH (kJ/mol)[1]	ΔC_P (kJ/(mol K))	ΔV (cm³/mol) [2]	$P\Delta V$ (kJ/mol)	ΔU (kJ/mol)	ΔS (J/mol/K)
0.1	327 ± 3	15.1 ± 0.9	0.1590 ± 0.0049	−2.99 ± 0.18	$(-2.99 \pm 0.18) \times 10^{-4}$	15.1 ± 0.9	46 ± 3
100	318 ± 3	12.8 ± 0.6	0.1184 ± 0.0026	−3.37 ± 0.16	−0.34 ± 0.02	13.2 ± 0.6	40 ± 2
200	308 ± 3	9.6 ± 0.7	0.1467 ± 0.0021	−3.76 ± 0.12	−0.75 ± 0.02	10.4 ± 0.7	31 ± 2
300	278 ± 2	2.5 ± 0.3	0.1662 ± 0.0005	−4.55 ± 0.04	−1.36 ± 0.01	3.9 ± 0.3	9 ± 1
400	270 ± 7	2.7 ± 1.0	0.1423 ± 0.0008	−4.87 ± 0.05	−1.95 ± 0.02	4.7 ± 1.0	10 ± 4

[1] ΔH were calculated at $T_0 = 300$ K and $P_0 = P$. [2] ΔV were calculated at $T_0 = 300$ K and $P_0 = 0.1$ MPa.

Table 5.5: The partial molar volume difference ΔV between the folded and unfolded state and the denaturation pressure P_{d}.

T (K)	ΔV (cm^3/mol)	P_{d} (MPa)
300	-4.0 ± 0.2	242 ± 24
350	-1.8 ± 0.5	-387 ± 76

The difference in the partial molar volume of unfolding $\Delta V^{\mathrm{unfold}}$ are listed in Table 5.5. According to Imamura *et al.* the difference in the partial molar volume was estimated as $\Delta V^{\mathrm{unfold}} = -8.8$ cm^3/mol by the Fourier transform infrared spectrometer experiment at $T = 298$ K [6]. The difference in the partial molar volume of unfolding $\Delta V^{\mathrm{unfold}}$ estimated from the simulation are lower than the experimental data, but they are still in good agreement with the experimental data. The absolute value of $\Delta V^{\mathrm{unfold}}$ becomes small at high temperatures.

5.3.4 Thermodynamics Quantities between Folded and Misfolded States

Figure 5.7 shows differences in the Gibss free energy between folded and misfolded state, $\Delta G^{\mathrm{misfold}} = G_{\mathrm{misfold}} - G_{\mathrm{fold}}$. Temperature and pressure dependence of $\Delta G^{\mathrm{misfold}}$ were fitted using Eq. (5.5) and Eq. (5.6), respectively. Thermodynamic parameters obtained from fitting of $\Delta G^{\mathrm{misfold}}(T)$ and $\Delta G^{\mathrm{misfold}}(P)$ are listed in Tables 5.1 and 5.2. The thermodynamic quantities for the folded and misfolded states at 300 K are summarized in Table 5.6. The author found that the misfolded state is enthalpically stable, while the folded state is entropically stable. The contribution of $T\Delta S^{\mathrm{misfold}}$ to $\Delta G^{\mathrm{misfold}}$ is greater than that of $\Delta H^{\mathrm{misfold}}$ at $P = 0.1 - 400$ MPa. However, the contribution of $\Delta H^{\mathrm{misfold}}$ to $\Delta G^{\mathrm{misfold}}$ is slightly greater than that of $T\Delta S^{\mathrm{misfold}}$ at $P = 500$ MPa. The isobaric heat capacity was evaluated as $\Delta C_p^{\mathrm{misfold}} = C_p^{\mathrm{misfold}} - C_p^{\mathrm{fold}} = 0.17 \pm 0.01$ kJ/mol/K, meaning that the fluctuation of enthalpy of the misfolded state is larger than that of the folded state. This is because the β-strand of the misfolded state is shifted by one residue, and the C-terminus has a large fluctuation.
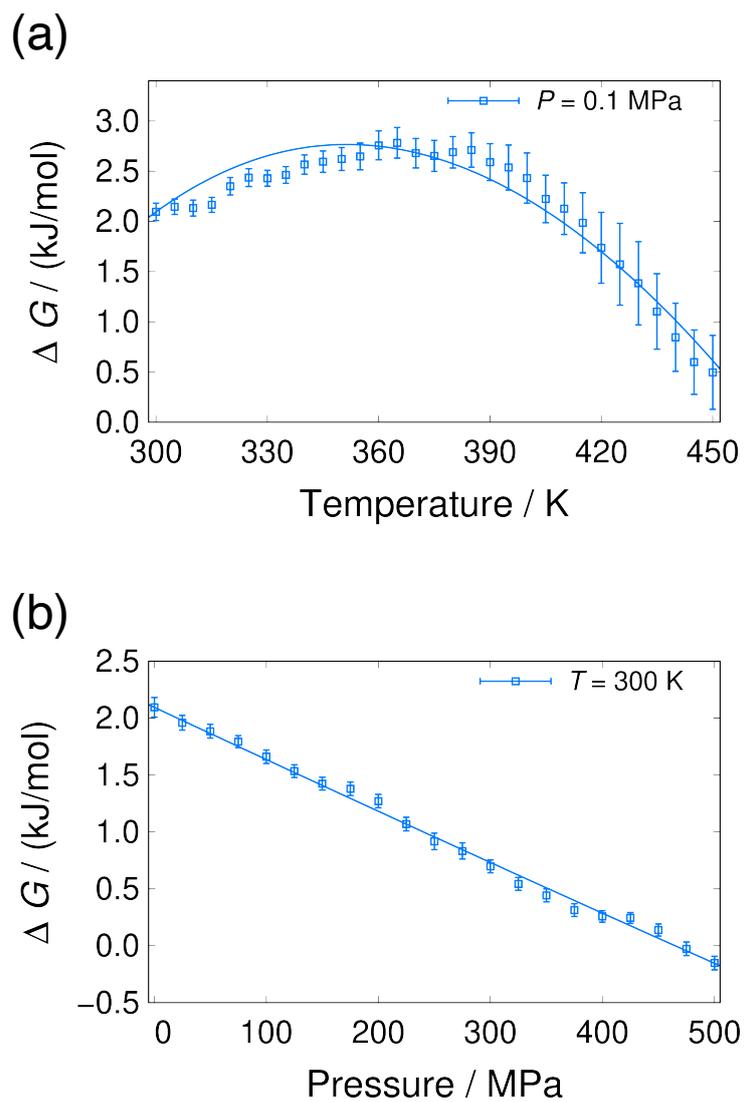
(a)

(b)

Figure 5.7: The differences in Gibbs free energy between folded and misfolded states. (a) Temperature and (b) pressure dependence.

Table 5.6: Differences in thermodynamics quantities between the folded and misfolded states at 300 K.

P (MPa)	ΔG (kJ/mol) [1]	ΔH (kJ/mol) [1]	ΔV (cm^3/mol) [2]	$P\Delta V$ (kJ/mol)	ΔU (kJ/mol)	$T\Delta S$ (kJ/mol)	ΔS (J/mol/K)
0.1	2.10 ± 0.09	-5.9 ± 0.2	-4.61 ± 0.22	$(-4.61 \pm 0.22) \times 10^{-4}$	-5.9 ± 0.2	-8.0 ± 0.2	-26.8 ± 0.7
100	1.66 ± 0.06	-3.6 ± 0.2	-4.57 ± 0.15	-0.46 ± 0.01	-3.1 ± 0.2	-5.2 ± 0.2	-17.4 ± 0.6
200	1.27 ± 0.06	-4.3 ± 0.1	-4.52 ± 0.08	-0.90 ± 0.02	-3.4 ± 0.1	-5.6 ± 0.1	-18.6 ± 0.4
300	0.70 ± 0.06	-4.7 ± 0.1	-4.48 ± 0.03	-1.34 ± 0.01	-3.4 ± 0.1	-5.4 ± 0.1	-18.1 ± 0.4
400	0.26 ± 0.05	-8.2 ± 0.2	-4.43 ± 0.07	-1.77 ± 0.03	-6.4 ± 0.2	-8.5 ± 0.2	-28.2 ± 0.7
500	-0.15 ± 0.06	-10.0 ± 0.2	-4.39 ± 0.14	-2.19 ± 0.07	-7.8 ± 0.2	-9.9 ± 0.2	-32.9 ± 0.6

[1]ΔH were calculated at $T_0 = 300$ K and $P_0 = P$. [2]ΔV were calculated at $T_0 = 300$ K and $P_0 = 0.1$ MPa.

Moreover, the author calculated the difference in the partial molar volume ΔV between the folded and misfolded states by using Eq. (5.11): $\Delta V^{\text{misfold}} = V_{\text{misfolded}} - V_{\text{folded}} = -4.6 \pm 0.2$ cm^3/mol at $T = 300$ K. This result means that the partial molar volume of the misfolded structure is smaller than that of the folded structure.

5.3.5 Stabilization Mechanism of the Misfolded State under High Pressure Conditions

Let us discuss the reason why the fraction of the folded state decrease while that of the misfolded state increases as pressure increases. Figure 5.8 presents the average distance between the hydrogen bonds (Asp3O–Gly7N, Asp3O–Thr8N, and Asp3N–Thr8O for the folded state, and Asp3O–Gly7N and Asp3N–Gly7O for the misfolded state) and the closest water molecule. Here, the distance between the hydrogen bonds and the closest water molecule was defined as follows:

$$r = \frac{r_{\text{N}-\text{H}_2\text{O}} + r_{\text{O}-\text{H}_2\text{O}}}{2}, \tag{5.12}$$

where $r_{\text{N}-\text{H}_2\text{O}}$ is the distance between the backbone N atom and the O atom of the closest water molecule, and $r_{\text{O}-\text{H}_2\text{O}}$ is the distance between the backbone O atom and the O atom of the water molecule. Fig. 5.8 shows that the water molecules were located closer to the hydrogen bond Asp3N–Thr8O of the folded state than to the other hydrogen bonds. Since the folded state has a hydrophobic core between Tyr2 and Trp9 at the back side, the hydrogen bond Asp3N–Thr8O is exposed to the solvent at the front of chignolin. In Fig. 5.9(a), there is space around the hydrogen bond Asp3N–Thr8O at the front side of the folded structure. The other hydrogen bonds, Asp3O–Gly7N and Asp3O–Thr8N, are hidden inside chignolin. Thus, water molecules can approach Asp3N–Thr8O from the front side and break the hydrogen bonds although they cannot reach close to Asp3O–Gly7N and Asp3O–Thr8N. In contrast, Fig. 5.9(b) shows that Tyr2 and Trp9 of the misfolded state are oriented in the opposite direction; and thus, the hydrogen bonds are covered by the side chains on both sides in the misfolded state and are hidden from the solvent. The Hydrogen bond Asp3O–Gly7N of the misfolded state is the second closest to the
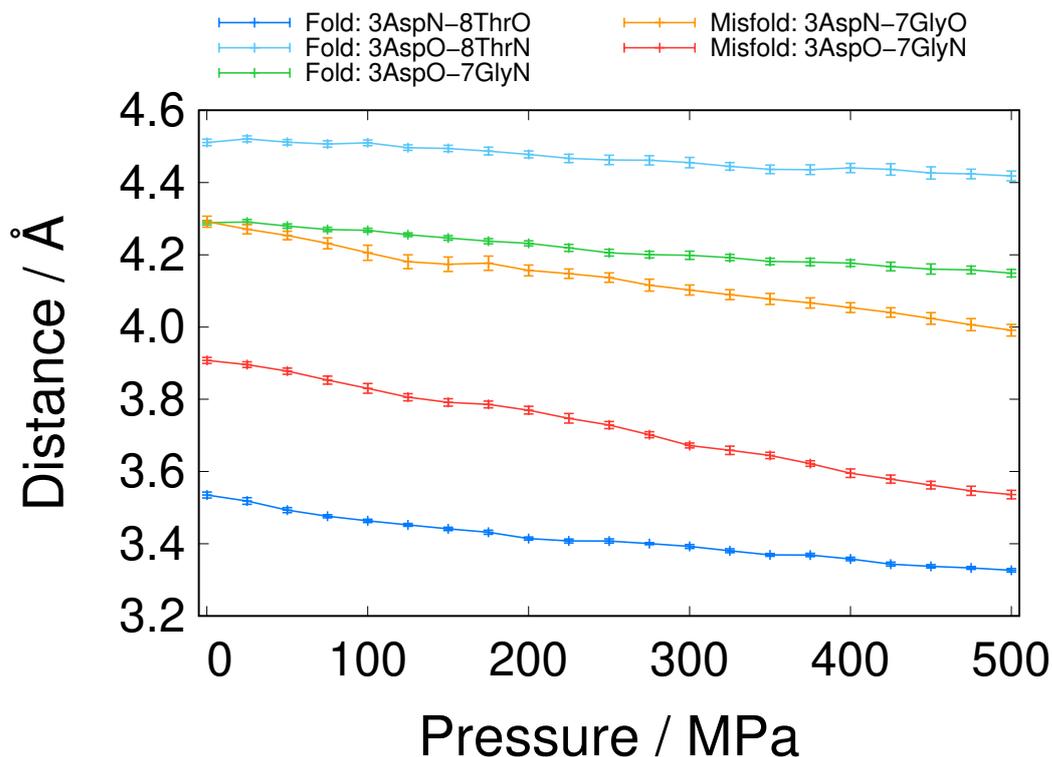
Figure 5.8: Pressure dependence of the average distance between hydrogen bonds and the closest water molecule.

water molecules, as shown in Fig. 5.8. This is because Gly7 has only a hydrogen atom as its side chain, and a part of the hydrogen bond is exposed to the solvent.

The stability of the chignolin can be explained based on the structural characteristics and thermodynamic quantities of chignolin. Water molecules approach the protein as the pressure increase. Water breaks the hydrogen bond Asp3N–Thr8O in the folded state. This result is consistent with the fact that $\Delta U^{\mathrm{unfold}}$ between the folded and unfolded states decrease as pressure increase. Thus, the folded state is broken and changes into the unfolded or misfolded state. Furthermore, the misfolded state becomes more enthalpically stable than the folded state under high-pressure condition. This is because the misfolded state has smaller partial molar volume, and the compact conformations of the misfolded state have lower values of internal energy at high pressure, as evident from the data presented in Table 5.6. Therefore, the fraction of the folded state decreases as pressure increases, while the fractions of the unfolded and misfolded states increase.
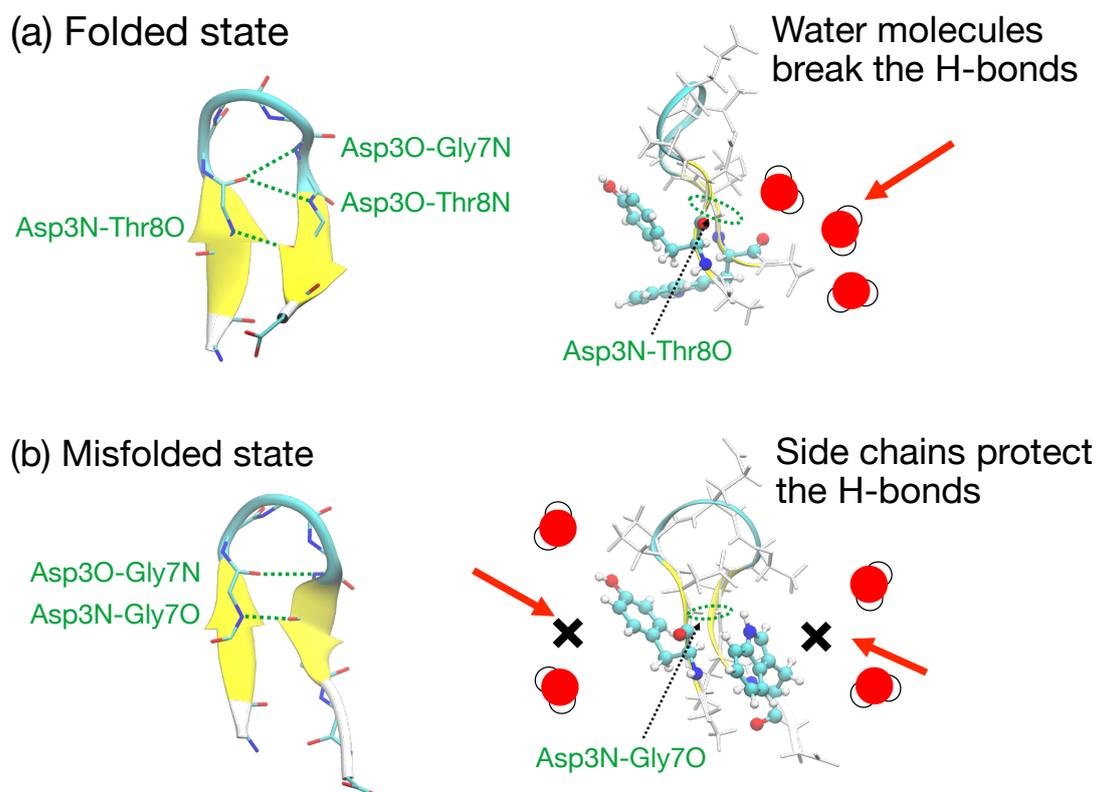
Figure 5.9: Schematic figure of the mechanism by which folded chignolin becomes unstable while misfolded chignolin becomes stable under high-pressure conditions.

5.4 Conclusions

In this study, the author applied the isothermal–isobaric replica-permutation method to a β-hairpin miniprotein, chignolin. Thanks to the efficient sampling by the isothermal–isobaric replica-permutation simulation, the author succeeded in evaluating the thermodynamics quantities such as the differences in the Gibbs free energy ΔG, partial molar enthalpy ΔH, partila molar volume ΔV, and difference in the isobaric heat capacity ΔC_p among folded, misfolded and unfolded states. The obtained thermodynamics quantities are in good agreement with the experimental results.

In this simulation, the author found a new phenomenon; the misfolded structure of the chignolin becomes stable under high pressure conditions, whereas that of the folded structure of the chignolin denature. The author revealed that the molecular mechanism of the different stability between the folded and misfolded struc-

tures under high pressure conditions arise from different orientation of the Tyr2 and Trp9 side chains. In the folded structures, Tyr2 and Trp9 form hydrophobic core. Consequently, the hydrogen bond Asp3N–Thr8O, which is important for the β-hairpin formation, is exposed to the solvents in the opposite side of the hydrophobic core. Water molecules approach the hydrogen bond as pressure increases, and water molecules break the hydrogen bond. Therefore, the folded chignolin transits to the unfolded state. On the other hand, in the misfolded structure, the hydrogen bonds that are important to form the misfolded structure are covered with the Try2 and Trp9 side chains. The side chains protect the hydrogen bonds from the approaching water molecules. Therefore, the misfolded structure becomes more stable with compressing its structure under high-pressure conditions.

References

[1] Jerson L Silva, Andrea C. Oliveira, Tuane C R G Vieira, Guilherme A. P. de Oliveira, Marisa C Suarez, and Debora Foguel. High-Pressure Chemical Biology and Biotechnology. *Chemical Reviews*, 114(14):7239–7267, 2014.

[2] Tigran V. Chalikian and Kenneth J. Breslauer. On volume changes accompanying conformational transitions of biopolymers. *Biopolymers*, 39(5):619–626, 1998.

[3] Julien Roche, Jose A. Caro, Douglas R. Norberto, Philippe Barthe, Christian Roumestand, Jamie L. Schlessman, Angel E. Garcia, Bertrand E. García-Moreno, and Catherine A. Royer. Cavities determine the pressure unfolding of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 109(18):6945–6950, 2012.

[4] Takahiro Takekiyo, Takashi Imai, Minoru Kato, and Yoshihiro Taniguchi. Temperature and Pressure Effects on Conformational Equilibria of Alanine Dipeptide in Aqueous Solution. *Biopolymers*, 73(2):283–290, 2004.

[5] Hiroshi Imamura and Minoru Kato. Effect of pressure on helix-coil transition of an alanine-based peptide: An FTIR study. *Proteins: Structure, Function and Bioinformatics*, 75(4):911–918, 2009.

[6] H Imamura and M Katao. Unfolding of beta-hairpin peptides by pressure: FT-IR and FRET studies. In *The 48th Annual Meeting of the Biophysical Society of Japan*, volume 3P006, sendai, 2010.

[7] Sabine Neumaier, Maren Büttner, Annett Bachmann, and Thomas Kiefhaber. Transition state and ground state properties of the helix-coil transition in pep-

tides deduced from high-pressure studies. *Proceedings of the National Academy of Sciences of the United States of America*, 110(52):20988–20993, 2013.

[8] Shinya Honda, Kazuhiko Yamasaki, Yoshito Sawada, and Hisayuki Morii. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure*, 12(8):1507–1518, 2004.

[9] Hisashi Okumura. Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins: Structure, Function, and Bioinformatics*, 80(10):2397–2416, 2012.

[10] Masataka Yamauchi and Hisashi Okumura. Development of isothermal-isobaric replica-permutation method for molecular dynamics and Monte Carlo simulations and its application to reveal temperature and pressure dependence of folded, misfolded, and unfolded states of chignolin. *The Journal of Chemical Physics*, 147(18):184107, 2017.

[11] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[12] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[13] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[14] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[15] William G Hoover. Canonical dynamics: equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[16] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[17] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[18] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[19] M Tuckerman, Bruce J Berne, and Glenn J Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[20] Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[21] T. F. Miller III, M Eleftheriou, P Pattnaik, A Ndirango, D Newns, and G. J. Martyna. Symplectic quaternion scheme for biophysical molecular dynamics. *The Journal of Chemical Physics*, 116(20):8649–8659, 2002.

[22] Hisashi Okumura, Satoru G Itoh, and Yuko Okamoto. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric-isothermal, and related ensembles. *The Journal of Chemical Physics*, 126(8):084103, 2007.

[23] Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12):124105, 2008.

[24] Evangelos A. Coutsias, Chaok Seok, and Ken A. Dill. Using quaternions to calculate RMSD. *Journal of Computational Chemistry*, 25(15):1849–1857, 2004.

[25] Daisuke Satoh, Kentaro Shimizu, Shugo Nakamura, and Tohru Terada. Folding free-energy landscape of a 10-residue mini-protein, chignolin. *FEBS Letters*, 580(14):3422–3426, 2006.

[26] Ryuhei Harada and Akio Kitao. Exploring the folding free energy landscape of a beta-hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method. *Journal of Physical Chemistry B*, 115(27):8806–8812, 2011.

[27] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[28] S. A. Hawley. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry*, 10(13):2436–2442, 1971.

[29] M. H. Quenouille. Notes on Bias in Estimation. *Biometrika*, 43(3/4):353, 1956.

[30] Rupert G. Miller. The jackknife-a review. *Biometrika*, 61(1):1–15, 1974.

Chapter 6

Dimerization of α-synuclein Fragments

M. Yamauchi and H. Okumura. "Dimerization of α-synuclein fragments studied by isothermal–isobaric replica-permutation molecular dynamics simulation" *The Journal of Chemical Information and Modeling* (in press).

6.1 Introduction

Alpha-synuclein (α-Syn), an intrinsically disordered protein consisting of 140 amino acid residues, is abundantly distributed in the nucleus and presynaptic nerve terminal [1]. In 1997, α-Syn was identified as the major component of Lewy bodies and Lewy neurites, which are pathological hallmarks of Parkinson's disease, Parkinson's disease dementia, and dementia with Lewy bodies [2]. Although aggregates and amyloid fibrils of α-Syn are believed to be associated with Parkinson's disease, the mechanisms underlying the aggregation and fibril formation of α-Syn have yet to be elucidated. It is necessary to understand the aggregation and fibril formation processes of α-Syn to develop treatment for Parkinson's disease.

The amino acid sequence of α-Syn is typically divided into three regions based on its physicochemical properties: the amphipathic N-terminal domain (residues 1–60); the highly hydrophobic central domain (residues 61–95), which is referred to as the non-amyloid-β component [3]; and the acidic C-terminal domain (96–140). The hydrophobic central domain was originally identified in amyloid taken from the brains of Alzheimer's disease patients after amyloid-β (Aβ) and is referred to as the non-amyloid-β component (NAC). The NAC is an essential region for the aggregation and amyloid fibril formation of α-Syn [4–14] and responsible for cytotoxicity [6, 9, 10, 14]. A variety of fragments within the NAC domain have been investigated to determine an essential region for the amyloid formation and its cytotoxicity. These studies revealed that some fragments within the NAC show a tendency for amyloid formation and cytotoxicity, but the others do not [7, 9, 10, 13]. Among them, a segment from the 68th to 78th residue of α-Syn termed NACore, $_{68}$GAVVTGVTAVA$_{78}$, plays a critical role in both amyloid fibril formation and cytotoxicity [10, 15]. Analysis of the small crystal structure of NACore using micro-electron diffraction revealed that the NACore peptides form steric-zipper protofibrils, in which a pair of parallel β-sheets are stacked in an antiparallel orientation [15]. The NACore peptide lacks the residues that contribute to stabilizing the amyloid fold of α-Syn according to atomic-resolution fibril structures that were recently determined by solid-state NMR [16] and cryo-electron microscopy [17–20]. However, it is one of the best model peptides to investigate the mechanisms underlying aggregation and amyloid

fibril formation of amyloidogenic peptides in general. Although previous experimental and molecular dynamics studies have investigated the properties and stability of the NACore aggregates and fibril structures [21–24], the mechanisms involved in the early stages of aggregation and fibril formation, including the dimerization process, remain unclear.

In this study, the author investigated the dimerization process of the NACore peptides by molecular dynamics simulation. The author employed an isothermal–isobaric replica-permutation molecular dynamics method [25, 26] to achieve efficient sampling. The replica-permutation method [25–28], a generalized-ensemble method, is an improved alternative to the replica-exchange method [29, 30]. In the replica-permutation method, molecular simulations at different temperatures are performed in parallel, and the temperature permutation among more than two replicas is performed during the simulation. Instead of the Metropolis algorithm [31], the Suwa–Todo algorithm [32] is applied in the permutation trials so that a higher transition ratio from one temperature to another can be realized. The author analyzed stabilities and structural properties of the NACore peptides along the dimerization process and then identified possible mechanisms for the NACore dimer formation.

6.2 Materials and Methods

6.2.1 System Preparation

This study focused on the dimerization of the NACore peptides in the initial stage of aggregation and fibril formation. The amino acid sequence of the NACore peptide, a segment from the 68th to 78th residue of α-Syn, is $_{68}$GAVVTGVTAVA$_{78}$ [15]. An initial structure of the isothermal–isobaric replica-permutation simulation was prepared as follows: the author constructed an extended NACore peptide in which bonds and angles were set to equilibrium values and all dihedral angles were set to 180°. The N- and C-termini were not capped following aggregation and fibrillation experiments [15, 24]. Thus, the N- and C-termini of the NACore peptides have a positive ($+1$) and a negative (-1) charge, respectively, but the total charge of the NACore peptides was zero (i.e., neutralized). Then, the two extended NACore
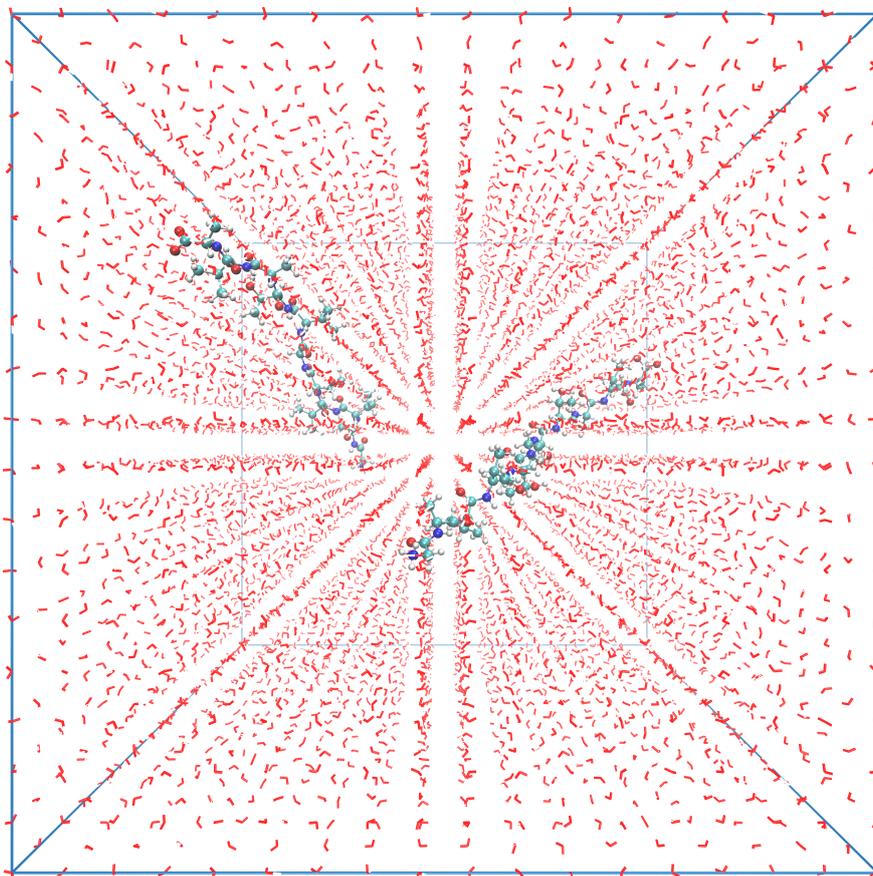
Figure 6.1: The initial structure of the isothermal–isobaric replica-permutation simulation. Two extended NACore peptides and water molecules were placed in a cubic box.

peptides and 5,720 water molecules were placed in a cubic box so that the distance between the two peptides was 16 Å. The orientation of the NACore peptides was randomly determined. No counter ion was contained because a mixture of fibrils and crystals of the NACore was obtained in sterile water in the experiments [15] although the counter ions would affect the low temperature trajectories. The total number of atoms was 17,438. The initial side length of the cubic cell was 55.20 Å. The initial structure of the simulation is illustrated in Fig. 6.1

6.2.2 Computational Conditions

To realize efficient conformation sampling for the NACore peptides, an isothermal–isobaric replica-permutation molecular dynamics method [25, 26] was employed. The isothermal–isobaric replica-permutation simulation was performed with the Generalized Ensemble Molecular Biophysics (GEMB) program developed by the authors. This program has been applied to several proteins [25, 28, 33–42]. A cubic unit cell with periodic boundary conditions was used. The assisted model building with energy refinement (AMBER) param14SB force field [43] was employed for the NACore peptides, and the TIP3P rigid-body model [44] was used for the water molecules. The temperature was controlled by the Nosé–Hoover thermostat [45–47], and the pressure was controlled by the Andersen barostat [48]. Electrostatic interactions were calculated using the particle mesh Ewald method (PME) [49, 50]. The cutoff distance was set to 12.0 Å for the van der Waals interactions and the electrostatic interactions in the direct space sum of the PME. A reversible multiple time-scale molecular dynamics integrator was applied [51]. The symplectic quaternion scheme was adopted for the rigid body water molecules [52–54]. In this integrator, the author supposed that a force F is composed of three contributions, $F = F_{\text{fast}} + F_{\text{intermed}} + F_{\text{slow}}$. F_{fast} is composed of forces derived from bonded interactions of the peptide atoms. F_{intermed} is composed of forces derived from non-bonded interactions (van der Waals and direct space sum of PME) between the peptide atoms and those between the peptide atoms and solvent molecules. F_{fast} is composed of forces from non-bonded interactions (van der Waals and direct space sum of PME) between solvent molecules, and those from the reciprocal space term of PME for all atoms. Time steps for F_{fast}, F_{intermed}, and F_{slow} were taken as $\Delta t = 0.5$ fs, $\Delta t = 2.0$ fs, and $\Delta t = 4.0$ fs, respectively. The time-step could be set as $\Delta t = 4.0$ fs because the symplectic rigid-body algorithm was used for the water molecules [54].

Eighty replicas were used in the isothermal–isobaric replica-permutation simulation. The temperatures of the replicas, T_1 to T_{80}, were distributed from 210.0 to 500.0 K, following $T_m = T_1 \gamma^{m-1}$, where $\gamma = (T_M/T_1)^{\frac{1}{M-1}}$, M is the number of replicas, $T_1 = 210.0$ K, and $T_M = 500.0$ K. The temperature values used in the simulation are listed in Table 6.1. The pressure was controlled at 1 atm. To reduce the

number of combinations between temperatures and replicas that should be considered in one parameter-permutation trial, the replicas and temperatures were divided into subsets [25, 27] so that one subset included eight replicas and temperatures as follows:

$$
X_\mu^{[1-10]} = \left\{
\begin{array}{ll}
X_\mu^1 = & [x_1^{i(1)},\ x_2^{i(2)},\ x_3^{i(3)},\ x_4^{i(4)},\ x_5^{i(5)},\ x_6^{i(6)},\ x_7^{i(7)},\ x_8^{i(8)}] \\
X_\mu^2 = & [x_9^{i(9)},\ x_{10}^{i(10)},\ x_{11}^{i(11)},\ x_{12}^{i(12)},\ x_{13}^{i(13)},\ x_{14}^{i(14)},\ x_{15}^{i(15)},\ x_{16}^{i(16)}] \\
\vdots & \\
X_\mu^{10} = & [x_{73}^{i(73)},\ x_{74}^{i(74)},\ x_{75}^{i(75)},\ x_{76}^{i(76)},\ x_{77}^{i(77)},\ x_{78}^{i(78)},\ x_{79}^{i(79)},\ x_{80}^{i(80)}]
\end{array}
\right\}
\tag{6.1}
$$

Note that the temperature indices in the subsets should be sequential and cyclic. The parameter permutation were performed among replicas in the same subset. Additional subsets are prepared because if the above subsets are used repeatedly, replicas cannot take all temperature values during a simulation. The additional subsets were defined by shifting the temperature indices in each subset cyclically from $X_\mu^{[1-10]}$ as follows:

$$
X_\mu^{[11-20]} = \left\{
\begin{array}{ll}
X_\mu^{11} = & [x_{77}^{i(77)},\ x_{78}^{i(78)},\ x_{79}^{i(79)},\ x_{80}^{i(80)},\ x_1^{i(1)},\ x_2^{i(2)},\ x_3^{i(3)},\ x_4^{i(4)}] \\
X_\mu^{12} = & [x_5^{i(5)},\ x_6^{i(6)},\ x_7^{i(7)},\ x_8^{i(8)},\ x_9^{i(9)},\ x_{10}^{i(10)},\ x_{11}^{i(11)},\ x_{12}^{i(12)}] \\
\vdots & \\
X_\mu^{20} = & [x_{69}^{i(69)},\ x_{70}^{i(70)},\ x_{71}^{i(71)},\ x_{72}^{i(72)},\ x_{73}^{i(73)},\ x_{74}^{i(74)},\ x_{75}^{i(75)},\ x_{76}^{i(76)}]
\end{array}
\right\}
\tag{6.2}
$$

Replica permutation in the subsets $X_\mu^{[1-10]}$ were carried out independently at odd-numbered permutation trials, and replica permutations in the subsets $X_\mu^{[11-20]}$ were carried out independently at even-numbered trials, which enable the replicas to take all temperature values during the simulation. The replica-permutation trials were performed every 2 ps. The trajectory data were stored every 2 ps. The simulation was performed for 1.2 μs, including 0.2 μs equilibrium per replica. The total simulation time for all replicas was 96 μs.

Table 6.1: Temperatures used in the isothermal–isobaric replica-permutation simulation.

indices	temperatures (K)
T_1 –T_{10}	210.0, 212.3, 214.7, 217.0, 219.4, 221.9, 224.3, 226.8, 229.3, 231.8
T_{11}–T_{20}	234.4, 237.0, 239.6, 242.2, 244.9, 247.6, 250.3, 253.1, 255.9, 258.7
T_{21}–T_{30}	261.6, 264.5, 267.4, 270.3, 273.3, 276.3, 279.4, 282.5, 285.6, 288.7
T_{31}–T_{40}	291.9, 295.2, 298.4, 301.7, 305.0, 308.4, 311.8, 315.3, 318.7, 322.3
T_{41}–T_{50}	325.8, 329.4, 333.1, 336.7, 340.5, 344.2, 348.0, 351.9, 355.7, 359.7
T_{51}–T_{60}	363.6, 367.7, 371.7, 375.8, 380.0, 384.2, 388.4, 392.7, 397.0, 401.4
T_{61}–T_{70}	405.8, 410.3, 414.9, 419.4, 424.1, 428.7, 433.5, 438.3, 443.1, 448.0
T_{71}–T_{80}	452.9, 457.9, 463.0, 468.1, 473.3, 478.5, 483.8, 489.1, 494.5, 500.0

6.2.3   A Measure of the Distance between Peptides

A reaction coordinate $d_{\alpha\alpha}$ is introduced as a measure of distance between two NA-Core peptides. The reaction coordinate $d_{\alpha\alpha}$ is defined as the shortest C_α–C_α distance between the first and second NACore peptides as follows:

$$d_{\alpha\alpha}(t_k) \equiv \min\{r_{ij}^{\alpha\alpha}(t_k) \mid \forall\, i \in \text{NACore 1}, \forall\, j \in \text{NACore 2}\}, \qquad (6.3)$$

where $r_{ij}^{\alpha\alpha}(t_k)$ is the distance between the C_α atoms of residues i and j at time-step t_k. It is assumed that i and j are residue indices in the first and second NACore peptides, respectively.

6.2.4   Error Estimation of the Physical Quantities

In the analyses, errors of physical quantities were estimated by the jackknife method [55–57]. To employ the jackknife method, the trajectories from the production run were divided into five segments. Therefore, the time length of the trajectories included in one segment was 200 ns.

6.3   Results and Discussion

6.3.1   Validation of Sampling Quality

Figure 6.2(a) shows the time series of temperatures of the replicas. To evaluate the sampling quality in the temperature space, the author computed the number

of tunneling events of the replicas. Here, one tunneling event in the temperature space was defined as a round trip between the low-temperature region (210.0 K–298.4K) and the high-temperature region (401.4K–500.0K). The calculation of the tunneling events was performed after the equilibration period of 200 ns (i.e., the time series from 200 to 1200 ns were used for the evaluation). The average number of tunneling events per replica was 21 ± 7. Here, the error was estimated by calculating the standard deviation over replicas. Almost all replicas took various temperatures from low to high during the simulation. However, some replicas were trapped in the low-temperature region for long time-steps, resulting in poor statistics for estimating physical quantities at low temperatures due to the correlation between snapshots. Therefore, the analyses in this study were carried out using a trajectory at $T_{34} = 301.7$ K if there is no special remark. The average side length of the simulation box at $T_{34} = 301.7$ K was 56.30 ± 0.07 Å.

Figure 6.2(b) shows the time series of distance between the NACore peptides, $d_{\alpha\alpha}$, of the replicas. The number of round trips between the association and dissociation states were computed to confirm the sampling quality in the conformational space. The author regarded the association state as that with $d_{\alpha\alpha} \leq 4.0$ Å and the dissociation state as that with $d_{\alpha\alpha} \geq 20.0$ Å. Here, the threshold $d_{\alpha\alpha} = 4.0$ Å was selected because an intermolecular hydrogen bond can be formed between main chains. The other threshold $d_{\alpha\alpha} = 20.0$ Å was selected as enough far distance in which the NACore peptides can be regarded as monomers. Note that the average side length of the simulation cells over all replicas and temperatures was 57.6 ± 2.3 Å. The average number of round trips was 125 ± 54, indicating that the simulation succeeded in sampling a sufficient number of association-dissociation events of the NACore peptides. To confirm convergence of the isothermal–isobaric replica-permutation simulation at $T_{34} = 301.7$ K, the author calculated probability distributions over the time intervals of 200–400 ns, 200–600 ns, 200–800 ns, 200–1000 ns, and 200–1200 ns in Fig. 6.2(c). The probability distributions for the time intervals over 200–800 ns, 200–1000 ns, and 200–1200 ns remain unchanged, meaning that the simulation reached equilibrium.
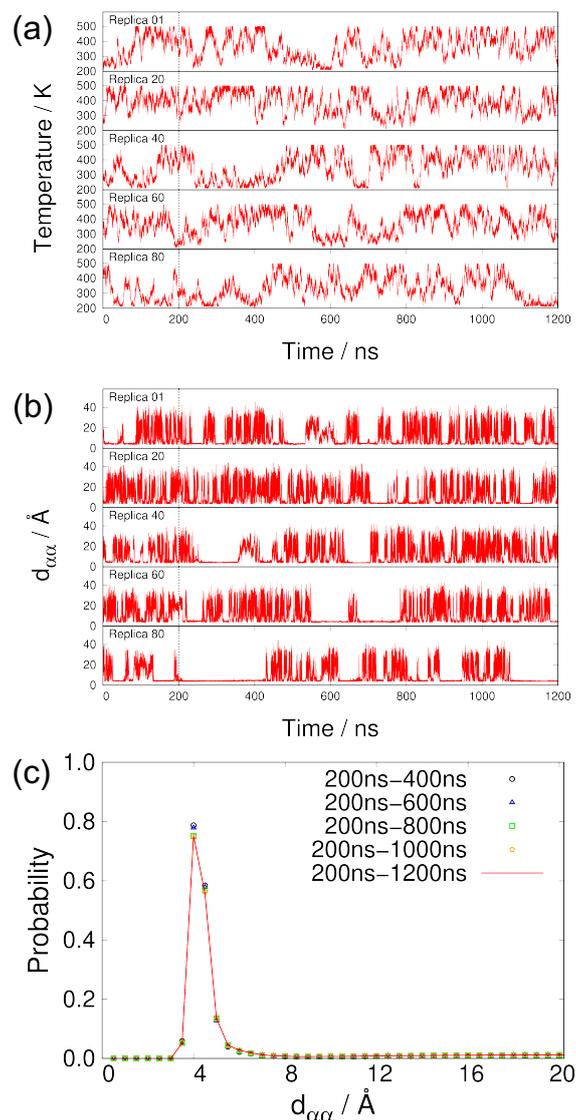
Figure 6.2: Time series of (a) temperatures and (b) distances between the NACore peptides, $d_{\alpha\alpha}$, for replicas 1, 20, 40, 60, and 80. $d_{\alpha\alpha}$ is defined as the shortest intermolecular C_α–C_α distance between the first and second NACore peptides. (c) Probability distributions of $d_{\alpha\alpha}$ at $T_{34} = 301.7$ K over the time intervals of 200–400 ns, 200–600 ns, 200–800 ns, 200–1000 ns, and 200–1200 ns.

6.3.2 Dimerization Tendency of the NACore Peptides

To investigate the tendency of dimerization of the NACore peptides, the author computed free energy as a function of $d_{\alpha\alpha}$ at $T_{34} = 301.7$ K. The free energy as a
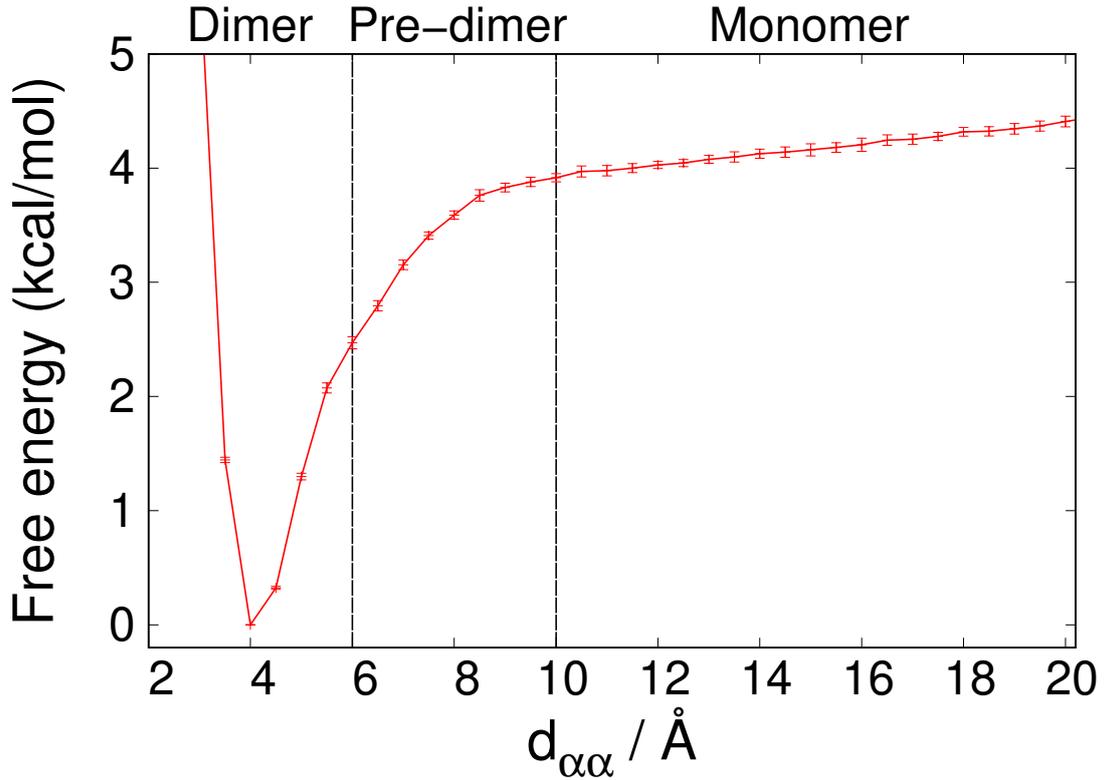
Figure 6.3: Free-energy landscape as a function of $d_{\alpha\alpha}$ at $T_{34} = 301.7$ K. The minimum value was set to zero.

function of $d_{\alpha\alpha}$ at temperature T_λ is given by

$$F_\lambda(d_{\alpha\alpha}) = -k_\mathrm{B}T_\lambda \ln \frac{Prob_\lambda(d_{\alpha\alpha})}{4\pi d_{\alpha\alpha}^2}, \qquad (6.4)$$

where λ is the temperature index, k_B is the Boltzmann constant, $Prob_\lambda$ is the probability distribution calculated using only snapshots at T_λ, and the denominator, $4\pi d_{\alpha\alpha}^2$, is the Jacobian of the coordinate transformation from Cartesian to polar coordinates. The result is presented in Fig. 6.3. Here, the bin size of $d_{\alpha\alpha}$ was 0.5 Å and the minimum value was set to zero in this free-energy landscape. For the convenience of discussion, the author define dimer, pre-dimer, and monomer states as when $d_{\alpha\alpha} < 6.0$ Å, 6.0 Å $\leq d_{\alpha\alpha} \leq 10.0$ Å, and 10.0 Å $< d_{\alpha\alpha}$, respectively. The free energy has a global-minimum at $d_{\alpha\alpha} = 4.0$ Å. The free energy difference between the dimer and monomer states is approximately 4 kcal/mol. This result indicates that the dimer state is the most stable for the NACore peptides.

The solvent accessible surface area (SASA) for the two NACore peptides were calculated to assess approaches of monomers into nascent dimers. Figure 6.4(a) presents a log scale probability distribution of the SASA as a function of $d_{\alpha\alpha}$. Here, a probe sphere of 1.4 Å was employed for the SASA calculation. In the monomer state, the SASA is between $2,100$ and $3,000$ Å^2. In the pre-dimer state, a smaller SASA is observed with a decrease of $d_{\alpha\alpha}$ from 10.0 Å to 6.0 Å. This result means that in the pre-dimer state, intermolecular contacts gradually increase as the two NACore peptides get close to each other. The distance between the two NACore peptides in the pre-dimer state is too far to form intermolecular main-chain contacts; therefore, a decrease in the SASA in the pre-dimer state results from the intermolecular side-chain contacts. In the dimer state, the SASA largely decreases as $d_{\alpha\alpha}$ changes from 6.0 Å to 4.0 Å.

The author then calculated solvent accessibility for each residue at corresponding $d_{\alpha\alpha}$ values to compare the area exposed to the solvent of amino-acid residues. The solvent accessibility is defined as the SASA normalized by the maximum SASA of the amino-acid residue. The maximum SASA of each residue was determined from the trajectory of SASA of each residue at $T = 301.7$ K. A decrease in the solvent accessibility means an increase in contact with other amino-acid residues, and a smaller solvent accessibility implies that the amino-acid residue forms contacts with multiple amino-acid residues. Figure 6.4(b) shows the average of solvent accessibility for each amino-acid residue at corresponding $d_{\alpha\alpha}$ values. The solvent accessibility for terminal residues is larger than that of amino-acid residues inside the NACore peptides, Ala69–Val77. The reason for the larger solvent accessibility at the terminal residues is that an -NH$_3$ group in the N-terminal and a -CO$_2$ group in the C-terminal are usually exposed to the solvent. The solvent accessibility of Gly73 is the smallest among the amino-acid residues at any $d_{\alpha\alpha}$ values and largely decreases as $d_{\alpha\alpha}$ changes from 5.0 Å to 4.0Å in the dimer state. This is because Gly73 has only a hydrogen atom in the side chain, and is easily covered with other amino acid residues when in the dimer state. The solvent accessibility gradually decreases as the two NACore peptides get close to each other. The decrease in the solvent accessibility for residues Ala69–Thr72 and Val74–Val77 was larger than that for the terminal residues, Gly68
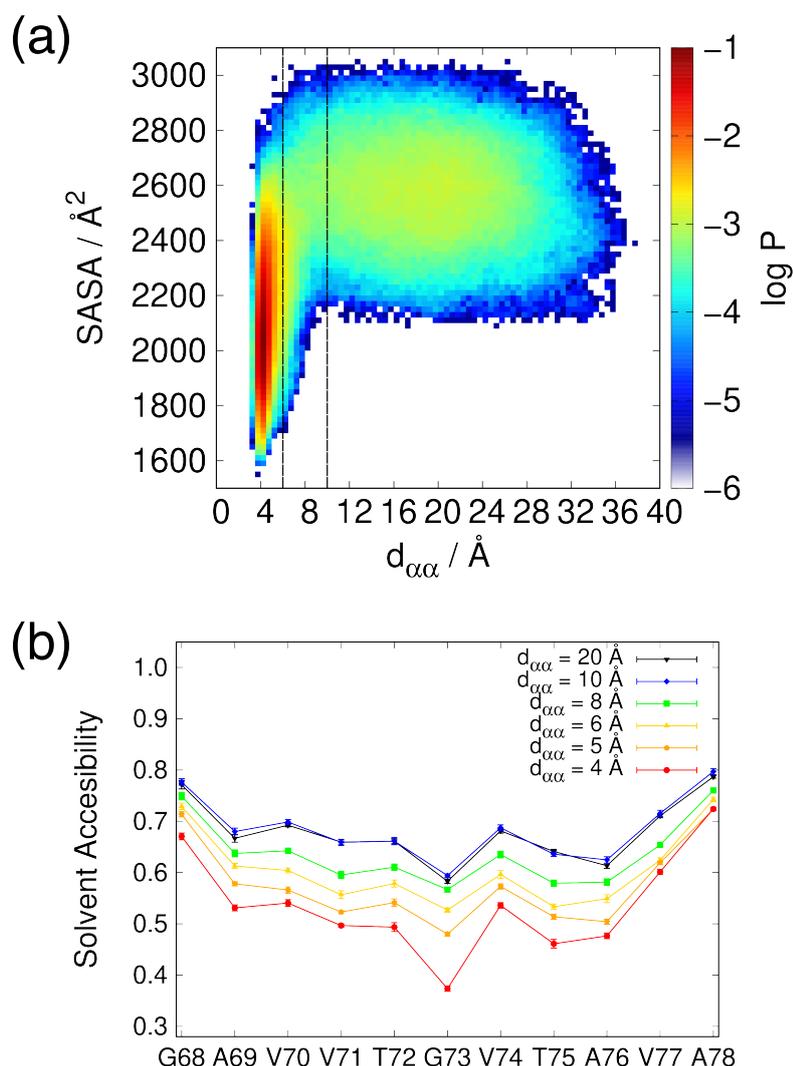
Figure 6.4: (a) Probability distribution of SASA in a logarithmic scale as a function of $d_{\alpha\alpha}$. The presented SASA is the sum of the two NACore peptides. The dashed lines at $d_{\alpha\alpha} = 6$ Å and 10 Å are the boundary between the dimer ($d_{\alpha\alpha} < 6$ Å) and pre-dimer (6 Å $\leq d_{\alpha\alpha} \leq 10$ Å) states and that between the pre-dimer and monomer (10 Å $< d_{\alpha\alpha}$) states, respectively. (b) Solvent accessibility of each residue at corresponding $d_{\alpha\alpha}$ values.

and Ala78. The large decrease in the solvent accessibility of Ala69–Thr72 indicates that they form contacts with multiple amino-acid residues when the two NACore peptides get close to each other. On the other hand, the small decrease in the solvent accessibility of Gly68 and Ala78 implies that they tend to have less contact with other amino-acid residues.

137

6.3.3 Propensity of Secondary Structures along the Dimerization Process

Secondary structures at the corresponding $d_{\alpha\alpha}$ values were analyzed to assess the structural properties of the NACore peptides on the path of the dimerization process. Figure 6.5 presents the probabilities of forming secondary structures and the ensemble-average values of the number of residues or bridges regarding the secondary structure at corresponding $d_{\alpha\alpha}$ values. The secondary structures were assigned using the Define Secondary Structure of Proteins (DSSP) algorithm [58]. The probability of forming a specific secondary structure at each $d_{\alpha\alpha}$ value was calculated by dividing the number of samples that form the secondary structure by the total number of samples in the $d_{\alpha\alpha}$ value. The ensemble-average values of the number of residues or bridges regarding the specific secondary structures at each $d_{\alpha\alpha}$ value were also calculated. The ensemble-average values were defined as dividing accumulation of the number of residues or bridges with the specific secondary structure in the NACore peptide for each $d_{\alpha\alpha}$ value by the number of samples in the $d_{\alpha\alpha}$ values. Here, the two NACore peptides were regarded as two different samples when counting the intramolecular secondary structures, whereas the two NACore peptides were regarded as one sample when counting the intermolecular secondary structures.

In the monomer state, the NACore peptides form not only the random-coil structure but also a variety of intramolecular secondary structures, such as the 3_{10}-helix, α-helix, antiparallel β-bridge, and parallel β-bridge. The average number of residues or bridges forming the intramolecular secondary structures is always less than one at any $d_{\alpha\alpha}$ value, meaning that the NACore peptides do not always form the intramolecular secondary structures. These results indicate that the NACore peptides usually have no specific structure in the monomeric form and their behavior is intrinsically disordered; thus, the NACore peptides have a low propensity to form intramolecular secondary structures.

The low propensity of the NACore peptides to form intramolecular secondary structures is understood from the characteristics of composed amino-acid residues. The NACore peptides contain four kinds of amino-acid residues: Val, Ala, Gly, and
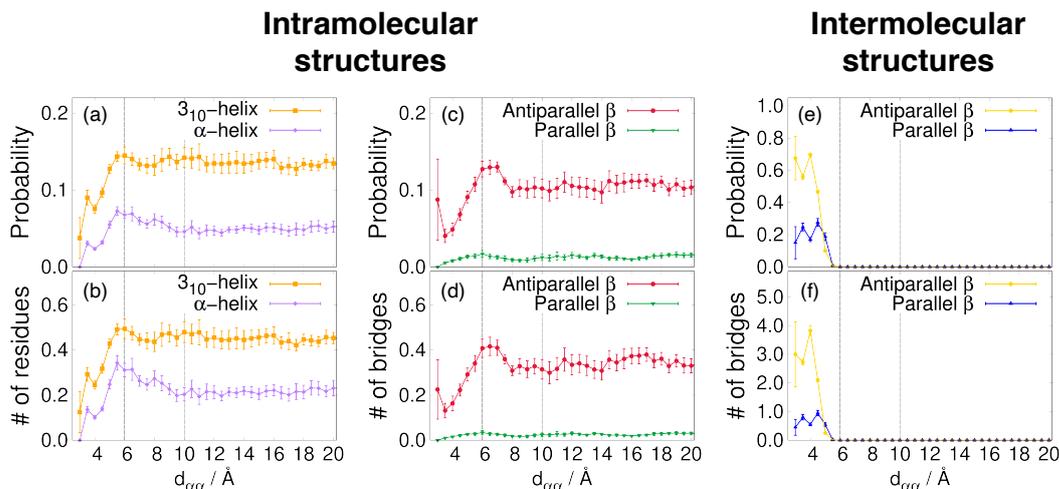
Intramolecular structures   **Intermolecular structures**

Figure 6.5: Probability and the ensemble-average value of the number of residues or bridges for the (a, b) intramolecular antiparallel and parallel β-bridge structures, (c, d) 3_{10}-helix and α-helix structures, and (e, f) intermolecular antiparallel and parallel β-bridge structures at the corresponding $d_{\alpha\alpha}$ values. The probability for each secondary structure at each $d_{\alpha\alpha}$ value was calculated by dividing the number of samples that have the secondary structure by the total number of samples in each $d_{\alpha\alpha}$ value. The ensemble-average values were calculated by dividing accumulation of the number of residues with the secondary structure in the NACore peptide for each $d_{\alpha\alpha}$ value by the number of samples in each $d_{\alpha\alpha}$ value. The dashed lines at $d_{\alpha\alpha} = 6$ Å and 10 Å are the boundary between the dimer ($d_{\alpha\alpha} < 6$ Å) and pre-dimer (6 Å $\leq d_{\alpha\alpha} \leq 10$ Å) states and that between the pre-dimer and monomer (10 Å $< d_{\alpha\alpha}$) states, respectively.

Thr. Val is known as a structurally indifferent residue because its homo-polypeptide show a low probability for helical or stretched conformation [59, 60]. Gly is known as a breaker of secondary structures because Gly has only one hydrogen atom as the side-chain and this enables dihedral angles to rotate easily. Gly73, which is located at the center of the amino-acid sequence, prevents the NACore peptides from forming a secondary structure. In addition, there are two polar residues, Thr72 and Thr75. The N- and C-termini have a positive and a negative charge, respectively, because the NACore peptides were not capped. These Thr and terminal residues prefer to interact with water molecules, which is unfavorable for the secondary structure formation. The intramolecular interaction between the terminal residues is not thought to be stable although the N- and C-termini have the opposite charges. The possible reason is that rotation of dihedral angles of Gly73, which is at the center

region of the NACore peptides, prevents the N- and C-termini from keeping them close to each other. Ilie *et al.* previously reported the conformational free-energy landscape of NACore monomer in explicit water by performing a metadynamics simulation and revealed that the conformational free-energy landscape of the NACore monomer has a common feature of intrinsically disordered proteins [22]; these results are consistent with these simulation results.

In the pre-dimer state, increases in the α-helix and intramolecular antiparallel β-bridge structures were found. Figures 6.5(a) and (b) show that the probability and average number of residues for the α-helix structure gradually increase as $d_{\alpha\alpha}$ changes from 10.0 Å to 6.0 Å. Furthermore, the author identified a small peak around $d_{\alpha\alpha} = 7.0$ Å in the probability and average number of bridges for the intramolecular antiparallel β-bridge structure in Figs. 6.5(c) and (d). These results indicate that intramolecular secondary structures become stable in the middle of the dimerization process.

In the dimer state, the intramolecular β-bridge and helix structures decrease as shown in Figs. 6.5(a)–(d). On the other hand, the dimer structures connected by β-bridges become dominant as shown in Figs. 6.5(e) and (f). Although the NACore dimer can form both intermolecular antiparallel and parallel β-bridges, the NACore dimer preferentially forms antiparallel β-bridges. Therefore, the NACore dimer with intermolecular antiparallel β-bridges is the most stable.

Figure 6.6 shows the probability distributions of the number of intermolecular β-bridges. The DSSP algorithm [58] was used to define the intermolecular antiparallel and parallel β-bridges. The distributions of the intermolecular parallel and antiparallel β-bridge structures were evaluated individually so that the summation of each distribution would be equal to one. Note that the probabilities for the zero number of bridges are the accumulation of snapshots that the NACore peptides do not form an intermolecular parallel or antiparallel β-bridge. The probability of antiparallel β-bridge formation is higher than that of parallel β-bridge formation, which is consistent with Fig. 6.5(f). Figure 6.6 also reveals that the intermolecular antiparallel β-bridges tend to form a longer β-strand structure than that of the intermolecular parallel β-bridges. The maximum number of intermolecular antiparallel
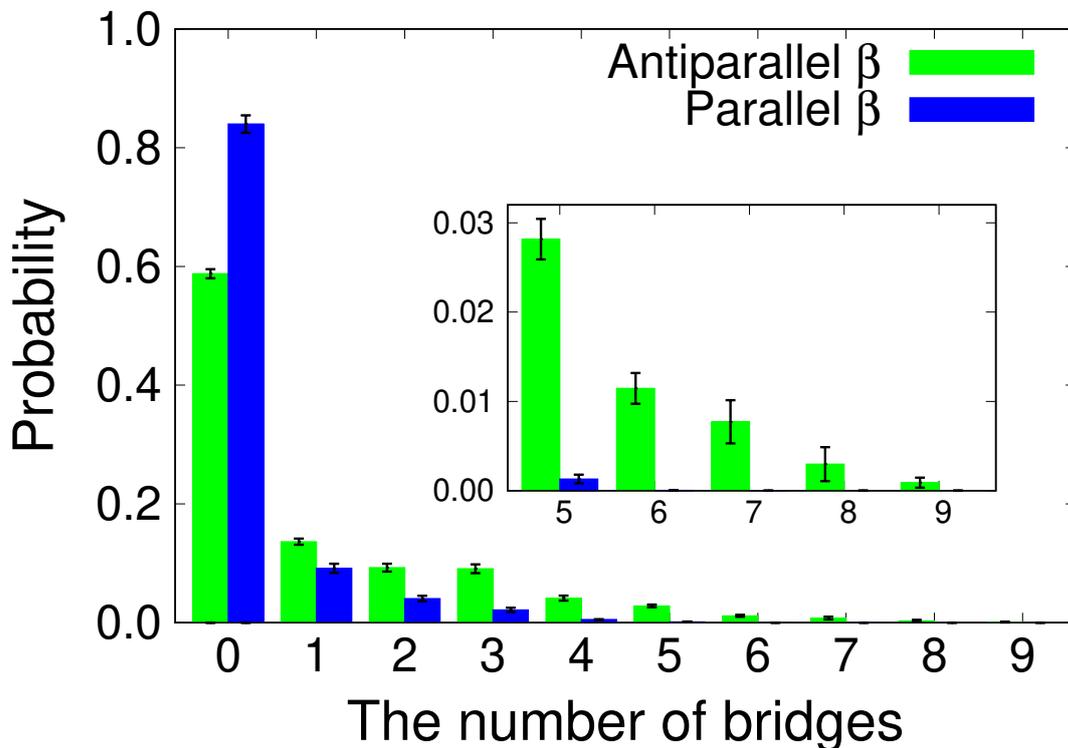
Figure 6.6: Probability distributions of the number of intermolecular β-bridges for the NACore dimer. The green and blue bars represent the antiparallel β-bridge and parallel β-bridge structures, respectively. The probability distributions were calculated for individual structures. The inset shows the same probability distributions with different ranges.

and parallel β-bridges observed in the trajectory at $T_{34} = 301.7$ K were nine and seven, respectively. These results indicate that the NACore dimer can form a fully extended antiparallel β-strand structure, whereas it rarely forms a fully extended parallel β-strand structure.

6.3.4 Mechanism of Increasing Intramolecular Secondary Structures in the Pre-Dimer State

The probabilities of intramolecular antiparallel β-bridge and α-helix structures increase in the pre-dimer state. To understand the stabilization of the intramolecular secondary structures in the pre-dimer state, the following thermodynamics quantities were estimated: the differences in the Gibbs-free energy ΔG, the partial molar

141

enthalpy ΔH, and the partial molar entropy ΔS. The difference in the Gibbs free energy between states A and B is calculated as follows:

$$\Delta G = G_{\mathrm{B}} - G_{\mathrm{A}} = -RT \ln\left(\frac{f_{\mathrm{B}}}{f_{\mathrm{A}}}\right), \tag{6.5}$$

where $R = 8.3145$ J/(mol K) is the gas constant, f_{A} and f_{B} are fractions of the states A and B, respectively. Here, state A is defined as the NACore structures that forms an intramolecular secondary structure in the monomer state. State B is defined as those in the pre-dimer state. The author considered the intramolecular antiparallel β-bridge and α-helix as the intramolecular secondary structures. The two NACore peptides were regarded as two different samples. The differences in the partial molar enthalpy and partial molar entropy are calculated as follows:

$$\Delta H = \left[\frac{\partial(\Delta G/T)}{\partial(1/T)}\right]_P, \tag{6.6}$$

$$\Delta S = \frac{\Delta H - \Delta G}{T}. \tag{6.7}$$

Because the difference in the partial molar enthalpy is derived from the temperature dependence of the difference in the Gibbs free energy, the author calculated the differences in the Gibbs free energy in the range of $T_{34} = 301.7$ K to $T_{80} = 500.0$ K. The first-order polynomial fitting was performed to evaluate the partial molar enthalpy. The temperature dependence of the differences in the Gibbs-free energy and the results of fitting are presented in Fig. 6.7. The obtained thermodynamics quantities are listed in Table 6.2. For the NACore structure with an intramolecular antiparallel β-bridge, ΔH and $T\Delta S$ are -4.1 ± 0.4 kJ/mol and -3.1 ± 0.3 kJ/mol, respectively. For the NACore structure with α-helix, on the other hand, ΔH and $T\Delta S$ are -3.5 ± 0.5 kJ/mol and -2.3 ± 0.3 kJ/mol, respectively. Hence, the differences in the Gibbs free energy are negative due to the contribution from the enthalpy term. The differences in the Gibbs free energy are negative, meaning that the intramolecular antiparallel β-bridge and α-helix structures in the pre-dimer state are more stable than those in the monomer state. As for the partial molar enthalpy and partial molar entropy terms in the Gibbs free energy, both the differences in the par-
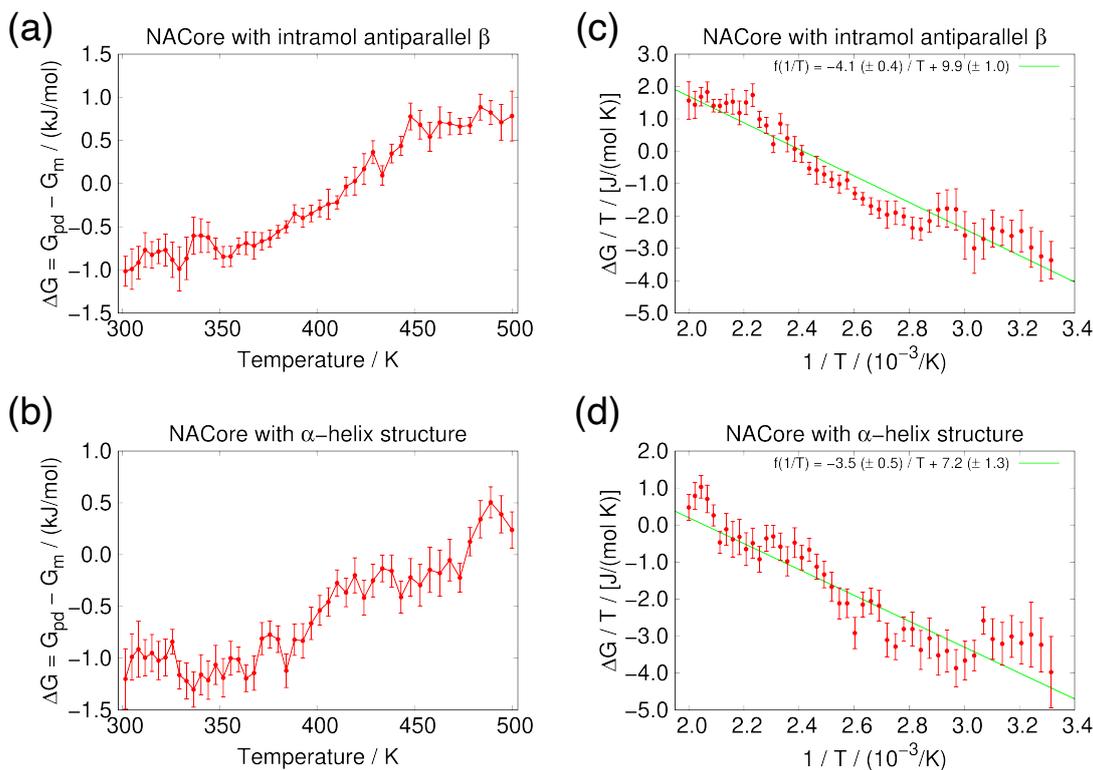
Figure 6.7: Temperature dependence of the differences in the Gibss free energy between the pre-dimer (pd) and monomer (m) states of the NACore with (a) intramolecular antiparallel β-bridge and (b) α-helix structures. $\Delta G/T$ as a function of $1/T$ for the NACore with (c) intramolecular antiparallel β-bridge and (d) α-helix structures. The results of the first-order polynomial fitting are shown with green lines.

Table 6.2: Differences in thermodynamics quantities between the pre-dimer and monomer states at $T = 301.7$ K.

Secondary structure	ΔG (kJ/mol)	ΔH (kJ/mol)	$T\Delta S$ (kJ/mol)
Intramolecular antiparallel β-bridge	-1.0 ± 0.2	-4.1 ± 0.4	-3.1 ± 0.3
α-helix	-1.2 ± 0.3	-3.5 ± 0.5	-2.3 ± 0.3

tial molar enthalpy and partial molar entropy are negative. These results indicate that the NACore structure with an intramolecular antiparallel β-bridge or α-helix in the pre-dimer state is enthalpically more stable but entropically unfavorable than that in the monomer state. The lower entropy of the NACore structures with an intramolecular antiparallel β-bridge and an α-helix in the pre-dimer state implies that their conformations are restricted when compared with those in the monomer state.

143

To investigate molecular mechanisms when the differences in the partial molar entropy become negative, The average number of intermolecular side-chain contacts was then computed when the NACore peptide had the intramolecular secondary structures. When the shortest distance between a pair of non-hydrogen atoms in the side chain was less than 5.0 Å, it was regarded as a side-chain contact. The average number of intermolecular side-chain contacts for a specific secondary structure was calculated over the snapshots that have specific secondary structure at each $d_{\alpha\alpha}$ value. Here, the author focused on the intramolecular antiparallel β-bridge and α-helix as specific secondary structure. For comparison, the average number of intermolecular side-chain contacts was also computed by considering all snapshots at each $d_{\alpha\alpha}$ value. The results are presented in Fig. 6.8. When the NACore peptides are in the monomer state, no intermolecular side-chain contact is observed. When the NACore peptides are in the pre-dimer states, the average number of intermolecular side-chain contacts increases as the NACore peptides get close to each other. In the pre-dimer state, there are more intermolecular side-chain contacts when the NACore peptides form an intramolecular antiparallel β-bridge or α-helix than the average values of intermolecular side-chain contacts over all snapshots at each $d_{\alpha\alpha}$ value. These results imply that intermolecular side-chain contacts play an important role in stabilizing the intramolecular secondary structures in the pre-dimer state.

The mechanism of increasing the intramolecular secondary structures by the intermolecular side-chain contacts can be explained as follows: Two NACore peptides have intermolecular side-chain contacts when they are close to each other and are in the pre-dimer state. The intermolecular side-chain contacts suppress the motion of the NACore peptides. This situation is entropically unfavorable for the NACore peptide and a lower enthalpy is preferred to compensate for the loss of entropy. Therefore, the NACore peptide forms hydrogen bonds to have a lower enthalpy conformation and, as a result, an intramolecular secondary structure, such as an antiparallel β-bridge or an α-helix is formed. Note that in the pre-dimer state, hydrogen bonds are formed only between intramolecular residues. This is because the distance between two NACore peptides in the pre-dimer state is too large to form intermolecular hydrogen bonds.
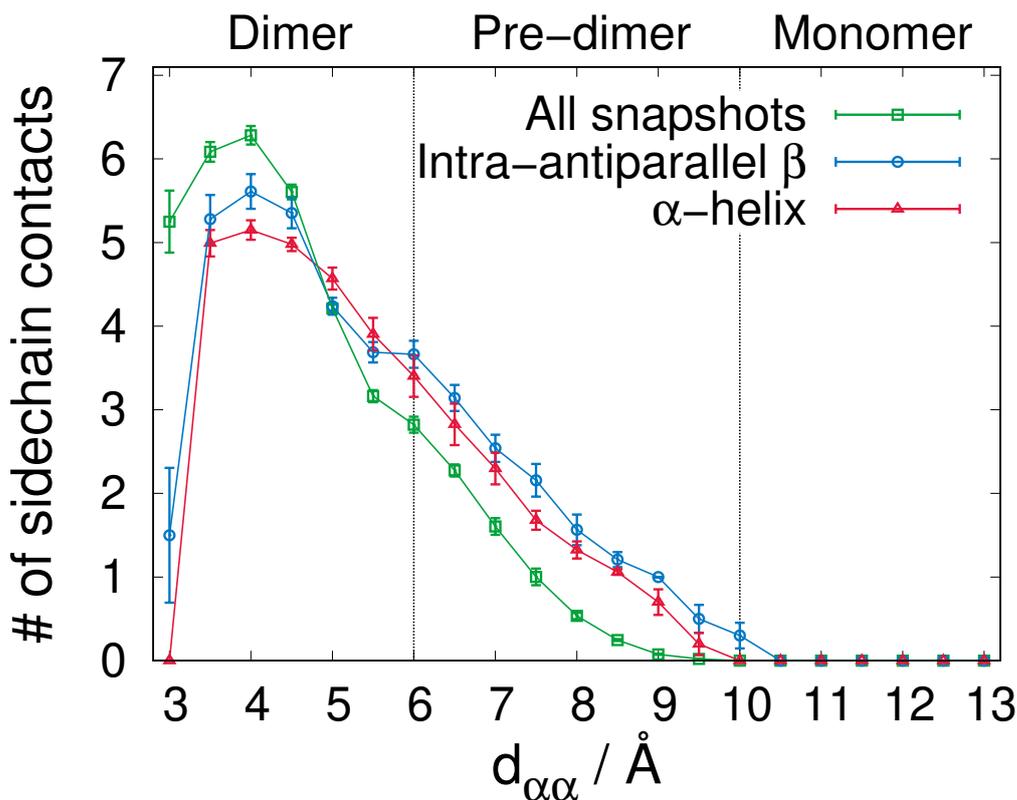
Figure 6.8: Average number of intermolecular side-chain contacts at corresponding $d_{\alpha\alpha}$ values. Data for all snapshots, intramolecular antiparallel β-bridge, and α-helix are shown in open green squares, blue open circles, and red open triangles, respectively.

The stabilization of the intramolecular secondary structures on the middle of dimerization was observed not only for the NACore but also for Aβ(29–42) peptides. Itoh *et al.* found that the β-hairpin structure increases when two Aβ(29–42) peptides come close to each other and have intermolecular side-chain contacts [61, 62]. Therefore, the stabilization of intramolecular secondary structures in the middle of the dimerization process appears to be a common property of amyloidogenic peptides.

6.3.5 Intermolecular Residue Contacts in the Dimer State

To investigate the molecular mechanism of how the NACore peptides form the intermolecular β-bridge structure in the dimer state, the author calculated the probability of intermolecular residue contacts at the corresponding $d_{\alpha\alpha}$ values. The probability

of intermolecular residue contact is defined as the arithmetic mean of the inter-molecular residue-contact maps at the target temperature, $T_{34} = 301.7$ K. The intermolecular residue-contact map between residues i and j, $Q_{ij}^{\text{intermol}}(t_k)$, is defined as

$$Q_{ij}^{\text{intermol}}(t_k) = \begin{cases} 1 & \text{for } r_{ij}(t_k) \leq 6.5 \text{ Å,} \\ 0 & \text{for } r_{ij}(t_k) > 6.5 \text{ Å,} \end{cases} \tag{6.8}$$

where $r_{ij}(t_k)$ is the distance between the C_α atoms in residues i and j at time-step t_k. It was assumed that residues i and j belonged to different NACore peptides. Figures 6.9 shows the probability of intermolecular residue contact at corresponding $d_{\alpha\alpha}$ values. Here, the bin size of $d_{\alpha\alpha}$ was set to 1.0 Å. The intermolecular residue contacts were observed only when $d_{\alpha\alpha} = 6.0$ Å or less due to the definition by Eq. (6.8). The highest intermolecular contact probability at $d_{\alpha\alpha} = 6.0$ Å is found between Gly68 (N-terminus) and Ala78 (C-terminus), as indicated by red arrows in Fig. 6.9(a). Gly68 at the N-terminus and Ala78 at the C-terminus are positively and negatively charged, respectively, because they were not capped. The attractive force due to the electrostatic interaction tends to draw the N- and C-termini closer to each other than other intermolecular residue pairs. Therefore, the electrostatic interaction between these partial charges plays an important role in capturing other NACore peptides. The probability of residue contacts also reveals that Gly68 and Ala78 tend to have one direct contact rather than to have contacts with multiple amino-acid residues, which is consistent with the small decreases in the solvent accessibility of Gly68 and Ala78 as shown in Fig. 6.4(b). When $d_{\alpha\alpha} = 5.0$ Å, increases in the probability of intermolecular contact are observed, not only between the N- and C-termini but also between residues corresponding to diagonal components on the contact map. Consecutive residue contacts on the diagonal components on the contact map are typically observed when there is a parallel or antiparallel β-strand structure. In addition, intermolecular antiparallel and parallel β-bridges begin to increase when $d_{\alpha\alpha}$ changes from 6.0 Å to 5.0 Å as shown in Figs. 6.5(e) and (f). Therefore, the increase in the probability of intermolecular contact on the diagonal components arises from the formation of intermolecular antiparallel and parallel
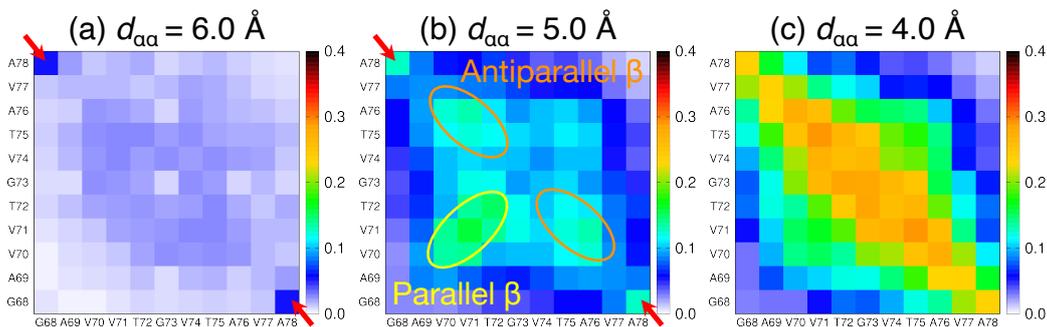
146

Figure 6.9: Probability of intermolecular residue contacts between the NACore peptides. A contact is considered when the distance between C_α atoms is less than 6.5 Å.

β-bridges. Intermolecular antiparallel β-bridges tend to form between the regions Val70–Thr72 and Val74–Ala76, while the intermolecular parallel β-bridges tend to form between Val70–Thr72 and Val70–Thr72. More intermolecular contacts are observed when $d_{\alpha\alpha} = 4.0$ Å. The probabilities of intermolecular contacts around the diagonal components, especially the contacts related to the antiparallel β-bridge, increase.

6.3.6 Intermolecular β-Bridge Formation in the Dimerization Process

As seen in Figs. 6.5(c) and (d), the intramolecular antiparallel β-bridge structure was stabilized in the pre-dimer state. To elucidate the role of the intramolecular β-bridge structure in forming the intermolecular β-bridge structure, the author analyzed the NACore dimer structures that have both intramolecular and intermolecular β-bridges. Figure 6.10 presents the probability distribution of the number of residues shared by both intramolecular and intermolecular β-bridges. The sum of probabilities that the NACore peptides have shared residues between the intramolecular and intermolecular β-bridges is higher than the probability that the NACore peptides do not so. In particular, intermolecular antiparallel β-bridges tend to be formed. These results suggest that intermolecular β-bridges are easily formed by binding intramolecular β-bridges if the intramolecular β-bridges have already been formed. Therefore, there is a possible pathway that an intermolecular β-bridge structure is
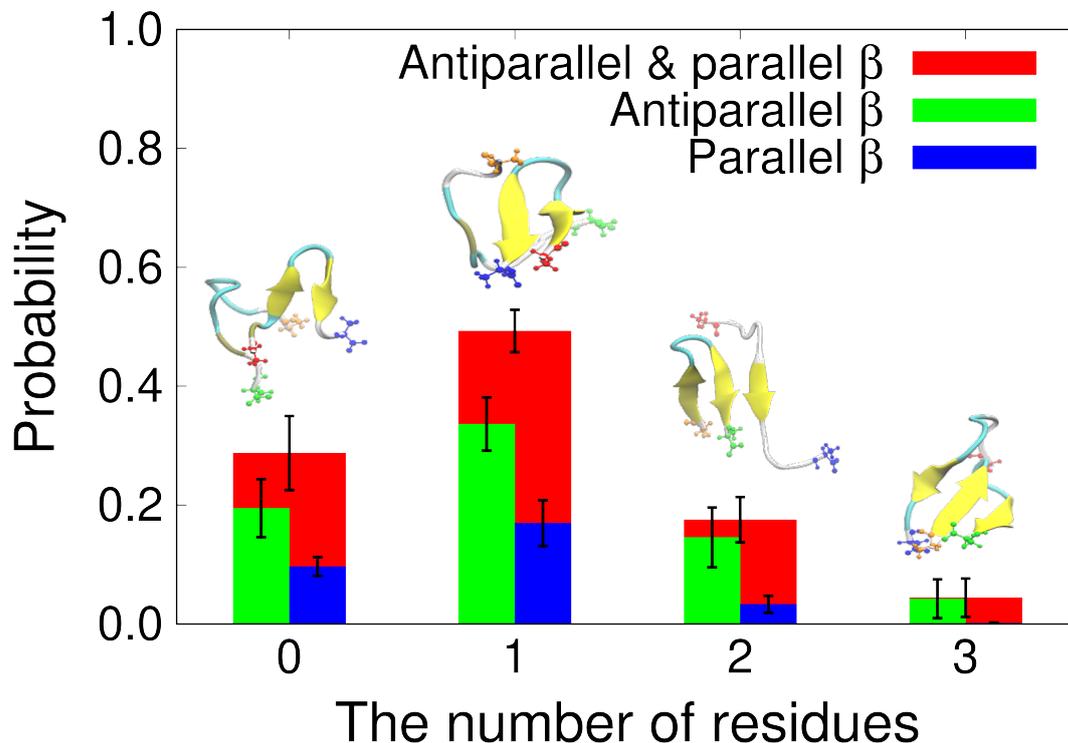
Figure 6.10: Probability distribution of the number of residues shared by intramolecular and intermolecular β-bridges. The blue, green, and red bars represent the antiparallel β-bridge, parallel β-bridge, and both of them, respectively. A corresponding structure is shown above each bar. The N- and C-termini are colored red and blue, respectively, for the first NACore peptide. The N- and C-termini are colored orange and green, respectively, for the second NACore peptide.

formed by binding with an intramolecular β-bridge that has already formed.

To investigate how intermolecular β-bridges are formed, the author calculated the two-dimensional probability distributions with respect to the number of intramolecular and intermolecular β-bridges at various $d_{\alpha\alpha}$ values. Here, the bin size of $d_{\alpha\alpha}$ was set to 1.0 Å and the probability distribution was normalized at each $d_{\alpha\alpha}$ value. Note that the number of intramolecular β-bridges was summed over the two NACore peptides. The results are presented in Figs. 6.11(a)–(c). When $d_{\alpha\alpha} = 6.0$ Å, there is no intermolecular β-bridge structure, and the NACore peptides form only intramolecular β-bridges or other intramolecular conformations. Intermolecular β-bridges are observed when $d_{\alpha\alpha} \leq 5.0$ Å. When $d_{\alpha\alpha} = 5.0$ Å, the NACore dimer that forms one intermolecular β-bridge with no intramolecular β-bridge has the highest
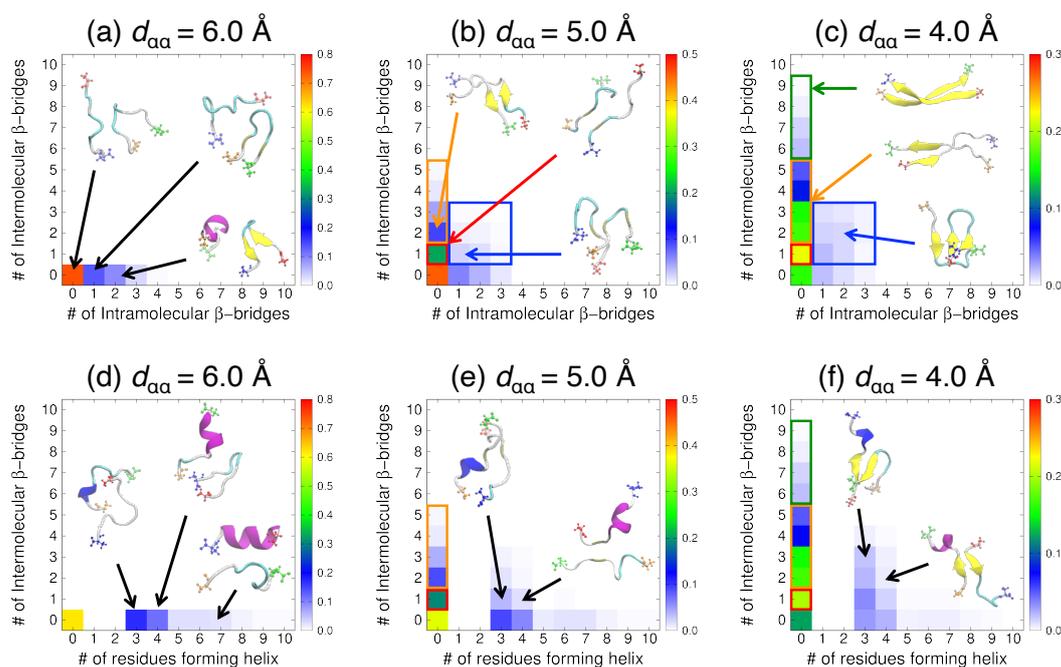
Figure 6.11: (a)–(c) Two-dimensional probability distributions with respect to the number of intramolecular β-bridges and intermolecular β-bridges at the corresponding $d_{\alpha\alpha}$ values. The number of intramolecular β-bridges was summed over two NACore peptides. (d)–(f) Two-dimensional probability distributions with respect to the number of residues forming helix and intermolecular β-bridges at the corresponding $d_{\alpha\alpha}$ values. Representative structures are also presented. The N- and C-termini are colored red and blue, respectively, for the first NACore peptide. The N- and C-termini are colored orange and green, respectively, for the second NACore peptide.

probability, and the NACore dimer that forms two intermolecular β-bridges with no intramolecular β-bridge has the second highest probability. Compared to the probabilities of NACore dimer structures that form only intermolecular β-bridges, the probabilities that the NACore dimer has both intermolecular and intramolecular β-bridges are much lower. When $d_{\alpha\alpha} = 4.0$ Å, longer β-strand structures without an intramolecular β-bridge were formed.

From the results in the above paragraph, another pathway of the intermolecular β-bridge formation is found; an intermolecular β-bridge is formed directly between residues that have no specific structure when the two NACore peptides are close to each other. If it tends to take the pathway by which an intermolecular β-bridge is formed via binding with an intramolecular β-bridge, the probabilities of the dimer structures that form both intermolecular and intramolecular β-bridges should be

higher than those of the dimer structures that only form intermolecular β-bridges with no intramolecular β-bridge. In other words, in Fig. 6.11(b), the probabilities in the region surrounded by the blue square should be higher than those in the region surrounded by the red square or orange rectangle. As Fig. 6.10 shoed, intermolecular β-bridges are easily formed by binding intramolecular β-bridges when the intramolecular β-bridges have already been formed. However, Figs. 6.11(b) and (c) indicate that the dimer structures that have both intermolecular and intramolecular β-bridges were minority in the structural ensemble in the dimer state. This is due to the low proportion of intramolecular β-bridges in the monomer and pre-dimer states. In fact, Fig. 6.5 indicates that the coil structure is dominant both in the monomer and pre-dimer states even though the stabilization of β-bridge and α-helix structures is observed in the pre-dimer states. Therefore, it is reasonable to conclude that there is a low tendency to take the pathway by which an intermolecular β-bridge structure is formed by binding with and intramolecular β-bridges. In contrast, the NACore peptides prefer to form intermolecular β-bridges without forming intramolecular β-bridges when $d_{\alpha\alpha} = 5.0$ Å. This result support that there is a high tendency to take the pathway by which an intermolecular β-bridge is formed directly between the residues that have no specific structures when the NACore peptides get close to each other. Figure 6.12 shows the probability distribution of the number of intermolecular β-bridges when the NACore dimer forms intermolecular β-bridges but no intramolecular β-bridge. An intermolecular antiparallel β-bridge is likely to form though both antiparallel and parallel β-bridges are observed. Therefore, antiparallel β-bridge is likely in the spontaneous intermolecular β-bridge formation.

The author further discuss how helix structures, which are stabilized in the pre-dimer state, play a role in the dimerization. Figures 6.11(d)–(f) present two-dimensional probability distributions with respect to the number of residues with helix structure and intermolecular β-bridges at various $d_{\alpha\alpha}$ values. When $d_{\alpha\alpha} =$ 6.0 Å, there is no intermolecular β-bridge structure, and the NACore peptides form only helix or other intramolecular structures. The intermolecular β-bridge begins to form when $d_{\alpha\alpha} = 5.0$ Å. The most likely dimer conformation is one inter molecular β-bridge with no helical structures. Longer intermolecular β-strand structures are
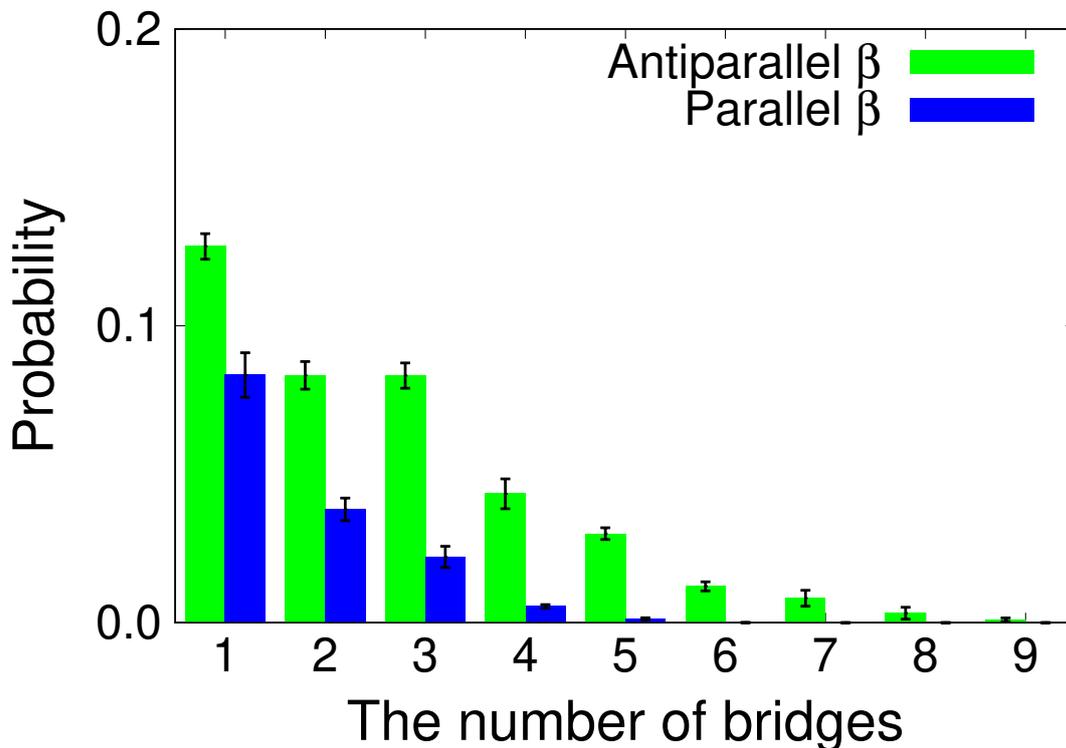
Figure 6.12: Probability distribution of the number of intermolecular β-bridges when the NACore dimer forms intermolecular β-bridges but no intramolecular β-bridge. The blue and green bars represent the antiparallel β-bridge and parallel β-bridge structures, respectively.

observed when $d_{\alpha\alpha} = 4.0$ Å. Although there are NACore dimers that form both helix and intermolecular β-bridges, its probabilities are low as shown in Figs. 6.11(e) and (f). The intermolecular β-bridges usually form between the residues that do not form helix structures.

In the aggregation of various amyloidogenic peptides, the transition from α-helical to β-rich conformations has been observed [63–72]. For instance, Qi *et al.* identified the following two mechanisms in the transition from α-helical to β-rich conformations of the human amylin fragment dimer: (i) the random coil part of the helical conformation directly transits into an intermolecular β-sheet and (ii) the helical conformation first unfolded to coil conformation and then the coil structures convert into an intermolecular β-sheet [65]. In the dimerization process of the NACore peptides, the spontaneous intermolecular β-bridge formation is more likely than the intermolecular β-bridge formation through helix structure considering from

Figs. 6.11(d) and (e). This is because, in the NACore peptides, the coil conformation is dominant in both the monomer and pre-dimer states even though the α-helical conformation is stabilized in the pre-dimer state as shown in Figs. 6.5(a) and (b). Thus, the helical conformations are unlikely to trigger the formation of intermolecular β-bridges.

The data indicate two possible pathways of intermolecular β-bridge formation in the dimerization process: (i) an intermolecular β-bridge structure is formed by binding with an intramolecular β-bridge that has already formed, or (ii) an intermolecular β-bridge is formed directly between residues that have no specific structure when the two NACore peptides are close to each other. Judging from Figs. 6.11(a)–(c) the latter pathway is more likely than the former one in the dimerization of the NACore peptides. This is because the NACore peptides have a low propensity to form intramolecular secondary structures and consequently the NACore peptides have less opportunity to take the former pathway. These observations can be regarded as an example of "inverse side chain effect" [73] or "natively disordered model" [74]: amyloid formation is primarily defined by main chain interactions, and is a distinct process from protein folding in which the side chain contacts play an important role (i.e., absence of folding propensity results in amyloid formation).

On the other hand, a number of experimental and computational studies have investigated the role of β-hairpin formation in the aggregation of amyloidogenic peptides such as α-Syn, Aβ peptide, β2-microglobulin, and islet amyloid polypeptide [61, 62, 75–98]. These studies have revealed that although these amyloidogenic peptides in monomeric form have an intrinsically disordered nature, an ordered conformation such as β-hairpin as well as random coil was found in the conformational ensembles [77–86]. Furthermore, the β-hairpin structure also participates in the dimer structures [61, 62, 82–84, 88–94]. It has been suggested that the β-hairpin structure intermediates the intermolecular β-bridge formation and accelerates the aggregation of amyloidogenic peptides [61, 62, 82, 84, 86, 87, 95–98]. These observations indicate that, in the aggregation process of amyloidogenic peptides, intermolecular β-bridges are easily formed on intramolecular β-bridges if the intramolecular β-bridges are already formed. Thus, if an amyloidogenic peptide has a high propen-

sity to form intramolecular β-bridges, intermolecular β-bridges tend to form on the intramolecular β-bridges. To support this idea, let us discuss the dimerization mechanism of amyloid-β(29–42) reported by Itoh *et al.* [61, 62], which is a comparable example to the present study. They revealed that the intermolecular β-bridges are readily formed when there are intramolecular β-bridges in the dimerization process of amyloid-β(29–42). The average number of residues forming the intramolecular β-bridges of amyloid-β(29–42) is approximately three when it is in the monomer state. In the pre-dimer state, more intramolecular β-bridges are observed. This means that amyloid-β(29–42) has a high propensity to form intramolecular β-bridges in the monomer and pre-dimer states. This propensity was observed by NMR experiments as well as molecular dynamics simulations [99]. Aβ(29–42) peptide, with amino-acid sequence Ace-GAIIGLMVGGVVIA-Nme, has only hydrophobic amino acid residues. Because of its high hydrophobicity, amyloid-β(29–42) tends to form the β-hairpin structure as a low-enthalpy conformation. Due to the high proportion of intramolecular β-bridges in the monomer and pre-dimer states, the probability of taking the pathway that forms the intermolecular β-bridges on the intramolecular β-bridges is also high in the dimerization of amyloid-β(29–42).

Therefore, the author propose that pathway of an intermolecular β-bridge formation of the amyloidogenic peptide depends on its propensity to form an intramolecular secondary structure, especially intramolecular β-bridges. If the amyloidogenic peptide has a high propensity to form intramolecular β-bridges, intermolecular β-bridges tend to form on the intramolecular β-bridges. This is because intermolecular β-bridges are easily formed on the intramolecular β-bridges. On the other hand, if an amyloidogenic peptide has a low propensity to form intramolecular β-bridges, the intermolecular β-bridges tend to be formed directly between residues with no specific structure. This mechanism could be applied not only for the NACore peptide but also for amyloidogenic peptides in general.

6.4 Conclusions

The author investigated the dimerization of NACore peptides by molecular dynamics simulation. The author employed the isothermal–isobaric replica-permutation method for the two NACore peptides in explicit water to achieve efficient conformational sampling. The free-energy landscape as a function of $d_{\alpha\alpha}$, which is a measure of the distance between the peptides, at $T_{34} = 301.7$ K revealed that the dimer state was the most stable for the NACore peptides. By analyzing the secondary structure, the author revealed that the NACore dimer with intermolecular β-bridges is dominant in the dimer state. More intermolecular antiparallel β-bridges are observed than the intermolecular parallel β-bridges in the NACore dimer structures.

The author also found that the intramolecular antiparallel β-bridge and α-helix structures are stabilized in the pre-dimer state. The author revealed that the NACore structures with an intramolecular antiparallel β-bridge or α-helix in the pre-dimer state are enthalpically stable but entropically unfavorable than those in the monomer state. This is because the intermolecular side-chain contacts between the NACore peptides suppress the motion of the peptides, which leads to entropy reduction. To compensate for the entropy loss, the NACore peptide forms intramolecular hydrogen-bonds to be of lower enthalpy conformation; as a result, increases in the intramolecular antiparallel β-bridges and α-helix structures are observed. The stabilization of the intramolecular secondary structures in the pre-dimer state has also been reported in the dimerization process of amyloid-β(29-42) peptides [61, 62]. Therefore, it is suggested that the stabilization of the intramolecular secondary structure in the middle of the dimerization process is a common property of amyloidogenic peptides.

In the dimerization of the NACore peptides, there are two possible pathways of intermolecular β-bridge formation: (i) an intermolecular β-bridge structure is formed by binding with an intramolecular β-bridge that has already formed, and (ii) an intermolecular β-bridge is formed directly between residues that have no specific structure when the two NACore peptides are close to each other. The author revealed that the latter pathway tends to be realized in the dimerization of the NACore peptides. This is because the random coil conformation is dominant in both

the monomer and pre-dimer states even though the intramolecular antiparallel β-bridge and α-helix structures are stabilized in the pre-dimer state. In other word, the high propensity of the NACore peptide to form a coil conformation in the monomer and pre-dimer states hide the former pathway.

On the other hand, when the NACore dimer has both intramolecular and intermolecular β-bridges, the intramolecular β-bridges and intermolecular β-bridges tend to share the same residues. This result suggests that an intermolecular β-bridge is easily formed when an intramolecular β-bridge has already formed. Moreover, the propensity of the intramolecular β-bridge formation in the monomeric form possibly changes the pathway of the intermolecular β-bridge formation. In fact, a role of the β-hairpin structure in the dimerization process has been investigated in other amyloidogenic peptides as well, suggesting that the β-hairpin structure intermediate the intermolecular β-bridge formation and accelerate the aggregation of amyloidogenic peptides [61, 62, 82, 84, 86, 95–98]. Therefore, the author propose a general view of intermolecular β-bridge formation in the dimerization process, which is as follows: the propensity of an amyloidogenic peptide to form intramolecular β-bridges determines the mechanism of forming intermolecular β-bridges. If the amyloidogenic peptide has a high propensity to form intramolecular β-bridges, intermolecular β-bridges tend to form on the intramolecular β-bridges. If the amyloidogenic peptide has a low propensity to form intramolecular β-bridges, the intermolecular β-bridges tend to form directly between residues with no specific structures.

Insights from this simulation provide useful information to assess the general mechanisms underlying the aggregation and fibril formation of amyloidogenic peptides. Further comparisons among amyloidogenic peptides are necessary to confirm the relationship between the propensity to form intramolecular structures and the fibrillation process. The author believe that these insights will also help to understand the polymorphism of amyloid fibril structures and their cytotoxicity.

References

[1] L. Maroteaux, JT Campanelli, and RH Scheller. Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *The Journal of Neuroscience*, 8(8):2804–2815, 1988.

[2] Maria Grazia Spillantini, Marie Luise Schmidt, Virginia M.Y. Lee, John Q. Trojanowski, Ross Jakes, and Michel Goedert. α-synuclein in Lewy bodies. *Nature*, 388(6645):839–840, 1997.

[3] K. Ueda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, D. A.C. Otero, J. Kondo, Y. Ihara, and T. Saitoh. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America*, 90(23):11282–11286, 1993.

[4] Hogyu Han, Paul H. Weinreb, and Peter T. Lansbury. The core Alzheimer's peptide NAC forms amyloid fibrils which seed and are seeded by β-amyloid: is NAC a common trigger or target in neurodegenerative disease? *Chemistry & Biology*, 2(3):163–169, 1995.

[5] Akihiko Iwai, Makoto Yoshimoto, Eliezer Masliah, and Tsunao Saitoh. Non-Aβ Component of Alzheimer's Disease Amyloid (NAC) is Amyloidogenic. *Biochemistry*, 34(32):10139–10145, 1995.

[6] Omar M.A. El-Agnaf, Ross Jakes, Martin D. Curran, Derek Middleton, Raffaele Ingenito, Elisabetta Bianchi, Antonello Pessi, David Neill, and Andrew Wallace. Aggregates from mutant and wild-type α-synuclein proteins and NAC peptide

induce apoptotic cell death in human neuroblastoma cells by formation of β-sheet and amyloid-like filaments. *FEBS Letters*, 440(1-2):71–75, 1998.

[7] Omar M. A. El-Agnaf, Angela M. Bodles, David J. S. Guthrie, Patrick Harriott, and G. Brent Irvine. The N-terminal region of non-Abeta component of Alzheimer's Disease amyloid is responsible for its tendency to assume beta-sheet and aggregate to form fibrils. *European Journal of Biochemistry*, 258(1):157–163, 1998.

[8] Anja Leona Biere, Stephen J. Wood, Jette Wypych, Shirley Steavenson, Yi-jia Jiang, Dan Anafi, Frederick W. Jacobsen, Mark A. Jarosinski, Gay-May Wu, Jean-Claude Louis, Francis Martin, Linda O. Narhi, and Martin Citron. Parkinson's Disease-associated α-Synuclein Is More Fibrillogenic than β- and γ-Synuclein and Cannot Cross-seed Its Homologs. *Journal of Biological Chemistry*, 275(44):34574–34579, 2000.

[9] Angela M. Bodles, David J. S. Guthrie, Patrick Harriott, Pamela Campbell, and G. Brent Irvine. Toxicity of non-Aβ component of Alzheimer's disease amyloid, and N-terminal fragments thereof, correlates to formation of β-sheet structure and fibrils. *European Journal of Biochemistry*, 267(8):2186–2194, 2000.

[10] Angela M. Bodles, David J. S. Guthrie, Brett Greer, and G. Brent Irvine. Identification of the region of non-Aβ component (NAC) of Alzheimer's disease amyloid responsible for its aggregation and toxicity. *Journal of Neurochemistry*, 78(2):384–395, 2001.

[11] Benoit I. Giasson, Ian V. J. Murray, John Q. Trojanowski, and Virginia M.Y. Lee. A Hydrophobic Stretch of 12 Amino Acid Residues in the Middle of α-Synuclein Is Essential for Filament Assembly. *Journal of Biological Chemistry*, 276(4):2380–2386, 2001.

[12] Hong-Tao Li, Hai-Ning Du, Lin Tang, Jun Hu, and Hong-Yu Hu. Structural transformation and aggregation of human α-synuclein in trifluoroethanol: Non-amyloid component sequence is essential and β-sheet formation is prerequisite to aggregation. *Biopolymers*, 64(4):221–226, 2002.

[13] Hai Ning Du, Lin Tang, Xiao Ying Luo, Hong Tao Li, Jun Hu, Jia Wei Zhou, and Hong Yu Hu. A peptide motif consisting of glycine, alanine, and valine is required for the fibrillization and cytotoxicity of human α-synuclein. *Biochemistry*, 42(29):8870–8878, 2003.

[14] Magali Periquet, Tudor Fulga, Liisa Myllykangas, Michael G. Schlossmacher, and Mel B. Feany. Aggregated-Synuclein Mediates Dopaminergic Neurotoxicity In Vivo. *Journal of Neuroscience*, 27(12):3338–3346, 2007.

[15] J. A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang, L. Jiang, M. A. Arbing, B. L. Nannenga, J. Hattne, J. Whitelegge, A. S. Brewster, M. Messerschmidt, S. Boutet, N. K. Sauter, T. Gonen, and D. S. Eisenberg. Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature*, 525(7570):486–90, 2015.

[16] Marcus D. Tuttle, Gemma Comellas, Andrew J. Nieuwkoop, Dustin J. Covell, Deborah A. Berthold, Kathryn D. Kloepper, Joseph M. Courtney, Jae K. Kim, Alexander M. Barclay, Amy Kendall, William Wan, Gerald Stubbs, Charles D. Schwieters, Virginia M Y Lee, Julia M. George, and Chad M. Rienstra. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nature Structural & Molecular Biology*, 23(5):409–415, 2016.

[17] Binsen Li, Peng Ge, Kevin A. Murray, Phorum Sheth, Meng Zhang, Gayatri Nair, Michael R. Sawaya, Woo Shik Shin, David R. Boyer, Shulin Ye, David S. Eisenberg, Z. Hong Zhou, and Lin Jiang. Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications*, 9(1):1–10, 2018.

[18] Yaowang Li, Chunyu Zhao, Feng Luo, Zhenying Liu, Xinrui Gui, Zhipu Luo, Xiang Zhang, Dan Li, Cong Liu, and Xueming Li. Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Research*, 28(9):897–903, 2018.

[19] Ricardo Guerrero-Ferreira, Nicholas M.I. Taylor, Daniel Mona, Philippe Ringler, Matthias E. Lauer, Roland Riek, Markus Britschgi, and Henning Stahlberg. Cryo-EM structure of alpha-synuclein fibrils. *eLife*, 7(i):1–18, 2018.

[20] Ricardo Guerrero-Ferreira, Nicholas M.I. Taylor, Ana-Andreea Arteni, Pratibha Kumari, Daniel Mona, Philippe Ringler, Markus Britschgi, Matthias E. Lauer, Ali Makky, Joeri Verasdonck, Roland Riek, Ronald Melki, Beat H. Meier, Anja Böckmann, Luc Bousset, and Henning Stahlberg. Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife*, 8:1–24, 2019.

[21] Yunxiang Sun, Xinwei Ge, Yanting Xing, Bo Wang, and Feng Ding. β-barrel Oligomers as Common Intermediates of Peptides Self-Assembling into Cross-β Aggregates. *Scientific Reports*, 8(1):10353, 2018.

[22] Ioana M. Ilie, Divya Nayar, Wouter K. den Otter, Nico F.A. A van der Vegt, and Wim J. Briels. Intrinsic Conformational Preferences and Interactions in α-Synuclein Fibrils: Insights from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*, 14(6):3298–3310, 2018.

[23] Keiichi Masutani, Yu Yamamori, Kang Kim, and Nobuyuki Matubayasi. Free-energy analysis of the hydration and cosolvent effects on the β-sheet aggregation through all-atom molecular dynamics simulation. *The Journal of Chemical Physics*, 150(14):145101, 2019.

[24] Jon Pallbo, Emma Sparr, and Ulf Olsson. Aggregation behavior of the amyloid model peptide NACore. *Quarterly Reviews of Biophysics*, 52(May):e4, 2019.

[25] Masataka Yamauchi and Hisashi Okumura. Development of isothermal–isobaric replica-permutation method for molecular dynamics and Monte Carlo simulations and its application to reveal temperature and pressure dependence of folded, misfolded, and unfolded states of chignolin. *Journal of Chemical Physics*, 147(18), 2017.

[26] Masataka Yamauchi, Yoshiharu Mori, and Hisashi Okumura. Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble. *Biophysical Reviews*, 11(3):457–469, 2019.

[27] Satoru G Itoh and Hisashi Okumura. Replica-Permutation Method with the Suwa–Todo Algorithm beyond the Replica-Exchange Method. *Journal of Chemical Theory and Computation*, 9(1):570–581, 2013.

[28] Masataka Yamauchi and Hisashi Okumura. Replica sub-permutation method for molecular dynamics and monte carlo simulations. *Journal of Computational Chemistry*, 40(31):2694–2711, 2019.

[29] K Hukushima and K Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65(6):1604–1608, 1996.

[30] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[31] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[32] Hidemaro Suwa and Synge Todo. Markov Chain Monte Carlo Method without Detailed Balance. *Physical Review Letters*, 105(12):120603, 2010.

[33] Hisashi Okumura and Yuko Okamoto. Multibaric-Multithermal Molecular Dynamics Simulation of Alanine Dipeptide in Explicit Water. *Bulletin of the Chemical Society of Japan*, 80(6):1114–1123, 2007.

[34] Hisashi Okumura. Partial multicanonical algorithm for molecular dynamics and Monte Carlo simulations. *The Journal of Chemical Physics*, 129(12):124116, 2008.

[35] Hisashi Okumura and Yuko Okamoto. Temperature and Pressure Dependence of Alanine Dipeptide Studied by Multibaric − Multithermal Molecular Dynam-

ics Simulations. *The Journal of Physical Chemistry B*, 112(38):12038–12049, 2008.

[36] Hisashi Okumura. Optimization of partial multicanonical molecular dynamics simulations applied to an alaninedipeptide in explicit water solvent. *Physical Chemistry Chemical Physics*, 13(1):114–126, 2011.

[37] Hisashi Okumura. Temperature and pressure denaturation of chignolin: Folding and unfolding simulation by multibaric-multithermal molecular dynamics method. *Proteins: Structure, Function, and Bioinformatics*, 80(10):2397–2416, 2012.

[38] Hisashi Okumura and Satoru G Itoh. Transformation of a design peptide between the α-helix and β-hairpin structures using a helix-strand replica-exchange molecular dynamics simulation. *Physical Chemistry Chemical Physics*, 15(33):13852, 2013.

[39] Hiroaki Nishizawa and Hisashi Okumura. Comparison of Replica-Permutation Molecular Dynamics Simulations with and without Detailed Balance Condition. *Journal of the Physical Society of Japan*, 84(7):074801, 2015.

[40] Rupali Gupta, Shigeru Saito, Yoshiharu Mori, Satoru G Itoh, Hisashi Okumura, and Makoto Tominaga. Structural basis of TRPA1 inhibition by HC-030031 utilizing species-specific differences. *Scientific Reports*, 6(1):37460, 2016.

[41] Hiroaki Nishizawa and Hisashi Okumura. Rapid QM/MM approach for biomolecular systems under periodic boundary conditions: Combination of the density-functional tight-binding theory and particle mesh Ewald method. *Journal of Computational Chemistry*, 37(31):2701–2711, 2016.

[42] Hisashi Okumura and Satoru G. Itoh. Molecular dynamics simulations of amyloid-β (16–22) peptide aggregation at air-water interfaces. *The Journal of Chemical Physics*, 152(9):095101, 2020.

[43] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of

Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.

[44] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[45] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[46] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511–519, 1984.

[47] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[48] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384–2393, 1980.

[49] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[50] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[51] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[52] Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

[53] T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna. Symplectic quaternion scheme for biophysical molecular dynamics. *The Journal of Chemical Physics*, 116(20):8649–8659, 2002.

[54] H. Okumura, S. G. Itoh, and Y. Okamoto. Explicit symplectic integrators of molecular dynamics algorithms for rigid-body molecules in the canonical, isobaric–isothermal, and related ensembles. *J. Chem. Phys.*, 126(8):084103, 2007.

[55] M. H. Quenouille. Notes on Bias in Estimation. *Biometrika*, 43(3/4):353, 1956.

[56] Rupert G. Miller. The jackknife-a review. *Biometrika*, 61(1):1–15, 1974.

[57] Bernd A. Berg. *Markov Chain Monte Carlo Simulations and Their Statistical Analysis.* World Scientific, 2004.

[58] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[59] Sara Furlan, Giovanni La Penna, and Angelo Perico. Modeling the free energy of polypeptides in different environments. *Macromolecules*, 41(8):2938–2948, 2008.

[60] Yuko Okamoto and Ulrich H E Hansmann. Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms. *The Journal of Physical Chemistry*, 99(28):11276–11287, 1995.

[61] Satoru G. Itoh and Hisashi Okumura. Dimerization process of amyloid-β(29-42) studied by the hamiltonian replica-permutation molecular dynamics simulations. *Journal of Physical Chemistry B*, 118(39):11428–11436, 2014.

[62] Satoru G Itoh and Hisashi Okumura. Oligomer Formation of Amyloid-β(29–42) from Its Monomers Using the Hamiltonian Replica-Permutation Molecular Dynamics Simulation. *The Journal of Physical Chemistry B*, 120(27):6555–6561, 2016.

[63] Hisakazu Mihara and Yuta Takahashi. Engineering peptides and proteins that undergo α-to-β transitions. *Current Opinion in Structural Biology*, 7(4):501–508, 1997.

[64] David R. Booth, Margaret Sunde, Vittorio Bellotti, Carol V. Robinson, Winston L. Hutchinson, Paul E. Fraser, Philip N. Hawkins, Christopher M. Dobson, Sheena E. Radford, Colin C.F. Blake, and Mark B. Pepys. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*, 385(6619):787–793, 1997.

[65] Ruxi Qi, Yin Luo, Buyong Ma, Ruth Nussinov, and Guanghong Wei. Conformational distribution and α-helix to β-sheet transition of human amylin fragment dimer. *Biomacromolecules*, 15(1):122–131, 2014.

[66] Qin Qiao, Ruxi Qi, Guanghong Wei, and Xuhui Huang. Dynamics of the conformational transitions during the dimerization of an intrinsically disordered peptide: A case study on the human islet amyloid polypeptide fragment. *Physical Chemistry Chemical Physics*, 18(43):29892–29904, 2016.

[67] Murali Jayaraman, Ravindra Kodali, Bankanidhi Sahoo, Ashwani K. Thakur, Anand Mayasundari, Rakesh Mishra, Cynthia B. Peterson, and Ronald Wetzel. Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *Journal of Molecular Biology*, 415(5):881–899, 2012.

[68] Mingchen Chen, Weihua Zheng, and Peter G. Wolynes. Energy landscapes of a mechanical prion and their implications for the molecular mechanism of long-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, 113(18):5006–5011, 2016.

[69] Marina D. Kirkitadze, Margaret M. Condron, and David B. Teplow. Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. *Journal of Molecular Biology*, 312(5):1103–1119, 2001.

[70] Dmitri K. Klimov and D. Thirumalai. Dissecting the assembly of Aβ16-22 amyloid peptides into antiparallel β sheets. *Structure*, 11(3):295–307, 2003.

[71] Ana V. Rojas, Adam Liwo, and Harold A. Scheraga. A study of the α-helical intermediate preceding the aggregation of the amino-terminal fragment of the

β amyloid peptide (Aβ 1-28). *Journal of Physical Chemistry B*, 115(44):12978–12983, 2011.

[72] Yu Shan Lin, Gregory R. Bowman, Kyle A. Beauchamp, and Vijay S. Pande. Investigating how peptide length and a pathogenic mutation modify the structural ensemble of amyloid beta monomer. *Biophysical Journal*, 102(2):315–324, 2012.

[73] M. Fandrich. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal*, 21(21):5682–5690, 2002.

[74] Rebecca Nelson and David Eisenberg. Structural Models of Amyloid-Like Fibrils. In *Advances in Protein Chemistry*, volume 73, pages 235–282. 2006.

[75] Ioana M. Ilie and Amedeo Caflisch. Simulation Studies of Amyloidogenic Polypeptides and Their Aggregates. *Chemical Reviews*, 119(12):6956–6993, 2019.

[76] Phuong H Nguyen and Philippe Derreumaux. Structures of the intrinsically disordered Aβ, tau and α-synuclein proteins in aqueous solution from computer simulations. *Biophysical Chemistry*, 264:106421, 2020.

[77] Hang Yu, Wei Han, Wen Ma, and Klaus Schulten. Transient β-hairpin formation in α-synuclein monomer revealed by coarse-grained molecular dynamics simulation. *The Journal of Chemical Physics*, 143(24):243142, 2015.

[78] Timo Graen, Reinhard Klement, Asaf Grupi, Elisha Haas, and Helmut Grubmüller. Transient Secondary and Tertiary Structure Formation Kinetics in the Intrinsically Disordered State of α-Synuclein from Atomistic Simulations. *ChemPhysChem*, 19(19):2507–2511, 2018.

[79] Rafael Ramis, Joaquín Ortega-Castro, Rodrigo Casasnovas, Laura Mariño, Bartolomé Vilanova, Miquel Adrover, and Juan Frau. A Coarse-Grained Molecular Dynamics Approach to the Study of the Intrinsically Disordered Protein

α-Synuclein. *Journal of Chemical Information and Modeling*, 59(4):1458–1471, 2019.

[80] Guanghong Wei and Joan Emma Shea. Effects of solvent on the structure of the alzheimer amyloid-β(25-35) peptide. *Biophysical Journal*, 91(5):1638–1647, 2006.

[81] Nicolas L. Fawzi, Aaron H. Phillips, Jory Z. Ruscio, Michaeleen Doucleff, David E. Wemmer, and Teresa Head-Gordon. Structure and dynamics of the Aβ21-30 peptide from the interplay of NMR experiments and molecular simulations. *Journal of the American Chemical Society*, 130(19):6145–6158, 2008.

[82] Govardhan Reddy, John E. Straub, and D. Thirumalai. Influence of preformed Asp23-Lys28 salt bridge on the conformational fluctuations of monomers and dimers of Aβ peptides with implications for rates of fibril formation. *Journal of Physical Chemistry B*, 113(4):1162–1172, 2009.

[83] Yassmine Chebaro, Normand Mousseau, and Philippe Derreumaux. Structures and thermodynamics of alzheimer's amyloid-β Aβ(16-35) monomer and dimer by replica exchange molecular dynamics simulations: Implication for full-length Aβ fibrillation. *Journal of Physical Chemistry B*, 113(21):7668–7675, 2009.

[84] Luca Larini and Joan-Emma Shea. Role of β-Hairpin Formation in Aggregation: The Self-Assembly of the Amyloid-β(25–35) Peptide. *Biophysical Journal*, 103(3):576–586, 2012.

[85] Masamichi Nishino, Yuji Sugita, Takao Yoda, and Yuko Okamoto. Structures of a peptide fragment of β 2 -microglobulin studied by replica-exchange molecular dynamics simulations - Towards the understanding of the mechanism of amyloid formation. *FEBS Letters*, 579(24):5425–5429, 2005.

[86] Sang Hee Shim, Ruchi Gupta, Yun L. Ling, David B. Strasfeld, Daniel P. Raleigh, and Martin T. Zanni. Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6614–6619, 2009.

[87] Nicholas F. Dupuis, Chun Wu, Joan-Emma Shea, and Michael T. Bowers. The Amyloid Formation Mechanism in Human IAPP: Dimers Have β-Strand Monomer − Monomer Interfaces. *Journal of the American Chemical Society*, 133(19):7240–7243, 2011.

[88] Tingting Zhang, Yuanxin Tian, Zhonghuang Li, Shuwen Liu, Siming Liu, Xiang Hu, Zichao Yang, Xiaotong Ling, and Jiajie Zhang. Molecular Dynamics Study to Investigate the Dimeric Structure of the Full-Length α-Synuclein in Aqueous Solution. *Journal of Chemical Information and Modeling*, 57(9):2281–2293, 2017.

[89] Guanghong Wei, Andrew I. Jewett, and Joan-Emma Shea. Structural diversity of dimers of the Alzheimer amyloid-β(25–35) peptide and polymorphism of the resulting fibrils. *Physical Chemistry Chemical Physics*, 12(14):3622, 2010.

[90] Bogdan Tarus, Thanh T. Tran, Jessica Nasica-Labouze, Fabio Sterpone, Phuong H. Nguyen, and Philippe Derreumaux. Structures of the Alzheimer's Wild-Type Aβ1-40 Dimer from Atomistic Simulations. *Journal of Physical Chemistry B*, 119(33):10478–10487, 2015.

[91] Phuong H. Nguyen, Fabio Sterpone, Josep M. Campanera, Jessica Nasica-Labouze, and Philippe Derreumaux. Impact of the A2V Mutation on the Heterozygous and Homozygous Aβ1-40 Dimer Structures from Atomistic Simulations. *ACS Chemical Neuroscience*, 7(6):823–832, 2016.

[92] Phuong H. Nguyen, Fabio Sterpone, Ramon Pouplana, Philippe Derreumaux, and Josep M. Campanera. Dimerization mechanism of Alzheimer Aβ40 peptides: The high content of intrapeptide-stabilized conformations in A2V and A2T heterozygous dimers retards amyloid fibril formation. *Journal of Physical Chemistry B*, 120(47):12111–12126, 2016.

[93] Payel Das, Anita R. Chacko, and Georges Belfort. Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters Aβ42 Dimer Structure. *ACS Chemical Neuroscience*, 8(3):606–618, 2017.

[94] Viet Hoang Man, Phuong H. Nguyen, and Philippe Derreumaux. High-Resolution Structures of the Amyloid-β 1-42 Dimers from the Comparison of Four Atomistic Force Fields. *Journal of Physical Chemistry B*, 121(24):5977–5987, 2017.

[95] Frisco Rose, Miroslav Hodak, and Jerzy Bernholc. Mechanism of copper(II)-induced misfolding of Parkinson's disease protein. *Scientific Reports*, 1(Ii):1–5, 2011.

[96] Kelly N.L. Huggins, Marco Bisaglia, Luigi Bubacco, Marianna Tatarek-Nossol, Aphrodite Kapurniotu, and Niels H. Andersen. Designed hairpin peptides interfere with amyloidogenesis pathways: Fibril formation and cytotoxicity inhibition, interception of the preamyloid state. *Biochemistry*, 50(38):8202–8212, 2011.

[97] Ewa A. Mirecka, Hamed Shaykhalishahi, Aziz Gauhar, Şerife Akgül, Justin Lecher, Dieter Willbold, Matthias Stoldt, and Wolfgang Hoyer. Sequestration of a β-hairpin for control of α-synuclein aggregation. *Angewandte Chemie - International Edition*, 53(16):4227–4230, 2014.

[98] Yang Cao, Xuehan Jiang, and Wei Han. Self-Assembly Pathways of β-Sheet-Rich Amyloid-β(1-40) Dimers: Markov State Model Analysis on Millisecond Hybrid-Resolution Simulations. *Journal of Chemical Theory and Computation*, 13(11):5731–5744, 2017.

[99] Satoru G Itoh, Maho Yagi-Utsumi, Koichi Kato, and Hisashi Okumura. Effects of a Hydrophilic/Hydrophobic Interface on Amyloid-β Peptides Studied by Molecular Dynamics Simulations and NMR Experiments. *The Journal of Physical Chemistry B*, 123(1):160–169, 2019.

Chapter 7

Conclusions

Proteins perform biological functions by folding into the native structures that are the minimum free-energy state. On the other hand, the misfolding of proteins results in aggregation and amyloid fibril formation [1]. These aggregates and fibrils are believed to be associated with neurodegenerative disorders such as Alzheimer's and Parkinson's diseases. Therefore, understanding protein stability, aggregation mechanism, fibril formation process are fundamental and significant issues in protein science. All-atom molecular dynamics simulation is a useful tool to reveal the behavior of biomolecules at atomic-resolution, which complements experiments and theories. However, the conformation of biomolecules tends to get trapped in the local-minimum free-energy state due to its complex free-energy landscape. This situation hampers the efficient conformational sampling of the biomolecules. In this thesis, the author investigated the following four subjects regarding the above issues: (i) Development of isothermal-isobaric replica-permutation method (RPM); (ii) Development of replica sub-permutation method (RSPM); (iii) Investigation of structural stability of a small protein, chignolin, under various temperature and pressure conditions; and (iv) Dimerization process of α-synuclein fragments as the initial stage of aggregation.

In Chapter 3, the author proposed the isothermal–isobaric RPM to realize an efficient simulation of the biomolecules under a wide range of temperature and pressure. The RPM was originally proposed in the canonical ensemble. This means that the original RPM cannot evaluate physical quantities such as the differences

in the partial molar enthalpy, ΔH, and the partial molar volume, ΔV, because it does not control the pressure. In this study, the author extended the RPM to the isothermal–isobaric ensemble. The isothermal–isobaric RPM performs not only temperature but also pressure permutation to realize two-dimensional random walks in temperature and pressure spaces. Therefore, we can obtain physical quantities at various temperature and pressure values after the simulation. This is the first realization of a multidimensional RPM. The author compared sampling efficiency among the RPM with the Suwa–Todo, heat bath, metropolis algorithms, and the replica-exchange method (REM). The RPM with the Suwa–Todo algorithm achieves the most efficient parameter transitions among them. The author also pointed out that introducing subsets affects the transition ratio of the parameters. Therefore, one subset should include as many replicas as possible to realize more efficient sampling; however, it is a trade-off between sampling efficiency and the computational costs in the permutation trials. The upper limit of the number of replicas in one subset is from eight to ten in practical.

In Chapter 4, the author developed the RSPM. This method introduces a new permutation algorithm termed "sub-permutation" to reduce the number of permutation candidates. The sub-permutation is a set of permutations that include transitions only to neighboring parameters. Therefore, we can avoid using subsets, which is not desirable in the sampling efficiency, to reduce the number of permutation candidates. The author showed that the parameter transition by RSPM is nearly identical to the replica permutation that uses all possible combinations between replicas and parameters. Thus, the RSPM succeeds in reducing the number of permutation candidates without the loss of sampling efficiency. The author also demonstrated that the RSPM achieves more efficient simulation than the RPM and REM when we need more than eight replicas or perform multidimensional replica simulations. The RSPM will be a useful technique when we need more replicas to tackle a large-scale molecular system.

In Chapter 5, the author described an application study of the isothermal–isobaric RPM to the β-hairpin mini protein, chignolin, that consists of ten amino-acid residues. The chignolin has folded and misfolded structures as the native and

metastable states, respectively. Both of them have β-hairpin structures and similar to each other. In this study, the author performed the isothermal–isobaric replica-permutation simulation to the chignolin. We succeeded in the conformational sampling of the chignolin from $T = 300$–450 K and $P = 0.1$–500 MPa by one simulation. The author evaluated the melting temperature T_m, differences in the partial molar enthalpy ΔH, the partial molar volume ΔV between folded and unfolded states. They are in good agreement with the experimental results. In addition, we gained the difference in isobaric heat capacity ΔC_p between folded and unfolded states that could not evaluate in the experiment.

The author also revealed that the folded and misfolded states show completely different stability under high-pressure conditions; the folded structure decreases as pressure increases, whereas the misfolded structure increases as pressure increases. The difference in the partial molar volume was estimated as $\Delta V = V_\mathrm{misfold} - V_\mathrm{fold} = -4.6 \pm 0.2$ cm^3/mol at $T = 300$ K. That is, the misfolded chignolin has a smaller partial molar volume than the folded chignolin. The different pressure dependence of folded and misfolded state is strange because both folded and misfolded states have β-hairpin structure. To explore this unusual pressure dependence of the folded and misfolded states, the author focused on the orientation of Tyr2 and Trp9 side chains.In the folded structures, Tyr2 and Trp9 form a hydrophobic core. Consequently, the hydrogen bond Asp3N–Thr8O, which is essential for the β-hairpin formation, is exposed to the solvents on the opposite side of the hydrophobic core. Water molecules approach and break the hydrogen bond as pressure increases. As a result, the folded states decrease under high-pressure conditions. In the misfolded state, on the other hand, the hydrogen bonds that are essential to form β-hairpin structure are covered with the Tyr2 and Trp9 side chains. Namely, the hydrogen bonds do not expose to the solvent. The side chains protect the hydrogen bonds from the approaching water molecules under high-pressure conditions. Therefore, the misfolded structure becomes more stable with further compressing its structure under the high-pressure conditions.

In Chapter 6, the author investigated the dimerization process of α-synuclein fragments by the isothermal–isobaric replica-permutation molecular dynamics sim-

ulation. In particular, we focused on a segment from 68th to 78th of α-synuclein called NACore because the NACore is an essential region for both fibril formation and cytotoxicity of the α-synuclein. Analyses of the free-energy landscape and secondary structure revealed that the dimer structure with the intermolecular antiparallel β-bridge is the most stable. The author also found that intramolecular secondary structure such as α-helix and antiparallel β-bridge were stabilized in the pre-dimer state. The evaluation of differences in the partial molar enthalpy ΔH and partial molar entropy ΔS revealed that the NACore structures with intramolecular antiparallel β-bridge or α-helix in the pre-dimer state are enthalpically stable but entropically unfavorable than those in the monomer state. As for the dimerization process, we identified that the intermolecular β-bridge tends to spontaneously form between residues that do not have a secondary structure. This is because the NACore peptides have low a propensity to form intramolecular secondary structures. From the simulation, the author proposed that the pathway of an intermolecular β-bridge formation of amyloidogenic peptide depends on its propensity to form an intramolecular secondary structure.

In the two application studies in Chapters 5 and 6, we have gained fundamental knowledge about the stabilization/destabilization mechanisms of the β-hairpin protein under high-pressure conditions, and aggregation process of amyloidogenic peptides. Furthermore, the application studies have demonstrated that the isothermal–isobaric RPM and RSPM are a powerful tool to investigate molecular insight such as conformational changes and the aggregation process of proteins. They also have demonstrated that the isothermal–isobaric RPM and RSPM enable accurate estimation of the free energy and thermodynamics quantities, which is important to compare the simulations with the experimental data and understand the characteristics of the proteins. Because the isothermal–isobaric RPM and RSPM can access higher pressure and lower temperature region, cold denaturation of biomolecules is also one of the targets of these methods [2–7]. Furthermore, not only focusing on temperature and pressure permutations but also introducing other permutation parameters such as Hamiltonian [8–12], chemical potential [13], and pH [14] are future applications to make the RPM more useful.

References

[1] Fabrizio Chiti and Christopher M. Dobson. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry*, 86(1):27–68, 2017.

[2] Ryo Kitahara, Akira Okuno, Minoru Kato, Yoshihiro Taniguchi, Shigeyuki Yokoyama, and Kazuyuki Akasaka. Cold denaturation of ubiquitin at high pressure. *Magnetic Resonance in Chemistry*, 44(S1):S108–S113, 2006.

[3] Carlos F. Lopez, Richard K. Darst, and Peter J. Rossky. Mechanistic elements of protein cold denaturation. *Journal of Physical Chemistry B*, 112(19):5961–5967, 2008.

[4] Tatsuya Ikenoue, Young Ho Lee, Jõzsef Kardos, Miyu Saiki, Hisashi Yagi, Yasushi Kawata, and Yuji Goto. Cold denaturation of α-synuclein amyloid fibrils. *Angewandte Chemie - International Edition*, 53(30):7799–7804, 2014.

[5] Changwon Yang, Soonmin Jang, and Youngshang Pak. A fully atomistic computer simulation study of cold denaturation of a beta-hairpin. *Nat Commun*, 5:5773, 2014.

[6] Sang Beom Kim, Jeremy C. Palmer, and Pablo G. Debenedetti. Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 113(32):8991–8996, 2016.

[7] Andrea Arsiccio, James McCarty, Roberto Pisano, and Joan Emma Shea. Heightened Cold-Denaturation of Proteins at the Ice-Water Interface. *Journal of the American Chemical Society*, 142(12):5722–5730, 2020.

[8] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *Journal of Chemical Physics*, 113(15):6042–6051, 2000.

[9] Hiroaki Fukunishi, Osamu Watanabe, and Shoji Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058–9067, 2002.

[10] Pu Liu, Byungchan Kim, Richard A Friesner, and J Berne, B. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences*, 102(39):13749–13754, 2005.

[11] Satoru G Itoh and Hisashi Okumura. Coulomb replica-exchange method: Handling electrostatic attractive and repulsive forces for biomolecules. *Journal of Computational Chemistry*, 34(8):622–639, 2013.

[12] Satoru G Itoh and Hisashi Okumura. Hamiltonian replica-permutation method and its applications to an alanine dipeptide and amyloid-β (29–42) peptides. *Journal of Computational Chemistry*, 34(29):2493–2497, 2013.

[13] Jie Ji, Tahir Çağin, and B. Montgomery Pettitt. Dynamic simulations of water at constant chemical potential. *The Journal of Chemical Physics*, 96(2):1333–1342, 1992.

[14] Satoru G Itoh, Ana Damjanović, and Bernard R Brooks. ph replica-exchange method based on discrete protonation states. *Proteins*, 79(12):3420–3436, 2011.

Acknowledgments