OUKA
Osaka University Knowledge Archive

Title	Theoretical Study for Binding Processes of Intrinsically Disordered Proteins
Author(s)	白井, 伸宙
Citation	大阪大学, 2015, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/52317
rights	
Note	

Osaka University Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

Osaka University

Theoretical Study for Binding Processes of Intrinsically Disordered Proteins

Nobu C. Shirai

Department of Physics, Graduate School of Science

Osaka University

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

February 5, 2015

Preface

Intrinsically disordered proteins (IDPs) are proteins without stable three-dimensional structures under physiological conditions. They are also called intrinsically unstructured proteins (IUPs). Fig. 1 shows the number of search results of Google Scholar (`http://scholar.google.com`) plotted against the year of publication. The search words are "intrinsically disordered" protein and "intrinsically unstructured" protein. Both curves begin to increase at the end of 20th century. It might not be a coincidence that the review paper of Wright and Dyson entitled "Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm" was published in 1999 (1). These numbers have also increased in recent years, and thousands of scholarly documents including these words have been published in 2013. Therefore, I can say that IDP has been recognized as a new research target in protein science.

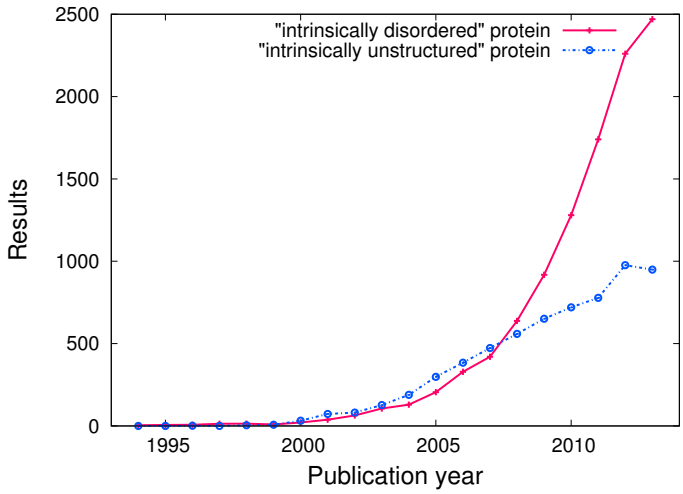

Figure 1: The number of the search results of Google Scholar plotted against publication year (1994–2013).

0. PREFACE

The new edition of "Molecular Biology of the Cell", which is one of the most famous textbooks in molecular biology, was published in December 2014 (2). In this textbook, there is a newly added section entitled "Proteins Contain a Surprisingly Large Amount of Intrinsically Disordered Polypeptide Chain" in Chapter 3, and this section explains IDPs with some examples. Thus, I can also say that the existence of IDPs has become common knowledge in molecular biology. In this thesis, I explain the basic properties of IDPs in comparison with those of folded proteins in Chap. 1 of Part I.

Although the existence of IDPs is now well established, there is a fundamental question about intrinsic disorder. Genome-wide analysis using a bioinformatic disorder predictor shows that higher organisms have more IDPs in their genomes, and about one-third of human proteins contain half or more disordered residues in their amino-acid sequences. Their persistence under selective pressure makes me think of functional advantages over globular proteins in some biological functions. Then, I would like to ask, "What is the advantage of IDPs?" In Part II, I discuss about the advantage of IDPs on binding processes by constructing and analyzing the lattice model of an IDP.

α-synuclein (α-syn) is one of IDPs which forms amyloid fibrils in highly concentrated solutions. It is known that the fibril formation is accelerated by macromolecular crowding, and White et al. suggested that this acceleration is the characteristic property of IDPs (3). In order to explore the relation between intrinsic disorder of α-syn and macromolecular crowding in the fibrillation processes I constructed and analyzed the lattice model of IDPs. In Part III, I describe the basic properties of α-syn, the construction of the model and its analysis.

Contents

Preface i

List of Figures vii

List of Tables xi

I Introduction to Intrinsically Disordered Proteins 1

1 Intrinsically Disordered Proteins 3

1.1 Functions and structures of folded proteins 3

1.2 Thermodynamics of protein folding 4

1.3 Thermodynamic models for protein folding 6

1.4 Disordered proteins are also functional 9

1.5 Macromolecular crowding 12

II Binding Processes of an Intrinsically Disordered Protein to a Specific Target Protein 13

2 Overview 15

3 Model and Method 17

3.1 Calculating free-energy landscapes by using the MSOE 23

4 Result and Discussion 25

4.1 Intrinsic disorder represented as a mixture of several states 25

4.2 Size dependence of the specific heat of the IDP 25

4.3 Thermal stability of the native bound state 27
4.4 Stepwise target recognition of the IDP due to multiform binding effect . 31
4.5 Coupled folding and binding of the flexible target 32
4.6 Other examples of the free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ 33

III Binding and Fibrillation Processes of α-Synuclein 37

5 Overview 39

6 Introduction to α-Synuclein 41

7 Models 45

8 Methods 49
8.1 Metropolis method for the lattice gas model of α-synuclein including crowding agents 49
8.2 Method of adding crowding agents to the snapshots produced by the simulation without explicit crowding agents 50

9 Results 53
9.1 The effect of macromolecular crowding is taken into account as the effective internal entropy of disordered states 53
9.2 Protofibril formation above the critical concentration of α-synuclein . . 55
9.3 Macromolecular crowding induces protofibril formation 61
9.4 α-state tetramers suppress protofibril formation 63
9.5 Size dependence of fibril formation and α-state-tetramer formation . . . 65

10 Discussion 69
10.1 Scenario of protofibril formation induced by macromolecular crowding . 69
10.2 An explanation for the controversy over the observation of helically folded tetramers 71
10.3 Solid–liquid equilibrium analogy 71
10.4 A coarse-grained model with internal states: funnel gas model 72

11 Summary 73

References **81**

List of Figures

1 The number of the search results of Google Scholar plotted against publication year (1994–2013). i

1.1 Examples of functional proteins. 4
1.2 Three types of representations of a tertiary structure of ribonuclease. . . 5
1.3 Typical behavior of the temperature dependence of specific heat. 6
1.4 The Gō model. (a) The native structure for the A and B models of the Gō model. (b) The definition of the native contact pairs for the A and B models. 7
1.5 Thermodynamic properties of the Gō model. (a) Energy E dependences of entropy S for the A, B and C models are shown as black curves. (b) The specific heat C for the A, B and C models. 8
1.6 The conceptual picture of the energy landscape given by the Gō model. 10
1.7 The conceptual picture of coupled folding and binding. 10

3.1 The lattice model of an IDP and a specific target. 18
3.2 Hydrophobic interactions in the native bound state. 19
3.3 The conformations of the left and right parts of the designed target. . . 20
3.4 The conceptual picture of an $L \times L \times L$ cubic system. 22

4.1 (A) Free-energy landscapes $F(N_\alpha, E_{\rm ID})$ of the IDP for $L = \infty$ at low, medium and high temperatures shown from left to right, respectively. (B) Two example conformations of the ground state. 26
4.2 α-helix contact pairs of the lattice model of an IDP. 26
4.3 Specific heat of the IDP ($C_{\rm ID}$) for $L = 16, 32, 64, 128$ and ∞. 27

4.4 Specific heat C for $L = 32$ with an IDP and the three types of binding surfaces: RR-target, FF-target and HH-target. 28
4.5 Thermal averages of the α-helix contact ($\langle N_\alpha \rangle$) for the four systems ($L = 32$): IDP only, IDP/RR-target, IDP/FF-target and IDP/HH-target. 28
4.6 Free-energy landscapes $F(N_\alpha, E_{\rm ID})$ of the IDP for $L = 32$ with and without the three types of binding surfaces: RR-target, FF-target and HH-target at the corresponding binding temperatures. 29
4.7 System size dependence ($L = 16 - 128$) of the binding temperature $T_{\rm B}$. 30
4.8 (A) Free-energy landscapes $F(E_{\text{ID-T}}, N_\alpha)$, $F(E_{\text{ID-T}}, E_{\rm ID})$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/FF-target, IDP/HF-target and IDP/HH-target at the binding temperature of each system. (B) Example structures of the IDP/target complexes. 31
4.9 The conceptual pictures of multiform binding of two different hidden binding sites. 33
4.10 Free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\rm L}, \xi_{\rm R}))$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/HH-target and IDP/FF-target at the binding temperature of each system. 34
4.11 Free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\rm L}, \xi_{\rm R}))$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/HR-target and IDP/HF-target at the binding temperature of each system. 35

6.1 The conceptual figure of the equilibrium of α-syn. 42

7.1 Lattice gas model of α-syn. (a) Definitions of α-syn and crowding agent. (b) Definitions of intermolecular interaction of two α-syns. (c) Snapshot of a small system ($L = 8$, $N = 16$). 46

9.1 Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 0$. 57
9.2 Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 0$. . . . 57
9.3 Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 1$. 58
9.4 Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 1$. . . . 58
9.5 Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 2$. 59
9.6 Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 2$. . . . 59

9.7 Equilibrium number of the β states which are components of β chains of length ℓ_β in the $L = 128$ systems. (a) N dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD and B^oD models. (b) μ dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD model. (c) N dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD and BDA models. . . . 60

9.8 N and μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 128$. 62

9.9 N and μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\bar{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 128$. 64

9.10 Snapshots for the BD and BDA models with $L = 64$, $N = 256$ and $\mu = 0, 1$ and 2. 66

9.11 N and μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 64$. 67

9.12 N and μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\bar{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 64$. 68

10.1 Stacked graphs of the equilibrium numbers of internal states for the BD and BDA models with $L = 128$, $\mu = 0$, 1, 2 and ∞, and $N = 431$. . . . 70

10.2 (a) $\langle n_\beta \rangle$ dependences of $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$ and $\langle n_d \rangle$ of the BD, B^oD, BDA and B^oDA models with $L = 128$. (b) $\langle n_\beta \rangle$ dependences of $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$ and $\langle n_\alpha^{(\bar{t})} \rangle$ for the BDA and B^oDA models with $L = 128$. 70

LIST OF FIGURES

List of Tables

3.1 Definitions of surface types for each side of the binding surface. 21
3.2 The list of physical quantities to calculate free energies. 23

9.1 The list of values for model parameters and k_B. 56
9.2 The list of values for L, N and μ. 56
9.3 The list of physical quantities for analysis. 61

Abbreviations and Acronyms

AMP Adenosine MonoPhosphate

ATP Adenosine TriPhosphate

Cdk2 Cyclin-dependent kinase 2

CREB Cyclic-AMP-Response-Element-Binding protein

DSC Differential Scanning Calorimetry

E. Coli Escherichia coli

EM Electron Microscopy

FPD Familial Parkinson's Disease

IDP Intrinsically Disordered Protein

ITC Isothermal Titration Calorimetry

IUP Intrinsically Unstructured Protein

KIX domain KID-binding domain

MSOE Multi-Self-Overlap Ensemble

NMR Nuclear Magnetic Resonance

PD Parkinson's Disease

PDB Protein Data Bank

pKID phosphorylated Kinase Inducible activation Domain

LIST OF TABLES

RNase RiboNuclease

SAXS Small Angle X-ray Scattering

SAW Self-Avoiding Walk

SPR Surface Plasmon Resonance

syn synuclein

Part I

Introduction to Intrinsically Disordered Proteins

1

Intrinsically Disordered Proteins

Proteins and protein fragments without secondary and tertiary structures under physiological conditions are called intrinsically disordered proteins (IDPs). Much attention has been paid to IDPs thanks to the development of techniques for experiments and bioinformatics to analyze structural and physicochemical properties of IDPs since the end of the 20th century. One of the reasons for the interest is that aggregates of IDPs are found in the pathologies of neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and Huntington's disease. The characteristic binding processes of IDPs, which are called "coupled folding and binding", also interest researchers including physicists and chemists. In this chapter, I compare the biological functions and physicochemical properties of IDPs with those of folded proteins to explain the specific properties of IDPs.

1.1 Functions and structures of folded proteins

Proteins are polypeptides with amino-acid sequence encoded by genes, and they have diverse functions in biological systems. Some examples are shown in Fig. 1.1. Enzymes accelerate the rate of chemical reactions. Structural proteins support a cell structure. Motor proteins convert the free energy of adenosine triphosphate (ATP) into power to move themselves. Transporters actively or inactively transport specific substances across biological membranes. This functional diversity is realized by protein structures.

The chemical structures of proteins are linear chains of naturally occurring amino acids, which are called primary structures. Some protein regions fold into α-helix or

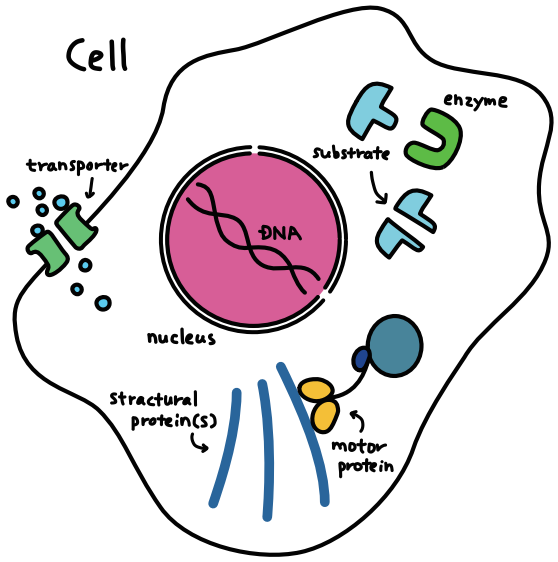

Figure 1.1: Examples of functional proteins.

β-sheet structures in a sequence-dependent manner. These local structures, which are called secondary structures, interact with each other to form thermally stable three-dimensional structures, which are called tertiary structures. Fig. 1.2 shows three types of representations of a tertiary structure of ribonuclease (RNase) as examples. Assemblies of protein subunits which have tertiary structures are called quaternary structures. The atomic coordinates of tertiary or quaternary structures are determined by X-ray crystallography, nuclear magnetic resonance (NMR) or electron microscopy (EM). Over 100,000 resolved structures of proteins and nucleic acids are stored in the Protein Data Bank (PDB) (4) today. In many cases, these structural data tell us about the biological functions of the proteins. These biologically functional structures are referred to as native structures. The processes for forming native structures are called protein folding.

1.2 Thermodynamics of protein folding

Anfinsen and his collaborators revealed that RNase spontaneously refolds into a native structure after removal of denaturant (6, 7). They hypothesized that the genetic sequences of amino acids determine the thermally stable native structures which are formed after spontaneous folding, and this hypothesis is known as Anfinsen's dogma.

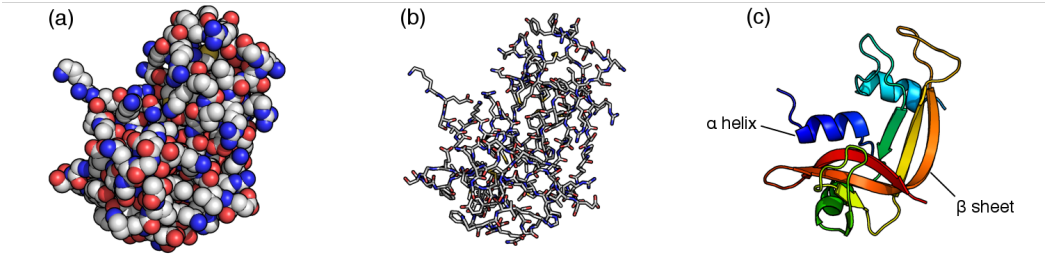

Figure 1.2: Three types of representations of a tertiary structure of ribonuclease (RNase; PDB code, 1DZA). (a) Atoms of RNase except hydrogen atoms are shown as spheres. Carbon, nitrogen, oxygen and sulfur atoms are colored white, blue, red and yellow, respectively. (b) Atoms and covalent bonds of RNase except hydrogen atoms are represented as branched sticks. Colors are the same as (a). (c) Cartoon representation of backbone atoms of RNase. α-helix structures are shown as spirals, and β-sheet structures are shown as arrows. NH_2-terminal (N-terminal) of the protein is colored blue, and COOH-terminal (C-terminal) is colored red. Other residues gradually change their colors along with the sequence. The images are created with PyMOL (5).

Since polypeptide chains with random amino-acid sequences never fold into a specific structure, it is considered that the structural properties of proteins are acquired during Darwinian evolution.

The characteristic thermodynamic properties of proteins are given by calorimetric experiments. Fig. 1.3 shows the typical behavior of temperature dependence of the specific heat. Proteins are in a native state in the low temperature region, and they are in denatured states above the peak temperature. There is a first-order-like transition around the peak temperature. In the early study of differential scanning calorimetry (DSC), Privalov and Khechinashvili measured the temperature dependence curves of the specific heat for the five proteins of which the X-ray structures had already been resolved, and they showed that the van't Hoff enthalpy calculated from two states model, which includes the native state and the denatured state, agrees well with the enthalpy measured by the experiments (8). Therefore, the heat denaturation transition shown above is called two-state transition. The two-state transition is generally found in single-domain globular proteins, and this transition suggest that the intramolecular interactions to form a native conformation are cooperative (9).

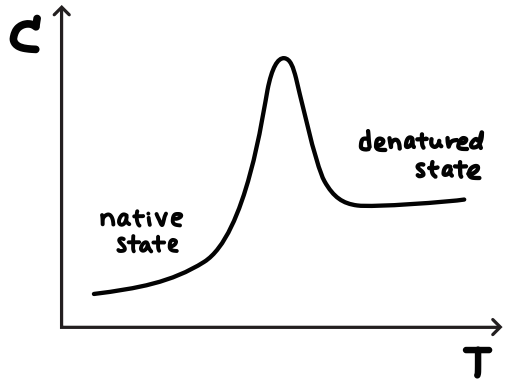

Figure 1.3: Typical behavior of the temperature dependence of specific heat.

1.3 Thermodynamic models for protein folding

The two-state transition of proteins is described by theoretical works of Gō and his collaborators based on statistical mechanics. In the second half of 1970, Taketomi and Gō constructed lattice protein models on two-dimensional lattices (10, 11). They modeled a protein as a self-avoiding walk (SAW) with intrachain interactions. A SAW is a path on a lattice with the spatial restriction that the path must not visit the same site more than once, which is the representation of the excluded volume effect of amino acid residues. Each amino acid residue is simplified as a single site on a two-dimensional lattice in this representation. They modeled three proteins labeled A, B and C. The amino-acid length of these models is 49.

The definition of the intrachain interactions of the A and B models are shown in Fig. 1.4 (b). In this figure, black squares specify the intrachain interaction pairs of the A model. These pairs correspond to the nearest-neighboring pairs in the native conformation shown in Fig. 1.4 (a) except the residue pairs connected by peptide bonds. These residue pairs are called native contact pairs. If the native contact pair residues are on the two adjacent sites in an arbitrary conformation of a lattice protein, each of these residue pairs have the negative energy $-\varepsilon$ ($\varepsilon > 0$). The entropic effect of solvent is implicitly included in the intrachain interaction, and thus the intrachain interactions defined above are free energy rather than energy. The effect of pressure exerted on a protein solution is not included in the model, and I do not distinguish enthalpy and energy, or the Gibbs free energy and the Helmholtz free energy. In the native conformation shown in Fig. 1.4 (a), there are 36 native contact pairs, and thus

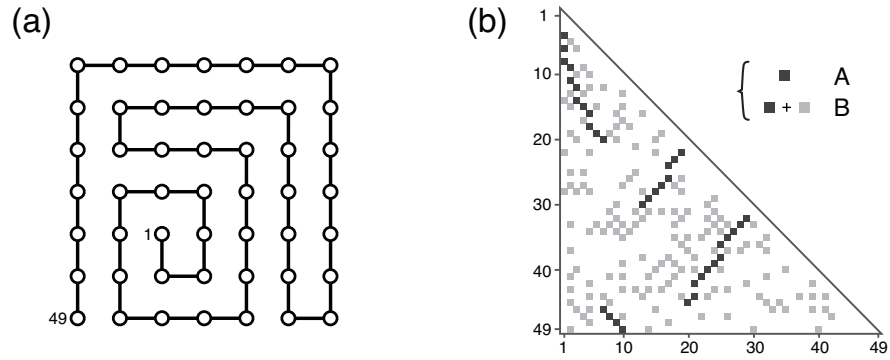

Figure 1.4: The Gō model. (a) The native structure for the A and B models of the Gō model. (b) The definition of the native contact pairs for the A (black) and B (black and gray) models. The axes are indices of amino acid residues. These figures are reproduced from Gō and Taketomi (1978), *Proc. Natl. Acad. Sci. USA* **75**, 559–563.

the ground-state energy E_0 is -36ε. This is the original version of the so-called Gō model. In the original Gō model, there is no frustration in the interaction pairs. In other words, the interactions of all the contact pairs are consistent with each other. This is one of the ideal protein models which represent the cooperative interactions of proteins.

The intrachain interaction pairs of the B model are defined by adding the pairs specified by the gray squares in Fig. 1.4 (b) to the pairs of the A model. These added pairs were randomly selected from residue pairs, and also have the negative energy $-\varepsilon$ ($\varepsilon > 0$). These newly added interaction pairs do not change the ground state from the native conformation shown in Fig. 1.4 (a), and thus the ground-state energy is the same as the A model. There are frustrations in the interaction pairs, and thus the interactions are not fully consistent.

The intrachain interactions of the C model is defined by all pairs of amino acid residues except the residue pairs connected by peptide bonds. Although the ground-state energy is the same value as the A and B models, there are many degenerated ground states. This is the inconsistent limit of intramolecular interactions.

I calculated the number of state $\Omega(E)$ of these models using the multi-self-overlap ensemble (MSOE) Monte Carlo method (12, 13), which is explained in Chap. 3.1. I take $\varepsilon = 1$ and the Boltzmann constant $k_B = 1$. In the calculation of $\Omega(E)$, I did not consider translational entropy, and thus the number of ground states of the A

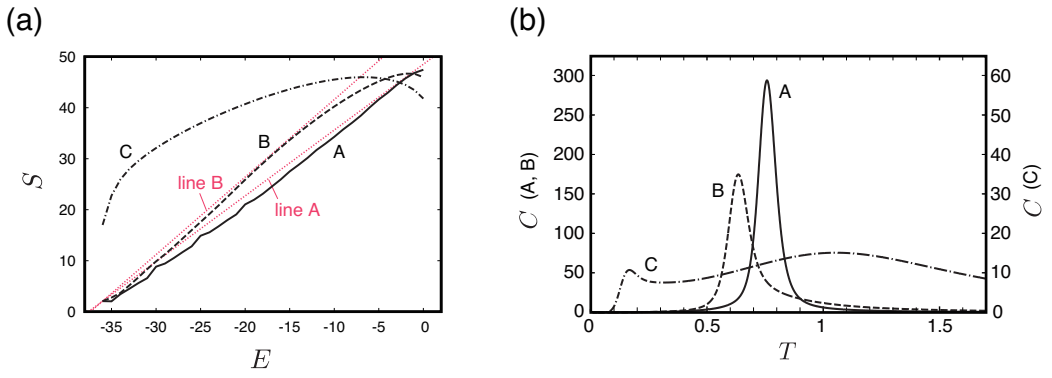

Figure 1.5: Thermodynamic properties of the Gō model. (a) Energy E dependences of entropy S for the A, B and C models are shown as black curves. The two red lines labeled as line A and line B are tangent at the two points of the E–S curves of the model A and the model B, respectively. The line A is written as $S = \frac{1}{T_f^{\mathrm{A}}}(E - E_0) + \log \Omega(E_0)$, and the line B is written as $S = \frac{1}{T_f^{\mathrm{B}}}(E - E_0) + \log \Omega(E_0)$, where $T_f^{\mathrm{A}} = 0.78$, $T_f^{\mathrm{B}} = 0.66$, $E_0 = -36$ and $\Omega(E_0) = 8$. (b) The specific heat C for the A, B and C models. The scale of the vertical axis on the left is shown for the A and B models, and that on the right is shown for the C model.

and B models is 8 including rotation-symmetric conformations and mirror-symmetric conformations. Both of the energy dependence of entropy S and partition function Z are calculated from $\Omega(E)$. The former one is written as the logarithm of the number of states $S = \log \Omega(E)$. The latter one is given by $Z = \sum_E \Omega(E) \exp(-E/T)$. Specific heat C is calculated by

$$C = \frac{\mathrm{d}\langle E \rangle}{\mathrm{d}T} = \langle E^2 \rangle - \langle E \rangle^2 = \frac{\sum_E E^2\, \Omega(E)\, \mathrm{e}^{-\frac{E}{T}}}{Z} - \frac{\left\{\sum_E E\, \Omega(E)\, \mathrm{e}^{-\frac{E}{T}}\right\}^2}{Z^2} \tag{1.1}$$

The energy dependence of S is shown in Fig. 1.5 (a), and the temperature dependence of C is shown in Fig. 1.5 (b). Thanks to the power of today's computer and the power of the MSOE, I obtained more accurate entropy and specific heat than those in the Taketomi and Gō's paper.

The specific heat of the A model in Fig. 1.5 (b) has a sharp peak, and shows a first-order-like transition. I can calculate the free energy $F = E - TS$ for each point on the curves of $S(E)$ in Fig. 1.5 (a). Energy state with smallest free energy is the most probable state in equilibrium. If I rewrite $F = E - TS$ as $S = \frac{1}{T}E - \frac{F}{T}$, I can determine $-\frac{F}{T}$ by $S(0)$. In the high temperature, the most probable state is in $E \sim -2$–0. The

large number of disordered conformations is included in these high-entropy states, and thus the A model is in denatured states. As T decreases, the most probable state switches to the ground state at T_f^{A}. Here, T_f^{A} is the inverse of the slope of the line A shown in Fig. 1.5 (a), which is tangent at the two points of the E–S curve of the A model, and this temperature corresponds to the peak temperature of the specific heat. There are no intermediate states in the transition, and the A model shows the two-state transition between the denatured state and the native state. Gō and Taketomi are the first researchers who discussed the characteristic energy landscape of proteins using a theoretical model with the simplified protein structures based on statistical mechanics (10, 11, 14).

Although the peak of the specific heat of the B model is less sharp than that of the A model, it also show a two-state transition at T_f^{B}, which is the inverse of the slope of the line B shown in Fig. 1.5 (a) and corresponds to the peak temperature of the specific heat. The specific heat of the C model, however, does not have a single sharp peak, and the most probable states gradually approaches to the native state. This is different from the behavior which I described in Fig. 1.2. Therefore, it might be true that consistency of intrachain interactions gives thermodynamic properties of proteins. After this paper, Gō started to argue that the consistency of interactions is important to the two-state transition, and he called it as “consistency principle“ in the thermodynamics of proteins (15).

Fig. 1.6 shows the conceptual picture of the characteristic energy landscape given by Taketomi and Gō. After the work by Gō and his collaborators, this energy landscape began to be known as a funnel-like energy landscape or a funnel landscape, which is named after the funnel-shaped object given by rotating the E–S curve.

1.4 Disordered proteins are also functional

I have discussed the functions and structural properties of folded proteins above. Here, I introduce the proteins which are unfolded and functional under physiological conditions. These proteins are called intrinsically disordered proteins (IDPs) or intrinsically unstructured proteins (IUPs). In 1999, Wright and Dyson reviewed NMR studies about the binding processes of IDPs to their biological targets (1). Since disordered regions of proteins were notorious as crystallization inhibitor, they were not considered as research

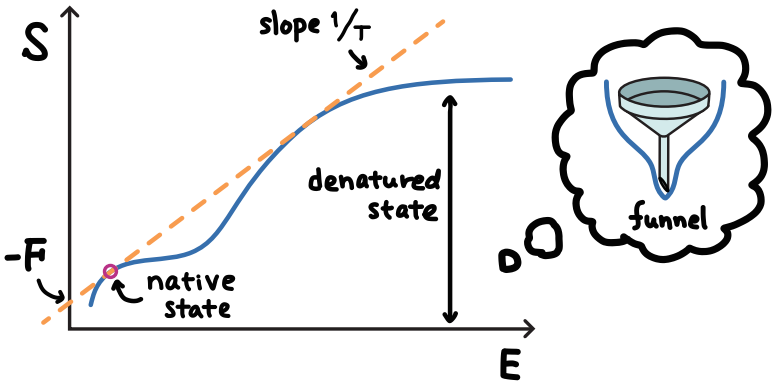

Figure 1.6: The conceptual picture of the energy landscape given by the Gō model. If I rotate the E–S curve around the E axis, it will become a funnel-shaped object. This shape is the word origin of the "funnel landscape".

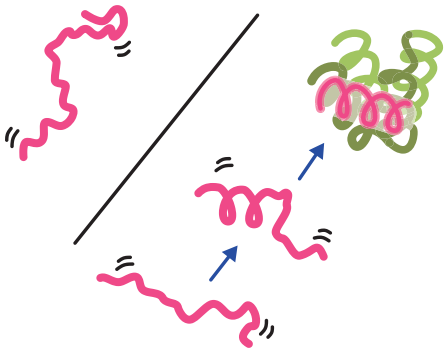

Figure 1.7: The conceptual picture of coupled folding and binding. The IDP is shown in pink and the target is shown in green.

targets for many researchers at that time. Their review, however, highlighted the functional aspects of IDPs, and brought IDPs to the attention of researchers. Therefore, the review is considered as one of the important milestones of IDPs. Wright and Dyson reported that many of IDPs have specific binding partners, and characteristic binding processes called coupled folding and binding have been observed by NMR. Fig. 1.7 shows the conceptual picture of the coupled folding and binding. The typical functions of IDPs are the following: binding to nucleic acids regulates transcription or translation, and binding to a signaling target leads to signal transduction. In these processes, IDPs bind to their partners and fold into a specific structure (16, 17, 18, 19, 20).

Wright and Dyson also mentioned the development of the bioinformatic analysis of protein databases in their 1999 review. Wootton and Federhen defined the complexity

of protein sequences, and analyzed protein sequence of PDB and Swiss-Prot (21) in their 1993 and 1994 papaers (22, 23, 24). Swiss-Prot is a protein-sequence database and is now a part of UniProt (25), Protein sequences of PDB are representative of structural proteins while those of Swiss-Prot include many proteins of which structures have not been resolved. They found that low complexity and highly repetitive sequences are only in Swiss-Prot. This work implicitly suggested the sequence properties of unstructured proteins.

After Wootton and Federhen's papers, many research groups started to develop tools to predict disordered regions based on sequence properties. Ward and his coworkers developed DISPRED2, and concluded that one-third of eukaryotic proteins are IDPs(26). Xue and his collaborators obtained similar results (27, 28).

Many IDPs are known as hub proteins in signaling networks (29, 30, 31, 32), and the evolutionary persistence of IDPs under selective pressure makes us think of their functional advantages over globular proteins in signaling processes. One of the proposed advantages of IDPs is adaptability of their shapes to several binding targets, which is a suitable property for signaling hubs (33, 34). Another proposed advantage of IDPs is the fast binding process due to its comparatively long capture radius (35), which is called the fly-casting mechanism. In Part. II, I propose a new advantage of IDPs, focusing on how they recognize their targets.

In the recognition process, an IDP transiently interacts with a tentative target and searches for a biologically functional bound form. If the transient complex is not sufficiently stabilized by the intermolecular interaction, the complex is finally dissociated by thermal fluctuations with a comparatively short lifetime. Since the lifetime depends on the free-energy barrier to dissociation, stabilization of the complex in the early stage of the binding process is important to successfully recognize the correct target at an accidental encounter.

In some of IDP-target complexes, however, not all of the interaction sites are exposed outside of the target molecule and easily accessible. The $p27^{Kip1}$/cyclin A/cyclin-dependent kinase 2 (Cdk2) complex is a good example. p27 is known to be an IDP (33, 36), and binds to the binary complex of cyclin A and Cdk2 (37). Cdk2 in the binary complex has intramolecular hydrophobic interaction sites (38, 39) which are finally exposed to p27 upon binding (40). As another example, the phosphorylated

kinase inducible activation domain (pKID) of the transcription factor cyclic-AMP-response-element-binding protein (CREB), which is known to be an IDP (18), binds to the KID-binding (KIX) domain of CREB binding protein inserting one of the hydrophobic residues deeply into the buried interaction pocket of KIX (18). In both examples of hidden binding sites, binding processes are not simple two-state transitions between a dissociated state and a bound state. It has been shown using isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) that p27 binds cyclin A before it binds to Cdk2 (36). In the binding process between pKID and KIX, the intermediate bound state has been observed by NMR titrations and ^{15}N relaxation dispersion (19). In this intermediate state, the buried interaction site of the target does not completely interact with pKID. I consider that the stepwise target recognition is a characteristic binding process between an IDP and its target with the "hidden" binding sites and it is a functional advantage of IDPs, that is, the stepwise target recognition enables IDPs to access these hidden binding sites. In Part. II, I construct a lattice model of an IDP based on the extended HP model (41), and investigate the target recognition processes using computer simulations of the model.

1.5 Macromolecular crowding

A cellular environment is crowded with macromolecules such as proteins and nucleic acids. Zimmerman and Trach estimated that the concentration of macromolecules is around 300–400 [g/l] in the cytoplasm of *Escherichia coli*, which is a type of bacteria, from their experiments (42). Minton pointed out the importance of the effect of this crowded environment, which is called macromolecular crowding effect, on structural stability and reactivity of proteins and other macromolecules (43). He theoretically discussed the excluded volume effect of macromolecules based on statistical thermodynamics, and suggested that macromolecular crowding stabilizes the tertiary or quaternary structures of proteins. The predictions of the theoretical models semi-quantitatively agree with experiments using crowding agents such as Ficoll, dextran and polyethylene glycol to mimic the effect of macromolecular crowding (44, 45). In Part III, I discuss a relation between macromolecular crowding and the stability of IDPs.

Part II

Binding Processes of an Intrinsically Disordered Protein to a Specific Target Protein

2

Overview

An IDP lacks a stable three-dimensional structure, while it folds into a specific structure when it binds to a target molecule. In some IDP-target complexes, not all target-binding surfaces are exposed on the outside, and intermediate states are observed in their binding processes. I consider that stepwise target recognition via intermediate states is a characteristic of IDP binding to targets with the "hidden" binding sites. To investigate IDP binding to the hidden target binding sites, I construct an IDP lattice model based on the HP model. In my model, the IDP is modeled as a chain and the target is modeled as a highly coarse-grained object. I introduce motion and internal interactions to the target to hide its binding sites. In the case of unhidden binding sites, a two-state transition between the free states and a bound state is observed, and I consider that this represents coupled folding and binding. Introducing the hidden binding sites, I find an intermediate bound state in which the IDP forms various structures to temporarily stabilize the complex. The intermediate state provides a scaffold for the IDP to access the hidden binding site. I call this process multiform binding. I conclude that the structural flexibility of IDPs enables them to access hidden binding sites, and this is a functional advantage of IDPs.

3

Model and Method

In the HP model, proteins are simplified as two-letter sequences of polar (P) and hydrophobic (H) residues and only hydrophobic interactions between H residues are considered. This model has been used to analyze the kinetics and thermodynamics of protein folding (46, 47, 48), in which the hydrophobic interactions play a major role. Although IDPs themselves do not contain sufficient hydrophobic residues to stabilize a specific structure (49), hydrophobic interactions are still important for IDPs to bind to their target molecules (16, 17, 18, 40, 50). I then consider that the HP model can be applied to a binding process of an IDP and a target.

I assume two important properties which a model for an IDP and its target should have. Firstly, I assume that IDPs do not form specific structures in free states. Secondly, I assume that the IDP and its target have a native bound state which is a non-degenerate ground state, and that the IDP adopts a specific structure in this bound state. The second assumption comes from the fact that many of the specific bound forms of IDP-target complexes were detected by crystal or NMR structures with sufficient stability. I introduce the second type of hydrophobic residues (H') to the HP model. In this extended model, a protein is modeled as a sequence of three types of residues: two types of hydrophobic residues (H, H') and one type of polar residue (P). I construct a model of IDP by using this model, in which the IDP is modeled as a chain and the target is modeled as a highly coarse-grained object designed as a combination of plates which represent the binding surface. I design a unique ground state of these two molecules as a bound state of an IDP and its target. The designed bound state is shown in Fig. 3.1A, and I call it the native bound state in this thesis. By separating the

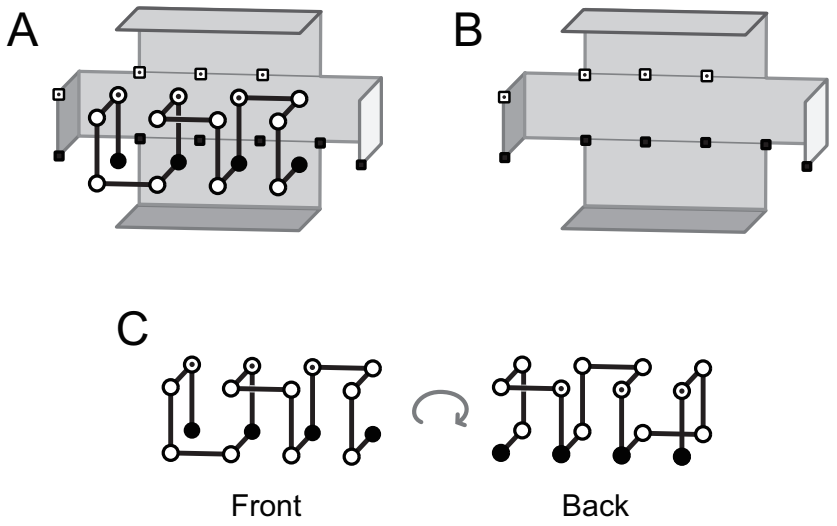

Figure 3.1: The lattice model of an IDP and a specific target. (A) The designed native bound structure is composed of the IDP which forms a helical structure and the binding surface of the target molecule. The IDP is modeled as a chain and the target is modeled as a combination of plates which represent the binding surface. H, H' and P residues of the IDP are denoted by filled circles, circles with a dot, and open circles, respectively. Filled squares on the binding surface denote H interaction sites and squares with a dot denote H' interaction sites. (B) Hydrophobic residues come together on one side of the helical structure of the IDP to form an amphipathic helix. (C) Target conformation in the native bound state.

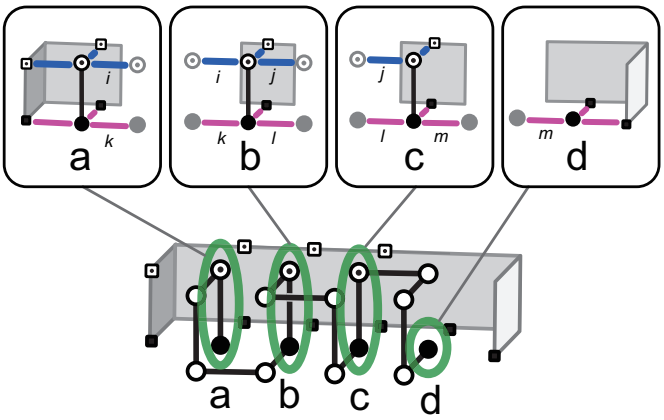

Figure 3.2: Hydrophobic interactions in the native bound state. The hydrophobic residues of the helical structure are divided into four parts (a-d), and interactions between hydrophobic residues are colored pink (*H*-*H* interactions) and blue (*H*′-*H*′ interactions). The internal interactions of the IDP are labeled *i*-*m*.

IDP and the target of this bound state, I get Fig. 3.1B (target) and Fig. 3.1C (IDP). In the native bound state, the IDP forms an amphipathic helix, while it lacks a specific structure in the ground state of the free states.

Each residue of an IDP is at a lattice point, and the whole protein is represented as a self-avoiding chain. I assume that there are three types of interactions between (i) *H*-*H*, (ii) *H*′-*H*′ and (iii) *H*-*H*′ pairs at nearest-neighboring sites, except pairs of consecutive residues along the chain. Contact energies of these interactions are denoted by ε_{HH}, $\varepsilon_{H'H'}$ and $\varepsilon_{HH'}$ ($\varepsilon_{HH}, \varepsilon_{H'H'}, \varepsilon_{HH'} < 0$), respectively. It should be noted that the hydrophobic interactions include solvent entropy, and thus the "contact energy" actually means "contact free energy". I use $(\varepsilon_{HH}, \varepsilon_{H'H'}, \varepsilon_{HH'}) = (-2, -1, -1)$ throughout this thesis. I assume that there are no interactions between *H*-*P*, *H*′-*P* and *P*-*P* pairs. Using the extended HP model, I design a 16-residue sequence of an IDP as $HH'P_3HH'P_3HH'P_3H$. In this sequence, the hydrophobic residues come together on one side of the helical structure to form an amphipathic helix (Fig. 3.1C). In the helical structure, there are three *H*-*H* interactions (*k*, *l* and *m* in Fig. 3.2) and two *H*′-*H*′ interactions (*i* and *j* in Fig. 3.2), and the total energy of the intramolecular interactions of the IDP (E_{ID}) is -8. Many different conformations also have the same energy. The ground state of the IDP has energy of -10 and is forty eight-fold degenerate. Two of the ground state conformations are shown in Fig. 4.1B as examples.

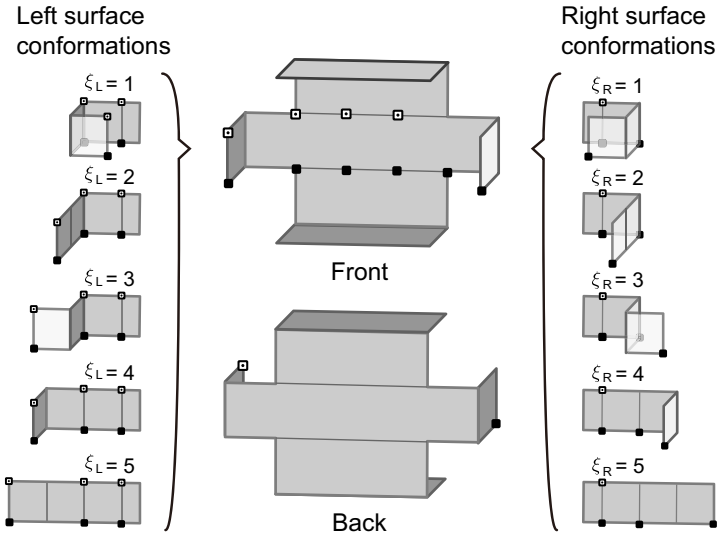

Figure 3.3: The conformations of the left and right parts of the designed target. The left part has five conformations labeled by ξ_{L} ($\xi_{\mathrm{L}} \in \{1, \cdots, 5\}$) and the right part has five conformations labeled by ξ_{R} ($\xi_{\mathrm{R}} \in \{1, \cdots, 5\}$). There are no interaction sites on the back of the binding surface.

I model only the binding surface of the target in order to focus on the recognition processes. Combining square plates and placing H and H' interaction sites on them, I design the binding surface of the native bound state shown in Fig. 3.1B. In this conformation, all the interaction sites are exposed. Interactions between the IDP and the interaction sites of the binding surface are also given by the contact energies ε_{HH}, $\varepsilon_{H'H'}$ and $\varepsilon_{HH'}$. The IDP cannot share the same site with the binding surface due to the excluded volume effect. As shown in Fig. 3.2, there are six *H*-*H* interactions and four *H*′-*H*′ interactions between the IDP and the binding surface in the native bound state, and the total energy of the intermolecular interactions ($E_{\text{ID-T}}$) is -16. Adding E_{ID} to $E_{\text{ID-T}}$ of the native bound state, I get the ground state energy $E = -24$.

In order to express the hidden binding sites of the target, I remodel the designed binding surface. Starting from the exposed binding surface shown in Fig. 3.1B, I introduce motion of both sides of the target to open or close the binding surface by adding five conformations for each side, which are shown in Fig. 3.3. The left and right sides of the conformations are labeled by ξ_{L} ($\xi_{\mathrm{L}} \in \{1, \cdots, 5\}$) and ξ_{R} ($\xi_{\mathrm{R}} \in \{1, \cdots, 5\}$), respectively, and a conformation of the whole binding surface is denoted by $(\xi_{\mathrm{L}}, \xi_{\mathrm{R}})$. An example of the conformation of the binding surface in the native bound state shown

Table 3.1: Definitions of surface types for each side of the binding surface. Conformations of the R surface are limited to $\xi_{\mathrm{L}} = 4$ (or $\xi_{\mathrm{R}} = 4$), while those of the F and H surfaces take $\xi_{\mathrm{L}} = 1 - 5$ (or $\xi_{\mathrm{R}} = 1 - 5$). Only the H surface has the internal interactions to close the binding surface, which come from non-zero λ_{L} (or λ_{R}), while the R and F surfaces do not have internal interactions.

Surface type	ξ_{L}	λ_{L}	Surface type	ξ_{R}	λ_{R}
R (rigid)	4	0	R (rigid)	4	0
F (flexible)	$1-5$	0	F (flexible)	$1-5$	0
H (hidden)	$1-5$	1	H (hidden)	$1-5$	1

in Fig. 3.1B is $(4, 4)$. In the conformations of $\xi_{\mathrm{L}} = 4$ or 5 (or $\xi_{\mathrm{R}} = 4$ and 5), all of the interaction sites are exposed outside, and in the other conformations, some of them are hidden. To energetically stabilize a conformation with hidden binding sites, I introduce intramolecular interactions between the interaction sites of the target by defining the energy of the target as

$$E_{\mathrm{T}} = \lambda_{\mathrm{L}}(\varepsilon_{HH} + \varepsilon_{H'H'})\delta_{\xi_{\mathrm{L}}1} + \lambda_{\mathrm{R}}\varepsilon_{HH}\delta_{\xi_{\mathrm{R}}1}, \tag{3.1}$$

where λ_{L} and λ_{R} are parameters which take 0 or 1 to control the intramolecular interactions. By setting $\lambda_{\mathrm{L}} = 1$ (or $\lambda_{\mathrm{R}} = 1$) , I can introduce the hidden binding sites. δ_{ij} is Kronecker's delta given by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Using the conformations and the intramolecular interactions, I define the three surface types, which are denoted by R (rigid), F (flexible) and H (hidden), for each side of the binding surface. The defined surface types are shown in Table 3.1. The R surface is fixed in the form of $\xi_{\mathrm{L}} = 4$ (or $\xi_{\mathrm{R}} = 4$), and the binding sites are always exposed. The F surface has flexibility to change its conformation ξ_{L} (or ξ_{R}) from 1 to 5. All of these conformations have the same energy. The H surface is also flexible, while the hidden conformation $\xi_{\mathrm{L}} = 1$ (or $\xi_{\mathrm{R}} = 1$) is stabilized by the internal interactions which come from non-zero λ_{L} (or λ_{R}). By selecting the surface types of both sides, I define the whole surface type of the target and denote it by two-letter prefix, such as "HF-target", which means that the left surface is hidden and the right surface is flexible.

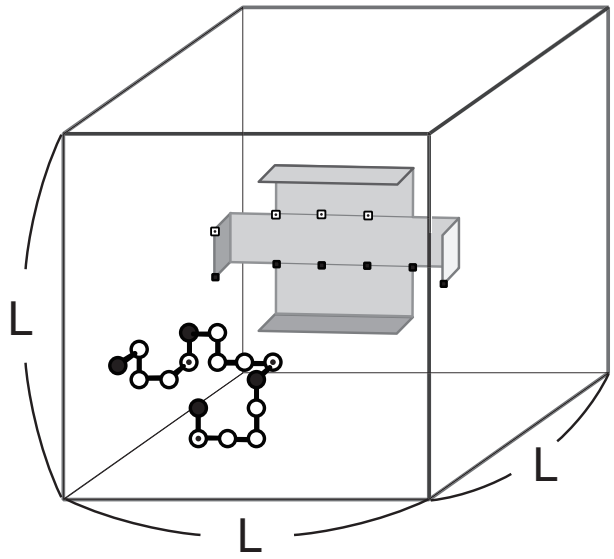

Figure 3.4: The conceptual picture of an $L \times L \times L$ cubic system.

I confine the IDP and the target in an $L \times L \times L$ cubic lattice space surrounded by walls. Fig. 3.4 shows the conceptual picture of the cubic system. The target is fixed at the center of the system, while the IDP can move around inside the system. The total energy of the system is given by

$$E = E_{\mathrm{ID}} + E_{\mathrm{T}} + E_{\mathrm{ID\text{-}T}}, \tag{3.2}$$

and the partition function is written as

$$Z = \sum_{E} \Omega(E) \mathrm{e}^{-E/T}, \tag{3.3}$$

where $\Omega(E)$ is the number of states with energy E. T is temperature (I take the Boltzmann constant $k_B = 1$). I also define the numbers of states $\Omega(O_1, E)$ and $\Omega(O_1, O_2, E)$ for the physical quantities O_1 and O_2. Further, $\Omega(O_1, E)$ and $\Omega(O_1, O_2, E)$ fulfill

$$\sum_{O_1} \Omega(O_1, E) = \sum_{O_1} \sum_{O_2} \Omega(O_1, O_2, E) = \Omega(E). \tag{3.4}$$

Using $\Omega(O_1, E)$ and $\Omega(O_1, O_2, E)$, I define free energies with one and two variables by

$$F(O_1) = -T \log \left\{ \sum_{E} \Omega(O_1, E) \mathrm{e}^{-E/T} \right\} + F_0, \tag{3.5}$$

$$F(O_1, O_2) = -T \log \left\{ \sum_E \Omega(O_1, O_2, E) \mathrm{e}^{-E/T} \right\} + F_0, \tag{3.6}$$

where F_0 is a constant reference value of the free energies. The physical quantities which I use to calculate free energy are listed in Table 3.2.

Table 3.2: The list of physical quantities to calculate free energies.

Total energy of the intermolecular interactions of the IDP	E_{ID}
Total energy of the target	E_{T}
Total energy of the intermolecular interaction between the IDP and the target	$E_{\mathrm{ID\text{-}T}}$
Structural similarity to a helical structure	N_α
Index for the conformation of the left surface of the target	ξ_{L}
Index for the conformation of the right surface of the target	ξ_{R}

I calculate the number of states by using the multi-self-overlap ensemble (MSOE) Monte Carlo method (12, 13), which is an extended version of the multicanonical Monte Carlo method (51, 52). The MSOE is a powerful tool in analyzing thermal properties of lattice polymers (48, 53), and is applied for statistical enumeration of self-avoiding walks (54). I use the Wang-Landau algorithm to produce a weight function for the MSOE (55, 56). The MSOE successfully attained thermal equilibrium between free states and the designed bound state.

3.1 Calculating free-energy landscapes by using the MSOE

In the MSOE simulation, I relax the excluded volume condition, and allow the residues of the IDP to overlap with themselves and with the target. The MSOE makes it possible to explore the configuration space faster than with the conventional multicanonical ensemble method. Let V be the number of overlaps. Introducing a weight $W(E, V)$ as a function of E and V, I define the transition probability from a state of the system s_a to another state s_b as

$$p(s_a \to s_b) = \min \left[\frac{W(E_b, V_b)}{W(E_a, V_a)}, 1 \right], \tag{3.7}$$

where E_a and E_b are energy of s_a and s_b, respectively. In the case of the IDP alone, I randomly select a residue and change the local conformation of the residue according to this transition probability. In the case of the system with both the IDP and the

target, I first select the IDP or the target with a fixed probability, and then change the state of the selected object according to the transition probability. I iteratively improve $W(E,V)$ to be approximately proportional to the inverse of the number of states $1/\Omega(E,V)$. Using the determined $W(E,V)$, the MSOE simulation produces a flat histogram $H(E,V)$. The $V=0$ subspace corresponds to the conventional multicanonical ensemble and the number of states $\Omega(E)$ is given by $H(E,0)/W(E,0)$ up to a normalization factor. In this thesis, I made the three- and four-dimensional histograms listed below to create free-energy landscapes.

- $H(E_{\mathrm{ID}},E,V)$ (for Fig. 4.8A, Fig. 4.10 and Fig. 4.11)
- $H(N_\alpha,E_{\mathrm{ID}},E,V)$ (for Fig. 4.1A and Fig. 4.6)
- $H(E_{\text{ID-T}},N_\alpha,E,V)$, $H(E_{\text{ID-T}},E_{\mathrm{ID}},E,V)$ (for Fig. 4.8A)
- $H(E_{\text{ID-T}},(\xi_{\mathrm{L}},\xi_{\mathrm{R}}),E,V)$, $H(E_{\text{ID-T}},(\xi_{\mathrm{L}},\xi_{\mathrm{R}}),E,V)$ (see $(\xi_{\mathrm{L}},\xi_{\mathrm{R}})$ as one variable; for Fig. 4.10 and Fig. 4.11)

From these histograms, I obtain free-energy landscapes using Eq. 3.5 and Eq. 3.6.

4

Result and Discussion

4.1 Intrinsic disorder represented as a mixture of several states

Free-energy landscapes $F(N_\alpha, E_{\mathrm{ID}})$ of the IDP at three temperatures are shown in Fig. 4.1A. N_α shown on the x-axis indicates structural similarity to a helical structure, and the definition of N_α is shown in Fig. 4.2. At the high temperature, the IDP is unfolded and extended, and it is denatured. As temperature decreases ($T = 0.80 \rightarrow 0.57 \rightarrow 0.35$), the distribution of the low free-energy states (shown as blue areas in Fig. 4.1A) moves from high E_{ID} states to low E_{ID} states. The states at point (p) in Fig. 4.1A, which includes the helical structure shown in Fig. 3.1B and its mirror image, are always less favorable. In this temperature range, the IDP is in a mixture of many states and it does not stay in a specific structure. Below these temperatures, the blue area of the landscape converges to the ground state of the IDP at the point (q) of Fig. 4.1A, which is a glassy state of many collapsed structures. Two of the ground state conformations of the IDP are shown in Fig. 4.1B as examples. I confirmed that the constructed model actually has the properties shown in the second paragraph of Chap. 3 and I successfully constructed a model of an IDP.

4.2 Size dependence of the specific heat of the IDP

The specific heat of the IDP (C_{ID}) for $L = 16, 32, 64, 128$ and ∞ is shown in Fig. 4.3. I denote the peak temperature of C_{ID} for $L = \infty$ by T_{ID}. As L increases, the curves

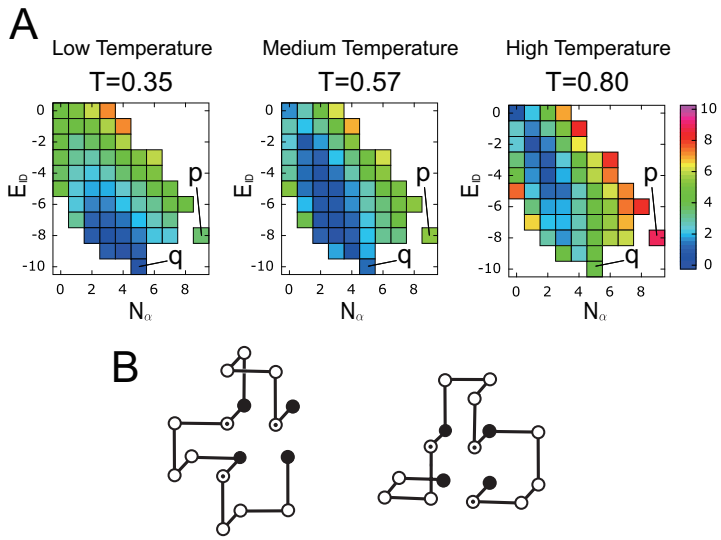

Figure 4.1: (A) Free-energy landscapes $F(N_\alpha, E_{\rm ID})$ of the IDP for $L = \infty$ at low, medium and high temperatures shown from left to right, respectively. The medium temperature is the peak temperature of the specific heat of the IDP for $L = \infty$ ($T_{\rm ID}$) shown in Fig. 4.3. Colors indicate the values of free energy. (B) Two example conformations of the ground state (q).

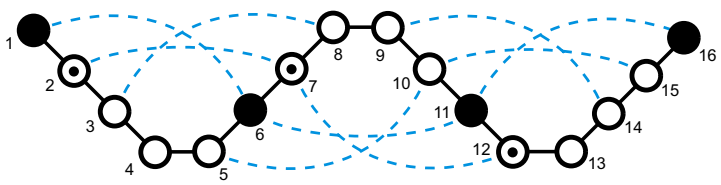

Figure 4.2: The α-helix contact pairs are defined by the residue pairs connected by blue dotted lines. The definition of the α-helix contact N_α is the number of the α-helix contact pairs which are at nearest-neighbor sites. In the helical structure shown in Fig. 3.1C and its mirror image, all these contact pairs are at nearest-neighbor sites and N_α takes the maximum value of 9.

converge to the curve of $L = \infty$. The shapes and peak temperatures of these curves do not strongly depend on the system size.

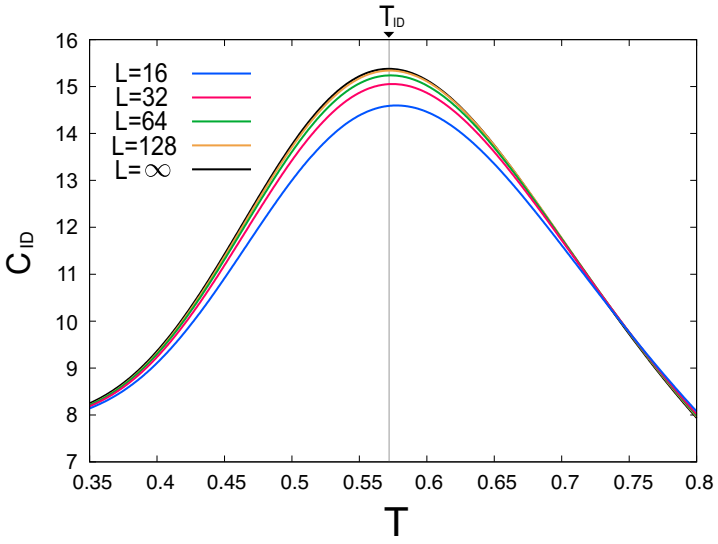

Figure 4.3: Specific heat of the IDP (C_{ID}) for $L = 16, 32, 64, 128$ and ∞. The peak temperature for $L = \infty$ (T_{ID}) is indicated by the vertical line.

4.3 Thermal stability of the native bound state

Fig. 4.4 shows the specific heat for $L = 32$ with the IDP and three types of the targets: RR-target, FF-target and HH-target. As temperature decreases, the change of $\langle E_{\text{ID-T}} \rangle$ occurs around the peak temperature of each system. At this temperature, the IDP binds to or dissociates from the target, and I call it binding temperature T_{B}. T_{B} reflects the thermal stability of the native bound state. Comparing T_{B} of the IDP/RR-target and IDP/FF-target systems, I find that the motion of the binding surface destabilizes the native bound state, and comparing T_{B} of the IDP/FF-target and IDP/HH-target systems, I find that the native bound state is destabilized by the hidden binding sites. These results suggest that a change in the surface types can induce binding of the IDP. For example, if post-translational modification or ligand binding causes a loss of degrees of freedom of the target, T_{B} increases. If T_{B} becomes larger than the cell temperature, the change of the target can induce binding of the IDP by enhancing the thermal stability of the bound state.

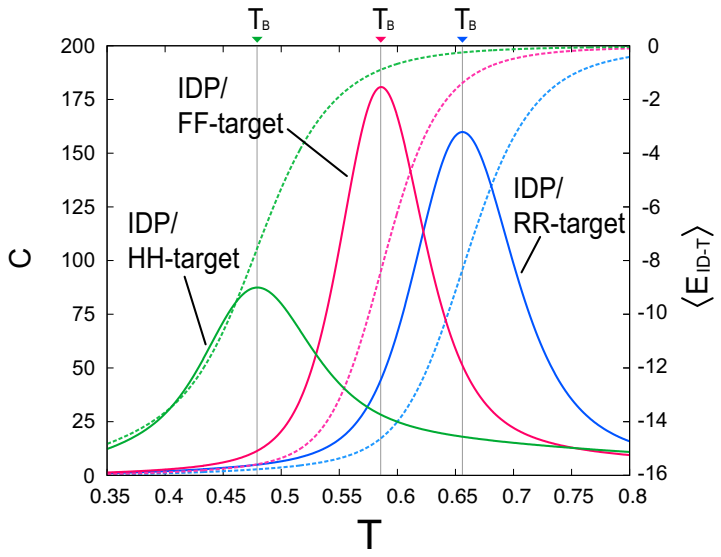

Figure 4.4: Specific heat C for $L = 32$ with an IDP and the three types of binding surfaces: RR-target (blue line), FF-target (pink line) and HH-target (green line). The binding temperature T_{B} of each system is indicated by the vertical line. I also show the thermal average of the contact energy between the IDP and the target ($\langle E_{\mathrm{ID\text{-}T}} \rangle$) of the same systems (broken lines in the same colors as the specific heats).

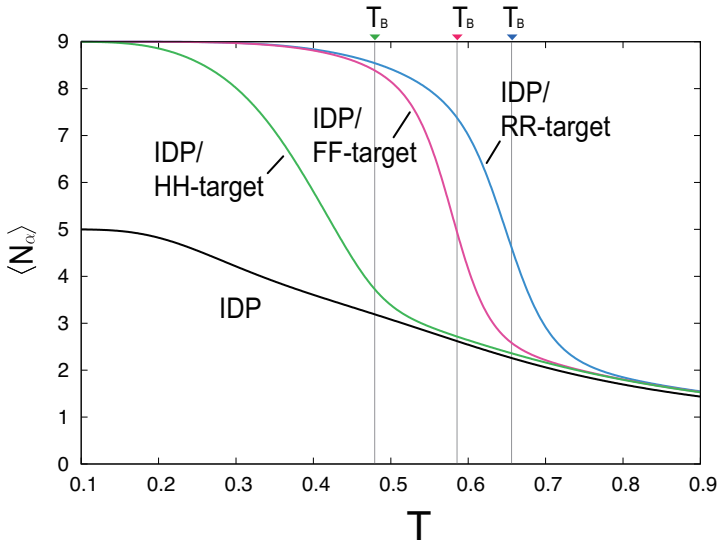

Figure 4.5: Thermal averages of the α-helix contact ($\langle N_\alpha \rangle$) for the four systems ($L = 32$): IDP only (black), IDP/RR-target (blue), IDP/FF-target (pink) and IDP/HH-target (green). T_{B} for the three IDP-target systems are indicated by the vertical lines.

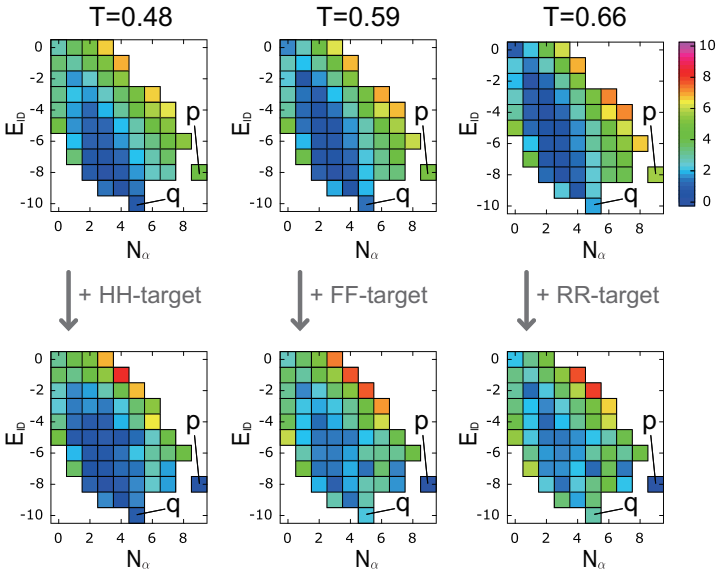

Figure 4.6: Free-energy landscapes $F(N_\alpha, E_{\mathrm{ID}})$ of the IDP for $L = 32$ with and without the three types of binding surfaces: RR-target (top and bottom left), FF-target (top and bottom center) and HH-target (top and bottom right) at the corresponding binding temperatures.

In order to follow structural changes of the IDP at the binding temperatures, I show the thermal averages of N_α in Fig. 4.5. As temperature decreases, $\langle N_\alpha \rangle$ of the IDP gradually converges to $N_\alpha = 5$, which is the value of the ground state conformations. By adding each of the three types of the targets to the system, the curve $\langle N_\alpha \rangle$ reaches the maximum value of 9 at the low temperature, and the IDP folds into the helical structure below T_{B} of each system. In Fig. 4.6, I show how the free-energy landscapes $F(N_\alpha, E_{\mathrm{ID}})$ of the IDP change according to the presence of the three types of binding surfaces at the corresponding binding temperatures. In all cases, the helical structure of the IDP is stabilized by the presence of the targets. Thus, I consider binding to the target induces folding processes and this is nothing but the coupled folding and binding. This is consistent with the discussion by Matsushita and Kikuchi (57) in which the presence of a target reduces the inconsistency of conformations of an IDP. In the systems of the IDP/RR-target and IDP/FF-target, the change of the curves occurs just around T_{B}, and this suggests the coupled folding and binding induced by temperature change. The change of the curve of the IDP/HH-system occurs at a lower temperature than T_{B}, and this might be caused by the multistep binding which I will

explain in Sec. 4.4.

The thermal stability of the complexes depends on the system size. System size dependence of T_{B} is shown in Fig. 4.7. As L decreases, the thermal stability of the complexes increases because of the decrease in the translational entropy of the free states. This result suggests that confinement in a small volume can also induce binding, which is basically the same as the confinement-induced dimerization discussed by Nakanishi and Kikuchi (53). Increasing the density of the two molecules can also reduce translational entropy of the free states, and stabilize bound states. It is known that IDPs of eukaryotes localize in the nucleus (26), and I consider that this localization stabilizes bound complexes of IDPs with nuclear proteins or DNAs by increasing their density. Macromolecular crowding will also stabilize the bound states by reducing free volume.

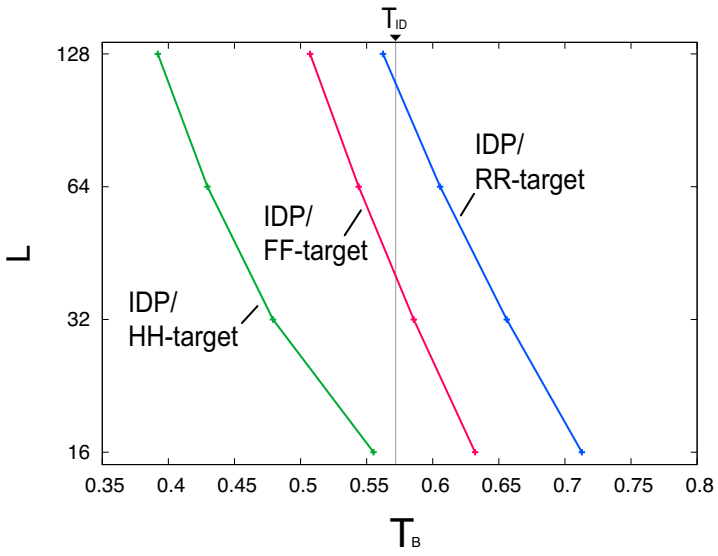

Figure 4.7: System size dependence ($L = 16 - 128$) of the binding temperature T_{B}. T_{ID}, which is indicated by the vertical line, is the peak temperature of the specific heat of the IDP for $L = \infty$ shown in Fig. 4.3.

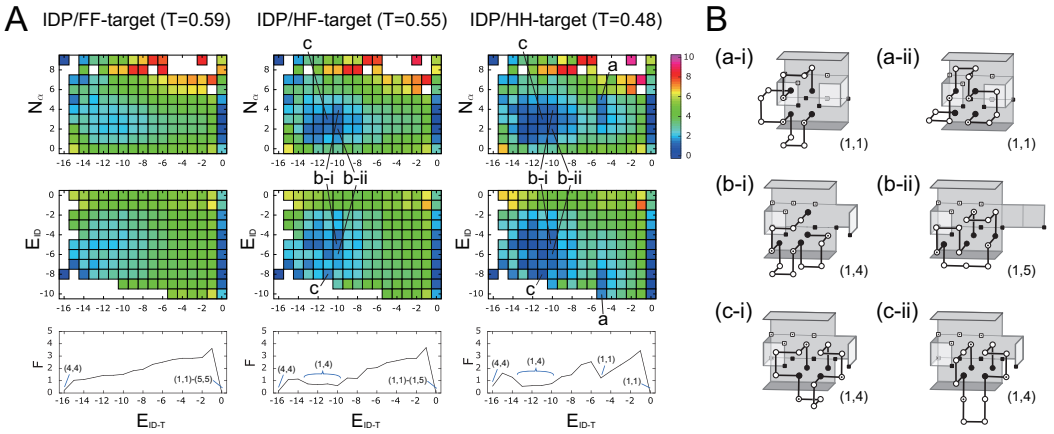

Figure 4.8: (A) Free-energy landscapes $F(E_{\text{ID-T}}, N_\alpha)$, $F(E_{\text{ID-T}}, E_{\text{ID}})$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/FF-target, IDP/HF-target and IDP/HH-target at the binding temperature of each system. In the graphs of $F(E_{\text{ID-T}})$, I show the most stable conformations $(\xi_{\text{L}}, \xi_{\text{R}})$ of the binding surface which is given in Figs. 4.10 and 4.11. Colors indicate the values of free energy. (B) Example structures of the IDP/target complexes at the points denoted by (a-c) shown in A. The target conformations $(\xi_{\text{L}}, \xi_{\text{R}})$ is shown in the bottom right of each figure.

4.4 Stepwise target recognition of the IDP due to multiform binding effect

Fig. 4.8A shows the free-energy landscapes of IDP/FF-target, IDP/HF-target and IDP/HH-target systems at the T_{B} of each system. All of the landscapes of $F(E_{\text{ID-T}})$ (bottom row) have the local minima at $E_{\text{ID-T}} = 0$ and -16, corresponding to the free states and the native bound state, respectively. These landscapes show that the closure of the left surface (FF→HF) stabilizes the non-native intermediate state around $E_{\text{ID-T}} = -12$, and further introduction of a hidden binding site (HF→HH) stabilizes another intermediate state around $E_{\text{ID-T}} = -5$. In these intermediate states, the IDP forms various structures to stabilize the encounter complexes, some of which are shown in Fig. 4.8B. Various non-native bound complexes have also been found in the detailed simulation of the N-terminal repressor domain of neural restrictive silencer factor (NRSF), which is known to be an IDP, and the paired amphipathic helix domain of mSin3 (target), using classical molecular dynamics in atomic resolution (58). I considered that these intermediate complexes are related to the advantages of IDPs.

In the first step to bind to the HH-target, the IDP interacts with the exposed

binding sites on the outside of the target by forming low-energy compact structures, two of which are shown in Fig. 4.8B (a) as examples ($-10 \leq E_{\mathrm{ID}} \leq -8$). These intermediate structures provide the first scaffold to open one side of the hidden binding sites. After opening the one side of the binding surfaces, the IDP interacts with the surface to stabilize the second scaffold to access the other hidden binding sites. Some of the complex structures of the second intermediate state are shown in Figs. 4.8B (b) and (c) as examples. The IDP is in a mixture of various states ($-8 \leq E_{\mathrm{ID}} \leq -3$) which can increase contacts with the unclosed interaction sites. Lastly, the IDP opens the hidden binding site and forms the native bound structure. Thus, binding to the HH-target has three steps via two intermediate states. In the case of the HF-target, the binding process starts with the second intermediate state shown above, and it involves two steps. These results suggest that the closure of the binding sites stabilizes the intermediate states and this mechanism simply explains the existence of the intermediate states of p27/cyclin A/Cdk2 and pKID/KIX systems.

These results give us a new insight into an advantage of IDPs shown in Fig. 4.9. Various forms of an IDP in transient encounter complexes can interact with exposed binding sites on the outside of the target and can provide a scaffold to search for hidden binding sites (Figs. 4.9A and 4.9B). I call this multiform binding. This is similar to the proposed advantage of structural adaptability of an IDP for several different targets. But in this case, I discuss the effect of multiple forms on binding to the same target.

4.5 Coupled folding and binding of the flexible target

The blue lines labeled "total" on the bottom-left and bottom-right of Fig. 4.10 show the free-energy landscapes $F(E_{\mathrm{ID\text{-}T}})$ of the IDP/HH-target and IDP/FF-target systems, respectively. I obtain the other lines in the same figure by decomposing these graphs into the free-energy landscapes $F(E_{\mathrm{ID\text{-}T}}, (\xi_{\mathrm{L}}, \xi_{\mathrm{R}}))$ which greatly contribute to reducing the value of $F(E_{\mathrm{ID\text{-}T}})$. Using these landscapes, I can deduce the dynamics of the binding and folding processes. Both sides of the HH-target are in the closed conformation $(1, 1)$ in the free states, and the IDP makes contact with this conformation. After the target opens the right side of the hidden binding sites, conformations are sequentially selected as $\{(1, 2), (1, 4), (1, 5)\} \rightarrow \{(1, 4), (1, 5)\} \rightarrow \{(1, 4)\}$ before switching to the open conformation $(4, 4)$. In this process, the conformational stabilization of the target

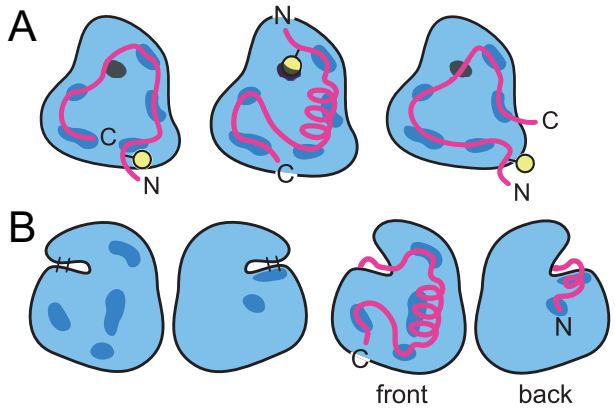

Figure 4.9: The conceptual pictures of multiform binding of two different hidden binding sites. (A) Several binding forms due to flexibility of an IDP (pink curves) stabilize an intermediate state with the target molecule (light blue objects with interaction sites shown as blue spots) and provide a scaffold to search for buried interaction sites (black spots) which interact with the hydrophobic residues of an IDP (yellow circle). (B) In other cases, several binding forms provide a scaffold to cut into intramolecular interactions of the target. The letters N and C shown in the figure indicate the N-terminus and C-terminus of IDPs, respectively.

is coupled with the binding process. This result suggests that the IDP and the target cooperatively reduce entropy of the complex and form a binding funnel. In the case of the IDP/FF-target system, all conformations of the target contribute to reducing the free-energy barrier of IDP binding, and after interacting with the IDP, the target reduces its conformational entropy and reaches $(4, 4)$. There are no free-energy barriers to the open conformation $(4, 4)$ in the final stage of the binding process. The IDP and the target also form a binding funnel in this process.

4.6 Other examples of the free-energy landscapes $F(E_{\mathbf{ID\text{-}T}}, (\xi_{\mathbf{L}}, \xi_{\mathbf{R}}))$

The blue lines labeled "total" on the bottom-left and bottom-right of Fig. 4.11 show the free-energy landscapes $F(E_{\text{ID-T}})$ of the IDP/HR-target and IDP/HF-target systems, respectively. I obtain the other lines in the same figure by decomposing these graphs into the free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ which greatly contribute to reducing the value of $F(E_{\text{ID-T}})$. In the case of the IDP/HR-target system, I observe only one

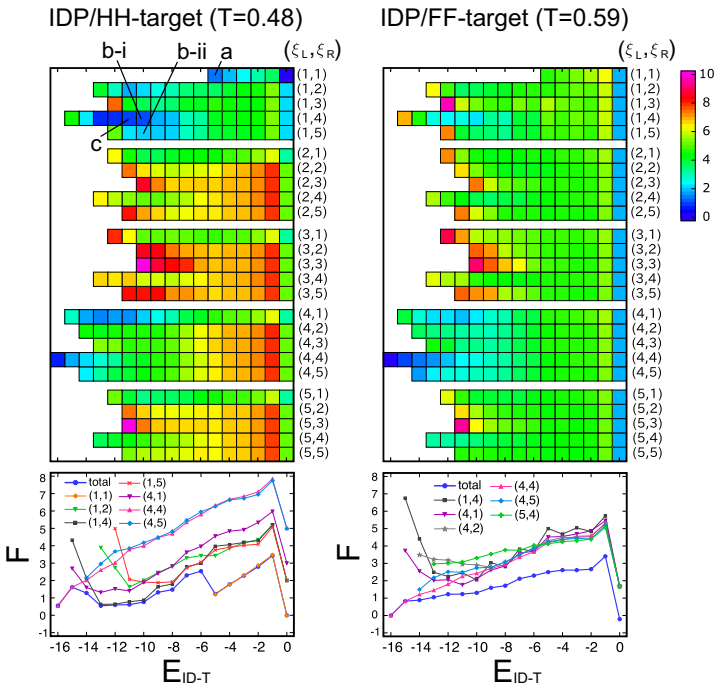

Figure 4.10: Free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/HH-target and IDP/FF-target at the binding temperature of each system. In the graph of $F(E_{\text{ID-T}})$ labeled "total", I superposed lines of $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ for selected $(\xi_{\text{L}}, \xi_{\text{R}})$ which greatly contribute to reducing the value of $F(E_{\text{ID-T}})$. Colors indicate the values of free energy. The letters (a), (b) and (c) correspond to the example structures of Fig. 4.8B.

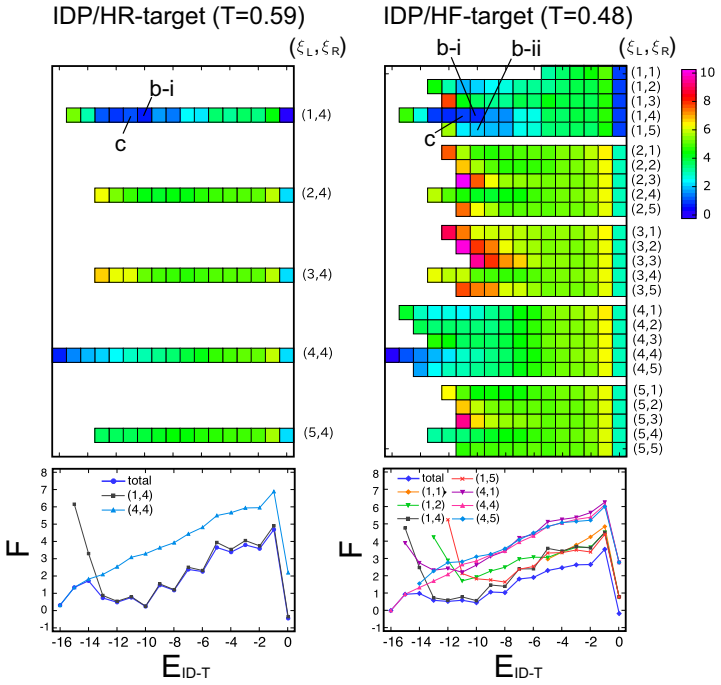

Figure 4.11: Free-energy landscapes $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ and $F(E_{\text{ID-T}})$ for $L = 32$ of IDP/HR-target and IDP/HF-target at the binding temperature of each system. In the graph of $F(E_{\text{ID-T}})$ labeled "total", I superposed lines of $F(E_{\text{ID-T}}, (\xi_{\text{L}}, \xi_{\text{R}}))$ for selected $(\xi_{\text{L}}, \xi_{\text{R}})$ which greatly contribute to reducing the value of $F(E_{\text{ID-T}})$. Colors indicate the values of free energy. The letters (b) and (c) correspond to the example structures of Fig. 4.8B.

route to the native bound state by switching from the closed conformation $(1, 4)$ to the open conformation $(4, 4)$. In this process, the folding of the IDP is coupled with the binding of the IDP to the target. By introducing flexibility into the right side of the target (HR→HF), I observe a conformational selection process of the target before switching to the open conformation $(4, 4)$. As $E_{\text{ID-T}}$ decreases, the target starts from the closed conformations $\{(1, 1) - (1, 5)\}$ in the free states and they are sequentially selected as $\{(1, 1), (1, 2), (1, 4), (1, 5)\} \to \{(1, 4), (1, 5)\} \to \{(1, 4)\}$. In this process, the folding of the IDP is also coupled with the binding process, and a flexible binding surface gradually leads to the native bound state by reducing fluctuation. This result suggests that the IDP and the target cooperatively reduce entropy of the complex and form a binding funnel.

Part III

Binding and Fibrillation Processes of α-Synuclein

5

Overview

α-synuclein (α-syn) is the intrinsically disordered protein which is considered to be one of the causes of Parkinson's disease. This protein forms amyloid fibrils when in a highly concentrated solution. The fibril formation of α-syn is induced not only by increases in α-syn concentration but also by macromolecular crowding. I focus on the relation between the intrinsic disorder of α-syn and macromolecular crowding, and construct a simplified model of α-syn including crowding agents based on statistical mechanics. The main assumption is that α-syn can be expressed as coarse-grained particles with internal states coupled with effective volume; and disordered states are modeled by larger particles with larger internal entropy than other states. It is found that the crowding effect is taken into account as the effective internal entropy; and the crowding effect reduces the effective internal entropy of disordered states. From a result of Monte Carlo simulation, I provide a scenario for the crowding-induced fibril formation. I also discuss the recent controversy over the existence of helically folded tetramers of α-syn, and suggest that macromolecular crowding is the key to resolving the controversy.

6

Introduction to α-Synuclein

α-synuclein (α-syn) is a 140-amino-acid protein (59) expressed in presynaptic terminals of the mammalian central nervous system (60, 61, 62, 63). α-syn is known as an intrinsically disordered protein (IDP), which lacks secondary or tertiary structures (64, 65, 66). Growing evidence suggests that α-syn has causative roles in Parkinson's disease (PD), which is the second most common neurodegenerative disease after Alzheimer's disease. Genetic mutations in α-syn (A30P, E46K and A53T) which have been identified in familial PD (FPD) directly link α-syn with PD (67, 68, 69, 70). Duplication and triplication of the α-syn gene are also found in FPD (71, 72, 73). The intracellular inclusions called Lewy bodies are the pathological hallmark of PD, and they are mainly composed of α-syn (74, 75). Lewy bodies also link α-syn with PD. Furthermore, *in vitro* studies revealed that β-sheet-rich amyloid fibrils are formed in high-concentration α-syn solutions after hours or days of incubation (76, 77), and that introduction of these fibrils into cultured cells with overexpression of α-syn leads to the formation of Lewy body-like inclusions (78). Inhibition of α-syn fibril formation is one of the suggested therapeutic approaches for PD (79, 80, 81, 82), and thus a good understanding of such fibril formation is demanded.

Fibril formation of α-syn is a nucleation-dependent process (83). Nucleation of α-syn fibrils occur above a certain α-syn concentration, which is called critical concentration. These nucleated α-syn fibrils, which are called "protofibrils", become seeds for subsequent elongation, and they grow to form amyloid fibrils (76, 77, 83). Kinetics of the fibril formation are affected by macromolecular crowding, and crowding agents induce and accelerate the fibril formation of α-syn (84, 85). White et al. observed that

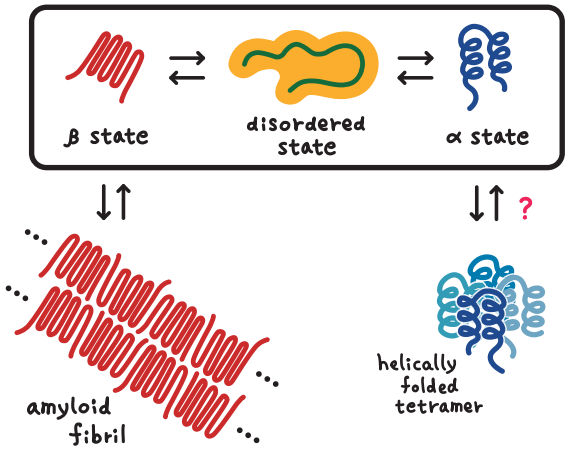

Figure 6.1: Equilibrium of α-synuclein. Monomeric α-synuclein (α-syn) is in equilibrium between three states: a β-sheet-rich state (β state), a disordered state and an α-helix-rich state (α state). α-syn aggregates into a β-sheet-rich amyloid fibril above a critical concentration. The existence of helically folded tetramer is still controversial.

macromolecular crowding accelerates the fibril formation of IDPs including α-syn, and it decelerates the fibril formation of folded proteins. From these results, they concluded that accelerated fibril formation is characteristic of IDPs (3). Macromolecular crowding is an important factor in considering the fibril formation of α-syn.

In 2011, Bartels et al. reported the existence of helically folded tetramers of α-syn (80). Although helically folded monomers of α-syn had already been identified as a lipid-bound form at that time (65, 86, 87, 88, 89, 90, 91), this was the first report of a tetrameric state of helically folded monomers. These tetramers resist amyloid fibril formation, and they suggested that stabilization of helically folded tetramers is one of the possible strategies for designing an amyloid inhibitor. There are three other reports which support the result of Bartels et al. (82, 92, 93). The existence of helically folded tetramers is, however, still controversial; Fauvet et al. (81) and Burré et al. (94) concluded that α-syn exists predominantly as a disordered monomer. Fauvet et al. suggested that stabilization of a disordered monomer of α-syn can be an alternative strategy to inhibit fibril formation. I summarize the equilibrium state of α-syn including the tetrameric states in Fig. 6.1.

In this thesis, I construct a highly simplified model of α-syn with crowding agents. Disordered states and fibril-forming β-sheet-rich states (β states) are assumed as the internal states of α-syn. Using this model, I investigate the coupled effect of crowding

and intrinsic disorder on fibril formation. I also investigate the thermal stability of the helically folded tetramers by adding tetramer-forming α-helix-rich states (α states) to the internal states of α-syn.

7

Models

I construct a model of α-syn using a two-dimensional lattice gas model. α-syns and the crowding agents are modeled as particles on a lattice. The definitions of these molecules are shown in Fig. 7.1 (a). α-syn has three types of internal states: two β states (β_H and β_V), two disordered states (d_H and d_V) and four α states (α_{NE}, α_{SE}, α_{SW} and α_{NW}). The variations of each type of the states correspond to orientations of α-syn.

I assume that the disordered states are larger than the other states based on the fact that disordered α-syn has a larger gyration radius than folded proteins with the same number of residues (64, 95, 96). The disordered states occupy two sites and all the other states and the crowding agents occupy a single site (Fig. 7.1 (a)). Thus, the unit volume ν is defined as the volume of a single site, the volume of disordered states v_d, the volume of β and α states v_f, and the volume of the crowding agents v_c are written as 2ν, ν and ν, respectively. I take $\nu = 1$ throughout this thesis. I also assume that the disordered states have internal entropy s based on the fact that α-syns exhibit significant entropy loss when they bind to a lipid bilayer surface and fold into an α-helical structure (97, 98). I simply introduce the difference of entropy between the disordered states and the folded states by s.

I define two types of interactions between two β_H and two β_V based on the model of amyloid fibril formation given by Zhang and Muthukumar (99). The definitions of these interactions are shown in Fig. 7.1 (b). Two horizontally-aligned β_H and two vertically-aligned β_V have β interactions. A linear chain of α-syns aligned by the β interactions defines the longitudinal direction of a fiber. Two β_H and two β_V have β' interactions

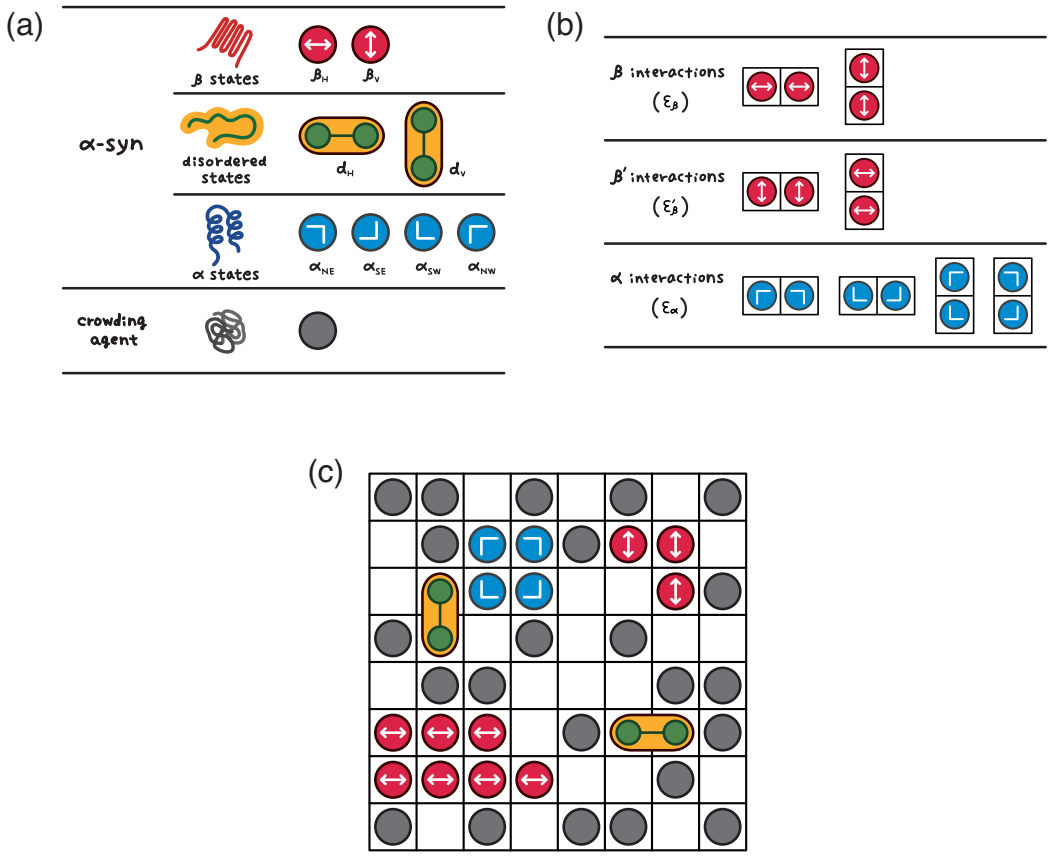

Figure 7.1: Lattice gas model of α-synuclein. (a) Definitions of two molecules. α-synuclein (α-syn) has three types of internal states: two β states (β_H and β_V), two disordered states (d_H and d_V) and four α states (α_{NE}, α_{SE}, α_{SW} and α_{NW}). The crowding agent has only one state. d_H and d_V occupy two sites while the other internal states of α-syn and the crowding agent occupy a single site. (b) There are three types of interactions between two α-syns at nearest-neighboring sites: two β interactions, two β' interactions and four α interactions. The interaction constants of these three interactions are given by ε_β, ε'_β and ε_α, respectively. (c) Snapshot of a small system ($L = 8$, $N = 16$) of the lattice gas model.

in the direction perpendicular to the longitudinal one. The interaction constants of the β and β' interactions are given by ε_β and ε'_β ($\varepsilon_\beta < 0, \varepsilon'_\beta \leq 0, |\varepsilon'_\beta| \leq |\varepsilon_\beta|$), respectively.

I define four α interactions between four pairs of α states, as shown in Fig. 7.1 (b). Dimers, trimers and tetramers of α states are formed by the α interactions. There are no conformations which include oligomers of α states larger than tetramers connected by the α interactions. The interaction constant of the α interactions is ε_α ($\varepsilon_\alpha < 0$). The crowding agents do not have intermolecular interaction except excluded volume interaction.

I define two models by including or excluding the internal states of α-syn: BD and BDA. The letters B, D and A correspond to β states, disordered states and α states, respectively. The BD model, for example, includes β states and disordered states. I also define B^oD and B^oDA, which correspond to the BD and BDA models without β' interaction, as controls.

I consider a canonical ensemble of the four models presented above combined with a grand canonical ensemble of crowding agents. I define a system by an $L \times L$ square lattice with N α-syns surrounded by walls in contact with a crowding agent reservoir at temperature T and chemical potential μ. μ controls the number of the crowding agents, and the $\mu = -\infty$ system corresponds to the system without the crowding agents. I define a state of the system by a configuration of N α-syns with a determined internal state and m crowding agents on the lattice. I show a snapshot of a small system ($L = 8$, $N = 16$) of the BDA model in Fig. 7.1 (c), which includes all the internal states and all of the β and α interactions, as an example. A short fiber and a helically folded tetramer are also shown in the snapshot. The numbers of the internal states of α-syn n_β, n_d and n_α are counted as $n_\beta = 10$, $n_d = 2$ and $n_\alpha = 4$ in the snapshot. The number of the crowding agents is $m = 23$. The total energy E is calculated as $E = 6\varepsilon_\beta + 4\varepsilon'_\beta + 4\varepsilon_\alpha$.

The grand partition function of the N α-syn system is given by

$$\Xi = \sum_{E,n_d,m} \Omega(E, n_d, m)\, \mathrm{e}^{-\frac{E}{T}} \mathrm{e}^{sn_d} \mathrm{e}^{\frac{\mu}{T}m}, \tag{7.1}$$

where $\Omega(O_1, O_2, O_3, \cdots)$ is the number of states of the system with physical quantities O_i ($i = 1, 2, 3, \cdots$). I took the Boltzmann constant $k_B = 1$.

8

Methods

8.1 Metropolis method for the lattice gas model of α-synuclein including crowding agents

I use the Metropolis method to produce an equilibrium. In the lattice model of α-syn, a state of the system is given by conformations of N α-syns and m crowding agents on the $L \times L$ lattice. The transition probability from state i to state j is given by

$$p(i \to j) = \min \left[\frac{\exp\left(-E^j/T + \sigma n_d^j\right)}{\exp\left(-E^i/T + \sigma n_d^i\right)}, 1 \right]. \tag{8.1}$$

I define the following four Monte Carlo moves for α-syns, which are performed with the transition probability in Eq. 8.1.

(i) Local displacement of a single α-syn: (a) If a selected α-syn is in β or α state, move it in a randomly selected direction (north, south, east or west) by one site. This move will be rejected if the destination site is covered with another α-syn. (b) If a selected α-syn is in a disordered state, move it to a randomly selected direction by one site. This move will be rejected if one or both of the two destination sites are covered with other α-syns.

(ii) Local displacement of a group of α-syns: (a) In the case that the selected α-syn is a part of a β-state linear chain connected by β interactions, move the whole linear chain to a randomly selected direction by one site if the destination sites

are empty. This move will be rejected if the number of β interactions increases. (b) In the case that the selected α-syn is a part of an α-state oligomer (dimer, trimer or tetramer) connected by α interactions, move the oligomer to a randomly selected direction by one site if the destination sites are empty. This move will be rejected if the number of α interactions increases. (c) In the case that a selected α-syn is not connected with other α-syns by β or α interaction, perform the move shown in (i), above. This move will be rejected if the number of β or α interactions increases.

(iii) Nonlocal displacement of a single α-syn: (a) If a selected α-syn is in a β or α state, move it to a randomly selected site of the system. This move will be rejected if the destination site is covered with other α-syns. (b) If a selected α-syn is in a disordered state, move it to two randomly selected adjacent sites of the system. This move will be rejected if one or both of the two destination sites are covered with other α-syns.

(iv) Internal state switching: Switch the internal state of a selected α-syn to another randomly selected state. This move will be rejected if a disordered state is selected as the next state and if it overlaps with another α-syn.

I randomly select an α-syn from the system, and randomly perform one of (i-iv) with prescribed probabilities. In all cases of (i), (ii) and (iii), the move will be rejected if the selected α-syns go outside the system.

8.2 Method of adding crowding agents to the snapshots produced by the simulation without explicit crowding agents

I explain the method of adding crowding agents based on μ to the snapshots produced by the method shown above. Here, I also use the Metropolis method to produce the grand canonical ensemble of the crowding agents. The transition probability from state

i to state j is given by

$$p(i \to j) = \min \left[\frac{\exp(\mu\, m^j/T)}{\exp(\mu\, m^i/T)}, 1 \right]. \tag{8.2}$$

I randomly select a site from the system and perform the move shown below.

- Insertion or removal of a crowding agent: (a) In the case that a selected site is empty, insert a new crowding agent into that site. (b) In the case that a selected site is covered by a crowding agent, remove it from the system. (c) In the case that a selected site is covered by an α-syn, this move will be rejected.

When I created the snapshots shown in Fig. 9.1 and 9.2, I first created the snapshots by the simulation without explicit crowding agents, and then added the crowding agents to the snapshots by using the method presented here.

9

Results

9.1 The effect of macromolecular crowding is taken into account as the effective internal entropy of disordered states

In the N α-syn system, there n_d disordered states occupying v_d $(= 2)$ sites, and there are $N - n_d$ other states occupying v_f $(= 1)$ site. Thus, a number of sites without α-syns is calculated as

$$L^2 - v_d\, n_d - v_f(N - n_d) = L^2 - v_f N - \Delta v\, n_d, \tag{9.1}$$

where $\Delta v = v_d - v_f$. Then, I can count configurations of m single-site crowding agents by the combination $\binom{L^2 - v_f N - \Delta v\, n_d}{m}$, where $0 \leq m \leq L^2 - v_f N - \Delta v\, n_d$. Using the identity

$$\sum_{m=0}^{L^2 - v_f N - \Delta v\, n_d} \binom{L^2 - v_f N - \Delta v\, n_d}{m} \mathrm{e}^{\frac{\mu}{T} m} = \left(1 + \mathrm{e}^{\frac{\mu}{T}}\right)^{L^2 - v_f N - \Delta v\, n_d},$$

I can rewrite the grand partition function in Eq. 7.1 as

$$\begin{aligned} \Xi &= \left(1 + \mathrm{e}^{\frac{\mu}{T}}\right)^{L^2 - v_f N} \sum_{E, n_d} w(E, n_d)\, \mathrm{e}^{-\frac{E}{T}} \mathrm{e}^{\sigma n_d} \\ &= \left(1 + \mathrm{e}^{\frac{\mu}{T}}\right)^{L^2 - v_f N} Z, \end{aligned} \tag{9.2}$$

where $w(O_1, O_2, O_3, \cdots)$ is the number of states of the system with $m = 0$ and physical quantities O_i $(i = 1, 2, 3, \cdots)$, and I defined the effective entropy σ and the canonical

partition function Z as

$$\sigma = s - \Delta v \log\left(1 + \mathrm{e}^{\frac{\mu}{T}}\right), \tag{9.3}$$

$$Z = \sum_{E,n_d} w(E,n_d)\,\mathrm{e}^{-\frac{E}{T}}\mathrm{e}^{\sigma n_d}. \tag{9.4}$$

A thermal average of a physical quantity O is calculated as

$$\langle O\rangle = \frac{1}{\Xi}\sum_{O,E,n_d,m} O\,\Omega(O,E,n_d,m)\,\mathrm{e}^{-\frac{E}{T}}\mathrm{e}^{sn_d}\mathrm{e}^{\frac{\mu}{T}m}. \tag{9.5}$$

In the case of $O = m$, Eq. 9.5 is written as

$$\begin{aligned}\langle m\rangle &= T\frac{\partial}{\partial\mu}\log\Xi \\ &= T\frac{\partial}{\partial\mu}\log\left(1+\mathrm{e}^{\frac{\mu}{T}}\right)^{L^2-v_fN} + T\frac{\partial}{\partial\mu}\log Z \\ &= \frac{1}{\mathrm{e}^{-\frac{\mu}{T}}+1}\Big\{L^2 - v_fN - \Delta v\,\langle n_d\rangle\Big\},\end{aligned} \tag{9.6}$$

using the Fermi distribution function $\frac{1}{\mathrm{e}^{-\mu/T}+1}$. If O is not a function of m, Eq. 9.5 is rewritten further using Z as

$$\langle O\rangle = \frac{1}{Z}\sum_{O,E,n_d} O\,w(O,E,n_d)\,\mathrm{e}^{-\frac{E}{T}}\mathrm{e}^{\sigma n_d}. \tag{9.7}$$

This means that the original ensemble with a single-site crowding agent reservoir is written as a canonical ensemble of an N α-syn system of which disordered states have the effective internal entropy σ. Thus, the effect of macromolecular crowding is absorbed in σ of disordered states.

Here, I define the concentration of α-syns ρ_{syn} and that of the crowding agents ρ_m as

$$\rho_{\mathrm{syn}} = \frac{v_fN + \Delta v\,\langle n_d\rangle}{L^2} \tag{9.8}$$

$$\rho_m = \frac{\langle m\rangle}{L^2}. \tag{9.9}$$

Using ρ_{syn} and ρ_m, Eq. 9.6 is rewritten as

$$\begin{aligned}\rho_m &= \frac{1}{\mathrm{e}^{-\frac{\mu}{T}}+1}\,(1-\rho_{\mathrm{syn}}) \\ \mathrm{e}^{\frac{\mu}{T}}+1 &= \frac{1-\rho_{\mathrm{syn}}}{1-\rho_{\mathrm{syn}}-\rho_m}.\end{aligned} \tag{9.10}$$

Therefore, Eq. 9.3 is rewritten as

$$\sigma = s - \Delta v \, \log\left(\frac{1-\rho_{\rm syn}}{1-\rho_{\rm syn}-\rho_m}\right) \tag{9.11}$$

In the limit of $\mu \to \infty$, that is, $\sigma \to -\infty$ from Eq. 9.3, Eq. 9.7 is rewritten as

$$\langle O \rangle = \frac{1}{Z} \sum_{O,E} O \, w(O,E) \, \mathrm{e}^{-\frac{E}{T}}, \tag{9.12}$$

and the models with the disordered states reduce to those without disordered states.

As long as $\omega(O_1, \cdots)$ is properly counted and the volume of crowding agent v_c is equal to 1, the same analysis is also valid for three-dimensional lattice models and models with different v_d and v_f.

9.2 Protofibril formation above the critical concentration of α-synuclein

I calculate thermal averages through a Monte Carlo simulation using the Metropolis method based on Eq. 9.4 and 9.7. I do not consider the crowding agents explicitly in the simulation; instead I introduced σ calculated from Eq. 9.3 with a given μ. Details of the method were given in the Methods section. I use the following values throughout this thesis: $T = 1$, $\varepsilon_\beta = -8$, $\varepsilon'_\beta = -0.5$, $\varepsilon_\alpha = \varepsilon_\beta/1.2 \simeq -6.67$ and $s = 5$. These values are summarized in Table 9.1. I use L, N and μ as parameters. I listed values for these parameters in Table 9.2. I also listed physical quantities for analysis in Table 9.3.

In the simulations for the BD and BDA models, α-syns form a two-dimensional β-state cluster, which involves both β and β' interactions, above a certain N. In this thesis, I only discuss the equilibrium behavior around the critical concentration, and I do not discuss the elongation process of amyloid fibril. Thus, I refer to the β-state cluster as a protofibril. I write the critical value of N as N_c. For $N \simeq N_c$, I observe formation and dissolution of a protofibril repeatedly in equilibrium. For $N > N_c$, even if there are multiple protofibrils in the early stage of the simulation, they merge into a single thermally stable protofibril. Examples of protofibrils are shown in Fig. 9.1, 9.3, 9.5, 9.2, 9.4 and 9.6 for the BD and BDA models with $L = 128$, $N = 1024$ and different μ. In order to analyze the protofibril formation, I define a β chain as a linear chain of β states connected by β interactions, and I define ℓ_β as the number of the β

Table 9.1: The list of values for model parameters and k_B.

Temperature	T	1
Energy constant for β interactions	ε_β	-8
Energy constant for β' interactions	ε'_β	-0.5
Energy constant for α interactions	ε_α	-6.67
Internal entropy for disordered states	s	5
Unit volume (volume of a single site)	ν	1
Volume of disordered states	v_d	2ν
Volume of β and α states	v_f	ν
Volume of crowding agents	v_c	ν
Difference between v_d and v_f	$\Delta v = v_d - v_f$	ν
Boltzmann constant	k_B	1

Table 9.2: The list of values for L, N and μ. The values for N was rounded off to the nearest integer.

Length of the system	L	64	128
Number of α-syn	N	$2^{i/4}$ ($i = 13$–32)	$2^{i/4}$ ($i = 21$–40)
Chemical potential of crowding agents	μ	$-\infty, 0, 1, 2, 3, \infty$	

states in a β chain. If a β state is not connected with other α-syns by β interactions, I count ℓ_β as 1. I count the number of β states in β chains of length ℓ_β for every ℓ_β.

N dependences of the $\langle n_\beta \rangle$ distributions for the BD and B^oD models with $L = 128$ and $\mu = 0$ are shown in Fig. 9.7 (a). The pairs of the curves of these models coincide for the $N = 512$ and 609 systems; they separate in the $N = 724$ and 861 systems, and the curves for the BD model have two peaks. The snapshots for the $N = 724$ and 861 systems of the BD model show that β chains with ℓ_β corresponding to the right peaks form a protofibril. I observed formation and dissolution of a protofibril in the BD model with $N = 724$, and thus N_c of the BD model is around 724.

N dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 128$ and fixed μ are shown in the left column of Fig. 9.8. μ was set as $-\infty$, 0, 1, 2 and ∞. Here, $n_\beta^{(f)}$ is the number of the β states connected to other β states by β' interactions, and $n_\beta^{(\overline{f})}$ is the number of the β states not connected to other β states by β' interactions. I counted the number of horizontally aligned β_V and vertically aligned β_H as $n_\beta^{(f)}$ also for the models without β' interactions. The sum of $n_\beta^{(f)}$ and $n_\beta^{(\overline{f})}$ is

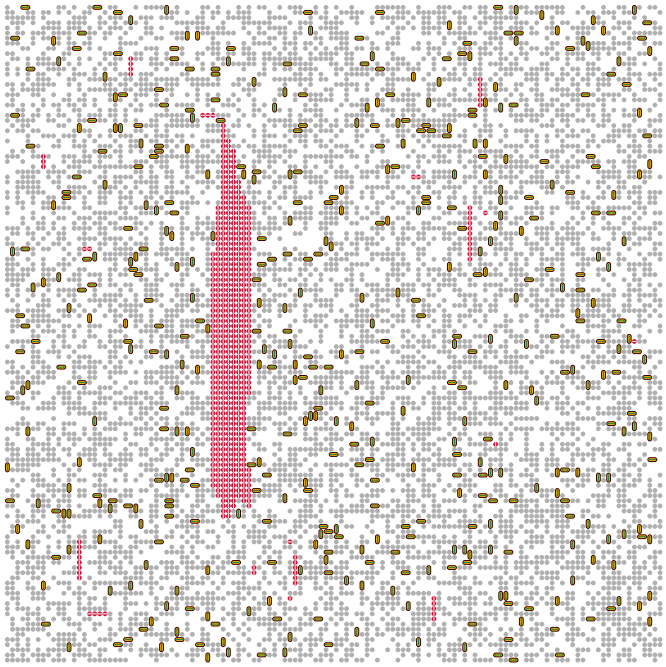

Figure 9.1: Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 0$.

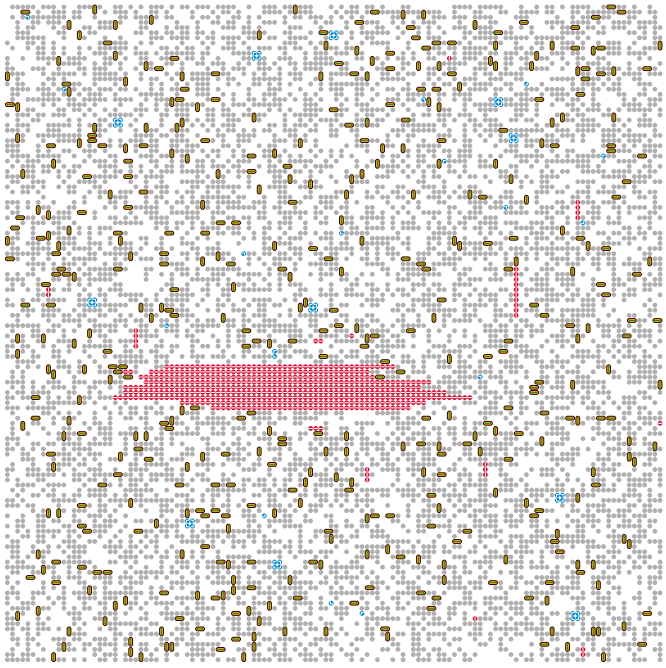

Figure 9.2: Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 0$.

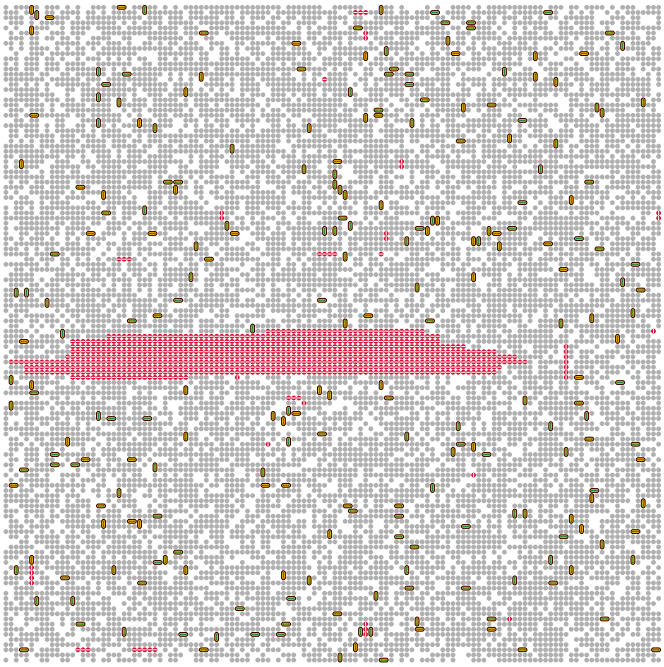

Figure 9.3: Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 1$.

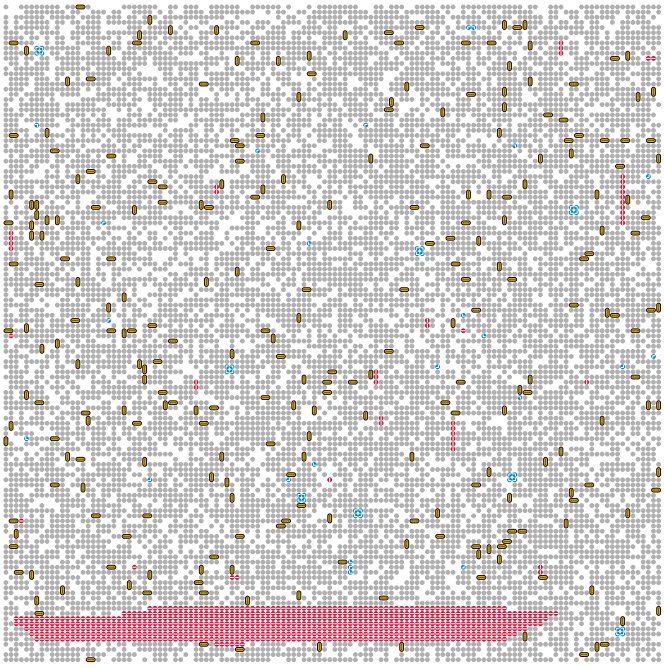

Figure 9.4: Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 1$.

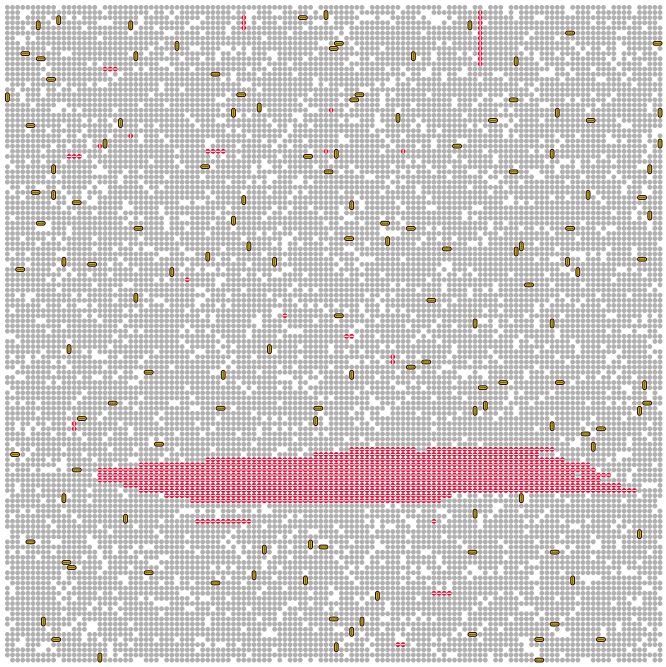

Figure 9.5: Snapshot for the BD model with $L = 128$, $N = 1024$ and $\mu = 2$.

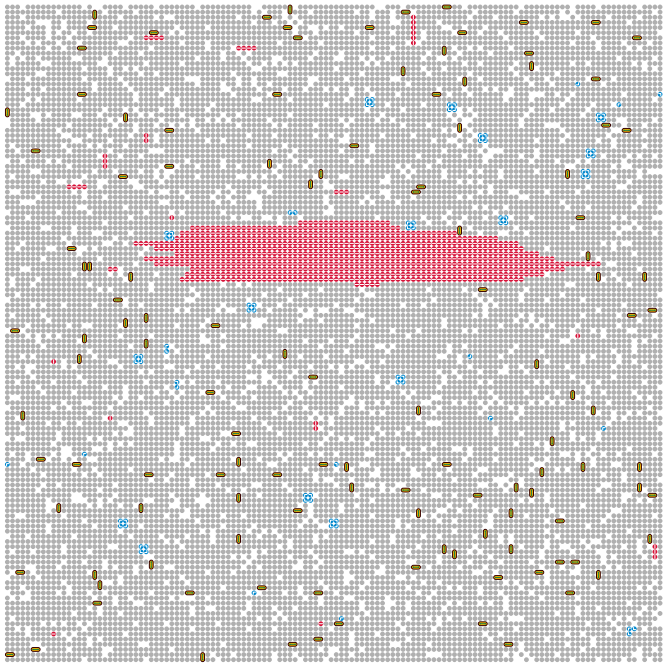

Figure 9.6: Snapshot for the BDA model with $L = 128$, $N = 1024$ and $\mu = 2$.

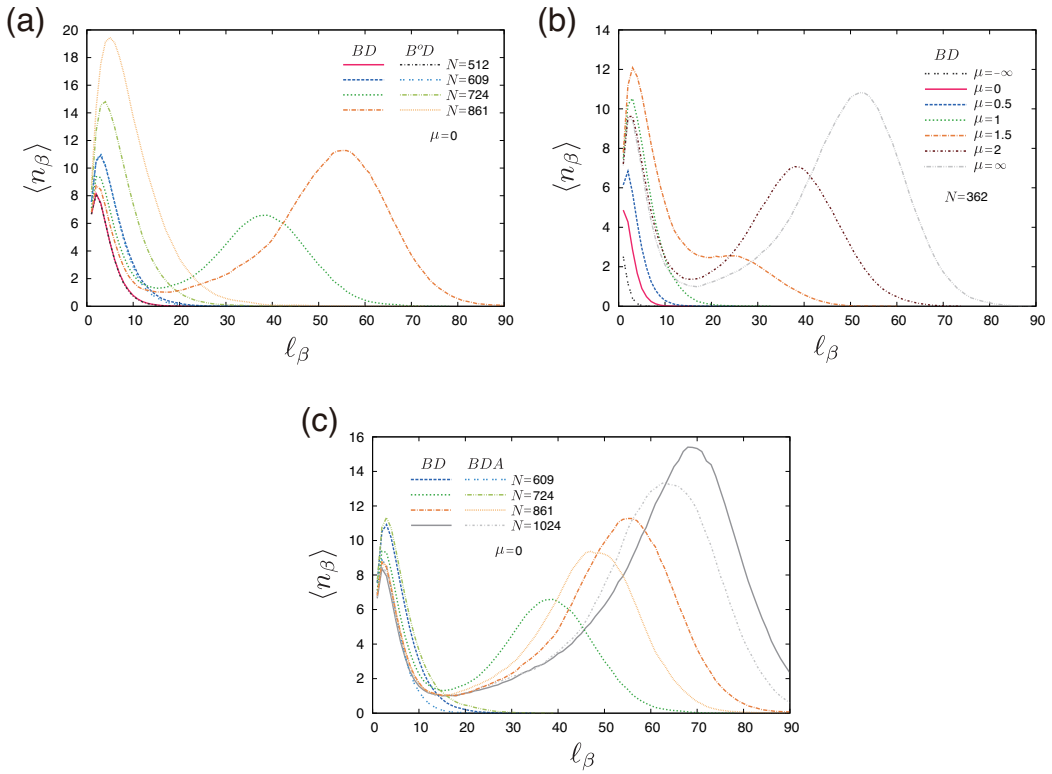

Figure 9.7: Equilibrium number of the β states which are components of β chains of length ℓ_β in the $L = 128$ systems. Here, I define a β chain as a linear chain of the β states connected by β interactions. (a) N dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD and B^oD models. (b) μ dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD model. (c) N dependences of the $\langle n_\beta \rangle$ distributions against ℓ_β for the BD and BDA models.

Table 9.3: The list of physical quantities for analysis.

Total energy	E
Number of disordered states	n_d
Number of crowding agents	m
Number of β states	n_β
Number of the β states connected to other β states by β' interactions	$n_\beta^{(f)}$
Number of the β states not connected to other β states by β' interactions	$n_\beta^{(\bar{f})}$
Number of the β states in a β chain	ℓ_β
Number of α states	n_α
Number of the α states forming tetramers	$n_\alpha^{(t)}$
Number of the α states not forming tetramers	$n_\alpha^{(\bar{t})}$

n_β, and thus $\langle n_\beta^{(f)} \rangle + \langle n_\beta^{(\bar{f})} \rangle = \langle n_\beta \rangle$. The sum of n_β and n_d is N in the BD and B^oD models, and thus $\langle n_\beta \rangle + \langle n_d \rangle = N$.

In all four graphs in the left column of Fig. 9.8, curves for the BD and B^oD models with the same μ separate at N_c except for $\mu = \infty$. The slopes of the N–$\langle n_\beta^{(f)} \rangle$ curves for the BD model change from ~ 0 to ~ 1 around N_c. Each N–$\langle n_\beta^{(\bar{f})} \rangle$ curve for the BD model has a maximum around N_c, and decreases to the constant value which does not depend on μ for large N. Each N–$\langle n_d \rangle$ curve for the BD model has a maximum around N_c, and decreases to a constant value. I found that N_c decreases as μ increases by comparing the separation points for different μ. This means that macromolecular crowding reduces the critical concentration of α-syn.

9.3 Macromolecular crowding induces protofibril formation

μ dependences of the $\langle n_\beta \rangle$ distributions for the BD model with $L = 128$ and $N = 362$ are shown in Fig. 9.7 (b). As μ increases, the right peak begins to grow at $\mu = 1.5$, and gets larger at $\mu = 2$. I observed formation and dissolution of a protofibril in these systems, and thus the critical value of the chemical potential μ_c is around 1.5 for this N. This result suggests that macromolecular crowding induces protofibril formation, and this agrees with the experiments (84, 85).

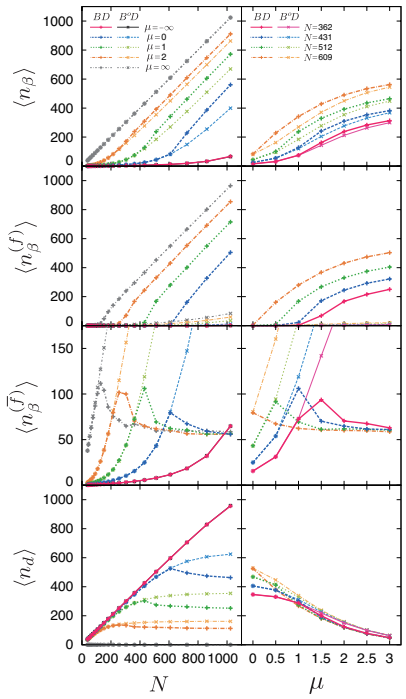

Figure 9.8: Left column: N dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 128$. Right column: μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 128$.

μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 128$ and fixed N are shown in the right column of Fig. 9.8. N was set as 362, 431, 512 and 609. In all four graphs in the right column of Fig. 9.8, curves for the BD and B^oD models with the same N separate at μ_c. Each μ–$\langle n_\beta^{(\bar{f})} \rangle$ curve for the BD model has a maximum around μ_c, and decreases to the constant value which does not depend on N for large μ. The μ–$\langle n_d \rangle$ curves for both models monotonically decrease along with μ. Thus, macromolecular crowding reduces disordered states. Since $\langle n_\beta \rangle$ fulfills $\langle n_\beta \rangle = N - \langle n_d \rangle$, the μ–$\langle n_\beta \rangle$ curves for both models monotonically increase along with μ. Thus, macromolecular crowding promotes β states. I found that μ_c decreases as N increases by comparing the separation points for different values of N.

9.4 α-state tetramers suppress protofibril formation

N dependence of the ℓ_β–$\langle n_\beta \rangle$ curves for the BD and BDA models with $L = 128$ and $\mu = 0$ is shown in Fig. 9.7 (c). Although N_c for the BD model is about 724, the corresponding curve of the BDA model has only a single peak, and thus N_c is higher than that of the BD model. This means that introduction of α states raises N_c. Snapshots for the BD and BDA models with $L = 128$, $N = 1024$ and $\mu = 0$–2 are shown in Fig. 9.1, 9.2, 9.3, 9.4, 9.5 and 9.6. A protofibril coexists with short β chains in both snapshots, and oligomers of α states are found in Fig. 9.2, 9.4 and 9.6.

N and μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$ and $\langle n_\alpha^{(\bar{t})} \rangle$ in the $L = 128$ systems are shown in the left and right columns of Fig. 9.9, respectively. Here, $\langle n_\alpha^{(t)} \rangle$ is the number of the α states forming tetramers, and $\langle n_\alpha^{(\bar{t})} \rangle$ is the number of the α states not forming tetramers. The sum of $n_\alpha^{(t)}$ and $n_\alpha^{(\bar{t})}$ is n_α, and thus $\langle n_\alpha^{(t)} \rangle + \langle n_\alpha^{(\bar{t})} \rangle = \langle n_\alpha \rangle$. The sum of n_β, n_α and n_d is N in the BDA and B^oDA models, and thus $\langle n_\beta \rangle + \langle n_\alpha \rangle + \langle n_d \rangle = N$.

I found that both N_c and μ_c decrease as μ increases by comparing the separation points. These behaviors are similar to those observed in the BD model. I also found that N_c and μ_c for the BDA model are larger than those for the BD model by comparing the separation points. This means that introduction of α states raises both the critical concentration of α-syn and the critical chemical potential of the crowding agents.

The overall shapes of the N–$\langle n_\beta \rangle$, N–$\langle n_\beta^{(f)} \rangle$, N–$\langle n_\beta^{(\bar{f})} \rangle$ and N–$\langle n_d \rangle$ curves in Fig. 9.9 are similar to those in Fig. 9.8. The N–$\langle n_\alpha \rangle$, N–$\langle n_\alpha^{(t)} \rangle$ and N–$\langle n_\alpha^{(\bar{t})} \rangle$ curves for the

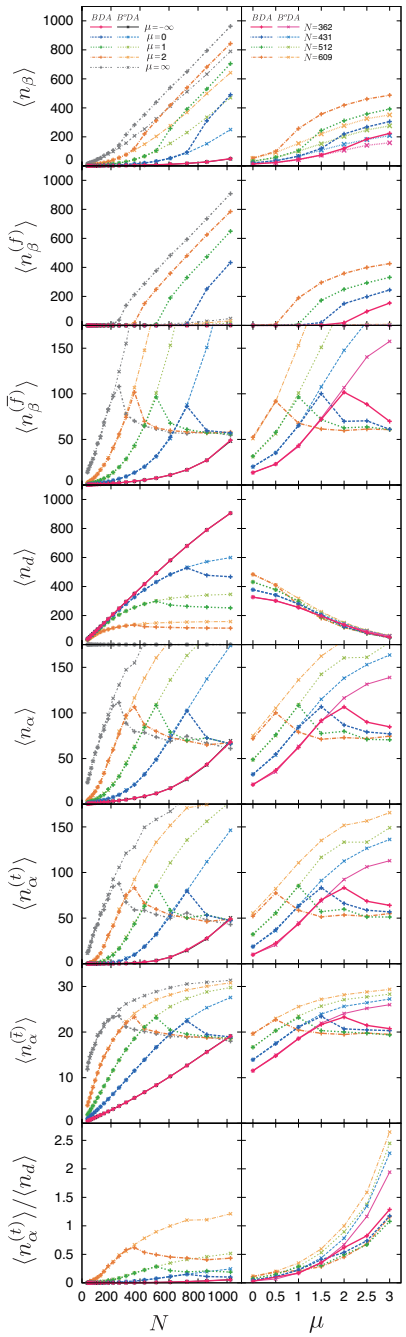

Figure 9.9: Left column: N dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\overline{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 128$. Right column: μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\overline{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 128$.

BDA and B^oDA models are all similar to the N–$\langle n_\beta^{(\bar{f})} \rangle$ curves. Each curve has a maximum around N_c, and decreases to the constant value which does not depend on μ. The majority of α states form tetramers around N_c, and thus I can say that α-state tetramers suppress the fibril formation. This is consistent with the report of Bartels et al. (80)

9.5 Size dependence of fibril formation and α-state-tetramer formation

The results which I have discussed above are obtained from the calculations for the $L = 128$ systems. Here, I show the results of the calculations for the $L = 64$ systems. Fig. 9.10 shows the snapshots of the BD and BDA models with $L = 64$, $N = 256$ and $\mu = 0, 1$ and 2. The α-syn concentration N/L^2 of these systems is the same as that of the systems shown in Fig. 9.1–9.6. In the snapshots of the $L = 64$ systems, there is a single protofibril. Fig. 9.11 and 9.12 show the same physical quantities for the $L = 64$ systems as shown in Fig. 9.8 and 9.9, respectively. The overall shapes of the curves for the $L = 64$ systems are similar to those of the curves for the $L = 128$ systems. The $L = 64$ systems also have N_c and μ_c, and the existence of α-state tetramers increases these values.

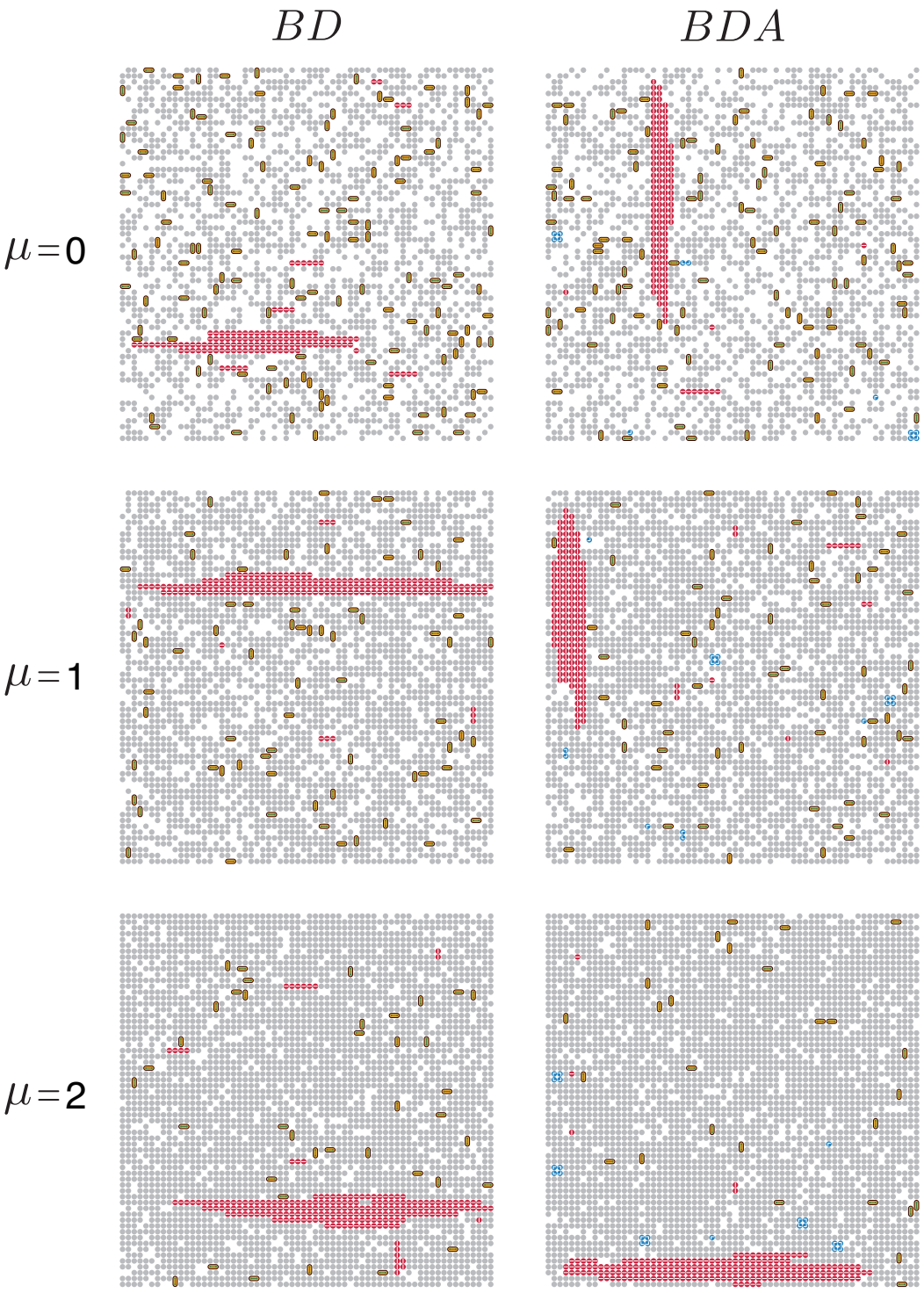

Figure 9.10: Snapshots for the BD and BDA models with $L = 64$, $N = 256$ and $\mu = 0, 1$ and 2.

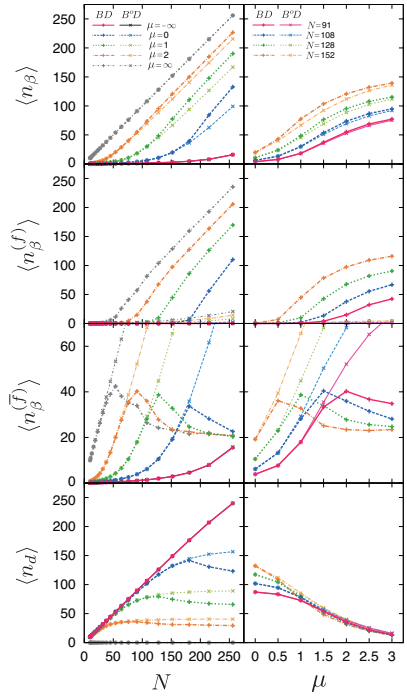

Figure 9.11: Left column: N dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 64$. Right column: μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$ and $\langle n_d \rangle$ for the BD and B^oD models with $L = 64$.

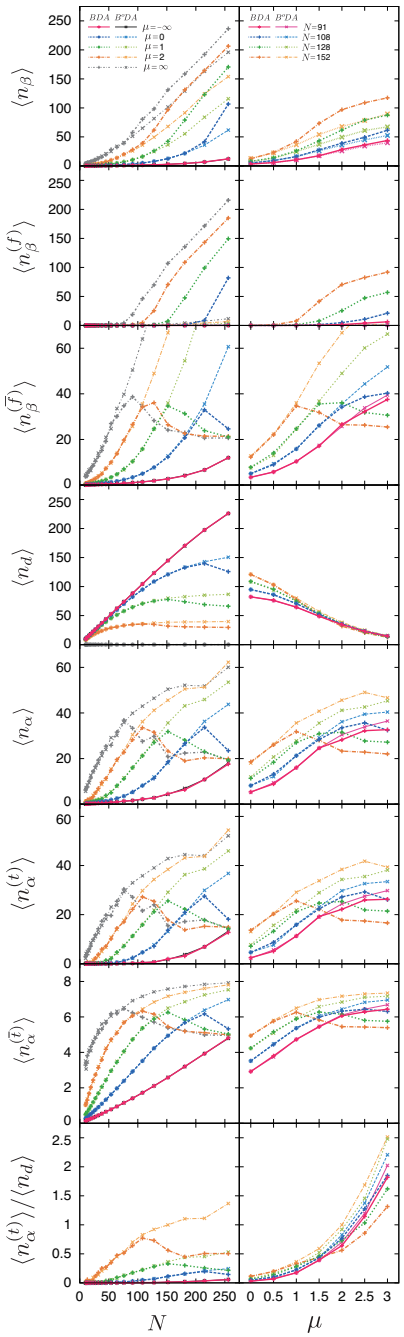

Figure 9.12: Left column: N dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\overline{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 64$. Right column: μ dependences of $\langle n_\beta \rangle$, $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\overline{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$, $\langle n_\alpha^{(\overline{t})} \rangle$ and $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models with $L = 64$.

10

Discussion

10.1 Scenario of protofibril formation induced by macromolecular crowding

The curves for the BD and BDA models with $L = 128$ in Fig. 9.8 and 9.9 are summarized in Fig. 10.1 as stacked graphs. N dependences are shown in the left, and μ dependences are shown in the right. The numbers of internal states fulfill $\langle n_\beta \rangle + \langle n_d \rangle = \langle n_\beta^{(f)} \rangle + \langle n_\beta^{(\bar{f})} \rangle + \langle n_d \rangle = N$ in the BD model, and they fulfill $\langle n_\beta \rangle + \langle n_\alpha \rangle + \langle n_d \rangle = \langle n_\beta^{(f)} \rangle + \langle n_\beta^{(\bar{f})} \rangle + \langle n_\alpha^{(t)} \rangle + \langle n_\alpha^{(\bar{t})} \rangle + \langle n_d \rangle = N$ in the BDA model. The $\langle n_\beta^{(f)} \rangle$ regions start to increase around N_c in the left graphs and μ_c in the right graphs, respectively. The existence of these critical values qualitatively agrees with the experiments of α-syn (76, 77, 84, 85).

$\langle n_\beta \rangle$–$\langle n_\beta^{(f)} \rangle$ and $\langle n_\beta \rangle$–$\langle n_\beta^{(\bar{f})} \rangle$ relations for all four models with $L = 128$ are plotted in Fig. 10.2 (a), and $\langle n_\beta \rangle$–$\langle n_\alpha \rangle$, $\langle n_\beta \rangle$–$\langle n_\alpha^{(t)} \rangle$ and $\langle n_\beta \rangle$–$\langle n_\alpha^{(\bar{t})} \rangle$ relations for the BDA and B^oDA models are plotted in Fig. 10.2 (b). These figures show that these relations do not depend on μ. Thus, the ratio of $\langle n_\beta \rangle$ to $\langle n_\alpha \rangle$ also does not depend on μ. I write the critical value of $\langle n_\beta \rangle$ as $\langle n_\beta \rangle_c$. Since $\langle n_d \rangle$ for the BD model with $\mu = \infty$ is zero for any N, and $\langle n_\beta \rangle$ is equal to N from the identity $\langle n_\beta \rangle + \langle n_d \rangle = N$, the graphs for the BD model with $\mu = \infty$ in Fig. 10.2 (a) are exactly the same as those in Fig. 9.8. Thus, $\langle n_\beta \rangle_c$ of the BD and BDA models are equal to N_c for the BD model with $\mu = \infty$ irrespective of μ.

I propose the following scenario for the protofibril formation induced by macromolecular crowding based on the stacked graphs for μ dependence of internal states

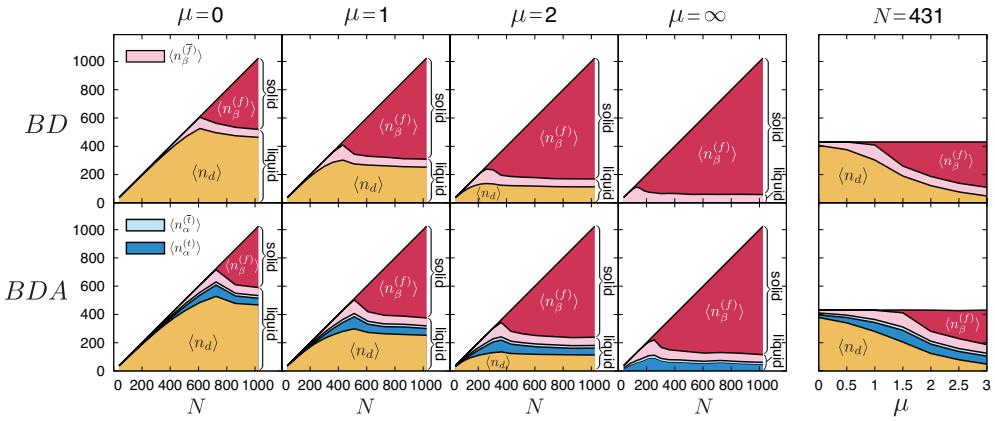

Figure 10.1: Stacked graphs of the equilibrium numbers of internal states for the BD and BDA models with $L = 128$, $\mu = 0$, 1, 2 and ∞, and $N = 431$. $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$, $\langle n_\alpha^{(\bar{t})} \rangle$, $\langle n_\alpha^{(t)} \rangle$ and $\langle n_d \rangle$ are colored red, pink, light blue, blue and yellow, respectively.

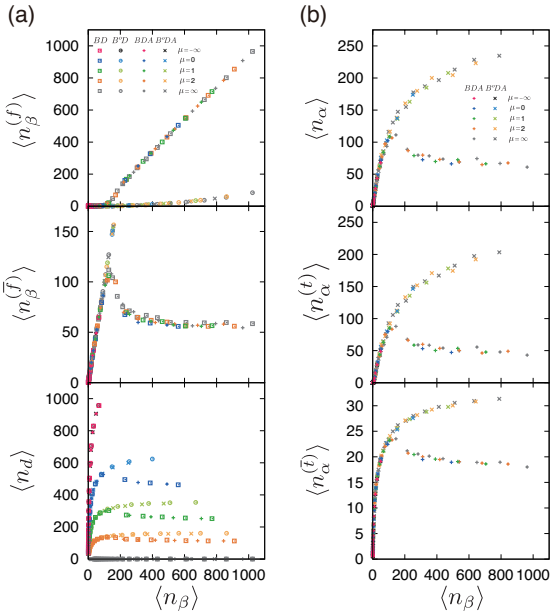

Figure 10.2: (a) $\langle n_\beta \rangle$ dependences of $\langle n_\beta^{(f)} \rangle$, $\langle n_\beta^{(\bar{f})} \rangle$ and $\langle n_d \rangle$ of the BD, B^oD, BDA and B^oDA models with $L = 128$. (b) $\langle n_\beta \rangle$ dependences of $\langle n_\alpha \rangle$, $\langle n_\alpha^{(t)} \rangle$ and $\langle n_\alpha^{(\bar{t})} \rangle$ for the BDA and B^oDA models with $L = 128$.

presented on the right of Fig. 10.1. (i) Since the stability of the disordered states is determined by σ, which is given by Eq. 9.3, the $\langle n_d \rangle$ regions decrease along with μ as a result of decreasing σ. In other words, macromolecular crowding destabilizes the disordered states by reducing the effective internal entropy. (ii) As I explained above, the $\langle n_\beta \rangle$ region increases as a result of the decrease in $\langle n_d \rangle$ from $\langle n_\beta \rangle = N - \langle n_d \rangle$ in the BD model. In the case of the BDA model, the ratio of $\langle n_\beta \rangle$ and $\langle n_\alpha \rangle$ does not depend on μ, and thus I can write $\langle n_\beta \rangle \simeq k \langle n_\alpha \rangle$, $(k > 0)$ irrespective of μ. Then, I can rewrite the identity $\langle n_\beta \rangle + \langle n_\alpha \rangle + \langle n_d \rangle = N$ as $\langle n_\beta \rangle \simeq \frac{1}{1+k} \{N - \langle n_d \rangle\}$. This also shows the increase in the $\langle n_\beta \rangle$ region as a result of the decrease in $\langle n_d \rangle$. (iii) If increased $\langle n_\beta \rangle$ is larger than $\langle n_\beta \rangle_c$, the $\langle n_\beta^{(f)} \rangle$ regions start to increase, and α-syns form a protofibril. At this point, $\mu \simeq \mu_c$. The factor $\frac{1}{1+k} (< 1)$ indicates that $\langle n_\beta \rangle$ of the BDA model is smaller than that of the BD model, and it explains why the μ_c of the BDA model is larger than that of the BD model.

10.2 An explanation for the controversy over the observation of helically folded tetramers

I plotted the N and μ dependences of $\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ for the BDA and B^oDA models at the bottom left and right of Fig. 9.9, respectively. Each N–$\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ curve for the BDA model has a maximum around N_c, and decreases to a constant value. This means that relative stability of helically folded tetramers depends on the concentration of α-syn. The μ–$\langle n_\alpha^{(t)} \rangle / \langle n_d \rangle$ curves for the BDA model monotonically increase along with μ. Thus, macromolecular crowding increases the relative number of the helically folded tetramers to that of the disordered monomers. The macromolecular crowding effect may provide a good explanation for the controversy over the existence of the helically folded tetramers. The present result that the density of such tetramers depends on crowdedness suggests that whether or not the tetramers are observed in experiments is determined by the crowdedness of the environment.

10.3 Solid–liquid equilibrium analogy

The left graphs in Fig. 10.1 show that $\langle n_\beta^{(f)} \rangle$ monotonically increases above N_c, while $\langle n_\beta^{(\bar{f})} \rangle$, $\langle n_d \rangle$, $\langle n_\alpha^{(t)} \rangle$ and $\langle n_\alpha^{(\bar{t})} \rangle$ each decrease to constant values. In other words, the

protofibril absorbs increased α-syns above N_c while the other non-protofibril components converge to a constant concentration. From this result, I can draw an analogy of protofibril–non-protofibril equilibrium with a solid–liquid equilibrium; the protofibril can be seen as solid, and the non-protofibril components can be seen as liquid. The words "solid" and "liquid" are written in the stacked graphs.

10.4 A coarse-grained model with internal states: funnel gas model

The definition of the disordered states and the derivation of Eq. 9.3 and 9.4 are applicable to other IDPs as long as the volume of the folded states is greater than or equal to that of the crowding agents. Eq. 9.3 and 9.4 will be a first approximation to discuss the relation between macromolecular crowding and the stability of the disordered states also for larger crowding agents. Eq. 9.3 provides a simple theoretical expression for the interplay of intrinsic disorder and macromolecular crowding which was suggested by White et al. (3).

The main assumption of my model is that proteins can be expressed as coarse-grained particles with some internal states coupled with the effective volume of each state. This simplification enabled me to derive the expression of effective internal entropy shown above. Here, I generalize this idea. Proteins have a funnel-like energy landscape (15, 47, 100). IDPs with a folded state in their bound complex with other proteins including α-syn also have a funnel-like energy landscape (35, 54). In this study, I assumed that α-syn has a triple-well free-energy landscape, and I constructed the coarse-grained particle with the three internal states based on the free-energy landscape of α-syn. In the same way, I can also construct coarse-grained particles with internal states based on the free-energy landscape of specific proteins. Then, I can produce equilibrium states of multiple molecular systems including the crowding effect. This is the spirit of the new coarse-grained model, and I call this type of models the "funnel gas model".

11

Summary

In this thesis, I discussed the binding processes of IDPs in two ways by constructing and analyzing two models with different spatial resolution. I focused on the binding processes between an IDP and its target molecule in Part II, while I focused on the self-assembly of hundreds of IDPs in Part III. In both cases, IDPs exhibit characteristic binding behavior.

In Part II, I constructed a lattice model of an IDP with its target molecule based on the HP model. I represented an IDP as a mixture of various states, without a specific structure formed in an equilibrium state. The IDP binds to the target below its binding temperature and forms a helical structure. This represents coupled folding and binding. By comparing free-energy landscapes of various types of the binding surface, I found that the closure of the binding surface produces a new intermediate state on the binding pathway. This behavior can simply explain the existence of an intermediate state of p27/cyclin A/Cdk2 and pKID/KIX systems. From this result, I conclude that flexibility of the IDP provides a scaffold to access the closed binding site, which I call multiform binding, and this is a functional advantage of an IDP. By decomposing the free-energy landscape by the conformations of the target, I observed the conformational selection process of the target. This result suggests that the IDP and the target form a binding funnel.

In Part III, I constructed a lattice gas model of α-syn, which is one of the IDPs, with the effect of macromolecular crowding. I derived a mathematical expression that relates macromolecular crowding to intrinsic disorder. In this expression, the effect of macromolecular crowding, which is represented as chemical potential of the crowding

agents μ, is taken into account as the effective internal entropy of the disordered states σ. Increase in μ reduces σ, and decrease in σ leads to the destabilization of the disordered states. Thanks to the derived expression, the original lattice gas model was reduced to a model with implicit crowding agents. This modification to the model reduced the computational cost of the Monte Carlo simulations. From the simulation and its analyses, I observed that macromolecular crowding reduces the critical concentration of α-syn. This agrees with *in vitro* experiments. I discussed the recent controversy over the existence of the helically folded tetramers of α-syn. The result suggests that macromolecular crowding raises the relative number of α-state tetramers to that of the disordered states. It also suggests that whether or not the tetramers are observed in experiments is determined by the crowdedness of the environment.

Acknowledgements

I would like to thank all members in Kikuchi laboratory, Odanaka laboratory, Furihata laboratory at Osaka University, Sasai laboratory at Nagoya University, Aizawa laboratory at Tokyo Institute of Technology and Takada laboratory at Kyoto University.

I enjoyed and struggled in the master and Ph.D. student days at Osaka University from 2009 to 2013. Kikuchi laboratory had a really free atmosphere. I did what I wanted to do, although all things were quite inefficient. Thanks to the struggling days, now I can say that the works in this thesis are my own works.

I talked a lot with members on the sixth floor of Cybermedia Center in the Toyonaka campus of Osaka University. On that floor, there is a shared space called "free media". Everyone there can freely talk with anyone there about anything they want to. It is the most creative space as far as I know. I first heard about this place from Associate Professor Tomotoshi Nishino at Kobe University, and he recommended me to be a member of Kikuchi laboratory. That was a great recommendation.

When I saw Akimasa Kitajima for the first time there, I was totally shocked by his unconventional character. I think I learned a lot from his unusual way of thinking and his fantastic talking skill. I remember Dr. Nen Saito, who is now an Assistant Professor of Kaneko laboratory at the University of Tokyo, said that he finally could have done his original work, after his defense. That word raised the bar of my thesis in my mind.

I discussed various topics with Associate Professor Daisuke Furihata. He gave me a lot of advices about my studies and my life. Professor Shinji Odanaka told me about his experiences, and he encouraged me to write papers. Dr. Tomoyuki Obuchi, who is now an Assistant Professor at Tokyo

Institute of Technology, has a wide range of interest, and he taught me a lot of things ranging from the studies of random systems to cool board games. When I was studying about self-avoiding walks, I came up with the idea to create a new board game based on self-avoiding polygon, and I created the board game which we call "Polygo" with Hidetoshi Sugihara and Hirokazu Koma. This game made Professor Macoto Kikuchi, who is my supervisor, more excited than any of my studies.

One day, Assistant Professor Masayuki Ohzeki at Kyoto University came to the free media with Assistant Professor Chihiro H. Nakajima at Tohoku University and Assistant Professor Keisuke Fujii at Kyoto University to discuss the research related to the self-avoiding walks. After that discussion, we spend more time to make and record short comedies played with the good low voice of Dr. Nakajima, which is very close to the dubbed-in voice of "Space Alien Jones" appeared in the Japanese famous TV commercial for the canned coffee "BOSS". These are good memories of the free media, and now I believe that researchers need such free spaces to produce creative works. I would like to express my gratitude to people who shared time with me in the free media.

In those days, I went to dinner almost all of weekdays with Dr. Shohiro Sho, who is a member of Odanaka laboratory. Everyday talks with him took me away from the stressful academic life. I often asked what he had done in that day, and I also talked about what I had done. These talks became a good check for tiny everyday progress. Thanks a lot Dr. Sho. Let's go to a sushi bar when I successfully finish the defense of this thesis.

I had a valuable experience of staying at Sasai laboratory in Nagoya University. I stayed there two months from June 2011 and three months from December 2011. I discussed a lot with Professor Masaki Sasai, who was a supervisor during the stay. Every evening, I went to a cafeteria with Associate Professor Tomoki P. Terada to have a dinner, and I asked many questions about experimental techniques to analyze thermodynamic properties of proteins. When I was writing this thesis, he helped me to find the original papers of differential scanning calorimetry (DSC) and isothermal

titration calorimetry (ITC). I also asked about the analysis of thermodynamic quantities given by these methods. Shintaro Minami and Kengo Sawada, who were Ph.D. students in Sasai laboratory, taught me basic knowledge of protein structures and algorithms of bioinformatic tools. I respect their attitude toward the research targets, and I learned that not only love for methodologies but also that for research targets make it possible to create useful research tools which produce interesting non-trivial results.

In Sasai laboratory, I was in the same room as Assistant Professor George Chikenji. At that time, I did not have enough basic knowledge on biophysics and protein science, and I asked a lot of questions about what was already known and what was unknown. He kindly answered my questions. His original background is the same as mine, that is physics, and thus he can explain every topic in terms of physics. It made all of his talks very clear, understandable and interesting to me. Sometimes, we talked all day. Now I think that if I had not stayed in Sasai laboratory, I would not have continued the studies about proteins until now. I would like to thank all stuffs and students there.

In my student days, I attended the meetings of the Physical Society of Japan (JPS), the Biophysical Society of Japan (BSJ) and the Protein Science Society of Japan (PSSJ). I also attended the several meetings of the Grant-in-Aid for scientific Research on Innovative Areas "Target recognition and expression mechanism of intrinsically disordered protein". These meetings provided great opportunities for me to meet experts on statistical physics, non-equilibrium systems, biophysics and protein science. I am grateful to them for the fruitful discussions.

Associate Professor Yukito Iba at the Institute of Statistical Mathematics gave me chances to talk about my studies in the seminars related to Monte Carlo simulations. I first met Associate Professor Koji Hukushima at the University of Tokyo in one of those seminars. When I saw him at the wedding party of Dr. Ohzeki in September 2014, he encouraged me not to give up getting a Ph.D. degree, and he set a time limit: March, 2015. Fortunately, it looks like I am going to make it.

In a JPS meeting, I first met Assistant Professor Masashi Fujii at the University of Tokyo and Dr. Hirofumi Niiya at Nagoya University. I was surprised to know their busy schedules of presentations, and I thought that I should be more active. Associate Professor Akinori Awazu at Hiroshima University and Associate Professor Yuichi Togashi at Hiroshima University gave me some comments and suggestions on my studies. Research Associate Masamichi J. Miyama at Tohoku University once came and stayed at my room in Kyoto after Dr. Ohzeki's wedding party, and we discussed what is important to do a good study on science.

Professor Junichi Higo in Institute for Protein Research at Osaka University, Professor Motonori Ota at Nagoya University and Associate Professor Munehito Arai at the University of Tokyo gave me chances to talk in a seminar, a workshop and a symposium. They gave me a lot of comments on my studies, and these helped me to find what I should do next.

Dr. Ryotaro Koike, who is an Assistant Professor of Ota laboratory at Nagoya University, told me about the story of his Ph.D. student days and postdoc days. When I was not confident in my study, he advised me to write a paper in order to clarify the purpose and meaning of the study. After that comment, I realized what I should do, and I could concentrate on writing a paper. Assistant Professor Sotaro Fuchigami at Yokohama City University gave me the advice that it is better to prepare a single take-home message for the audience in a short presentation so as not to spend a lot of time to add and modify the details which are not related to the main theme. That is the basic strategy to prepare for a presentation today.

Dr. Tomoshi Kameda, who is a Senior Research Scientist at National Institute of Advanced Industrial Science and Technology, gave me a lot of time to discuss my studies and my future career with me. Dr. Kota Kasahara in Institute for Protein Research at Osaka University told me about the story of his thesis defense, and he gave me his doctorial thesis. I learned how a doctorial thesis should be from his thesis. Dr. Kazuo Yamashita in Immunology Frontier Research Center at Osaka University told me about his experience in his Ph.D. student days and about changing his research field.

Assistant Professor Takayuki Amemiya at Nagoya University, Dr. Satoshi Omori at Yokohama City University and Dr. Shinji Matsushita at Kyoto University also told me their stories of Ph.D. student days. Dr. Hiroo Kenzaki at RIKEN and Assistant Professor Hiromitsu Shimoyama at Kitasato University taught me a lot about protein simulations. Assistant Professor Hafumi Nishi at Tohoku University told me not to care about failing to get a Ph.D. degree within a time frame because there are many such people. I would like to go to the BSJ and PSSJ meetings to see them, and I would like to say "I finally got a Ph.D. degree". Of course, if I succeed in getting it.

In the "Seminar on Simulations of Intrinsically Disordered Protein", I discussed the IDP related topics with Dr. Koji Umezawa at Waseda University and Dr. Jinzen Ikebe at Japan Atomic Energy Agency. Associate Professor Yasunori Aizawa at Tokyo Institute of Technology, who I first met at the seminar, invited me to his laboratory. After my presentation, he kindly planned for me to visit several laboratories in Graduate School of Bioscience and Biotechnology. I deeply appreciate his gracious hospitality.

After I failed to graduate from Osaka University, Professor Shoji Takada at Kyoto University gave me a job opportunity in 2014. I would like to offer my special thanks to Kouya Sakuma, who is a member of Takada laboratory. He read many parts of my thesis carefully, and gave me a lot of helpful comments about the contents and my English. He has a lot of knowledge about biology and chemistry, and he is also interested in physics. He taught me a lot about molecular biology and biochemistry. We discussed the basic theories and the applications of protein sciences. I'm looking forward to his future studies about protein design. I bet he will surprise us by designing wonderful proteins beyond our imagination.

I am grateful to all the reviewers of this thesis. Associate Professor Koichi Fujimoto in the Department of Biological Sciences at Osaka University, who is one of the reviewers, taught me how to explain theoretical works convincingly to experimentalist in the field of biology. I would like to follow his advice to tell them about my theoretical studies.

At the end, I would like to thank my supervisor Professor Macoto Kikuchi. He gave me a lot of time to discuss the results of simulations and to write papers. He commented on my poorly written English, and I learned a lot from his comments. Even after I dropped out from the Ph.D. course, he also gave me a lot of time. I am still not sure whether he is really interested in my research or not, but someday, I would like to please him by my study.

Nobu C. Shirai

References

[1] P. E. Wright and H. J. Dyson. **Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.** *Journal of molecular biology*, **293**(2):321–331, October 1999. i, 9

[2] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell 6th edition*. Garland Science, December 2014. ii

[3] D. A. White, A. K. Buell, T. P. J. Knowles, M. E. Welland, and C. M. Dobson. **Protein Aggregation in Crowded Environments**. *J. Am. Chem. Soc.*, **132**(14):5170–5175, March 2010. ii, 42, 72

[4] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. **The Protein Data Bank**. *Nucleic Acids Research*, **28**(1):235–242, January 2000. 4

[5] Schrödinger, LLC. **The PyMOL Molecular Graphics System, Version 1.7.1.1**. 2014. 5

[6] C. B. ANFINSEN, E. HABER, M. SELA, and F. H. WHITE. **The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain.** *Proceedings of the National Academy of Sciences of the United States of America*, **47**:1309–1314, September 1961. 4

[7] C. B. Anfinsen. **Principles that Govern the Folding of Protein Chains**. *Science*, **181**(4096):223–230, July 1973. 4

[8] P. L. Privalov and N. N. Khechinashvili. **A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study**. *Journal of Molecular Biology*, **86**(3):665–684, July 1974. 5

[9] P. L. Privalov and A. I. Dragan. **Microcalorimetry of biological macromolecules**. *Biophysical Chemistry*, **126**(1-3):16–24, March 2007. 5

[10] H. Taketomi, Y. Ueda, and N. Go. **STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION**. *International Journal of Peptide and Protein Research*, **7**(6):445–459, November 1975. 6, 9

[11] N. Go and H. Taketomi. **Respective roles of short- and long-range interactions in protein folding**. *Proceedings of the National Academy of Sciences*, **75**(2):559–563, February 1978. 6, 9

[12] Y. Iba, G. Chikenji, and M. Kikuchi. **Simulation of Lattice Polymers with Multi-Self-Overlap Ensemble**. *Journal of the Physical Society of Japan*, **67**(10):3327–3330, 1998. 7, 23

[13] G. Chikenji, M. Kikuchi, and Y. Iba. **Multi-Self-Overlap Ensemble for Protein Folding: Ground State Search and Thermodynamics**. *Physical Review Letters*, **83**(9):1886–1889, August 1999. 7, 23

[14] N. Go. **THEORY OF REVERSIBLE DENATURATION OF GLOBULAR PROTEINS**. *International Journal of Peptide and Protein Research*, **7**(4):313–323, 1975. 9

[15] N. Go. **Theoretical Studies of Protein Folding**. *Annual Review of Biophysics and Bioengineering*, **12**(1):183–210, 1983. 9, 72

[16] P. H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine, and N. P. Pavletich. **Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain**. *Science*, **274**(5289):948–953, November 1996. 10, 17

[17] M. Uesugi, O. Nyanguile, H. Lu, A. J. Levine, and G. L. Verdine. **Induced Helix in the VP16 Activation Domain upon Binding to a Human TAF**. *Science*, **277**(5330):1310–1313, August 1997. 10, 17

[18] I. Radhakrishnan, G. C. Pérez-Alvarado, D. Parker, Dyson, M. R. Montminy, and P. E. Wright. **Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions**. *Cell*, **91**(6):741–752, December 1997. 10, 12, 17

[19] K. Sugase, H. J. Dyson, and P. E. Wright. **Mechanism of coupled folding and binding of an intrinsically disordered protein**. *Nature*, **447**(7147):1021–1025, May 2007. 10, 12

[20] G. A. Papoian and P. G. Wolynes. **The physics and bioinformatics of binding and folding—an energy landscape perspective**. *Biopolymers*, **68**(3):333–349, March 2003. 10

[21] A. Bairoch and B. Boeckmann. **The SWISS-PROT protein sequence data bank.** *Nucleic acids research*, **19 Suppl**:2247–2249, April 1991. 11

[22] J. C. Wootton and S. Federhen. **Statistics of local complexity in amino acid sequences and sequence databases**. *Computers & Chemistry*, **17**(2):149–163, June 1993. 11

[23] J. C. Wootton. **Sequences with 'unusual' amino acid compositions**. *Current Opinion in Structural Biology*, **4**(3):413–421, June 1994. 11

[24] J. C. Wootton. **Non-globular domains in protein sequences: Automated segmentation using complexity measures**. *Computers & Chemistry*, **18**(3):269–285, September 1994. 11

[25] T. U. Consortium. **Activities at the Universal Protein Resource (UniProt)**. *Nucleic Acids Research*, **42**(D1):D191–D198, January 2014. 11

REFERENCES

[26] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. **Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life**. *Journal of Molecular Biology*, **337**(3):635–645, March 2004. 11, 30

[27] B. Xue, A. K. Dunker, and V. N. Uversky. **Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life**. *Journal of Biomolecular Structure and Dynamics*, **30**(2):137–149, June 2012. 11

[28] J. Yan, M. J. Mizianty, P. L. Filipow, V. N. Uversky, and L. Kurgan. **RAPID: Fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale**. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1834**(8):1671–1680, August 2013. 11

[29] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky. **Flexible nets**. *FEBS Journal*, **272**(20):5129–5148, October 2005. 11

[30] C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, and L. M. Iakoucheva. **Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.** *PLoS computational biology*, **2**(8):e100+, August 2006. 11

[31] M. Higurashi, T. Ishida, and K. Kinoshita. **Identification of transient hub proteins and the possible structural basis for their multiple interactions**. *Protein Science*, **17**(1):72–78, January 2008. 11

[32] A. Patil, K. Kinoshita, and H. Nakamura. **Domain distribution and intrinsic disorder in hubs in the human protein–protein interaction network**. *Protein Science*, **19**(8):1461–1468, August 2010. 11

[33] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. **Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity**. *Proceedings of the National Academy of Sciences*, **93**(21):11504–11509, October 1996. 11

[34] P. Tompa. **Intrinsically unstructured proteins.** *Trends in biochemical sciences*, **27**(10):527–533, October 2002. 11

[35] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes. **Speeding molecular recognition by using the folding funnel: the fly-casting mechanism.** *Proceedings of the National Academy of Sciences of the United States of America*, **97**(16):8868–8873, August 2000. 11, 72

[36] E. R. Lacy, I. Filippov, W. S. Lewis, S. Otieno, L. Xiao, S. Weiss, L. Hengst, and R. W. Kriwacki. **p27 binds cyclin–CDK complexes through a sequential mechanism involving binding-induced protein folding**. *Nature Structural & Molecular Biology*, **11**(4):358–364, March 2004. 11, 12

[37] H. Toyoshima and T. Hunter. **p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21.** *Cell*, **78**(1):67–74, July 1994. 11

[38] P. D. Jeffrey, A. Russo, A., K. Polyak, E. Gibbs, J. Hurwitz, J. Massague, N. Pavletich, and P. **Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex**. *Nature*, **376**(6538):313–320, July 1995. 11

[39] A. A. Russo, P. D. Jeffrey, and N. P. Pavletich. **Structural basis of cyclin-dependent kinase activation by phosphorylation**. *Nature Structural & Molecular Biology*, **3**(8):696–700, August 1996. 11

[40] A. A. Russo, P. D. Jeffrey, A. K. Patten, J. Massague, and N. P. Pavletich. **Crystal structure of the p27Kip1 cyclin-dependent-kinase inibitor bound to the cyclin A–Cdk2 complex**. *Nature*, **382**(6589):325–331, July 1996. 11, 17

[41] K. F. Lau and K. A. Dill. **A lattice statistical mechanics model of the conformational and sequence spaces of proteins**. *Macromolecules*, **22**(10):3986–3997, October 1989. 12

[42] S. B. Zimmerman and S. O. Trach. **Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli.** *Journal of molecular biology*, **222**(3):599–620, December 1991. 12

[43] A. P. Minton. **Excluded volume as a determinant of macromolecular structure and reactivity**. *Biopolymers*, **20**(10):2093–2120, October 1981. 12

[44] E. R. John. **Macromolecular crowding: obvious but underappreciated**. *Trends in Biochemical Sciences*, **26**(10):597–604, October 2001. 12

[45] H.-X. X. Zhou, G. Rivas, and A. P. Minton. **Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences.** *Annual review of biophysics*, **37**(1):375–397, June 2008. 12

[46] C. J. Camacho and D. Thirumalai. **Kinetics and thermodynamics of folding in model proteins**. *Proceedings of the National Academy of Sciences*, **90**(13):6369–6372, July 1993. 17

[47] K. A. Dill and H. S. Chan. **From Levinthal to pathways to funnels**. *Nature Structural & Molecular Biology*, **4**(1):10–19, January 1997. 17, 72

[48] G. Chikenji and M. Kikuchi. **What is the role of non-native intermediates of -lactoglobulin in protein folding?** *Proceedings of the National Academy of Sciences*, **97**(26):14273–14277, December 2000. 17, 23

[49] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps et al. **Intrinsically disordered protein.** *Journal of molecular graphics & modelling*, **19**(1):26–59, 2001. 17

[50] M. Nomura, H. Uda-Tochio, K. Murai, N. Mori, and Y. Nishimura. **The Neural Repressor NRSF/REST Binds the PAH1 Domain of the Sin3 Corepressor by Using its Distinct Short Hydrophobic Helix**. *Journal of Molecular Biology*, **354**(4):903–915, December 2005. 17

[51] B. A. Berg and T. Neuhaus. **Multicanonical algorithms for first order phase transitions**. *Physics Letters B*, **267**(2):249–253, September 1991. 23

[52] B. A. Berg and T. Neuhaus. **Multicanonical ensemble: A new approach to simulate first-order phase transitions**. *Physical Review Letters*, **68**(1):9–12, January 1992. 23

[53] K. Nakanishi and M. Kikuchi. **Thermodynamics of Aggregation of Two Proteins**. *Journal of the Physical Society of Japan*, **75**(6):064803+, 2006. 23, 30

[54] N. C. Shirai and M. Kikuchi. **Structural flexibility of intrinsically disordered proteins induces stepwise target recognition**. *The Journal of Chemical Physics*, **139**(22):225103+, December 2013. 23, 72

[55] F. Wang and D. P. Landau. **Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States**. *Physical Review Letters*, **86**(10):2050–2053, March 2001. 23

[56] F. Wang and D. P. Landau. **Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram**. *Physical Review E*, **64**(5):056101+, October 2001. 23

[57] K. Matsushita and M. Kikuchi. **Frustration-induced protein intrinsic disorder**. *The Journal of Chemical Physics*, **138**(10):105101+, 2013. 29

[58] J. Higo, Y. Nishimura, and H. Nakamura. **A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations.** *Journal of the American Chemical Society*, **133**(27):10448–10458, July 2011. 31

[59] K. Uéda, H. Fukushima, E. Masliah, Y. Xia, A. Iwai, M. Yoshimoto, D. A. 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*, **90**(23):11282–11286, December 1993. 41

[60] R. Jakes, M. G. Spillantini, and M. Goedert. **Identification of two distinct synucleins from human brain**. *FEBS Letters*, **345**(1):27–32, May 1994. 41

[61] A. Iwai, E. Masliah, M. Yoshimoto, N. Ge, L. Flanagan, H. A. d. Silva, A. Kittel, and T. Saitoh. **The precursor protein of non-A beta component of Alzheimer's disease amyloid is a presynaptic protein of the central nervous system.** *Neuron*, **14**(2):467–475, February 1995. 41

[62] P. J. Kahle, M. Neumann, L. Ozmen, V. Müller, H. Jacobsen, A. Schindzielorz, M. Okochi, U. Leimer, H. v. d. Putten, A. Probst et al. **Subcellular Localization of Wild-Type and Parkinson's Disease-Associated Mutant α-Synuclein in Human and Transgenic Mouse Brain**. *The Journal of Neuroscience*, **20**(17):6365–6373, September 2000. 41

[63] G. Liu, C. Zhang, J. Yin, X. Li, F. Cheng, Y. Li, H. Yang, K. Uéda, P. Chan, and S. Yu. **α-Synuclein is differentially expressed in mitochondria from different rat brain regions and dose-dependently down-regulates complex I activity**. *Neuroscience Letters*, **454**(3):187–192, May 2009. 41

[64] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, and P. T. Lansbury. **NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded.** *Biochemistry*, **35**(43):13709–13715, October 1996. 41, 45

[65] D. Eliezer, E. Kutluay, R. Bussell, and G. Browne. **Conformational properties of α-synuclein in its free and lipid-associated states**. *Journal of Molecular Biology*, **307**(4):1061–1073, April 2001. 41, 42

[66] V. N. Uversky, C. J. Oldfield, and A. K. Dunker. **Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept**. *Annual Review of Biophysics*, **37**(1):215–246, June 2008. 41

[67] R. Kruger, W. Kuhn, T. Muller, D. Woitalla, M. Graeber, S. Kosel, H. Przuntek, J. T. Epplen, L. Schols, and O. Riess. **Ala30Pro mutation in the gene encoding α-synuclein in Parkinson's disease**. *Nature Genetics*, **18**(2):106–108, February 1998. 41

[68] J. J. Zarranz, J. Alegre, J. C. Gómez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. n. Atarés et al. **The new mutation, E46K, of α-synuclein causes parkinson and Lewy body dementia**. *Ann Neurol.*, **55**(2):164–173, February 2004. 41

[69] M. H. Polymeropoulos, C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer et al. **Mutation in the α-Synuclein Gene Identified in Families with Parkinson's Disease**. *Science*, **276**(5321):2045–2047, June 1997. 41

[70] E. Muñoz, R. Oliva, V. Obach, M. J. Martí, P. Pastor, F. Ballesta, and E. Tolosa. **Identification of Spanish familial Parkinson's disease and screening for the Ala53Thr mutation of the α-synuclein gene in early onset patients.** *Neuroscience letters*, **235**(1-2):57–60, October 1997. 41

[71] M.-C. C. Chartier-Harlin, J. Kachergus, C. Roumier, V. Mouroux, X. Douay, S. Lincoln, C. Levecque, L. Larvor, J. Andrieux, M. Hulihan et al. **α-synuclein locus duplication as a cause of familial Parkinson's disease.** *Lancet*, **364**(9440):1167–1169, 2004. 41

[72] P. Ibáñez, A. M. Bonnet, B. Débarges, E. Lohmann, F. Tison, Y. Agid, A. Dürr, A. Brice, and P. Pollak. **Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease**. *The Lancet*, **364**(9440):1169–1171, September 2004. 41

[73] A. B. Singleton, M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum et al. M. Baptista, D. Miller, J. Blancato, J. Hardy, and K. Gwinn-Hardy et al. **α-Synuclein locus triplication causes Parkinson's disease.** *Science*, **302**(5646):841+, October 2003. 41

[74] M. G. Spillantini, M. L. Schmidt, V. M. Y. Lee, J. Q. Trojanowski, R. Jakes, and M. Goedert. **α-Synuclein in Lewy bodies**. *Nature*, **388**(6645):839–840, August 1997. 41

[75] M. G. Spillantini, R. A. Crowther, R. Jakes, M. Hasegawa, and M. Goedert. **α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies**. *Proceedings of the National Academy of Sciences*, **95**(11):6469–6473, May 1998. 41

REFERENCES

[76] M. Hashimoto, L. J. Hsu, A. Sisk, Y. Xia, A. Takeda, M. Sundsmo, and E. Masliah. **Human recombinant NACP/α-synuclein is aggregated and fibrillated in vitro: Relevance for Lewy body disease**. *Brain Research*, **799**(2):301–306, July 1998. 41, 69

[77] M. E. v. Raaij, J. v. Gestel, I. M. J. Segers-Nolten, S. W. d. Leeuw, and V. Subramaniam. **Concentration Dependence of α-Synuclein Fibril Length Assessed by Quantitative Atomic Force Microscopy and Statistical-Mechanical Theory**. *Biophysical Journal*, **95**(10):4871–4878, November 2008. 41, 69

[78] K. C. Luk, C. Song, P. O'Brien, A. Stieber, J. R. Branch, K. R. Brunden, J. Q. Trojanowski, and V. M. Y. Lee. **Exogenous α-synuclein fibrils seed the formation of Lewy body-like intracellular inclusions in cultured cells**. *Proceedings of the National Academy of Sciences*, **106**(47):20051–20056, November 2009. 41

[79] H. A. Lashuel, C. R. Overk, A. Oueslati, and E. Masliah. **The many faces of α-synuclein: from structure and toxicity to therapeutic target**. *Nature Reviews Neuroscience*, **14**(1):38–48, January 2013. 41

[80] T. Bartels, J. G. Choi, and D. J. Selkoe. **α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation.** *Nature*, **477**(7362):107–110, September 2011. 41, 42, 65

[81] B. Fauvet, M. K. Mbefo, M.-B. Fares, C. Desobry, S. Michael, M. T. Ardah, E. Tsika, P. Coune, M. Prudent, N. Lion et al. **α-Synuclein in Central Nervous System and from Erythrocytes, Mammalian Cells, and Escherichia coli Exists Predominantly as Disordered Monomer**. *Journal of Biological Chemistry*, **287**(19):15345–15364, May 2012. 41, 42

[82] N. Gould, D. E. Mor, R. Lightfoot, K. Malkus, B. Giasson, and H. Ischiropoulos. **Evidence of Native α-Synuclein Conformers in the Human Brain**. *Journal of Biological Chemistry*, **289**(11):7929–7934, March 2014. 41, 42

[83] S. J. Wood, J. Wypych, S. Steavenson, J.-C. Louis, M. Citron, and A. L. Biere. **α-Synuclein Fibrillogenesis Is Nucleation-dependent**. *Journal of Biological Chemistry*, **274**(28):19509–19512, July 1999. 41

[84] M. D. Shtilerman, T. T. Ding, and P. T. Lansbury. **Molecular Crowding Accelerates Fibrillization of α-Synuclein: Could an Increase in the Cytoplasmic Protein Concentration Induce Parkinson's Disease?**. *Biochemistry*, **41**(12):3855–3860, February 2002. 41, 61, 69

[85] V. N. Uversky, Cooper, K. S. Bower, J. Li, and A. L. Fink. **Accelerated α-synuclein fibrillation in crowded milieu**. *FEBS Letters*, **515**(1-3):99–103, March 2002. 41, 61, 69

[86] W. S. Davidson, A. Jonas, D. F. Clayton, and J. M. George. **Stabilization of α-Synuclein Secondary Structure upon Binding to Synthetic Membranes**. *Journal of Biological Chemistry*, **273**(16):9443–9449, April 1998. 42

[87] P. H. Jensen, M. S. Nielsen, R. Jakes, C. G. Dotti, and M. Goedert. **Binding of α-synuclein to brain vesicles is abolished by familial Parkinson's disease mutation.** *The Journal of biological chemistry*, **273**(41):26292–26294, October 1998. 42

[88] C. C. Jao, A. Der-Sarkissian, J. Chen, and R. Langen. **Structure of membrane-bound α-synuclein studied by site-directed spin labeling**. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(22):8331–8336, June 2004. 42

[89] C. C. Jao, B. G. Hegde, J. Chen, I. S. Haworth, and R. Langen. **Structure of membrane-bound α-synuclein from site-directed spin labeling and computational refinement**. *Proceedings of the National Academy of Sciences*, **105**(50):19666–19671, December 2008. 42

[90] T. S. Ulmer, A. Bax, N. B. Cole, and R. L. Nussbaum. **Structure and Dynamics of Micelle-bound Human α-Synuclein**. *Journal of Biological Chemistry*, **280**(10):9595–9603, March 2005. 42

[91] J. N. Rao, C. C. Jao, B. G. Hegde, R. Langen, and T. S. Ulmer. **A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins**. *J. Am. Chem. Soc.*, **132**(25):8657–8668, June 2010. 42

[92] W. Wang, I. Perovic, J. Chittuluru, A. Kaganovich, L. T. T. Nguyen, J. Liao, J. R. Auclair, D. Johnson, A. Landeru, A. K. Simorellis et al. **A soluble α-synuclein construct forms a dynamic tetramer**. *Proceedings of the National Academy of Sciences*, **108**(43):17797–17802, October 2011. 42

[93] U. Dettmer, A. J. Newman, E. S. Luth, T. Bartels, and D. Selkoe. **In Vivo Cross-linking Reveals Principally Oligomeric Forms of α-Synuclein and β-Synuclein in Neurons and Non-neural Cells**. *Journal of Biological Chemistry*, **288**(9):6371–6385, March 2013. 42

[94] J. Burre, S. Vivona, J. Diao, M. Sharma, A. T. Brunger, and T. C. Sudhof. **Properties of native brain α-synuclein**. *Nature*, **498**(7453):E4–E6, June 2013. 42

[95] T. D. Kim, S. R. Paik, and C.-H. Yang. **Structural and Functional Implications of C-Terminal Regions of α-Synuclein**. *Biochemistry*, **41**(46):13782–13790, October 2002. 45

[96] V. N. Uversky, J. Li, and A. L. Fink. **Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation**. *Journal of Biological Chemistry*, **276**(14):10737–10744, April 2001. 45

[97] B. Nuscher, F. Kamp, T. Mehnert, S. Odoy, C. Haass, P. J. Kahle, and K. Beyer. **α-Synuclein Has a High Affinity for Packing Defects in a Bilayer Membrane**. *Journal of Biological Chemistry*, **279**(21):21966–21975, May 2004. 45

[98] T. Bartels, L. S. Ahlstrom, A. Leftin, F. Kamp, C. Haass, M. F. Brown, and K. Beyer. **The N-Terminus of the Intrinsically Disordered Protein α-Synuclein Triggers Membrane Binding and Helix Folding**. *Biophysical Journal*, **99**(7):2116–2124, October 2010. 45

[99] J. Zhang and M. Muthukumar. **Simulations of nucleation and elongation of amyloid fibrils**. *The Journal of Chemical Physics*, **130**(3):035102+, January 2009. 45

[100] J. N. Onuchic, Z. L. Schulten, and P. G. Wolynes. **THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective**. *Annual Review of Physical Chemistry*, **48**(1):545–600, 1997. 72