JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations

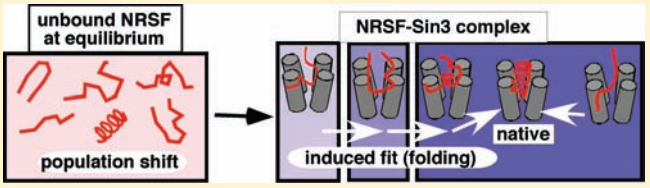
Junichi Higo,*[,†] Yoshifumi Nishimura,[‡] and Haruki Nakamura[†]

[†]Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

[‡]Graduate School of Supramolecular Biology, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku Yokohama 230-0045, Japan

Supporting Information

ABSTRACT: The N-terminal repressor domain of neural restrictive silencer factor (NRSF) is an intrinsically disordered protein (IDP) that binds to the paired amphipathic helix (PAH) domain of mSin3. An NMR experiment revealed that the minimal binding unit of NRSF is a 15-residue segment that adopts a helical structure upon binding to a cleft of mSin3. We computed a free-energy landscape of this system by an enhanced conformational sampling method, all-atom multicanonical molecular dynamics. The simulation started from a configuration where the NRSF segment was fully disordered and distant from mSin3 in explicit solvent. In the absence of mSin3, the disordered NRSF segment thermally fluctuated between hairpins, helices, and bent structures. In the presence of mSin3, the segment bound to mSin3 by adopting the structures involved in the isolated state, and non-native and native complexes were formed. The free-energy landscape comprised three superclusters, and free-energy barriers separated the superclusters. The native complex was located at the center of the lowest free-energy cluster. When NRSF landed in the largest supercluster, the generated non-native complex moved on the landscape to fold into the native complex, by increasing the interfacial hydrophobic contacts and the helix content. When NRSF landed in other superclusters, the non-native complex overcame the free-energy barriers between the various segment orientations in the binding cleft of mSin3. Both population-shift and induced-fit (or induced-folding) mechanisms work cooperatively in the coupled folding and binding. The diverse structural adaptability of NRSF may be related to the hub properties of the IDP.



INTRODUCTION

Protein function is a central subject in protein chemistry. A protein folds into a tertiary structure specific to its amino-acid sequence, and the folded structure provides a scaffold for the protein function. Thus, folding and binding are the main subjects in this research area. An intrinsically disordered protein (IDP) is structurally disordered in an isolated state and adopts a well-defined tertiary structure upon binding to its partner molecules.[1−3] A remarkable feature of IDPs is the indivisible coupling of folding and binding to exert their functions. This mechanism is referred to as "coupled folding and binding",[2] a new scheme for the structure−activity relationship, unlike the well-known "lock and key" mechanism. IDPs exist in a huge variety of eukaryotic genomes,[4,5] and they play important roles in physiological processes, such as cellular signal transduction, protein phosphorylation, molecular assemblies, transcription, and translation regulation.[2,6,7] A single IDP, as a hub protein, can interact with various targets with different biological functions.[8−10]

Neural restrictive silencer factor (NRSF),[11] also known as repressor element 1 (RE1) silencing transcription factor (REST),[12] is an essential transcriptional repressor for neuron-specific genes in non-neuronal cells and neuronal progenitors. NRSF/REST mediates transcriptional repression through the association of its N-terminal repressor domain with the mSin3/histone deacetylase-1/2 (HDAC1/2) complex[13] and directly interacts with TATA-binding protein (TBP).[14] Several neurological diseases, such as Down's syndrome,[15] medulloblastoma,[16,17] Huntington disease,[18] cardiomyopathy,[19] and neuropathic pain,[20] are related to the dysregulation of NRSF/REST and its target genes. Therefore, it is very important to elucidate the detailed interaction modes of NRSF/REST, because an inhibitor of the interaction between NRSF/REST and mSin3/HDAC could be a drug candidate for these severe diseases.

Sin3 contains four paired amphipathic helix (PAH) domains: PAH1−PAH4. Nomura et al. reported that PAH1 of mSin3B (an isoform of mammalian Sin3) binds to the repressor domain of NRSF/REST, and that the minimal repressor domain of NRSF/REST is a short region composed of amino-acid residues 44−55. In addition, they determined the complex structure of mSin3B PAH1 bound to a 15-residue peptide fragment (amino-acid residues 43−57, containing the minimal repressor domain) of NRSF/REST by NMR (PDB IB: 2CZY).[21] The NMR structure revealed that the 15-residue segment adopts a helical structure upon binding to a deep cleft of mSin3B, which is composed of four long helices (H1−H4) of the PAH1 domain

Received: November 18, 2010
Published: May 31, 2011

(Figure 1A). The complex is stabilized by hydrophobic interactions, and the interface lacks hydrogen bonds and salt bridges between the two molecules. The 83-residue N-terminal repressor domain of NRSF/REST containing the 15-residue segment also has the IDP, and the NMR complex structure is a minimal system for investigating the coupled folding and binding of NRSF/REST.[21]

"Fly-casting" has been proposed as a binding mechanism by IDP, based on a theoretical study.[22] The unfolded IDP conformation has a greater interaction radius than a well-structured conformation to catch the target molecule. If we assume that the binding free-energy landscape has a funnel-like enthalpy component, whose bottom corresponds to the native complex form, then the binding kinetics is accelerated considerably. Although some experiments supported the fly casting mechanism,[23−25] no definitive conclusion has been obtained experimentally.[26−28] On the basis of simulation work with a simplified model protein, another binding scheme for IDP was proposed.[29] In this scheme, the unfolded conformation has slower translational diffusion than the folded one, which is a trade-off for the swollen interaction radius. This natural result suggested that fly casting is not the dominant mechanism, but that the large conformational flexibility of IDP in the encounter complex reduces the free-energy barrier between the encounter complex and the final native complex. In other words, the coupled folding and binding belong to an induced-fit (or induced-folding) mechanism.[30,31] Okazaki and Takada[32] studied the coupled folding and binding and compared the induced-fit and population-shift mechanisms,[33−35] using a simplified protein model. In the induced-fit scheme, the folding of IDP occurs in the encounter complex, where the unbound IDP is in the unfolded state. In the population-shift scheme, in contrast, the unbound IDP fluctuates between a major unfolded state and a minor structured state, which resembles the tertiary structure in the native complex. The binding of IDP to its partner molecule occurs using the structured form. They showed that the dominant mechanism (induced-fit or population-shift) is determined by the interaction strength between IDP and its partner.

These computational/theoretical efforts introduced a knowledge-based Hamiltonian or simplified protein/solvent models to study IDPs.[22,29,32] In the knowledge-based Hamiltonian, the intra- and interchain interactions are modulated in advance, to highlight the factors that presumably play an important role in the coupled folding and binding. In the simplified protein/solvent model, the computational cost is considerably reduced. On the other hand, in exchange for these advantages, these studies may sustain disadvantages: The knowledge-based Hamiltonian may skip over some important intermediates in the coupled folding and binding, and the simple expression for the protein/solvent system may introduce inaccuracies in the results. Therefore, an all-atom protein model in explicit solvent would be useful to link the experiments and the simplified models.

A multicanonical algorithm is one of the enhanced conformational sampling methods.[36,37] This algorithm was developed originally to study the statistical behavior of a two-dimensional (2D) Potts model on lattices, using Monte Carlo (MC) sampling.[38] The MC-based method was applied to biological systems[39,40] and extended to molecular dynamics (MD).[41−44] The multicanonical MD (McMD) simulation developed by Nakajima et al.[42] executes the conformational sampling in the Cartesian coordinate space, by which the applicability of sampling was extended readily to a flexible multimolecular system[45]

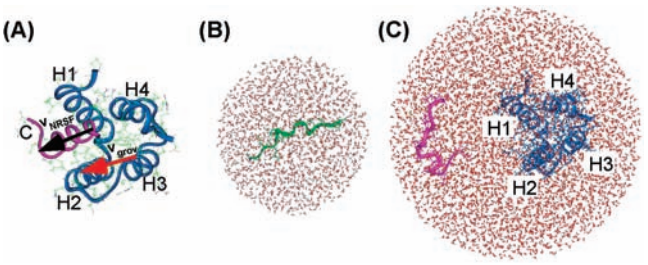


Figure 1. (A) Native complex structure (NMR model 1). Magenta and blue ribbons represent NRSF and Sin3, respectively. The black arrow, ν_{NRSF}, indicates the chain orientation (i.e., helix axis) of NRSF, and the red arrow, ν_{grov}, shows the orientation of the Sin3 groove (i.e., Sin3 binding cleft). See the main text for the strict definitions of ν_{grov} and ν_{NRSF}. The C-terminus of NRSF is labeled "C". (B) Initial conformation of the McMD simulation for the single-chain NRSF system with the solvent sphere. (C) Initial conformation of McMD simulations for the NRSF−Sin3 system with the solvent sphere. In (A) and (C), the four helices of Sin3 are labeled as H1−H4.

and to peptides in explicit solvent.[46,47] So far, we have performed folding McMD simulations of a 40-residue $\alpha+\beta$ protein[48] and a 57-residue protein consisting of two long helices,[49] starting from the fully disordered structures in explicit solvent. An McMD flexible docking of lysozyme and sugar in explicit solvent produced a free-energy landscape, where the lowest free-energy basin, which was separated by free-energy barriers from the other minor basins, was assigned to the native complex structure.[50] To expand the applicability of McMD to a large system, we have recently developed trajectory-parallelization methods.[51]

We have studied the free-energy landscape of the disordered state of a peptide, which was taken from the recognition helix of the DNA-binding protein cMyb, by the all-atom McMD in explicit solvent.[47] A CD measurement confirmed that the peptide is disordered in solution at 300 K. The computed free-energy landscape consisted of various structural clusters characterized by different secondary-structure elements. Similar landscapes were found for other peptides.[52] If the disordered state of an IDP is similar to those of the disordered peptides, then McMD may provide some clues to identify the mechanism of coupled folding and binding.

In the present study, we performed all-atom McMD simulations of the 15-residue segment of NRSF/REST interacting with its partner protein, the PAH1 domain of mSin3B, in explicit solvent. Starting from a fully disordered conformation, which is distant from the PAH1 domain, we obtained the free-energy landscape. The native complex structure was assigned to the most thermodynamically stable cluster. Furthermore, we performed the all-atom McMD of a single-chain NRSF/REST in explicit water starting from a disordered conformation and found that the single chain is disordered in solution. On the basis of these analyses, we propose a mechanism of coupled folding and binding.

METHODS

Although the NMR experiment was performed on an NRSF/REST segment consisting of 20 amino-acid residues (residues 38−57 of NRSF/REST), its N-terminal five residues were not deposited in the PDB, because these residues are exposed to the solvent and are quite flexible in the complex.[21] It is likely that the effect of this undetermined region on the binding is small. Thus, we used the determined 15-residue

segment for the current simulations. We refer to this NRSF/REST segment simply as "NRSF" and to the partner protein, the PAH1 domain of mSin3B, as "Sin3" in this Article. The amino-acid residue numbering is according to the PDB data: residues A31−A107 for Sin3 and B43−B57 for NRSF, where the characters "A" and "B" are the chain identifiers for Sin3 and NRSF, respectively. NRSF is rich in hydrophobic amino-acid residues (sequence: APQLIMLANVALTGE), and the binding site of Sin3 forms a groove (Figure 1A) with walls composed of hydrophobic amino-acid residues. Thus, no hydrogen bonds and salt bridges are formed between NRSF and Sin3 in the NMR complex structure. This suggests that the complex structure is stabilized by hydrophobic interactions.

We performed McMD simulations of two systems: the "single-chain NRSF" and "NRSF−Sin3" systems. To initiate the simulation of the former system, a disordered NRSF structure was immersed in explicit solvent, as described below (Figure 1B). In the NRSF−Sin3 system, NRSF and Sin3 were far apart and immersed in explicit solvent (Figure 1C). Because the same force field was used in both systems, as explained below, the resultant conformational differences of NRSF between the two systems are simply due to the presence/absence of Sin3. The preparation of the systems is described below.

Single-Chain NRSF System. To prepare the initial structure for the single-chain NRSF system, we first generated a random conformation. We then embedded it in a solvent sphere (called sphere 1; diameter = 50 Å) so that the geometrical center of NRSF was set to the center of sphere 1 (Figure 1B). In the embedding, the water molecules overlapping NRSF were removed. The spherical water buffer was configurationally equilibrated in advance at 300 K, with a density of 1 g/cm^3. The size of sphere 1 is sufficient for NRSF to adopt various conformations, such as α-helix, β-hairpin, and others, during the simulation, as described later. Next, we randomly picked a water molecule and replaced it with a chloride ion, to neutralize the net charge (+e) of NRSF. Finally, the system consisted of 6476 atoms (223 peptide atoms, one chloride ion, and 2084 water molecules). We used this configuration for the initial structure of the simulations.

In the simulations, the water molecules and the ion were confined in sphere 1, by applying a harmonic restoring force on the atoms only when they were outside sphere 1. Similarly, to avoid the exposure of NRSF atoms on the surface of sphere 1, we applied another harmonic force on the NRSF heavy atoms only when they were outside a sphere, which is concentric to sphere 1 and with a radius 3 Å smaller than that of sphere 1. Thus, NRSF was not affected by the boundary of sphere 1. Furthermore, the momentum and angular momentum of NRSF were set to zero during the simulation, to maintain NRSF in the middle of sphere 1. Because these momenta were dynamically constrained, no mechanical forces were applied to NRSF.

NRSF−Sin3 System. Sin3 (i.e., the PAH1 domain of mSin3B) is a four-helix bundle protein consisting of 77 amino-acid residues. The four helices, H1−H4, form a groove-shaped cleft to bind to NRSF (Figure 1A). In the complex, NRSF adopts a helical structure, and the helix axis is approximately parallel to the orientation of the groove (see arrows in Figure 1A). Three amino-acid residues at the C-terminal of Sin3 are exposed to the solvent and are distant from the binding site. We removed these three residues to reduce the system size, and thus Sin3 and NRSF consist of 74 (residues A31−A104) and 15 residues, respectively. Next, we prepared a sphere (sphere 2) with a 70 Å diameter, which was filled with water molecules equilibrated in advance at 300 K, with a density of 1 g/cm^3. We embedded the native complex structure (NMR model 1) in sphere 2, where the center of sphere 2 was set at the geometrical center of NRSF. We then removed NRSF and placed a random structure of NRSF at a position in sphere 2, where it was distant from Sin3 (Figure 1C). We removed the water molecules that overlapped with Sin3 and the relocated NRSF, and then to neutralize the system, we randomly replaced seven water molecules with three chloride and four sodium ions. The ionic concentration was approximately 1 mM. The final generated structure, consisting of 17 705 atoms (1194 atoms for Sin3; 223 for NRSF, 3 Cl^-, 4 Na^+, and 5427 water molecules), was used for the initial structure of the simulations. A solvent sphere that is too small may cause a serious artifact in the simulation results. As shown later, however, NRSF was spread throughout the current solvent sphere during the simulation. Thus, we concluded that the spherical boundary did not cause a serious artifact in the simulation results.

We confined the water molecules and ions in sphere 2 by applying a harmonic restoring force on the atoms only when they were outside sphere 2. Similarly, to avoid the exposure of NRSF on the surface of sphere 2, we applied another harmonic force on the NRSF heavy atoms only when they were outside a sphere, which is concentric to sphere 2 and with a radius 3 Å smaller than that of sphere 2. To maintain the tertiary structure of Sin3, we applied weak restraints on the intra-Sin3 Cα−Cα distances (see the section "Inter-Cα Atomic Restraints" in the Supporting Information for details). In the absence of these restraints, Sin3 unfolds during the McMD simulations, which prevents the NRSF−Sin3 docking in the computer simulation. However, these restraints were weak, and the Sin3 groove was able to open and close during the simulations (see the section "Non-restrained Canonical MD of Single Sin3" and Figure S6A in the Supporting Information). The momentum and angular momentum of Sin3 were set to zero during the simulation, so Sin3 was not exposed on the surface of sphere 2.

Multicanonical Simulations. The McMD method[42] is an enhanced conformational sampling method.[36,37] Here, we will briefly explain McMD (see the section "McMD and TTP-McMD" in the Supporting Information for details). McMD can explore a broad conformational space in a wide temperature range. The benefit of McMD is that a thermodynamic conformational ensemble $Q(T)$ at any temperature T is reconstructed from the sampled conformations, by using a reweighting technique. We designate the ensemble from the single-chain NRSF system as $Q_{sng1}(T)$ and that from the NRSF−Sin3 system as $Q_{N-S}(T)$. The numbers of conformations stored in $Q_{sng1}(300\ K)$ and $Q_{N-S}(300\ K)$ were 1582 and 3611, respectively.

TTP-McMD is a technique to increase the sampling efficiency of McMD, in which a number of multiple McMD runs are performed from different initial conformations.[51] The obtained trajectories are simply integrated. We performed 64 and 512 multiple runs for the single-chain NRSF and NRSF−Sin3 systems, respectively. The initial conformations of these multiple runs originated from the conformation in Figure 1B for the single-chain NRSF system and that in Figure 1C for the NRSF−Sin3 system (see the Supporting Information for details). In this report, the term "McMD simulation" means "TTP-McMD simulation", because TTP-McMD is a derivative of McMD.

We used the computer program PRESTO version 3[53] for the McMD simulations (the time step was 1 fs; SHAKE[54] was employed to constrain the covalent bonds between heavy atoms and hydrogen atoms; the cell-multipole expansion[55] was used to compute the long-range electrostatic interactions; and a constant-temperature method[56] was utilized to control the temperature). The force field parameters for the polypeptides are those from an AMBER-based hybrid force field,[57] $E(w) = (1-w)E_{94} + wE_{96}$, where E_{94} and E_{96} are the AMBER parm94[58] and parm96[59] force fields, respectively, and w is the mixture rate. We previously assessed the physicochemical validity of $E(w)$, by comparing the free-energy landscapes of short peptides between McMD simulations and quantum chemical calculations, and found that a w value of 0.75 is optimal.[57] Furthermore, McMD simulations with $E(0.75)$ revealed that a peptide with a helical propensity folds into a helix, while a peptide with a β-hairpin propensity forms a β-hairpin.[57] Thus, we used $E(0.75)$ for the current study. We used a TIP3P water model[60] for the water molecules.

Conformational Distribution: Free-Energy Landscape. The conformations in $Q_{sng1}(T)$ or $Q_{N-S}(T)$ are projected in a conformational

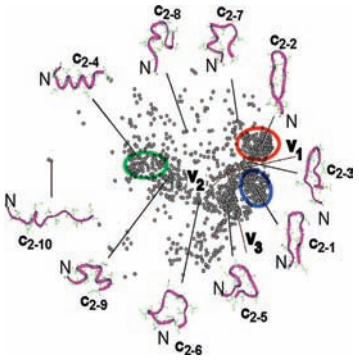


Figure 2. Conformational distribution of $Q_{sngl}(300\ K)$ for the single-chain NRSF system. The first three major principal component axes v_1, v_2, and v_3 construct the 3DPC subspace, and each dot is a projection of an NRSF structure on the subspace. The displayed tertiary structures (c_{2-1}−c_{2-10}) are taken from the positions indicated by the lines. The three colored circles are mentioned in the main text. "N" indicates the N-terminus of NRSF.

space, to provide a free-energy landscape. The conformational space is generated by a principal component analysis (PCA).[46,47,50,52,57,61−64] We will briefly explain the construction of the conformational space (see the section "PCA" in the Supporting Information for details).

First, we expressed the protein conformation by the Cα−Cα atomic distances (eqs S13 and S14 for $Q_{sng1}(T)$ and $Q_{N-S}(T)$, respectively) and calculated the variance−covariance matrix (eq S9) from each of the $Q_{sng1}(T)$ and $Q_{N-S}(T)$ ensembles. By diagonalizing the matrix, we obtained the eigenvectors and eigenvalues. We generated a three-dimensional (3D) conformational space by the first three eigenvectors, $v_1(T)$, $v_2(T)$, and $v_3(T)$, which are assigned to the largest, second largest, and third largest eigenvalues, respectively. Thus, the conformational space is an abstract space, where $v_1(T)$, $v_2(T)$, and $v_3(T)$ correspond to the *x*-, *y*-, and *z*-axes, respectively. Generally, in PCA, the conformational space constructed by the first three eigenvectors most effectively discriminates the conformational differences among the protein conformations (see the Supporting Information for details). We refer to the eigenvectors and the 3D space as "PC axes" and "3DPC subspace", respectively. Finally, we projected the conformations of $Q_{sng1}(T)$ or $Q_{N-S}(T)$ on the 3DPC subspace (see eq S10 of the Supporting Information) to produce a conformational distribution. Two conformations that are close to each other in the 3DPC subspace are structurally similar. A low free energy is assigned to the regions where the conformations are distributed densely.

The full-dimensional space constructed by all of the eigenvectors is a high-dimensional space. Although the full-dimensional expression is a perfect description of the conformational space, the high dimensionality interferes with understanding the conformational distribution intuitively. This is why the low-dimensional expression is used in the PCA study. To compensate for the drawback of the low-dimensional expression, we used a cluster analysis, in which the protein conformational differences are defined in the full-dimensional space.

RESULTS

Flat Energy Distribution. The McMD simulations generated a flat energy distribution, $P_{mc}(E,T_0)$, covering temperature ranges of 285−700 and 295−700 K for the single-chain NRSF and NRSF−Sin3 systems, respectively (Figure S1). This flatness guarantees that the McMD simulations sampled a sufficiently wide conformational space for both systems (see eq S7 in the Supporting Information). Figure S1 also demonstrates the canonical energy

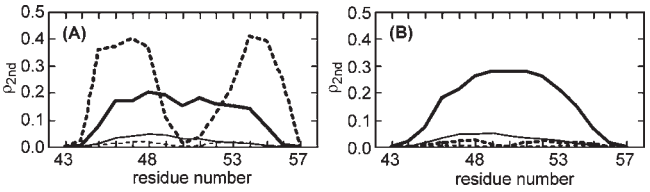


Figure 3. Secondary-structure content rates (ρ_{2nd}) at each residue site of NRSF from $Q_{sngl}(T)$ (A) and $Q_{N-S}(T)$ (B). The residue numbering corresponds to that of the original PDB file[21] for chain B (i.e., NRSF). Solid and dashed lines represent the rates for α-helix and β-strand, respectively. Thick and thin lines are the content rates at 300 and 600 K, respectively.

distributions $P_c(E,300\ K)$ and $P_c(E,700\ K)$, which are the energy distributions at 300 and 700 K, respectively (see eq S3 in the Supporting Information). The generated ensembles $Q_{sng1}(T)$ and $Q_{N-S}(T)$ were used for the analyses below.

Single-Chain NRSF. We computed the PC axes from the ensemble $Q_{sng1}(300\ K)$ of the single-chain NRSF system and constructed the 3DPC subspace. The contribution ratios from v_1, v_2, and v_3 to the entire structural variety were 46.6%, 17.8%, and 9.9%, respectively (see eq S11). Thus, the total contribution from the first three PC axes was 74.3% (see eq S12).

Figure 2 demonstrates the 3D conformational distribution of $Q_{sng1}(300\ K)$, where each dot corresponds to an NRSF conformation. This figure is a 2D projection of the 3D distribution. A 3D version (stereo view) of Figure 2 is presented in Figure S2. Figure 2 also displays some tertiary structures (labeled c_{2-1}−c_{2-10}) picked from the distribution. We found various structures: β-hairpins (c_{2-1}−c_{2-3}), α-helices (c_{2-4}), bent structures (c_{2-5}−c_{2-9}), and extended conformations (c_{2-10}). The β-hairpins c_{2-1} and c_{2-2} have different hydrogen-bond patterns between the strands. These β-hairpins are concentrated in the blue and red circles of Figure 2. The structure c_{2-3} is a distorted β-hairpin with hydrogen-bond patterns similar to those of c_{2-1}. The helices are concentrated in the green circle. Some bent structures (see c_{2-7}−c_{2-9} for instance) involved a short helical turn, although the position of the turn was irregular in the sequence. Extended conformations were rarely involved in $Q_{sng1}(300\ K)$. Because $Q_{sng1}(300\ K)$ is composed of thermodynamically probable structures at 300 K (see the section "McMD and TTP-McMD" of the Supporting Information), Figure 2 shows that the single-chain NRSF is intrinsically disordered in solution at 300 K and lacks a predominant structure. This result conforms to the experimental observation that the N-terminal repressor domain of NRSF/REST is an IDP.[21] Later, we will compare Q_{sng1} (300 K) with $Q_{N-S}(300\ K)$ from the NRSF−Sin3 system.

In this Article, we display many tertiary structures and have labeled them as c_{i-j}, where *i* specifies the figure number and *j* is the running number in the figure. Thus, for instance, c_{2-5} (the fifth structure in Figure 2) and c_{5A-5} (the fifth one in Figure 5A) are different structures.

We computed the content rates for the α and β secondary structures in $Q_{sng1}(T)$ at each residue site by the computer program DSSP,[65] which analyzes hydrogen-bond patterns in a given tertiary structure and assigns a secondary structure to each residue. Figure 3A shows the content rates at 300 and 600 K. Two β-strand regions (residues B45−B48 and B53−B56) are identified at 300 K, where hydrogen bonds bridge the strands to form β hairpins. Later, we will show that the β structure almost vanishes in the presence of Sin3. Figure 3A also shows that

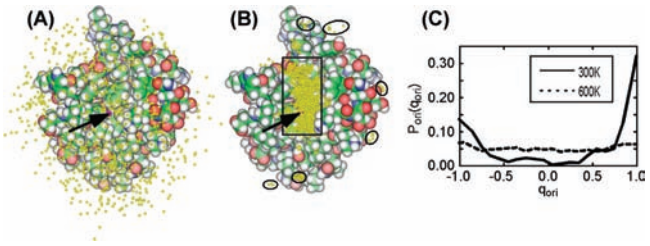


Figure 4. Spatial distributions of the mass center of NRSF around Sin3 at 600 K (A) and 300 K (B). Yellow dots represent the mass-center positions of NRSF. Arrows indicate the mass-center position of NRSF in the native complex structure (NMR model 1). The rectangle in panel B represents the Sin3 groove. The circles are mentioned in the main text. (C) Distribution function $P_{ori}(q_{ori})$ at 300 and 600 K. See the main text for the definition of q_{ori}.

residues B46−B54 have a helix rate above 10%. Therefore, we found again that the single-chain NRSF is in a mixture of secondary structures and that NRSF is thermally disordered, as a whole. At 600 K, the secondary structures are drastically diminished.

NRSF−Sin3 System and Two Free-Energy Barriers. Figure 4 demonstrates the mass-center of NRSF distributed around Sin3. At 600 K, NRSF was spread widely throughout the solvent sphere (Figure 4A). This figure shows that the volume of the sphere is large enough to randomize the NRSF conformation. In contrast, at 300 K NRSF converged in the Sin3 groove (rectangle in Figure 4B), although some NRSF conformations were out of the groove (see the dots in the circles in Figure 4B).

We discriminated the sampled structures by the orientation of NRSF relative to the Sin3 groove. To quantify the relative orientation, we first introduced two inter-Cα atomic vectors v_{NRSF} and v_{grov}: v_{NRSF} is from residue B46 (the fourth residue of NRSF) to residue B53 (the 11th one) of NRSF, and v_{grov} is from residue A75 to residue A60 of Sin3. In the NMR complex structure, v_{NRSF} and v_{grov} are almost parallel (see the black and red arrows in Figure 1A). Next, we defined the unit vectors e_{NRSF} and e_{grov} that are parallel to v_{NRSF} and v_{grov}, respectively. Finally, the relative orientation is defined as $q_{ori} = e_{NRSF} \cdot e_{grov}$, which ranges from $-1 \le q_{ori} \le +1$, because e_{NRSF} and e_{grov} are unit vectors. In a complex with $q_{ori} \approx 1.0$, NRSF is parallel to the Sin3 groove, and the complex is called the "parallel NRSF−Sin3 complex" (or simply the "parallel complex"). In fact, q_{ori} is 0.97 for the NMR structure (NMR model 1). In contrast, in a complex with $q_{ori} \approx -1.0$, NRSF is antiparallel to the groove, and the complex is called the "anti-parallel NRSF−Sin3 complex" (or the "anti-parallel complex"). Figure 4C shows plots of the distribution function $P_{ori}(q_{ori})$ at 300 and 600 K. If the relative orientation is random, then the distribution should be flat, as shown at 600 K. In contrast, at 300 K, $P_{ori}(q_{ori})$ has remarkable peaks at $q_{ori} = \pm 1$ and is small in the intermediate range of $-0.7 < q_{ori} < 0.7$. Thus, the parallel and antiparallel NRSF−Sin3 complexes have free energies lower than those of the other complexes; that is, the parallel and antiparallel complexes are thermodynamically more stable than the others. The highest free energy (i.e., the minimum of $P_{ori}(q_{ori})$) was assigned at $q_{ori} = 0.15$. Theoretically, the free energy at q_{ori} measured from the highest free energy is given as: $\Delta G(q_{ori}) = -RT \ln[P_{ori}(q_{ori})/P_{ori}(0.15)]$. The parallel NRSF−Sin3 complex is the lowest free-energy state ($\Delta G(+1) = -6.4$ kcal/mol), and the antiparallel one is the second lowest ($\Delta G(-1) = -1.9$ kcal/mol). Thus, the parallel complex is 4.5 kcal/mol more stable than the antiparallel one: $\Delta\Delta G = \Delta G(+1) - \Delta G(-1) = -4.5$ kcal/mol.

The lowering of $P_{ori}(q_{ori})$ in $-0.7 < q_{ori} < 0.7$ implies the existence of a free-energy barrier somewhere in this range of q_{ori}. However, we note a drawback of the quantity q_{ori} in specifying the free-energy barrier, because a variety of structures can have a single value of q_{ori}, or the same NRSF structures can have different q_{ori} values when the structures are oriented differently relative to the Sin3 groove. Thus, q_{ori} has low resolution to differentiate complex structures. Kamiya et al. have shown that a meaningful free-energy barrier is passed over when the conformational space is constructed by ambiguous quantities, such as solvent-accessible surface area, radius of gyration, and structural root-mean-square deviation.[66]

To avoid the oversight, we generated the 3DPC subspace, where the conformations of Q_{N-S}(300 K) are projected. Figure 5A shows the conformational distribution in the 3DPC subspace, where the sampled conformations (dots) are colored depending on q_{ori}. An overview of the distribution shows two volumes with crowded dots, labeled SC1 and SC2, and a smaller volume, labeled SC3. In the subsequent subsection, we perform a cluster analysis on the conformations of Q_{N-S}(300 K) and show that each volume consist of some clusters. Thus, we call these volumes "superclusters". The supercluster SC3 is isolated from SC1 and SC2 by two low-density trenches, which are highlighted by the black dashed lines labeled B1 and B2 in Figure 5A. Theoretically, a high free energy is assigned to a low-density region, where the dots are sparsely distributed. Thus, the low-density trenches found in Figure 5A are free-energy barriers. Apparently, the yellow dots (i.e., $-0.7 < q_{ori} < 0.7$) are distributed around the barriers B1 and B2. In Figure 4C, the existence of one or more free-energy barriers was implied. Figure 5A substantially illustrates that there are two free-energy barriers in the conformational space. A 3D version (stereo view) of Figure 5A is presented in Figure S3.

The contribution ratios ($R_{PCA}(k)$ defined by eq S11 in the Supporting Information) for the three major PC axes v_1, v_2, and v_3 were 22.4%, 19.0%, and 9.6%, respectively, and the total contribution (R_{PCA}^{1-3} defined by eq S12 in the Supporting Information) from the three was 51.0%. One may consider this ratio to be too small for identifying the free-energy barriers. However, we note that, in general, a free-energy barrier detected in a low-dimensional PC subspace remains in the full-dimensional PC space. In fact, we will later show that the free-energy barriers remain in the full-dimensional space.

Next, we show the sampled structures in SC1 that involve the native complex. Figure 5B illustrates the sampled conformations, colored depending on the number, N_{HB}, of intra-NRSF helical hydrogen bonds. The figure is shown from a viewpoint where the native complex is at the front. The helix-rich structures of NRSF ($N_{HB} \ge 5$) surround the native complex; for instance, see the three structures labeled c_{5B-1}−c_{5B-3}. Structures with $3 \le N_{HB} \le 4$ are distributed sparsely not only in SC1 but also in SC3, and those with $1 \le N_{HB} \le 2$ are spread entirely over the distribution. Figure 5B also shows that bent structures, c_{5B-4}−c_{5B-7}, lie near the free-energy barrier B2, where the bent portion of NRSF fits into the Sin3 groove. The structure c_{5B-7} involves a short helical turn. In contrast, the extended structures c_{5B-8} and c_{5B-9} are farther away from B2 in the 3DPC subspace. The extended structures may involve a short helix, as exemplified in c_{5B-8} and c_{5B-9}. Figure 6A−C illustrates the helix-rich NRSF structures binding to the Sin3 groove, where the N_{HB} values are

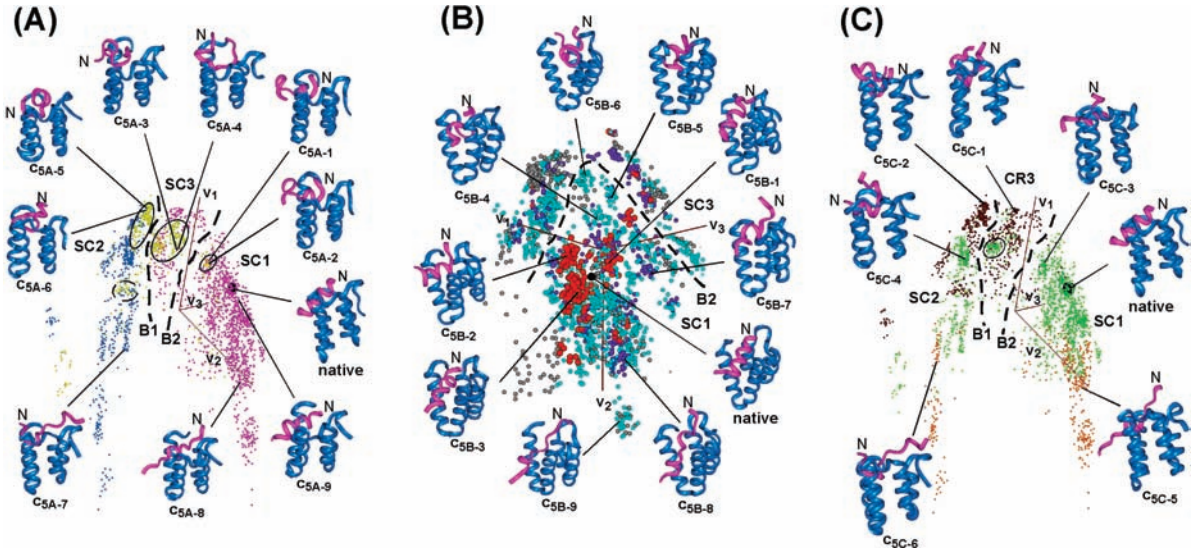


Figure 5. Conformational distributions for the NRSF−Sin3 system at 300 K. Dots are projections of structures of Q_{N-S}(300 K) on the 3DPC subspace constructed by the first three major PC axes v_1, v_2, and v_3. The black sphere corresponds to the native complex (NMR model 1). Black dashed lines highlight the free-energy barriers (labeled B1 and B2) separating superclusters SC1, SC2, and SC3. Circled regions are mentioned in the main text. Tertiary structures picked from the distributions are displayed, where "N" indicates the N-terminus of NRSF. The native complex (NMR model 1) is labeled "native". (A) Dots are colored depending on q_{ori}. Magenta, cyan, and yellow dots represent parallel NRSF−Sin3 complexes (i.e., $q_{ori} \geq 0.7$), antiparallel NRSF−Sin3 complexes (i.e., $q_{ori} \leq -0.7$), and the other complexes with intermediate orientations (i.e., $0.7 > q_{ori} > -0.7$), respectively. (B) Dots are colored depending on the number of intra-NRSF helical hydrogen bonds, N_{HB}. A hydrogen bond is defined as one formed between the carbonyl oxygen atom of residue i and the amide nitrogen atom of residue $i + 4$ in NRSF. Red, purple, and light cyan dots are conformations with $N_{HB} \geq 5$, $3 \leq N_{HB} \leq 4$ and $1 \leq N_{HB} \leq 2$, respectively. Gray dots are those with no helical hydrogen bonds ($N_{HB} = 0$). Barrier B1 and supercluster SC2 are behind the front dot distribution. (C) Dots are colored depending on d_{3-13}. See the main text for the definition of d_{3-13}. The view is from the same direction as panel A. Dark brown, light green, and ocher dots are conformations with $d_{3-13} < 12$ Å, 12 Å $\leq d_{3-13} \leq 18$ Å, and $d_{3-13} > 18$ Å, respectively.

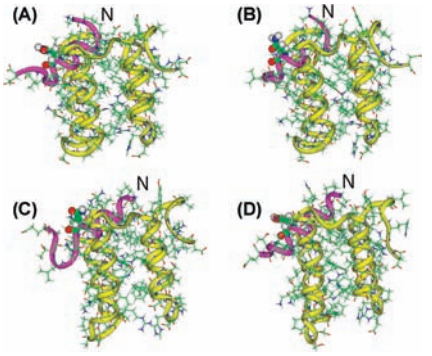


Figure 6. (A−C) Helix-rich NRSF structures (magenta) in the NRSF−Sin3 complex and (D) the native complex (NMR model 1). The amino-acid Asn B51 is represented by a ball-and-stick model. The figure is drawn so that helices H2 (left) and H3 (right) of Sin3 are in the front. "N" indicates the N-terminus of NRSF.

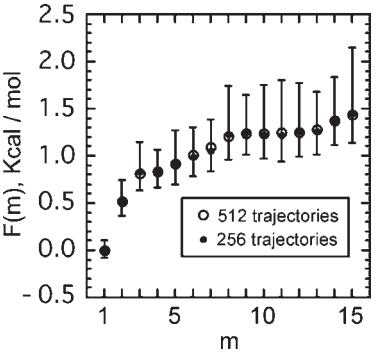


Figure 7. Free-energy $F(m)$ at 300 K assigned to the mth cluster. The "○" is computed from the full trajectories (512 trajectories), and the "●" is from 256 trajectories. The error bars are explained in the main text.

8, 7, and 6, respectively. Figure 6D is the native complex. A hydrophilic amino acid, Asn B51, is represented by a ball-and-stick model to show the solvent-exposed surface of NRSF. This residue is exposed to the solvent, as in the native complex, and no hydrogen bonds were formed between NRSF and Sin3.

To analyze the opening/closing motions of the Sin3 groove, we performed additional canonical MD simulations of a single Sin3 at 300 K, without the inter-Cα atomic distance restraints (see "Non-restrained Canonical MD of Single Sin3" in the Supporting Information). Remember that we introduced the restraints to prevent Sin3 from unfolding during McMD. We found that the groove opening/closing motions from Q_{N-S} (300 K) were compatible to those from the nonrestrained canonical MD at 300 K (see the Supporting Information for details). Thus, the inter-Cα atomic distance restraints were weak enough to allow the groove motions.

Clusters. The structure analysis in the 3DPC subspace may skip over important information, which is detectable in a full-dimensional space. We defined the structural similarity among the conformations in the full-dimensional space and classified the sampled conformations of Q_{N-S}(300 K) into clusters with structural similarity (see the section "Cluster Analysis" in the Supporting Information for details). We obtained 57 clusters and arranged them in the descending order of the number of

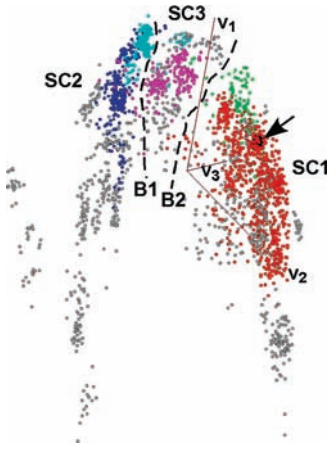


Figure 8. The first five clusters in 3DPC subspace. The first cluster, that is, the lowest free-energy cluster, is shown in red. The second, third, fourth, and fifth ones are colored blue, magenta, green, and cyan, respectively. The colors used in this figure have no relation to the same colors in other figures. The black sphere with the arrow corresponds to the native complex (NMR model 1).

conformations involved in the clusters. We refer to the content rate of cluster m to the entire ensemble $Q_{N-S}(300\ K)$ as $R_{occ}(m)$ (see the Supporting Information for the detailed definition of $R_{occ}(m)$). The resultant values of $R_{occ}(m)$ from the first to fifth clusters were 21.4%, 8.8%, 5.4%, 5.2%, and 4.6%, respectively. Thus, these five clusters occupied 45.4% of $Q_{N-S}(300\ K)$. The free energy of cluster m, measured from the lowest free energy, is formally defined as: $F(m) = -RT \ln[R_{occ}(m)/R_{occ}(1)]$. Figure 7 shows a plot of the cluster dependence of the free energy.

To assess the convergence of the free energy of the clusters obtained from the full (512) trajectories, we randomly picked 256 trajectories from the full trajectories and recalculated the content rate, $R_{occ}^{256}(m)$, from the 256 trajectories, where $R_{occ}^{256}(m)$ involves structures that are sampled in the 256 trajectories and are involved in cluster m defined for the full trajectories. We then recomputed the free energy of the clusters: $F(m) = -RT \ln[R_{occ}^{256}(m)] + c_{set}$, where c_{set} is a constant defined as $c_{set} = RT \ln[R_{occ}^{512}(1)]$ and $R_{occ}^{512}(1)$ is the content rate for the largest cluster calculated from the full trajectories. The term c_{set} was introduced to set $F(1)$ around zero ($F(1) \approx 0$). We repeated this procedure 300 times for different sets of 256 trajectories and calculated the standard deviation (error bar in Figure 7) for each cluster. Importantly, the lowest free-energy cluster (i.e., the first cluster) was maintained over the 300 sets, and the error bar for the first cluster did not overlap with those for the other clusters. Thus, the lowest free-energy cluster is reproducible in the current study. Remember that the multiple trajectories are independent of one another (see "McMD and TTP-McMD" in the Supporting Information). The order ranks for the other clusters may change, because the error bars for those clusters overlapped.

Figure 8 demonstrates the positions of the first five clusters in the 3DPC subspace. The first cluster dominated SC1, and the native complex was located at the center of this cluster. The second, third, fourth, and fifth clusters were involved in SC2, SC3, SC1, and SC2, respectively. There were no clusters that crossed the free-energy barriers B1 and B2. Thus, the barriers, which were identified in the 3DPC subspace, substantially differentiate SC1, SC2, and SC3 in the full-dimensional space. It may appear, for instance, that a few red dots (first cluster) are spattered in SC3. However, these dots are not embedded in SC3, but are popping up from SC3. Figure 8 indicates that SC1, SC2, and SC3 are superclusters consisting of clusters: The first cluster involves a large fraction of SC1, and minor clusters surround the first cluster. In contrast, there were no dominant clusters in SC2 and SC3.

Discriminators for the Free-Energy Barriers. We have shown that the free-energy barriers B1 and B2 separate the superclusters SC1−SC3. Here, we show the structural characteristics specifying the barriers. Referring to Figure 5A, the parallel NRSF−Sin3 complexes ($q_{ori} \geq 0.7$) dominated SC1 and SC3, while the antiparallel ones ($q_{ori} \leq -0.7$) dominated SC2. A closer look at the regions around the free-energy barriers reveals zones (circled regions in Figure 5A) where the intermediate-orientation complexes (i.e., $-0.7 < q_{ori} < 0.7$) are distributed. Even so, it is clear that SC1 and SC3 involve a few antiparallel complexes (cyan dots) and that SC2 involves a few parallel ones (magenta dots). Thus, the discriminator for the barrier B1 is the orientation q_{ori}. On the other hand, the discriminator for the barrier B2 is ambiguous from q_{ori}, because both SC1 and SC3 considerably involve the parallel NRSF−Sin3 complexes. This ambiguity remained even when we changed the coloring thresholds for q_{ori} in Figure 5A. We will characterize B2 by a different structural property at the end of this subsection.

As shown above, differently colored dots coexist in the circles in Figure 5A. To analyze this coexistence, we looked at the pairs of differently colored dots that are close in the 3DPC subspace. The pairing structures c_{5A-1} (a magenta dot) and c_{5A-2} (yellow) had q_{ori} values of 0.72 and 0.50, respectively, with an rmsd = 0.71 Å. Here, the rmsd is the mutual root-mean-square difference between the two NRSF structures, calculated using the main-chain heavy atoms of residues B46−B53. Remember that the Cα atomic positions of residues B46 and B53 defined q_{ori}. The structures c_{5A-3} (magenta) and c_{5A-4} (yellow) had q_{ori} values of 0.71 and 0.65, respectively, with an rmsd = 3.27 Å; and c_{5A-5} (cyan) and c_{5A-6} (yellow) had q_{ori} values of −0.71 and −0.41, respectively, with an rmsd = 3.46 Å. In common, these NRSF structures were bent. It is likely that the bent NRSF structures alter the relative orientation q_{ori} via small conformational changes in the Sin3 groove. In contrast, the tertiary structures c_{5A-7} and c_{5A-8} are extended in the Sin3 groove. Those extended conformations probably cannot significantly change q_{ori} without extensive structural motions.

In Figure 5A, the conformations with intermediate orientations (yellow dots) are concentrated in the dashed-line circle in SC2, although antiparallel conformations (cyan dots) dominated SC2. However, these yellow dots are well isolated from the cyan dots in SC2, when we rotate the viewpoint in the 3DPC subspace. Thus, the yellow dots are on a dead end in the free-energy landscape, and thus the direct transition from these yellow dots to cluster SC3, by passing through the free-energy barrier B1, is negligibly small. These yellow dots are not important to characterize the barrier B1.

Figure 5A and B showed that the bent NRSF structures are distributed near the free-energy barriers B1 and B2. In contrast, the extended NRSF structures were distant from the barriers. To relate this structural property of NRSF to the free-energy landscape, we introduced the quantity d_{3-13}: the Cα-atomic distance between residues B45 (the third residue of NRSF) and B55 (the 13th). An NRSF structure with a small d_{3-13} is possibly bent, whereas one with a large d_{3-13} is extended. In Figure 5C, the dots are colored depending on d_{3-13}. The bent structures

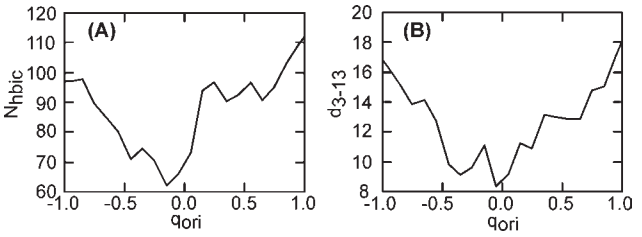


Figure 9. Distributions of the number (N_{hbic}) of hydrophobic atomic contacts at the NRSF−Sin3 interface (A) and d_{3-13} (B) as a function of q_{ori} at 300 K. See the text for the definition of q_{ori}.

(d_{3-13} < 12 Å) are clearly distributed around the free-energy barriers, as well as in the tertiary structures c_{5C-1} and c_{5C-2}. The native complex structure, with a distance d_{3-13}^{native} of 15.5 Å, belongs to regions for 12 ≤ d_{3-13} ≤ 18 Å, where the two tertiary structures c_{5C-3} and c_{5C-4} are exemplified. The extended structures (d_{3-13} > 18 Å) are distributed in regions distant from the free-energy barriers, as also seen in structures c_{5C-5} and c_{5C-6}. Both c_{5C-5} and c_{5C-6} are involved a short helical turn fitting into the Sin3 groove, which is a common structural feature for the extended NRSF structure, as shown in Figure 5B.

We emphasize that the fringe of supercluster SC1, facing the barrier B2, is occupied by tertiary structures with 12 ≤ d_{3-13} ≤ 18 Å (Figure 5C). In contrast, the majority of SC3 consists of conformations with d_{3-13} < 12 Å. Therefore, we conclude that the discriminator to identify B2 is d_{3-13}. Exceptionally, conformations with 12 ≤ d_{3-13} ≤ 18 Å densely populate the circle in SC3. However, these structures were well isolated from the other dots in SC3, when the viewpoint was rotated in the 3DPC subspace. Thus, these structures are on a dead end in the free-energy landscape and do not characterize the barrier B2. In summary, q_{ori} and d_{3-13} characterize the barriers B1 and B2, respectively.

Stability of the Native Complex Structure. Figure 5B showed that the complexes surrounding the native complex are native-like. Here we propose a question: What is the factor stabilizing the native-complex structure? Remember that no hydrogen bonds and salt bridges are formed at the NRSF−Sin3 interface in the NMR structure.[21] This suggests that the interface is stabilized by the hydrophobic contacts. Meanwhile, the intra-NRSF hydrogen bonds facilitate the helix formation. In the isolated state of NRSF, the α-helix is one of the semistable states (Figure 2). In the presence of Sin3, the β hairpin vanished and the α helix became the predominant secondary structure (Figure 3B). In the sampled native-like complexes, NRSF had an interface similar to that of the Sin3 grooves (Figure 6), which facilitates the hydrophobic contacts. These results suggest that both the interfacial hydrophobic contacts and the intra-NRSF helical hydrogen bonds stabilize the native-complex structure concertedly.

Figure 9A shows a plot of the number (N_{hbic}) of hydrophobic atomic contacts at the NRSF−Sin3 interface as a function of q_{ori}. See also the section "Hydrophobic Atoms" in the Supporting Information. The quantity N_{hbic} is largest at the parallel orientation (q_{ori} = −1). This result is reasonable because the parallel NRSF fits well in the Sin3 groove, as realized in the native complex. The antiparallel orientation (q_{ori} = −1) also had a relatively large N_{hbic} value. However, the N_{hbic} value was comparable to those in 0.2 < q_{ori} < 0.7. This is one reason why the antiparallel NRSF is less stable than the parallel NRSF. The

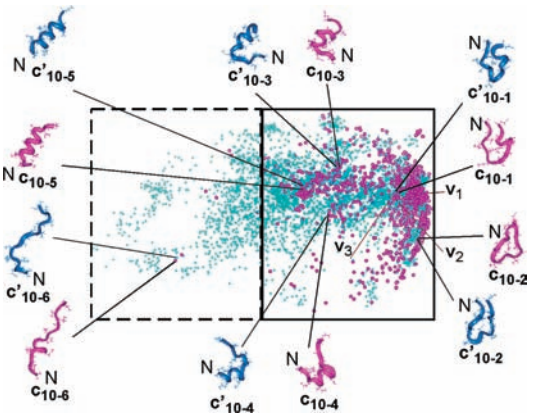


Figure 10. Distribution of NRSF structures from the NRSF−Sin3 system (cyan dots) and the single-chain NRSF system (magenta dots). The three major PC axes, v_1, v_2, and v_3, are those from Figure 2. Tertiary structures in magenta (labeled as c_{10-i}) are from the single-chain NRSF system, and those in cyan (labeled as c'_{10-j}) are from the NRSF−Sin3 system. Solid- and dashed-line rectangles are mentioned in the main text. This figure is presented from a viewpoint slightly different from that in Figure 2. "N" indicates the N-terminus of NRSF.

N_{hbic} value was small for −0.5 ≤ q_{ori} ≤ 0.1. This result is also understandable, because the bent NRSF structure provides fewer contacts to the Sin3 groove than do the parallel and antiparallel ones. Figure 9B shows a plot of the relationship between q_{ori} and d_{3-13}, which correlates well with the relationship between q_{ori} and N_{hbic} (Figure 9A). Therefore, as NRSF deviates from the parallel or antiparallel orientation, the NRSF structure bends and the interfacial hydrophobic contacts decrease.

Furthermore, we analyzed the relationship between q_{ori} and the number of intra-NRSF hydrogen bonds, N_{HB}. See "Relationship between q_{ori} and the Number of Intra-NRSF Hydrogen Bonds" in the Supporting Information. Figure S8B indicates that the intra-NRSF hydrogen bonds provide more stabilization of the parallel complex than the antiparallel complex. This result again supports the proposal that the parallel NRSF−Sin3 complex is stabilized by both the interfacial hydrophobic contacts and the intra-NRSF hydrogen bonds.

Comparison between the Single-Chain and NRSF−Sin3 Systems. In the last part of the Results, we compare the NRSF conformations between the single-chain NRSF and NRSF−Sin3 systems. The ensemble Q_{sng1}(300 K) involves only NRSF, while Q_{N-S}(300 K) involves both NRSF and Sin3. Thus, for the comparison, we discarded Sin3 from Q_{N-S}(300 K) and generated a new ensemble, Q'_{N-S}(300 K), which only consists of NRSF. Figure 10 shows a projection of the NRSF structures of Q'_{N-S}(300 K) on the distribution of Figure 2.

Figure 10 displays some pairs of structures that are close to each other in the 3DPC subspace, but taken from different systems. Apparently, the pairing structures are similar to each other. A majority of the structures from Q_{sngl}(300 K) are distributed in the solid-line rectangle in Figure 10. This rectangle involves bent structures (c_{10-1} and c'_{10-1}; c_{10-2} and c'_{10-2}; c_{10-3} and c'_{10-3}; c_{10-4} and c'_{10-4}) and helices (c_{10-5} and c'_{10-5}). Structures c_{10-3} and c'_{10-3} as well as c_{10-4} and c'_{10-4} involve a short helix element. As shown previously in structures c_{5A-1}, c_{5A-2}, c_{5B-7}, c_{5C-3}, and c_{5C-4}, the short helix fits in the Sin3 groove. Figure 3 indicated that NRSF does not adopt a regular β hairpin in the presence of Sin3, although the single-chain NRSF

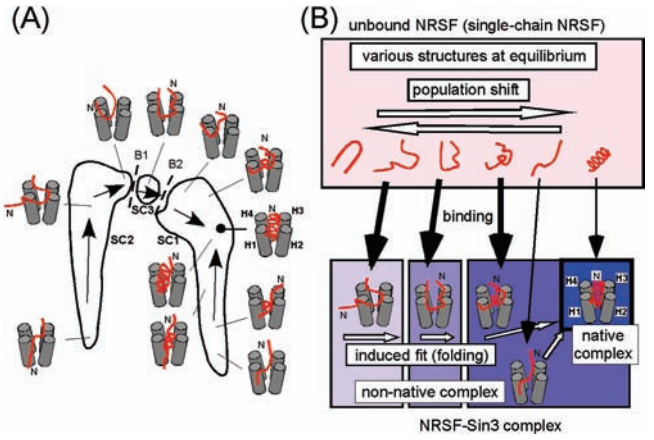


Figure 11. (A) Schematic overview of the conformational distributions for the NRSF−Sin3 system in 3DPC subspace at 300 K. The free-energy barriers B1 and B2 discriminate superclusters SC1, SC2, and SC3. Arrows indicate conformational changes from the non-native to native complex. Some tertiary structures are schematically shown, where the cylinders represent the four helices (H1−H4) and the red solid line represents NRSF. "N" indicates the N-terminus of NRSF. The "●" is the native complex. (B) Schematic overview of coupled folding and binding, proposed from the current study. White arrows represent conformational motions in each of the single-chain NRSF and NRSF−Sin3 systems. Black arrows represent NRSF binding to Sin3. The arrow thickness represents the transition probability assigned to the binding.

had an inherent propensity for β hairpins. It is likely that the regular hairpin experiences steric hindrances from the walls of the Sin3 groove. Structures c'_{10-1} and c'_{10-2} may be distorted hairpins, because a few interstrand hydrogen bonds are formed near the termini of NRSF.

The NRSF structures in the dashed-line rectangle in Figure 10 are extended, and most of them originated from the NRSF−Sin3 system. The single-chain NRSF rarely adopted extended structures (Figure 2). The extended NRSF structures involve fewer intra-NRSF hydrogen bonds. Therefore, the extended NRSF structures are stabilized only in the presence of Sin3, where they can contact Sin3.

■ DISCUSSION

The single-chain NRSF ensemble $Q_{sngl}(300\ K)$ consists of a variety of structures, including α and β structures as well as bent structures, and NRSF thermally fluctuates among them (Figure 2). Thus, the single-chain NRSF is disordered in solution at room temperature. However, $Q_{sngl}(300\ K)$ is different from a plain ensemble composed of random conformations. This result is consistent with our previous studies,[47] where the polypeptide fluctuated thermally among a variety of ordered structures. The current study has shown that the parallel ($q_{ori} \approx +1$) and antiparallel ($q_{ori} \approx -1$) NRSF−Sin3 complexes are more stable than the other complexes and that the parallel complex gains more interfacial hydrophobic contacts (Figure 9A) and more intra-NRSF hydrogen bonds (Figure S8B) than does the antiparallel one. Consequently, the parallel complex is most stable in $Q_{N-S}(300\ K)$. These natural results prompted us to analyze the biophysical quantities that are not experimentally obtainable.

The most important quantity in the current work is the free-energy landscape (i.e., the distribution of conformations in the 3DPC subspace). The free-energy landscape for the NRSF−Sin3 system consisted of non-native complexes as well as the native-like complex (Figure 5). The structures from the single-chain NRSF system couple with those from the NRSF−Sin3 system (Figure 10). Thus, the various single-chain NRSF structures are adaptable to the Sin3 groove. Because the helix is not the most stable structure in the single-chain NRSF system (Figures 2 and 3A), it is statistically rare for the native complex to be formed in the first approach to the Sin3 groove. We conclude that multiple encounter complexes (not a single encounter complex) are probably formed prior to the native complex. Figure 11A shows a schematic free-energy landscape of the NRSF−Sin3 system, summarized from the current study.

The arrows in Figure 11A indicate the conformational changes from the non-native to native complexes. When NRSF lands in a position within supercluster SC1, the orientation q_{ori} of NRSF is parallel to the Sin3 groove in the first generated non-native complex. Next, the NRSF structure moves to the native complex, without overcoming a free-energy barrier. The conformational changes from the non-native to native complex depend on the landing site in SC1. An extended NRSF structure, which rarely emerges during thermal fluctuations of the single-chain NRSF, can bind to the Sin3 groove via the interfacial hydrophobic contacts and reach the native complex with increasing N_{HB} and decreasing d_{3-13}. When a bent structure, which frequently emerges in the single-chain NRSF, is caught in the Sin3 groove, the structure moves toward the native complex with increasing N_{hbic}, N_{HB}, and d_{3-13}.

When NRSF lands in a position within SC3, NRSF is bent in the first generated non-native complex, as shown Figure 5A. The free-energy barrier B2 is then overcome by increasing d_{3-13}, as shown in Figure 5C. After crossing to SC1, the non-native complex reaches the native form, as described above.

When NRSF lands in a position within SC2, the antiparallel complex is generated. NRSF may reside in SC2 for awhile, because SC2 is a large supercluster in the free-energy landscape. However, because the antiparallel complex is less stable than the parallel complex, the complex finally becomes parallel. While ascending B1, NRSF alters q_{ori} largely, and a non-native parallel complex is formed. The non-native complex then overrides the second barrier B2 with increasing d_{3-13}. Of course, we do not exclude the possibility of NRSF dissociation from the Sin3 groove during this process. This point is discussed later.

One may presume that Figure S8B suggests another free-energy barrier than B1 and B2 between $N_{HB} = 2$ and $N_{HB} = 8$. However, this barrier is false, as shown in the section "Relationship between q_{ori} and the Number of Intra-NRSF Hydrogen Bonds" in the Supporting Information. A free-energy landscape expressed by a quantity with less structural resolution misreads an artificial free-energy barrier, as pointed out previously.[66]

The single-chain NRSF provides various structures adaptable to the Sin3 groove in the thermal fluctuations (Figure 10). This suggests a population-shift mechanism.[33−35] In parallel, Figure 5 indicates the conformational changes of NRSF in the bound state. This supports the induced-fit (or induced-folding) mechanism.[29−31] Consequently, the current McMD simulation proposes an integrated mechanism, where the population-shift and the induced-fit work sequentially and cooperatively to enhance the complex formation, unlike other studies[29,32] where the population-shift and the induced-fit are alternatively discussed to explain the coupled folding and binding. We summarize this cooperative mechanism in Figure 11B. We emphasize that our conclusion is derived naturally from the all-atom protein

model in explicit solvent, which is free from a knowledge-based bias (a knowledge-based Hamiltonian).

The large conformational flexibility of NRSF increases the interaction radius in both the single-chain and the complex states. Thus, the current results are consistent with the fly casting mechanism. However, the current simulation has been done on the fishhook (the 15-residue segment), by removing the fishing line (the entire disordered N-terminal repressor domain of NRSF/REST). We consider that not only the fishhook but also the fishing line contribute to the increase in the interaction radius. A more extensive all-atom analysis should be performed to judge the validity of the fly casting mechanism for the current system.

The structural diversity in the single-chain NRSF potentially provides wide adaptability to various receptors. In particular, helices (or extended conformations) and hairpins can result in completely different complex forms. This structural diversity may be a reason for the functional diversity (i.e., hub property) of NRSF/REST.[67]

Two quantities, q_{ori} and d_{3-13}, played important roles in characterizing the free-energy barriers B1 and B2, and the barriers were identified in the 3DPC subspace. Thus, one may expect that there is a relationship between the PC axes (v_1, v_2, and v_3) and either q_{ori} or d_{3-13}. For instance, the end-to-end distance (a quantity similar to d_{3-13}) correlates well with the motions along v_1 when the PCA axes are computed only from the coordinates of a single polypeptide.[68] However, we found no clear correlation in the NRSF−Sin3 system. This is because the PC axes for the NRSF−Sin3 system were computed using both the intra-NRSF and the inter-NRSF−Sin3 coordinates (see eq S14). This mixture of coordinates was important to specify equally the NRSF structure and the NRSF−Sin3 mutual positioning in the conformational space. In preliminary work, we examined several coordinate sets for PCA and finally reached the current set.

Do the motions of NRSF correlate with the Sin3 groove opening/closing? This question is important for understanding the kinetics of coupled folding and binding for the current system. However, we could find no correlation in Q_{N-S} (300 K). This correlation should be observed in precise moments of NRSF−Sin3 binding at 300 K. In the NRSF−Sin3 system, most of the unbound NRSF conformations and association/dissociation events were sampled at a high temperature. This is why McMD has high sampling efficiency. On the other hand, this advantage of McMD shifts the low-probability events (i.e., the association/dissociation events at 300 K) away from Q_{N-S}(300 K). With reference to some previous work,[69,70] we have examined the free-energy landscape at 300 K on a plane of the number of intra-NRSF hydrogen bonds (N_{HB}) and the number of inter-residue contacts between NRSF and Sin3 (N_{cont}) (see the section "Free-energy Landscape with Different Structural Measures" in the Supporting Information). Although Figure S9 supports the currently proposed mechanism in Figure 11B, we note that the free energy at N_{cont} = 0 is presented inaccurately. Therefore, we compared the NRSF−Sin3 system and the single-chain NRSF system to study the coupled folding and binding.

To detect the precise association/dissociation moments at 300 K, the solvent volume can be increased so that the single-chain NRSF and the bound NRSF coexist at 300 K. However, this increment is not practical because of the rapid increase in the computation time. The precise moments may be studied by incorporating another computational technique that focuses on the kinetics of complex formation. For instance, a study computed flux flows along the induced-fit and population-shift pathways.[71] Because this approach is based on a reaction equation, the binding kinetics may be determined, if the rate constants are relevantly introduced among some reaction steps. Canonical MD simulations at 300 K, which start from some configurations picked from McMD simulations, may be also useful. The incorporation of these approaches may highlight the association/dissociation kinetics.

■ ASSOCIATED CONTENT

Supporting Information. Inter-Cα atomic restraints; McMD and TTP-McMD; PCA; Figure S1; Figure S2; Figure S3; nonrestrained canonical MD of single Sin3; cluster analysis; hydrophobic atoms; relation between q_{ori} and the number of intra-NRSF hydrogen bonds; free-energy landscape with different structural measures; and complete ref 19. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
higo@protein.osaka-u.ac.jp

■ ACKNOWLEDGMENT

H.N. was supported by Grant-in-Aids for Scientific Research (B) (20370061) from the Japan Society for the Promotion of Science, and for Scientific Research on Priority Areas "Structures of Biological Macromolecular Assemblies" (513-20051013) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. J.H. was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (21113006) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. J.H. and H.N. were supported by grants from the New Energy and Industrial Technology Development Organization (NEDO), Japan. Y.N. was supported by a Grant-in-Aid for Scientific Research (S) (20227009) and a target proteins research program grant (o7067019) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Computations were partly supported by the Human Genome Center, Institute of Medical Science, The University of Tokyo, Japan.

■ REFERENCES

(1) Wright, P. E.; Dyson, H. J. *J. Mol. Biol.* **1999**, *293*, 321–331.
(2) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(3) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.
(4) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. *J. Mol. Biol.* **2004**, *337*, 635–645.
(5) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Brown, C. J.; Uversky, V. N.; Dunker, A. K. *Biochemistry* **2005**, *44*, 1989–2000.
(6) Fuxreiter, M.; Tompa, P.; Simon, I.; Uversky, V. N.; Hansen, J. C.; Asturias, F. J. *Nat. Chem. Biol.* **2008**, *4*, 728–737.
(7) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. *J. Mol. Biol.* **2002**, *323*, 573–584.
(8) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.* **2005**, *272*, 5129–5148.

(9) Patil, A.; Kinoshita, K.; Nakamura, H. *Int. J. Mol. Sci.* **2010**, *11*, 1930–1943.
(10) Patil, A.; Kinoshita, K.; Nakamura, H. *Protein Sci.* **2010**, *19*, 1461–1468.
(11) Schoenherr, C. J.; Anderson, D. J. *Science* **1995**, *267*, 1360–1363.
(12) Chong, J. A.; Tapia-Ramirez, J.; Kim, S.; Toledo-Aral, J. J.; Zheng, Y.; Boutros, M. C.; Altshuller, Y. M.; Frohman, M. A.; Kraner, S. D.; Mandel, G. *Cell* **1995**, *80*, 949–957.
(13) Naruse, Y.; Aoki, T.; Kojima, T.; Mori, N. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13691–13696.
(14) Murai, K.; Naruse, Y.; Shaul, Y.; Agata, Y.; Mori, N. *Nucleic Acids Res.* **2004**, *32*, 3180–3189.
(15) Bahn, S.; Mimmack, M.; Ryan, M.; Caldwell, M. A.; Jauniaux, E.; Starkey, M.; Svendsen, C. N.; Emson, P. *Lancet* **2002**, *359*, 310–315.
(16) Fuller, G. N.; Su, X.; Price, R. E.; Cohen, Z. R.; Lang, F. F.; Sawaya, R.; Majumder, S. *Mol. Cancer Ther.* **2005**, *4*, 343–349.
(17) Lawinger, P.; Venugopal, R.; Guo, Z.-S.; Immaneni, A.; Sengupta, D.; Lu, W.; Rastelli, L.; Carneiro, A. M. D.; Levin, V.; Fuller, G. N.; Echelard, Y.; Majumder, S. *Nat. Med.* **2000**, *6*, 826–831.
(18) Zuccato, C.; Tartari, M.; Crotti, A.; Goffredo, D.; Valenza, M.; Conti, L.; Cataudella, T.; Leavitt, B. R.; Hayden, M. R.; Timmusk, T.; Rigamontio, D.; Cattaneo, E. *Nat. Genet.* **2003**, *35*, 76–83.
(19) Kuwahara; et al. *EMBO J.* **2003**, *22*, 6310–6321.
(20) Uchida, H.; Ma, L.; Ueda, H. *J. Neurosci.* **2010**, *30*, 4806–4814.
(21) Nomura, M.; Uda-Tochio, H.; Murai, K.; Mori, N.; Nishimura, Y. *J. Mol. Biol.* **2005**, *354*, 903–915.
(22) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
(23) Vamvaca, K.; Jelesarov, I.; Hilvert, D. *J. Mol. Biol.* **2008**, *382*, 971–977.
(24) Lengyel, C. S. E.; Willis, L. J.; Mann, P.; Baker, D.; Kortemme, T.; Strong, R. K.; McFarland, B. J. *J. Biol. Chem.* **2007**, *282*, 30658–30666.
(25) Hoffman, R. M. B.; Blumenschein, T. M. A.; Sykes, B. D. *J. Mol. Biol.* **2006**, *361*, 625–633.
(26) Narayanan, R.; Ganesh, O. K.; Edison, A. S.; Hagen, S. J. *J. Am. Chem. Soc.* **2008**, *130*, 11477–11485.
(27) Crespin, M. O.; Boys, B. L.; Konermann, L. *FEBS Lett.* **2005**, *579*, 271–274.
(28) Onitsuka, M.; Kamikubo, H.; Yamazaki, Y.; Kataoka, M. *Proteins* **2008**, *72*, 837–847.
(29) Huang, Y.; Liu, Z. *J. Mol. Biol.* **2009**, *393*, 1143–1159.
(30) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88–118.
(31) Spolar, R. S.; Record, M. T., Jr. *Science* **1994**, *263*, 777–784.
(32) Okazaki, K.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11182–11187.
(33) Bosshard, H. R. *News Physiol. Sci.* **2001**, *16*, 171–173.
(34) James, L. C.; Tawfik, D. S. *Trends Biochem. Sci.* **2003**, *28*, 361–368.
(35) Yamane, T.; Okamura, H.; Nishimura, Y.; Kidera, A.; Ikeguchi, M. *J. Am. Chem. Soc.* **2010**, *132*, 12653–12659.
(36) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers* **2001**, *60*, 96–123.
(37) Iba, Y.; Chikenji, G.; Kikuchi, M. *J. Phys. Soc. Jpn.* **1998**, *67*, 3327–3330.
(38) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.
(39) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333–1338.
(40) Kidera, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9886–9889.
(41) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321–330.
(42) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817–824.
(43) Terada, T.; Matsuo, Y.; Kidera, A. *J. Chem. Phys.* **2003**, *118*, 4306–4311.
(44) Hori, N.; Chikenji, G.; Berry, R.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 73–78.
(45) Nakajima, N.; Higo, J.; Kiedra, A.; Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297–301.
(46) Higo, J.; Galzitskaya, O. V.; Ono, S.; Nakamura, H. *Chem. Phys. Lett.* **2001**, *337*, 169–175.
(47) Higo, J.; Ito, N.; Kuroda, M.; Ono, S.; Nakajima, N.; Nakamura, H. *Protein Sci.* **2001**, *10*, 1160–1171.
(48) Ikebe, J.; Kamiya, N.; Shindo, H.; Nakamura, H.; Higo, J. *Chem. Phys. Lett.* **2007**, *443*, 364–368.
(49) Ikebe, J.; Standley, D. M.; Nakamura, H.; Higo, J. *Protein Sci.* **2011**, *20*, 187–196.
(50) Kamiya, N.; Yonezawa, Y.; Nakamura, H.; Higo, J. *Proteins* **2008**, *70*, 41–53.
(51) Ikebe, J.; Umezawa, K.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Takano, Y.; Nakamura, H.; Higo, J. *J. Comput. Chem.* **2011**, *32*, 1286–1297.
(52) Yagisawa, R.; Kamiya, N.; Ikebe, J.; Umezawa, K.; Higo, J. *Chem. Phys. Lett.* **2008**, *455*, 293–296.
(53) Morikami, K.; Nakai, T.; Kidera, A.; Saito, M.; Nakamura, H. *Comput. Chem.* **1992**, *16*, 243–248.
(54) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(55) Ding, H.-Q.; Karasawa, N.; Goddard, W. A., III. *J. Chem. Phys.* **1992**, *97*, 4309–4315.
(56) Evans, D. J.; Morriss, G. P. *Phys. Lett. A* **1983**, *98*, 433–436.
(57) Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. *Chem. Phys. Lett.* **2005**, *401*, 312–317.
(58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(59) Kollman, P. A.; Dixon, R. W.; Cornell, W. D.; Chipot, C.; Pohorille, A. In *Computer Simulations of Biological Systems*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer/ESCOM: The Netherlands, 1997; p 83.
(60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(61) Kamiya, N.; Mitomo, D.; Shea, J.-E.; Higo, J. *J. Phys. Chem. B* **2007**, *111*, 5351–5356.
(62) Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005**, *58*, 45–52.
(63) Spiwok, V.; Lipovová, P.; Králová, B. *J. Phys. Chem. B* **2007**, *111*, 3073–3076.
(64) Sutto, L.; D'Abramo, M.; Gervasio, F. L. *J. Chem. Theory Comput.* **2010**, *6*, 3640–3646.
(65) Brahms, S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149–178.
(66) Kamiya, N.; Higo, J.; Nakamura, H. *Protein Sci.* **2002**, *11*, 2297–2307.
(67) Bruce, A. W.; Donaldson, I. J.; Wood, I. C.; Yerbury, S. A.; Sadowski, M. I.; Chapman, M.; Gottgens, B.; Buckley, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10458–10463.
(68) Ikeda, K.; Tomii, K.; Yokomizo, T.; Mitomo, D.; Maruyama, K.; Suzuki, S.; Higo, J. *Protein Sci.* **2005**, *14*, 1253–1265.
(69) Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 5214–5223.
(70) Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 2088–2089.
(71) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 13737–13741.