Supporting Information

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

Junichi Higo,[a,*] *Yoshifumi Nishimura,*[b] *and Haruki Nakamura*[a]

[a]Institute for Protein Research, Osaka University, Suita, Osaka, 565-0871, Japan, and [b]Graduate School of Supramolecular Biology, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku Yokohama 230-0045, Japan

Inter-Cα Atomic Restraints.

We applied the inter-Cα atomic distance restraints, $E_{\rm res}$, only to Sin3. The function form is:

$$E_{\rm res}=\begin{cases} 0.5\sum_{i,j}[r_{i,j}-(r^0_{i,j}-r_{\rm low})]^2 & (for \quad r_{i,j}\le r^0_{i,j}-r_{\rm low}) \\ 0 & (for \quad r^0_{i,j}-r_{\rm low}>r_{i,j}>r^0_{i,j}+r_{\rm high}) \\ 0.5\sum_{i,j}[r_{i,j}-(r^0_{i,j}+r_{\rm high})]^2 & (for \quad r_{i,j}\ge r^0_{i,j}+r_{\rm high}) \end{cases} \qquad \text{(S1)}$$

where $r_{i,j}$ and $r^0_{i,j}$ are the Cα atomic distance between residues *i* and *j* in a simulation snapshot and the native complex structure (NMR model 1), respectively, and $r_{\rm low}$ and $r_{\rm high}$ are set to 1.5 Å and 3.5 Å, respectively. Thus, no restraint ($E_{\rm rst}=0$) is applied to an atom pair when $r^0_{i,j}-r_{\rm low}<r_{i,j}<r^0_{i,j}+r_{\rm high}$. The atom pairs with $|i-j|\le 3$ were eliminated from the summation. Furthermore, the summation is taken over the Cα atomic pairs with distances $r^0_{i,j}$ smaller than 7.0 Å. Thus, Sin3 exhibits large-scale conformational fluctuations during the McMD simulation, as shown later.

McMD and TTP-McMD.

The conformational sampling of two systems, the single-chain NRSF and NRSF-Sin3 systems, was achieved by the multicanonical molecular dynamics (McMD) simulation.[1] In McMD, a modified potential energy, E_{mc}, is introduced as:

$$E_{\mathrm{mc}}\left(E\right) = RT_0 \ln[n(E)] = E + RT_0 \ln P_{\mathrm{c}}\left(E, T_0\right), \tag{S2}$$

where E is the original potential energy, $n(E)$ is the density of states of the system, R is the gas constant, and T_0 is the simulation temperature. The function $P_{\mathrm{c}}(E, T_0)$ is a canonical energy distribution at T_0, defined formally as:

$$P_{\mathrm{c}}\left(E, T_0\right) = \frac{n\left(E\right)}{Z_{\mathrm{c}}\left(T_0\right)} \exp\left[-\frac{E}{RT_0}\right], \tag{S3}$$

where $Z_{\mathrm{c}}(T_0)$ is the partition function: $Z_{\mathrm{c}}\left(T_0\right) = \int n\left(E\right) \exp\left[-E/RT_0\right] dE$. McMD is a canonical MD simulation at T_0 using E_{mc} to derive forces acting on atoms: $force = -\nabla E_{\mathrm{mc}}$. This simulation provides an energy distribution, $P_{\mathrm{mc}}(E, T_0)$, formally derived as:

$$P_{\mathrm{mc}}\left(E, T_0\right) = \frac{n\left(E\right)}{Z_{\mathrm{mc}}\left(T_0\right)} \exp\left[-\frac{E_{\mathrm{mc}}}{RT_0}\right], \tag{S4}$$

where $Z_{\mathrm{mc}}(T_0)$ is a partition function for the modified potential E_{mc}, defined as: $Z_{\mathrm{mc}}\left(T_0\right) = \int n\left(E\right) \exp\left[-E_{\mathrm{mc}}/RT_0\right] dE$. Equation S4 is transformed using Equation S2 as:

$$P_{\mathrm{mc}}(E, T_0) = \frac{n(E)}{Z_{\mathrm{mc}}(T_0)} \frac{1}{n(E)} = const. \tag{S5}$$

Since the temperature T_0 is constant during the simulation, the partition function $Z_{\mathrm{mc}}(T_0)$ is also constant. Thus, we do not need to calculate the absolute value of $Z_{\mathrm{mc}}(T_0)$. Instead, we simply regard $Z_{\mathrm{mc}}(T_0)$ as a constant to normalize $P_{\mathrm{mc}}(E, T_0)$.

Since the density of states $n(E)$ is unknown *a priori*, iterative simulations are required where $n(E)$ converges to an accurate function. A tentative density of state $n_t(E)$ obtained from an iterative simulation is used to define E_{mc} for the next iteration as:

$$E_{mc}(E) = RT_0 \ln[n_t(E)]. \quad \text{(S6)}$$

Then, the simulation generates an energy distribution:

$$P_{mc}(E,T_0) = \frac{n(E)}{n_t(E)} Z_{mc}(T_0)^{-1}. \quad \text{(S7)}$$

The deviation of $P_{mc}(E,T_0)$ from flatness is directly related to the accuracy of the estimated density of states, i.e., the flatness ensures $n_t(E) \approx n(E)$. Therefore, the iteration is continued until $P_{mc}(E,T_0)$ is flat enough in the desired energy range. We usually set the simulation temperature T_0 to a high value (700 K for the current study). The desired energy range is set from T_0 to T_{room}, where T_{room} is a temperature slightly lower than room temperature. The conformation readily overcomes energy barriers in the conformational space during the simulation, because T_0 is 700 K.

After obtaining the flat distribution for $P_{mc}(E,T_0)$, we performed a production run using the converged density of states to obtain a conformational ensemble for analyses. Note that the canonical energy distribution $P_c(E,T)$ at an arbitrary temperature *T* is derived from Equation S3, by replacing T_0 with *T*. A conformational ensemble $Q(T)$ is constructed by assigning the probability $P_c(E,T)$ to each conformation sampled in the production run. This procedure is called a reweighting technique.

To increase the sampling efficiency, we achieved trivial trajectory parallelization of the multicanonical molecular dynamics (TTP-McMD).[2] In this method, we perform multiple McMD runs starting from different initial conformations, and simply connect the multiple trajectories. The generated long trajectory can be regarded as a single multicanonical trajectory, because the connection of the multiple trajectories automatically satisfies a detailed balance.[3]

We performed 64 and 512 multiple McMD runs for the single-chain NRSF and NRSF-Sin3 systems, respectively. The actual procedure of TTP-McMD is as follows: In advance, we performed 64 or 512 high-temperature (1,000 K) canonical MD runs. The initial conformations for the high-temperature simulations are shown in Figures 1B and 1C of the main text, and each run was performed for 10 ns. Different initial atomic velocities were assigned to these multiple runs, to randomize the conformations of NRSF. The structure of Sin3 in the NRSF-Sin3 system was maintained well around the NMR structure

even at 1,000 K, because of the distance restraints mentioned above. Then, the first McMD iteration was started from those randomized conformations, and all of the multiple trajectories were integrated into one to compute a tentative E_{mc} (see Equation S2 or S6). In the second iteration, all of the multiple runs were executed using the same E_{mc} computed from the first iteration, but with different initial conformations: The initial conformation for the *j*-th run of the second iteration is the last snapshot of the *j*-th run of the first iteration, and so on. In other words, E_{mc} computed from the *i*-th iteration was used for all multiple runs of the (*i* + 1)-th iteration, and the last snapshot of the *j*-th run of the *i*-th iteration was used for the initial conformation for the *j*-th run of the (*i* + 1)-th iteration. The iterations were executed until a flat energy distribution $P_{\mathrm{mc}}(E, T_0)$ was obtained in the desired energy range. The potential E_{mc} and the last snapshots from the final iteration were used for the production runs. The 64 or 512 trajectories from the production runs are integrated to generate a conformational ensemble for analyses. We repeated 14 and 44 iterations for the single-NRSF and NRSF-Sin3 systems, respectively. The integrated trajectory lengths for the production runs were 0.64×10^9 steps (0.64 μs) for the single-NRSF system and 1.17×10^9 steps (1.17 μs) for the NRSF-Sin3 system. However, the time lengths are not equivalent to those from conventional canonical MD simulations, because the conformational motions are enhanced in TTP-McMD.

PCA.

We used principal component analysis (PCA) to analyze the distribution of polypeptides in an abstract conformational space. Imagine a polypeptide conformation specified by N general coordinates. We express the coordinates as a vector $\boldsymbol{q}$:

$$\boldsymbol{q} = [q_1, q_2, ..., q_N], \tag{S8}$$

where q_i is the *i*-th coordinate in an *N*-dimensional space. To obtain coordinate axes (PC axes) for constructing the conformational space, the PCA is processed as follows: First, we calculate a variance-covariance matrix *C* from the conformational ensemble $\mathrm{Q}(T)$ as:

$$C_{mn} = < q_m q_n >_T - < q_m >_T < q_n >_T, \tag{S9}$$

where C_{mn} is the matrix element (*m*,*n*), and $<...>_T$ is the ensemble average over the conformations in $\mathrm{Q}(T)$. We obtained a set of eigenvectors and eigenvalues by diagonalizing *C*. We arranged the eigenvectors in descending order of eigenvalues. The *k*-th eigenvector and eigenvalue are denoted as

$\boldsymbol{v}_k(T)$ and $\lambda_k(T)$, respectively, which satisfy equations $C\boldsymbol{v}_k = \lambda_k(T)\boldsymbol{v}_k$ and $\boldsymbol{v}_i(T)\cdot\boldsymbol{v}_j(T) = \delta_{ij}$. We use the eigenvectors $[\boldsymbol{v}_1,\boldsymbol{v}_2,...,\boldsymbol{v}_N]$ for the coordinate axes of the *N*-dimensional conformational space, where the coordinates $[x_1, x_2,..., x_N]$ of a conformation $\boldsymbol{q}$ are given by a projection:

$$x_i = \boldsymbol{v}_i(T)\cdot(\boldsymbol{q} - <\boldsymbol{q}>_T). \tag{S10}$$

We refer to the conformational space constructed by the PC axes as "PC space". Projecting the conformations in $\mathrm{Q}(T)$ on the *N*-dimensional PC space, we generate a conformational distribution, which is an image of a free-energy landscape. Furthermore, we can project any polypeptide conformation, which may not be a member of $\mathrm{Q}(T)$, on the PC space by using Equation S10. When the conformations in $\mathrm{Q}(T)$ are projected on an eigenvector, a one-dimensional distribution is generated. The magnitude of the standard deviation (SD) of the one-dimensional distribution depends on the selected eigenvector. As a general nature of PCA, the larger the eigenvalue assigned to an eigenvector, the larger the SD on the eigenvector. Thus, the first eigenvalue can be a coordinate axis that most effectively discriminates the structural variety of the conformations in $\mathrm{Q}(T)$. Similarly, the second and third eigenvectors can be effective coordinate axes to discriminate the structural variety. To express how the structural variety is served by an eigenvector, a contribution ratio is used. The ratio of the *k*-th eigenvector is given by:

$$R_{\mathrm{PCA}}(k) = \lambda_k(T) / \sum_i^{\mathrm{all}} \lambda_i(T). \tag{S11}$$

The contribution ratio from the first three eigenvectors, $\boldsymbol{v}_1(300\mathrm{K})$, $\boldsymbol{v}_2(300\mathrm{K})$ and $\boldsymbol{v}_3(300\mathrm{K})$, is then given by:

$$R_{\mathrm{PCA}}^{1-3} = R_{\mathrm{PCA}}(1) + R_{\mathrm{PCA}}(2) + R_{\mathrm{PCA}}(3). \tag{S12}$$

Up to this point, q_i can be a general coordinate. Below, for the single-chain NRSF system, we use an intra-NRSF Cα-Cα atomic distance, d_{ij}, between residues *i* and *j* to express the coordinate. Here, the residue number of NRSF starts from 1 (i.e., $i, j = 1,...,15$), unlike the original residue number in the PDB. Then, we obtained a 105-dimensional vector to describe an NRSF conformation as:

$$\boldsymbol{q} = [d_{12}, d_{13}, ..., d_{1415}] = [q_1, q_2, ..., q_{105}]. \quad \text{(S13)}$$

For the NRSF-Sin3 system, we expressed a conformation $\boldsymbol{q}$ by a 225-dimensional vector as:

$$\boldsymbol{q} = [q_1, ..., q_{105}; q_{106}, ..., q_{225}], \quad \text{(S14)}$$

where the former 105 elements (i.e., $[q_1, ..., q_{105}]$) are the intra-NRSF Cα atomic distances used for the single-chain NRSF system. The latter 120 elements (i.e., $[q_{106}, ..., q_{225}]$) are the inter-Cα atomic distances between NRSF and eight amino-acid residues of Sin3: $15 \times 8 = 120$. The eight residues are A34 (Val), A37 (Ala), A59 (Leu), A62 (Met), A72 (Thr), Val75 (Val), A96 (Phe), and A97 (Leu). These residues are located in the walls of the Sin3 groove in the NMR complex. We selected two residues from each helix to maintain the balance as: A34 and A37 from H1, A59 and A62 from H2, A72 and A75 from H3, and A96 and A97 from H4.

To visualize the distribution of Q(300K), we picked the three major PC axes, $\boldsymbol{v}_1(300\text{K})$, $\boldsymbol{v}_2(300\text{K})$, and $\boldsymbol{v}_3(300\text{K})$, which contribute most to the conformational varieties of the system at 300 K, and constructed a three-dimensional (3D) subspace. This subspace is referred as the "3DPC subspace" in this report. The position $\boldsymbol{x}$ of a structure in the 3DPC subspace is expressed as $\boldsymbol{x} = [x_1, x_2, x_3]$, where x_i is defined by Equation S10. We present the conformational distribution by dots projected on the 3DPC subspace, where a dot corresponds to a conformation.

Figure S1.

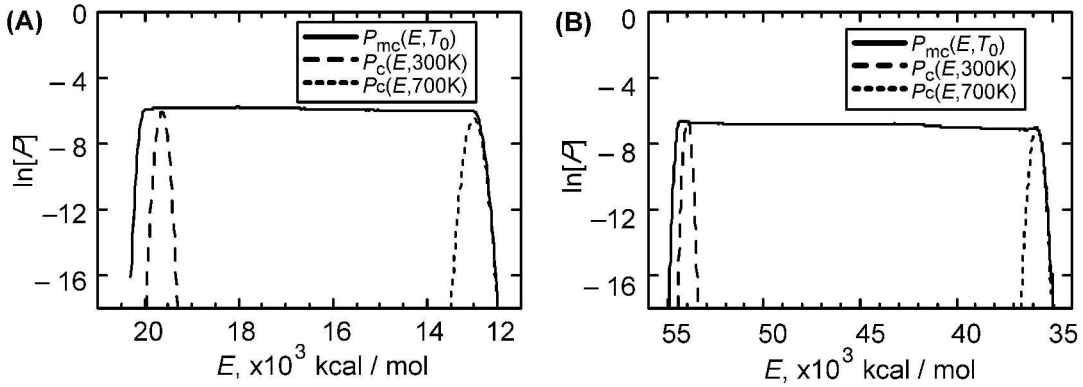

Figure S1. Flat energy (E) distributions $P_{mc}(E,T_0)$ in log scale for the single-chain NRSF system (A) and the NRSF-Sin3 system (B). Canonical energy distributions $P_c(E,300\text{K})$ and $P_c(E,600\text{K})$ are also shown.

Figure S2.

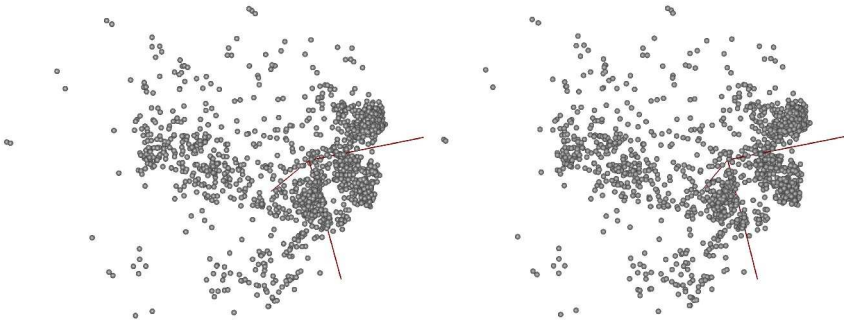

Figure S2. Stereo view (parallel) of Figure 2 of the main text. Since the view is from the same direction as Figure 2, the non-labeled coordinate axes correspond to those in Figure 2.

Figure S3.

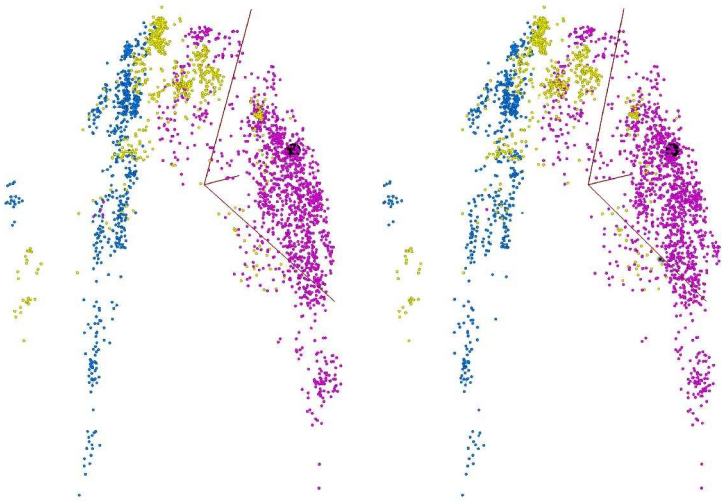

Figure S3. Stereo view (parallel) of Figure 5A of the main text. Since the view is from the same direction as Figure 5A, the non-labeled coordinate axes correspond to those in Figure 5A. The coloring of dots is also the same as that in Figure 5A.

Non-restrained Canonical MD of Single Sin3.

We assessed the effect of the inter-Cα atomic distance restraints (Equation S1) on the opening/closing motions of the Sin3 groove, as follows: First, we performed canonical MD simulations of a single Sin3 at 300 K, where the restraints were not applied. The initial structure of Sin3 was the native structure

(NMR model 1), after removing NRSF. This molecule was immersed in a solvent sphere (diameter = 70 Å), using the same method as that for the Sin3-NRSF system. The final generated system involved 5,503 water molecules, 4 Cl^- and 4 Na^+ (~ 1 mM ionic concentration), where the net charge of the entire system was neutralized. The other condition for the system was exactly the same as that for the NRSF-Sin3 system. After energy minimization, we performed six canonical MD runs at 300 K, each for 11.0 ns: heating for 1.0 ns from 1 to 300 K, 1.0 ns for equilibrium, and 9.0 ns for sampling. Figure S4 plots the fluctuations of *rmsd* in three of the six trajectories, where the *rmsd* is computed for the main-chain heavy atoms between a simulation snapshot and the NMR model 1.

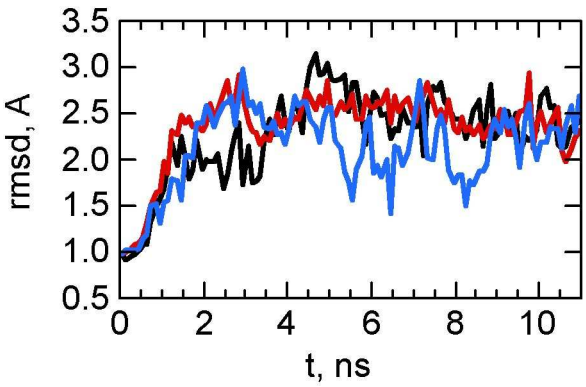

Figure S4. (A) Fluctuations of *rmsd* between the snapshot and NMR model 1. Trajectories are shown in different colors. The x- and y-axes are simulation time and *rmsd*, respectively. The initial *rmsd* value is not zero, because energy minimization was performed prior to the simulation.

To highlight the opening/closing motions of the Sin3 groove, we picked two pairs of Cα atoms: pair 1 between Val A34 and Lys A66, and pair 2 between Thr A72 and Phe A96. Their positions are shown in Figure S5. The distance r_{pair1} for pair 1 indicates the distance between helices H1 and H2, and r_{pair2} for pair 2 represents the distance between H3 and H4.

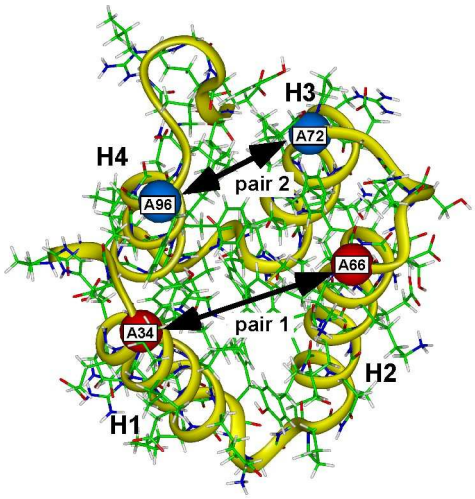

Figure S5. Inter-Cα atomic pairs, pair 1 (red spheres) and pair 2 (blue spheres), to monitor groove opening/closing motions of single Sin3. Arrows indicate the pairs. The four helices are labeled "H1" to "H4". The values of r_{pair1} and r_{pair2} in the native structure (NMR model 1) are 15.5 Å and 12.9 Å, respectively.

Next, we computed the distribution function on a plane of r_{pair1} and r_{pair2}, and converted it to the potential of mean force: $PMF_{\text{pair}} = -RT \ln[P_{\text{pair}}(r_{\text{pair1}}, r_{\text{pair2}}) / P_{\text{pair}}^{\max}]$, where the temperature T is set to 300 K, $P_{\text{pair}}(r_{\text{pair1}}, r_{\text{pair2}})$ is the distribution function projected on the plane, and $P_{\text{pair}}^{\max}$ is the maximum probability. We also computed PMF_{pair} from $Q_{\text{N-S}}(300\text{K})$ of the NRSF-Sin3 system. Figures S6A and S6B demonstrate PMF_{pair} from the non-restrained canonical MD simulations of single Sin3 and $Q_{\text{N-S}}(300\text{K})$, respectively. These figures indicate that the McMD with the inter-Cα atomic restraints provides the groove motions that are compatible with those of the non-restrained MD. The inter-Cα atomic restraint was weak and applied only on the Cα atomic pairs with $r_{i,j}^{0} < 7.0$ Å (see the section "Inter-Cα Atomic Restraints." above). This is why the groove motions are allowed in the McMD simulations.

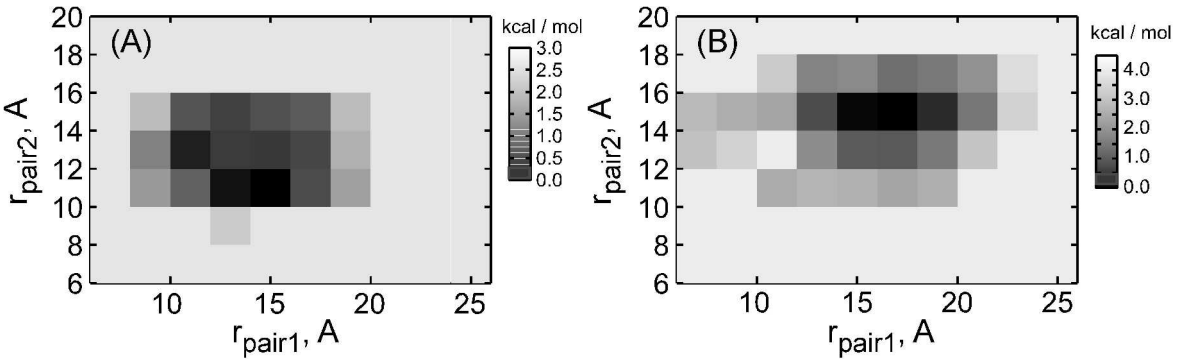

Figure S6. Potential of mean force PMF_{pair} on a 2D plane [r_{pair1}, r_{pair2}]. (A) PMF_{pair} from non-restrained canonical MD simulations of a single Sin3 at 300 K, and (B) that from $Q_{N\text{-}S}(300K)$ of McMD simulations of the NRSF-Sin3 system.

Cluster Analysis.

We classified the conformations of $Q_{N\text{-}S}(300K)$ into clusters using a structure-based clustering method, the average linkage method. Suppose we have two conformations, *i* and *j*, with coordinates expressed as $\boldsymbol{q}_i = [q_1(i),...,q_{225}(i)]$ and $\boldsymbol{q}_i = [q_1(j),...,q_{225}(j)]$ in the format of Equation S14. First, we defined the structural similarity D_{ij} between the two conformations:

$$D_{ij} = [\sum_{k=1}^{225} (q_k(i) - q_k(j))^2]^{1/2}. \quad \text{(S15)}$$

A smaller value of D_{ij} indicates greater similarity between the two conformations. Since D_{ij} is computed using all elements of Equation S14, this similarity score is defined not in the 3DPC subspace, but in the full-dimensional space.

The average linkage clustering was performed for the conformations of $Q_{N\text{-}S}(300K)$, as follows: In step 1, each conformation was treated as a cluster. We refer to the number of clusters as N_{clust}. There are $N_{clust}(N_{clust}-1)/2$ inter-cluster pairs. In step 2, the nearest-cluster pair (i.e., the pair with the smallest D_{ij}) in the inter-cluster pairs was merged as a new cluster. The new inter-cluster distance was defined as the average of D_{ij} between the conformations belonging to the two clusters. Here, we denote the inter-cluster distance for the nearest clusters as $r_{nearest}$. By repeating this procedure, the value of N_{clust} decreased by one. We must terminate the clustering at a step, because N_{clust} is not set in advance. If the clustering is terminated at a step where $r_{nearest} \geq D_{cut}$, then the resultant N_{clust} is expressed as a function of D_{cut}. Figure S7A demonstrates that N_{clust} has an inflection point at $D_{cut} = 3.5$ Å. We denote the number of conformations involved in the *m*-th cluster as $N(m)$, and arranged the clusters in descending order of $N(m)$: the first cluster is the largest cluster. Then, we computed the ratio of $N(m)$ to the number of conformations in $Q_{N\text{-}S}(300K)$ as: $R_{occ}(m) = N(m)/N_{tot}$, where N_{tot} is 3,611. Figure S7B demonstrates the relationship between D_{cut} and $R_{occ}(1)$; i.e., the ratio of the largest clusters.

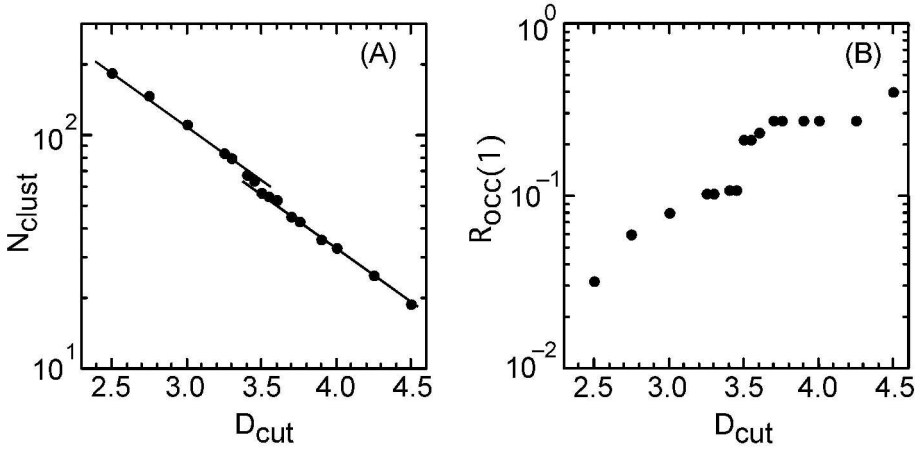

Figure S7. (A) Relationship between D_{cut} and N_{clust}. The two lines are data fitting for $D_{cut} \geq 3.5$ Å and $D_{cut} < 3.5$ Å. (B) Relationship between D_{cut} and $R_{occ}(1)$.

When the threshold D_{cut} is decreased, the number of clusters N_{clust} increases, and $Q_{N\text{-}S}(300K)$ is finally disassembled into singletons (clusters consisting of a single conformation). In contrast, with increasing D_{cut}, N_{clust} decreases, where small clusters are combined to a larger cluster and finally N_{clust} becomes 1. The existence of the inflection point ($D_{cut} = 3.5$ Å) indicates that the cluster distribution settles into a stable state above this point. In fact, Figure S7B indicates that $R_{occ}(1)$ exhibits discontinuity at $D_{cut} = 3.5$ Å, and is approximately constant for $D_{cut} \geq 3.5$ Å. In this study, we analyzed the clusters computed at $D_{cut} = 3.5$ Å, where $N_{clust} = 57$.

Hydrophobic Atoms.

To count the hydrophobic atomic contacts, N_{hbic}, in the NRSF-Sin3 interface, we defined the hydrophobic atoms as follows: all side-chain heavy atoms of eight hydrophobic amino-acids (Ala, Val, Leu, Ile, Phe, Pro, Met, and Trp); Cβ of Asn and Asp; Cβ and Cγ of Gln and Glu; Cβ, Cγ, Cδ of Arg and Lys; Cγ2 of Thr; and Cβ, Cγ, Cδ1, Cδ□, Cε1, and Cε□ of Tyr.

Relationship between q_{ori} and the Number of Intra-NRSF Hydrogen Bonds.

To analyze the relationship between the NRSF orientation q_{ori} and the number of intra-NRSF helical hydrogen bonds N_{HB}, we computed the potential of mean force $PMF_{HB}(q_{ori}, N_{HB})$ on a plane $[q_{ori}, N_{HB}]$. An intra-NRSF helical hydrogen bond is one formed between the carbonyl oxygen atom of residue *i* and the amide nitrogen atom of residue *i* + 4 in NRSF.

Figure S8A demonstrates $PMF_{HB}(q_{ori}, N_{HB})$ computed from $Q_{N\text{-}S}(300K)$. NRSF in the NMR complex structure involves ten helical hydrogen bonds, and the maximum of N_{HB} in $Q_{N\text{-}S}(300K)$ was

eight. Here, we focus on the parallel ($q_{ori} \approx +1$) and anti-parallel ($q_{ori} \approx -1$) complexes. In the parallel ones, the lowest free-energy state is at $N_{HB} = 2$ and the second lowest is at $N_{HB} = 8$. This second lowest state corresponds to the native-like complex structure. In the anti-parallel ones, the lowest free-energy state is at $N_{HB} = 0$ and the second lowest is at $N_{HB} = 5$. Since Figure S8A was computed from all conformations in $Q_{N\text{-}S}(300K)$, many extended NRSF conformations are involved in $PMF_{HB}(q_{ori}=\pm 1, N_{HB})$. Then, to focus on the structures with $d_{3\text{–}13}$ values similar to that of the native complex ($d_{3\text{–}13}^{native} = 15.5$ Å), we recomputed $PMF_{HB}(q_{ori}, N_{HB})$ from the structures only in the range of $d_{3\text{–}13}^{native} - 2.0 < d_{3\text{–}13} < d_{3\text{–}13}^{native} + 2.0$ Å (Figure S8B). In this recomputed $PMF_{HB}(q_{ori}, N_{HB})$, the stability of the second lowest free-energy state ($N_{HB} = 8$) for the parallel complexes increased, to be compatible with that of the lowest free-energy state ($N_{HB} = 2$). This is because the extended conformations with a few helical hydrogen bonds were eliminated from $PMF_{HB}(q_{ori}, N_{HB})$. In contrast, for the anti-parallel complexes, the stability of the second lowest free-energy state ($N_{HB} = 5$) did not increase.

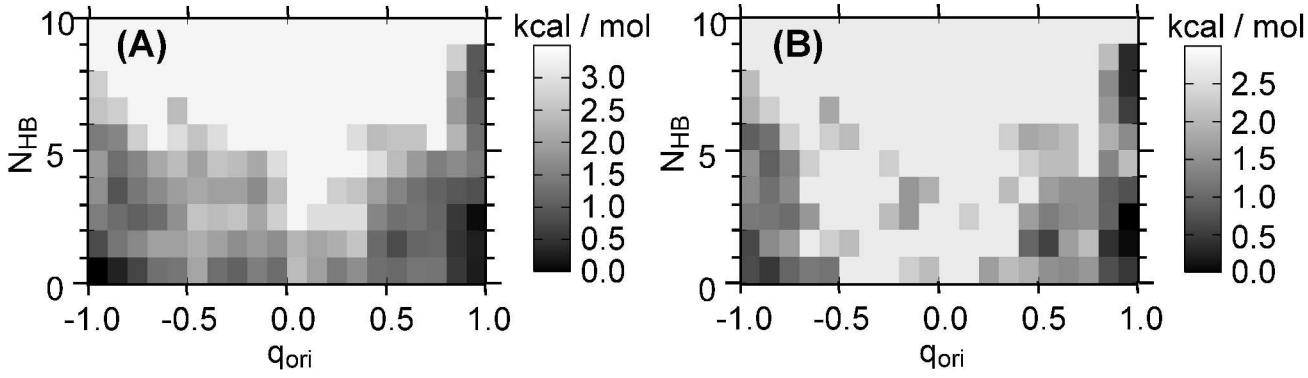

Figure S8. (A) Potential of mean force, PMF_{HB}, on the plane of q_{ori} and N_{HB} computed from all conformations in $Q_{NS}(300K)$. PMF_{HB} is defined as: $PMF_{HB}(q_{ori}, N_{HB}) = -RT \ln[P_{HB}(q_{ori}, N_{HB}) / P_{HBmax}]$, where $P_{HB}(q_{ori}, N_{HB})$ is the number of conformations detected in the range from $[q_{ori}, N_{HB}]$ to $[q_{ori} + \Delta q_{ori}, N_{HB}]$, and P_{HBmax} is the maximum of $P_{HB}(q_{ori}, N_{HB})$. N_{HB} is a positive integer. The temperature T is set to 300 K, and the bin size Δq_{ori} is set to 0.1. The value of PMF_{HB}, which is presented by a tone, is assigned to a tile centered at $[q_{ori} + 0.5\Delta q_{ori}, N_{HB} + 0.5]$. (B) PMF_{HB} computed only from the conformations in the range of $d_{3\text{–}13}^{native} - 2.0 < d_{3\text{–}13} < d_{3\text{–}13}^{native} + 2.0$ Å of $Q_{N\text{-}S}(300K)$, where $d_{3\text{–}13}^{native}$ is $d_{3\text{-}13}$ for the NMR model 1 (15.5 Å).

As mentioned above, two stable states exist at $N_{HB} = 2$ and $N_{HB} = 8$ in the parallel ($q_{ori} \approx +1$) NRSF-Sin3 complexes in Figure S8B. Therefore, one may imagine that there is another free-energy barrier besides B1 and B2. However, we note that N_{HB} has less resolution to differentiate structures than the 3DPC subspace does, because a variety of structures can have the same value of N_{HB}. In the 3DPC distribution (Figure 5B), the light cyan dots ($1 \le N_{HB} \le 2$) are found readily in the vicinity of the

red dots ($N_{HB} \geq 5$), and, more importantly, the purple dots ($3 \leq N_{HB} \leq 4$) do not separate the light cyan and red dots, but are spattered over the distribution. Therefore, no free-energy barrier exists between $N_{HB} = 2$ and $N_{HB} = 8$. This exemplifies that free-energy expression by a quantity with less structural resolution misreads an artificial free-energy barrier, as mentioned previously.[4]

Free-energy Landscape with Different Structural Measures.

Other types of free-energy landscapes may be useful for studying coupled folding and binding.[5,6] We computed the potential of mean force on a plane of the number of intra-NRSF hydrogen bonds, N_{HB}, and the number of residue-residue contacts between NRSF and Sin3, N_{cont} : $PMF_{cont} = -RT \ln[P_{cont}(N_{HB}, N_{cont}) / P_{cont}^{max}]$, where $P_{cont}(N_{HB}, N_{cont})$ and P_{cont}^{max} are the distribution function and the maximum distribution, respectively. Figure S9 demonstrates PMF_{cont} at 300 K, computed from the sampled conformations of the NRSF-Sin3 system.

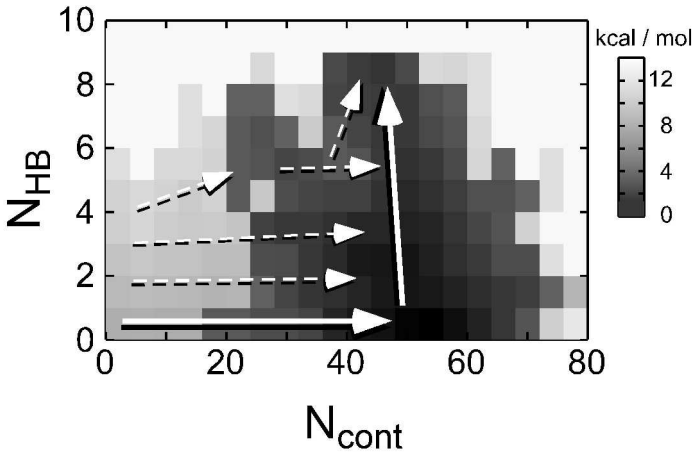

Figure S9. Potential of mean force, PMF_{cont}, at 300 K on the plane of N_{cont} and N_{HB}, computed from $Q_{N\text{-}S}(T)$. Arrows are mentioned in text.

One may have the impression that Figure S9 supports the mechanism proposed in Figure 11B: The solid-line arrows present the induced fit (induced folding), and the dashed-line arrows depict the population shift. However, as discussed in the DISCUSSION section, most unbound NRSF conformations were sampled in a high-energy region (a high-temperature region). Therefore, the free energy at $N_{cont} = 0$ in Figure S9 was contributed from conformations sampled in the high-energy region: The contribution from a high-energy conformation is given by $P_c\left(E_{high}, 300\text{K}\right)$, where E_{high} is the energy of the conformation. Consequently, the free energy at $N_{cont} = 0$ was presented inaccurately. To compensate for the statistics on the low probability events at 300 K, another computational approach

should be combined with McMD sampling, as discussed in the DISCUSSION section.

(1) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* 1997, 101, 817–824.
(2) Higo, J.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Nakamura, H. *Chem. Phys. Lett.* 2009, 473, 326–329.
(3) Ikebe, J.; Umezawa, K.; Kamiya, N.; Sugihara, T.; Yonzawa, Y.; Takano, Y.; Nakamura, H.; Higo, J. *J. Comput. Chem.* 2011, 32, 1286–1297.
(4) Kamiya, N.; Higo, J.; Nakamura, H. *Protein Sci.* 2002, 11, 2297–2307.
(5) Ganguly, D.; Chen, J. *J. Am. Chem. Soc.* 2009, 131, 5214–5223.
(6) Chen, J. *J. Am. Chem. Soc.* 2009, 131, 2088–2089.

Complete reference 19

Kuwahara, K.; Saito, Y.; Takano, M.; Arai, Y.; Yasuno, S.; Nakagawa, Y.; Takahashi, N.; Adachi, Y.; Takemura, G.; Horie, M.; Miyamaoto, Y.; Morisaki, T.; Kuratomi, S.; Noma, A.; Fujiwara, H.; Yoshimasa, Y.; Kinoshita, H.; Kawakami, R.; Kishimoto, I.; Nakanishi, M.; Usami, S.; Saito, Y.; Harada, M.; Nakao, K. *EMBO J.* 2003, 22, 6310–6321.