Supporting Information

Capturing a dynamic chaperone-substrate interaction using NMR-informed molecular modeling

Loïc Salmon,[*,†,#] Logan S. Ahlstrom,[†,‡,#] Scott Horowitz,[†] Alex Dickson,[‡,^] Charles L. Brooks III,[*,‡,§] and James C. A. Bardwell[*,†]

[†]Department of Molecular, Cellular and Developmental Biology, and the Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan 48109, United States

[‡]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

[§]Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, United States

[^]Present addresss: Department of Biochemistry and Molecular Biology and Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, United States

[#]These authors contributed equally.
*Correspondence should be sent to: *jbardwel@umich.edu, lsalmon@umich.edu, brookscl@umich.edu

Contents:

Supporting Methods
Figures S1-S19
Tables S1-S5
Test of the approach on known complexes (contains Figs. S15-19)
References

Supporting Methods

Protein Expression and Purification. The gene for *spy* 29-124 was amplified from plasmid pET28sumo-spy[1] with primer 1 (5'-CGC GGG ATC CTT CAA AGA CCT GAA CCT GAC CG-3') and primer 2 (5'- CGC GCT CGA GTT ATG TCA GAC GCT TCT CAA AAT TAG C-3'), and was cloned into pET28sumo via BamHI and XhoI sites. Spy 29-124 was expressed and purified as described previously[1] with the exception that Ni-HisTrap columns (GE Healthcare) were utilized instead of the Ni-NTA beads and mini-chromatography column. ULP1 cleavage occurred following elution from the Ni-HisTrap column overnight at 4°C while dialyzing to 40 mM Tris, 300 mM NaCl, pH 8.0. After dialysis, Spy was passed over the HisTrap column to remove the cleaved SUMO tag (20 mM imidazole was left over from the dialysis). The flow-through was then concentrated and diluted 5 times with 20 mM Tris, pH 8 for further purification on a HiTrap Q column. Spy has an isoelectric point of 9.5 and therefore was collected in the flow-through. The flow-through containing Spy was concentrated and diluted 5-fold with 50 mM sodium phosphate at pH 6.5 and passed over a HiTrap SP column. Spy was then eluted with a gradient from 0-1 M NaCl. Re-buffering to the final reaction buffer was accomplished by gel filtration, passing the pooled and concentrated fractions containing Spy over a HiLoad 75 column in 40 mM HEPES, 100 mM NaCl, pH 7.5. Fractions containing Spy were then concentrated, frozen in liquid nitrogen, and stored at –80°C. $\text{Im7}_{6\text{-}45}$ was expressed and purified as previously described.[2]

Circular Dichroism (CD) Spectroscopy. $\text{Im7}_{6\text{-}45}$ and Spy samples were buffer-exchanged via Amicon centrifugal filters to 10 mM sodium phosphate, pH 7. Samples were then diluted to 100 µM each, and CD spectra from 260 to 178 nm were recorded with a Jasco J-810 using a 0.1 mm cuvette. The buffer spectrum was subtracted and each spectrum was analyzed by VARSLC.[3] VARSLC was chosen as it gave the closest prediction to the crystallographic helicity of Spy alone as calculated by the DiChroWeb server.[4]

Contact Maps. For a given snapshot from the binding simulations, we assigned an intermolecular contact to a given residue pair if the $C\alpha$-$C\alpha$ distance of the two residues was less than or equal to $\gamma . r_{ij}^{min}$, where $r_{ij}^{\min}$ is the interaction radius for residues *i* and *j* in the inter-protein potential and $\gamma = 1.2$. For each intermolecular residue pair, we reported the contact probability averaged over all Spy-$\text{Im7}_{6\text{-}45}$ complexes sampled after the first 0.15 µs of simulation. The contact maps for the two Spy monomers were averaged into a single map. We also projected the contact maps onto a single dimension (i.e., the Spy or $\text{Im7}_{6\text{-}45}$ sequence; *Figure 2B*) by defining a contact score as the sum of the contact probabilities for a given residue across an entire column (Spy) or row (Im7). The scores are converted to a pseudo-energy by normalizing them by the maximum possible score and then taking the negative natural logarithm of this ratio.

Interaction Patterns. A given residue in Spy is defined as participating in binding if it is in contact with more than one residue in $\text{Im7}_{6\text{-}45}$ for at least 5% of the bound ensemble. The concave side of Spy was defined as the following residues: 28-36, 39, 42, 43, 46, 49, 50, 52-58, 61, 64, 65, 68, 85, 89, 92, 93, 96, 97, 100, 103, 118, 121-124.

Progress Variables. To characterize the global conformational sampling of the Spy dimer, we defined an inter-domain distance and dihedral angle that describe the relative positions of monomers A and B (*Figure 4B*). The distance is defined between the centers of mass of residues A:107–109 and B:107–109. The dihedral angle is defined over the centers of mass of the following four sites, in the same order: (*i*) A:42; (*ii*) B:61–63, B:80–82; (*iii*) A:61–63, A:80–82; and (*iv*) B:42.

Reproducibility and Convergence of the Trajectories. The 10 pooled Spy-Im7$_{6\text{-}45}$ binding trajectories are highly similar, exhibiting correlations (R-values) of their contact maps of >0.9 and binding affinities in the range of 0.04-0.31 μM. To further check the convergence and variability of the trajectories, 10 additional binding simulations were run**.** These additional simulations yielded a contact map R-value of 0.99 to the map obtained from the set of 10 original simulations, a binding affinity of 0.12 μM (compared to 0.11 μM for the original set of simulations), and an Im7$_{6\text{-}45}$ helicity profile that is essentially the same as the one obtained from the original set. Thus, the new set of binding simulations yields the same Spy-Im7$_{6\text{-}45}$ interaction mode as the original set of 10 simulations and underlines an adequate level of convergence of the calculations. For the Spy inter-domain dihedral angle and distance presented in Figure 4A-B of the main text, uncertainties were obtained by calculating the standard error across each of the 100 individual apo simulations and over 40 equal-length trajectory segments from the simulations of bound Spy (four from each of the 10 simulations).

Preferential Interaction of Spy with Helical or Unfolded Conformations of Im7. We evaluated the propensity of specific sites in Spy to interact with different structural elements in the Im7$_{6\text{-}45}$ substrate during the optimized binding simulations. For a given contact (c_{ij}) that forms between residue i in Spy and residue j in Im7$_{6\text{-}45}$, the interaction occurs with a helical element (H) or with an unstructured coil (C) in Im7$_{6\text{-}45}$ with the conditional probabilities $P(H_j|c_{ij})$ and $P(C_j|c_{ij})$, respectively. These probabilities are then normalized by the probability of a residue in Im7$_{6\text{-}45}$ being helical or coil at random: $P_{rand}(H_j)$ and $P_{rand}(C_j)$. $P_{rand}(H_j)$ is set to the average helicity of residue j in Im7$_{6\text{-}45}$ in the bound ensemble (as determined from the helicity profile shown in Figure 3C). $P_{rand}(C_j)$ is simply one minus this value. Thus, the likelihood scores for a given residue in Spy contacting a helical or an unstructured coil element in Im7$_{6\text{-}45}$ are $L_H = P(H_j|c_{ij})/P_{rand}(H_j)$ and $L_C = P(C_j|c_{ij})/P_{rand}(C_j)$, respectively. We then take the difference between these two scores, $\mathrm{L}_{\mathrm{H-C}} = L_H - L_C$, such that positive values indicate a propensity of a residue in Spy for helices and negative values denote a preference for coils. We then sum $\mathrm{L}_{\mathrm{H-C}}$ across all residues in Im7$_{6\text{-}45}$ in order to project this score onto the Spy crystal structure.

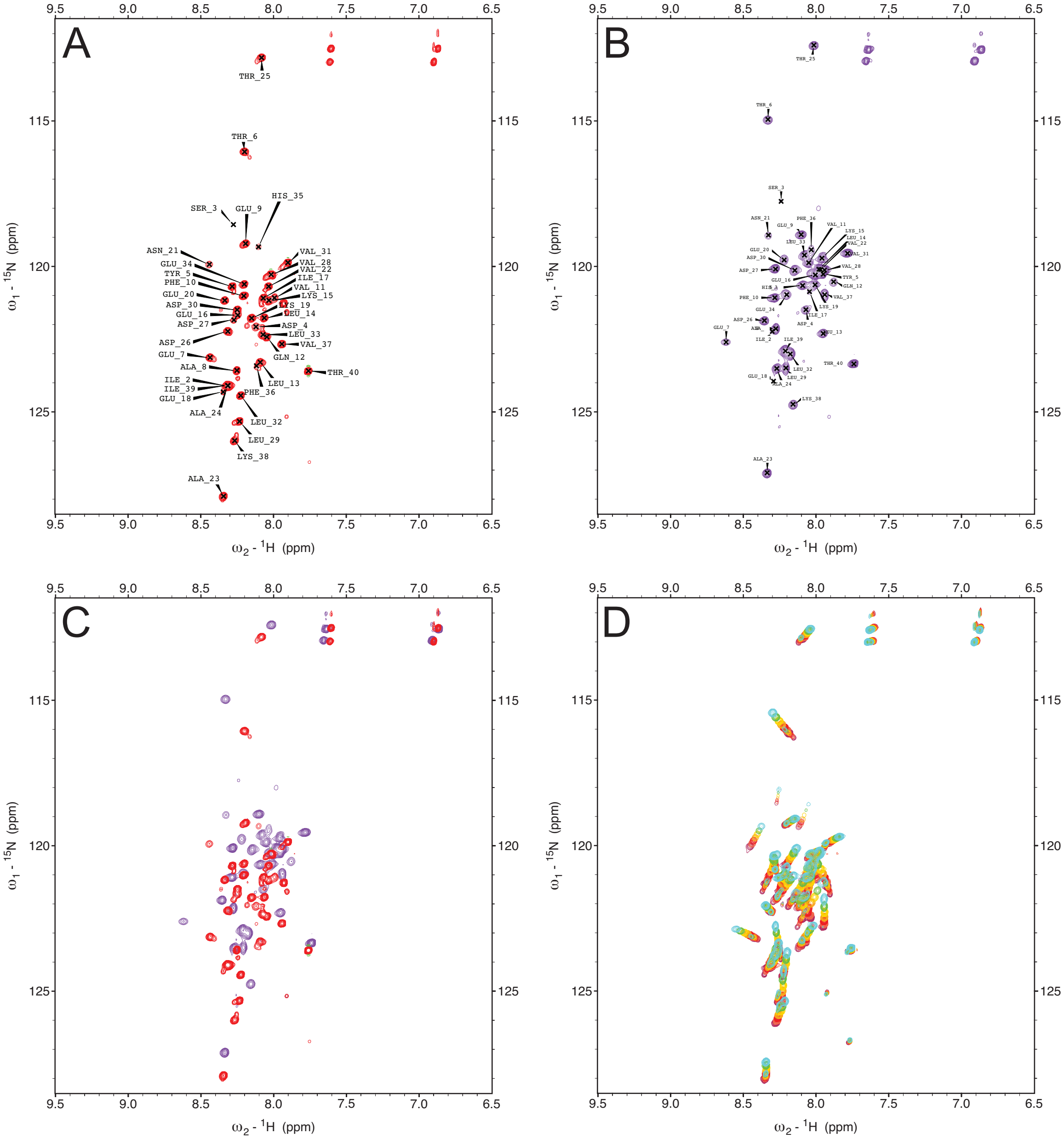

Figure S1. ^{15}N HSQC spectra of Im7$_{6\text{-}45}$ under various ionic strength conditions. (A) Assignment in a 4 mM sodium phosphate and 50 mM NaCl buffer. (B) Assignment in a 10 mM sodium phosphate and 150 mM NaCl buffer. (C) Comparison of the two spectra (A and B). (D) Spectra of Im7$_{6\text{-}45}$ in 4 mM sodium phosphate with increasing amounts of NaCl, from 0 (red) to 150 mM (blue).

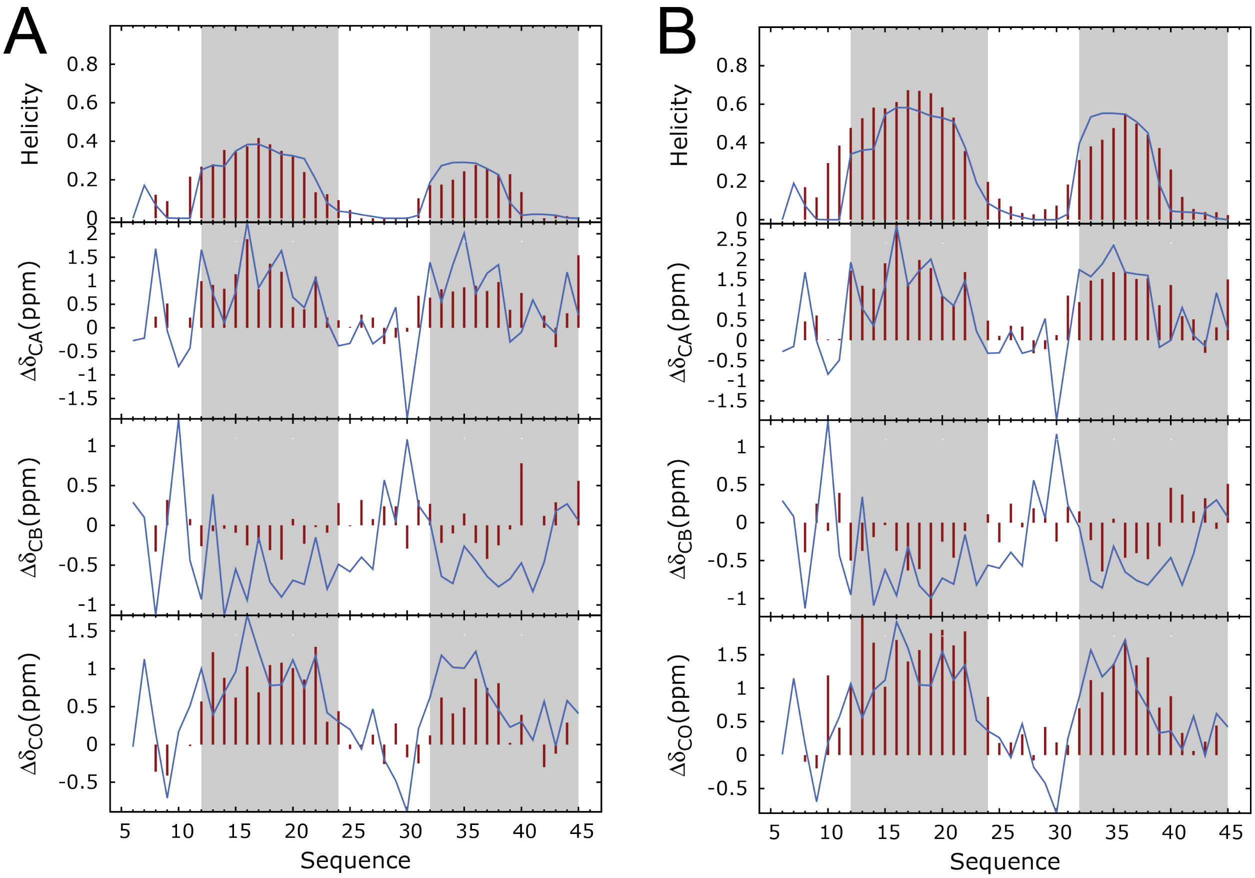

Figure S2. Reproduction of the helical propensity and secondary carbon chemical shifts after force-field optimization for Im7$_{6\text{-}45}$ under (A) low-salt conditions and (B) high-salt conditions (see Figure 1). The red bars represent experimental NMR data (SSP score and secondary chemical shift) and blue curves denote the corresponding back-calculated values predicted from the optimized simulations. The force-field optimization is done using the helical propensity. Chemical shifts were directly computed from the coarse-grained trajectory using LARMOR$^{C\alpha}$[5].

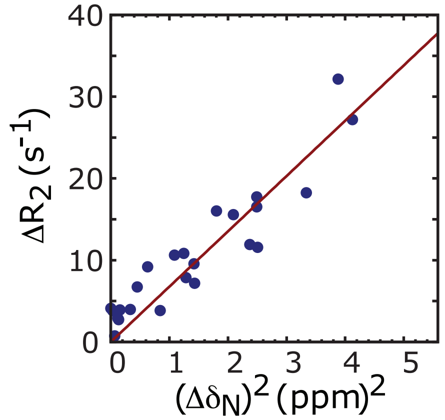

Figure S3. The correlation between chemical shift changes due to increasing ionic strength and differences in R_2 relaxation rates suggests that Im7$_{6\text{-}45}$ may be in exchange between a less folded state, under low salt conditions (4 mM sodium phosphate), and a more structured state, in a high salt environment (10 mM sodium phosphate and 150 mM NaCl).

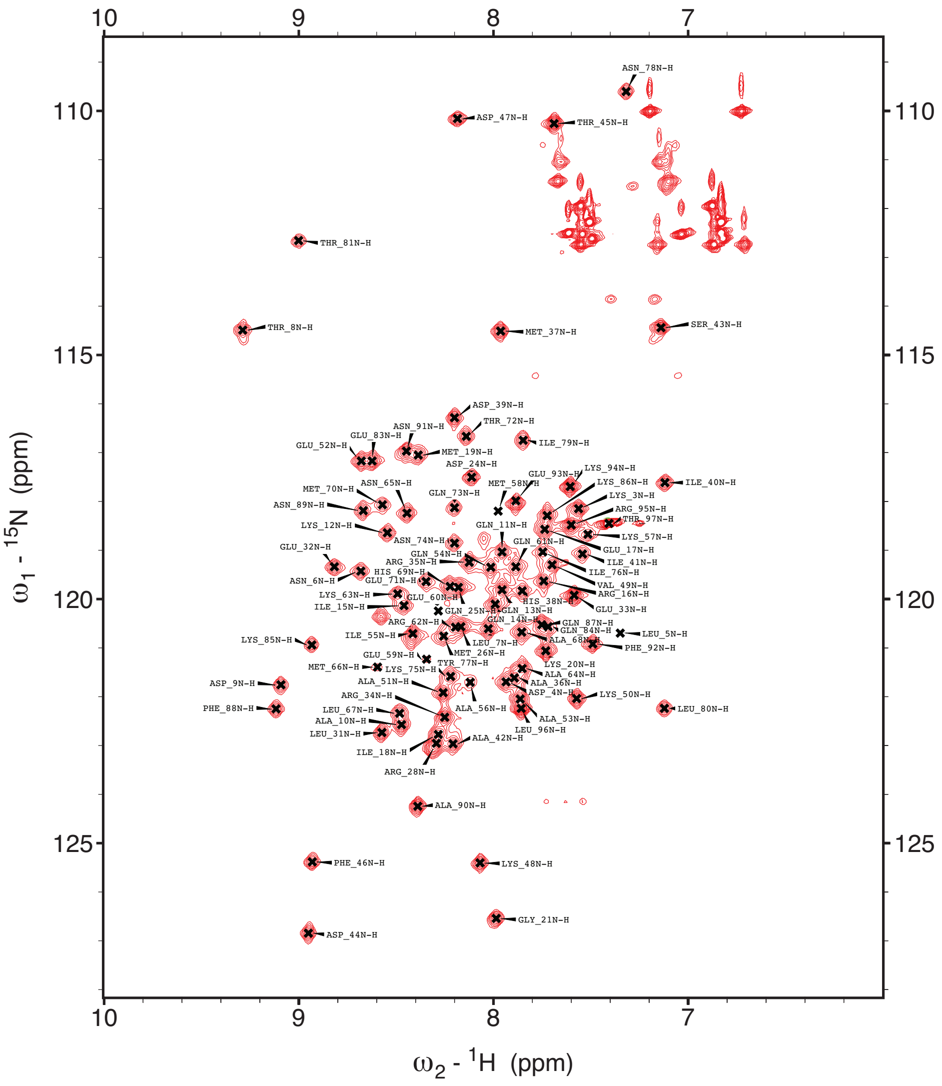

Figure S4. Assigned ^{15}N HSQC spectra of Spy.

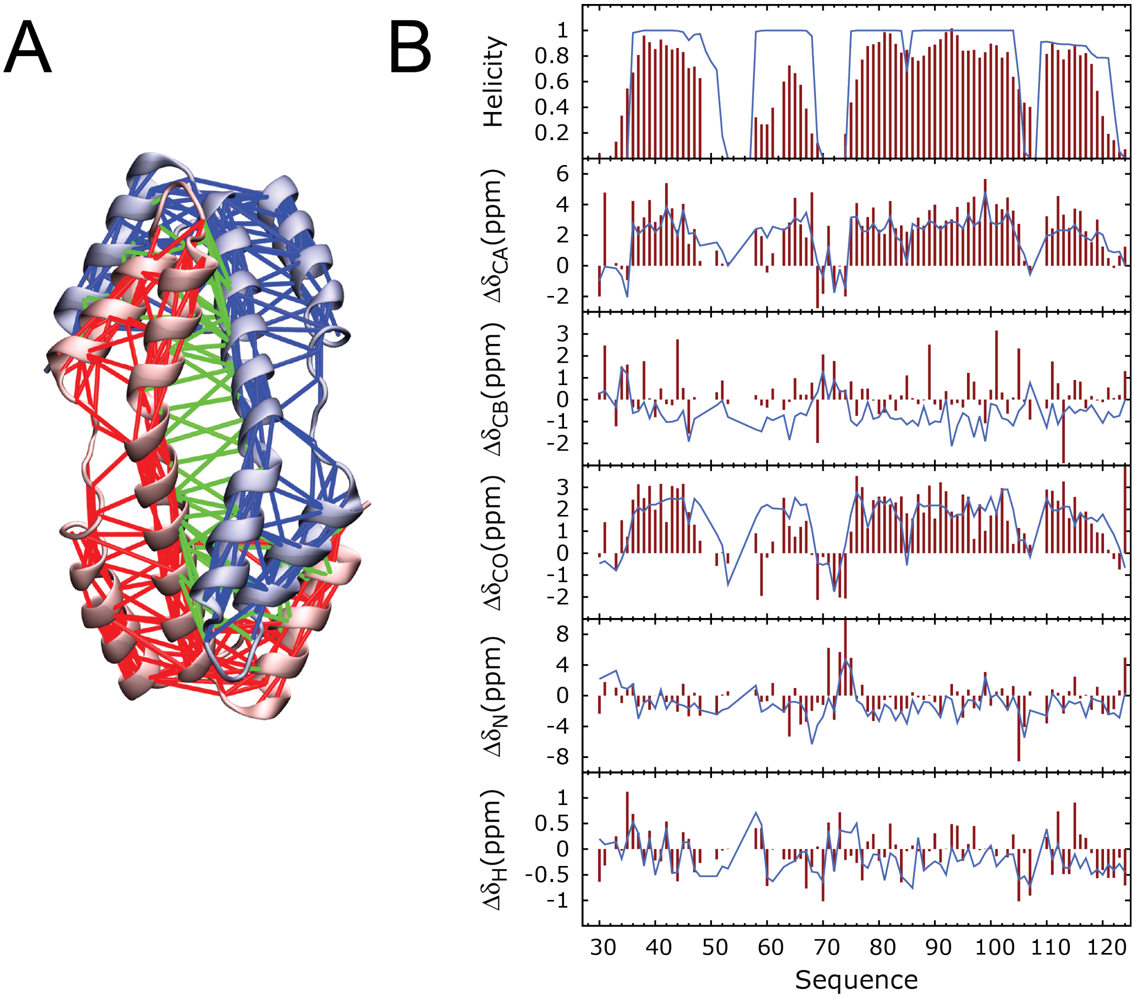

Figure S5. (A) Native interactions defined in the Gō-like model of Spy are represented as lines on the apo crystal structure of Spy; intra-monomer contacts in monomers A and B are shown in red and blue, respectively, and intermolecular contacts spanning the dimer interface are depicted in green. Only minor adjustments to the original Spy Gō-like model are needed to achieve a satisfying description of Spy. (B) From top to bottom: Spy helical propensity as measured by SSP scores and secondary chemical shifts CA, CB, CO, N, and H; experimental values (red bars) and predicted values (blue curves). Experimental measurements were performed at 313 K, and predicted values were derived from coarse-grained simulations at 300 K. NMR measurements of Spy were performed at a slightly elevated temperature (313K), as this yielded more amenable chemical shift spectra for assignment than at 300 K, and could account for differences in the experimental and simulated profiles. Parameterization of the Gō-like model of Spy was performed at 300 K to allow for incorporation into subsequent binding simulations with $\text{Im7}_{6\text{-}45}$.

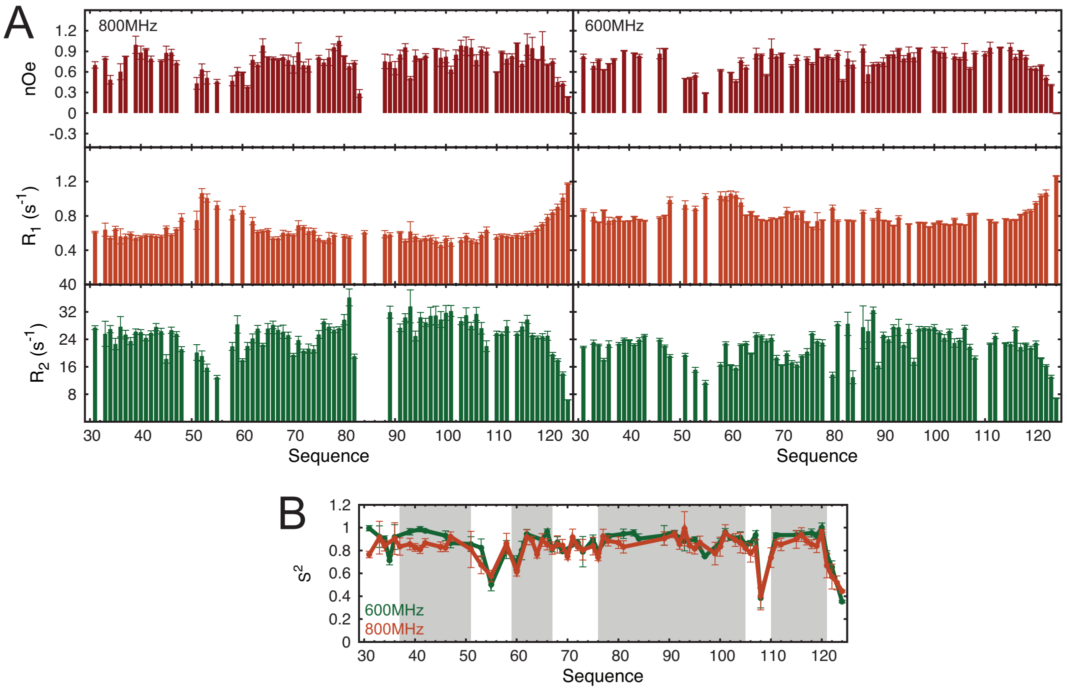

Figure S6. Conformational dynamics of the chaperone Spy determined by NMR. (A) Spy ^{15}N relaxation properties measured at 800 and 600 MHz to probe Spy fast (ps-ns) dynamics: (top) {^{1}H}-^{15}N heteronuclear Overhauser effect (nOe), (middle) longitudinal relaxation rates (R_1), and (bottom) transverse relaxation rates (R_2). For both magnetic field strengths, the rotational diffusion tensors of Spy exhibit significant anisotropy, as expected from Spy's elongated crystal structure. The rotational diffusion tensor also agrees well with the prediction based on the hydrodynamic properties of the Spy structure[6] (*Table S2*). The orientation of the rotational diffusion tensors is identical within experimental error from the two analyses, and the direction of the three axes of the tensors deviates on average by less than 15° compared to the inertia tensor.[7] (B) Order parameter (S^2) derived from a model-free analysis of the ^{15}N relaxation data at 600 MHz (green) and 800 MHz (orange). Order parameters are a measure of flexibility that range from 1 for rigid vectors and decrease towards 0 with an increased level of flexibility. Grey boxes correspond to the positions of the helices in the crystal structure of Spy. Overall, these data suggest that the structure of Spy in solution agrees well with the apo crystal structure and presents significant dynamics only for residues ~51-57 (which reside within the flexible linker region), the C terminus, and Thr108.

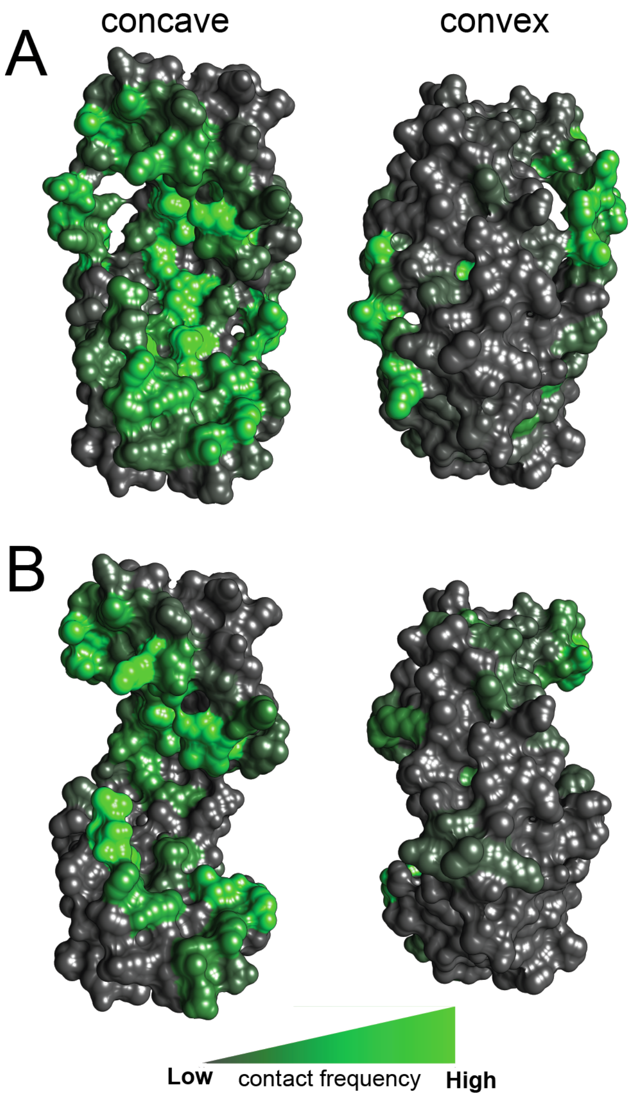

Figure S7. Comparison of the Spy contact pattern from (A) the binding simulations and (B) X-ray crystallography.[8] In both panels, the concave and convex surfaces of Spy are shown at left and at right, respectively. In the crystal structure, the flexible loops at the edges of the Spy cradle are highly flexible and not represented. Residues colored in green contact the Im7 substrate frequently, whereas the residues colored in grey do not.

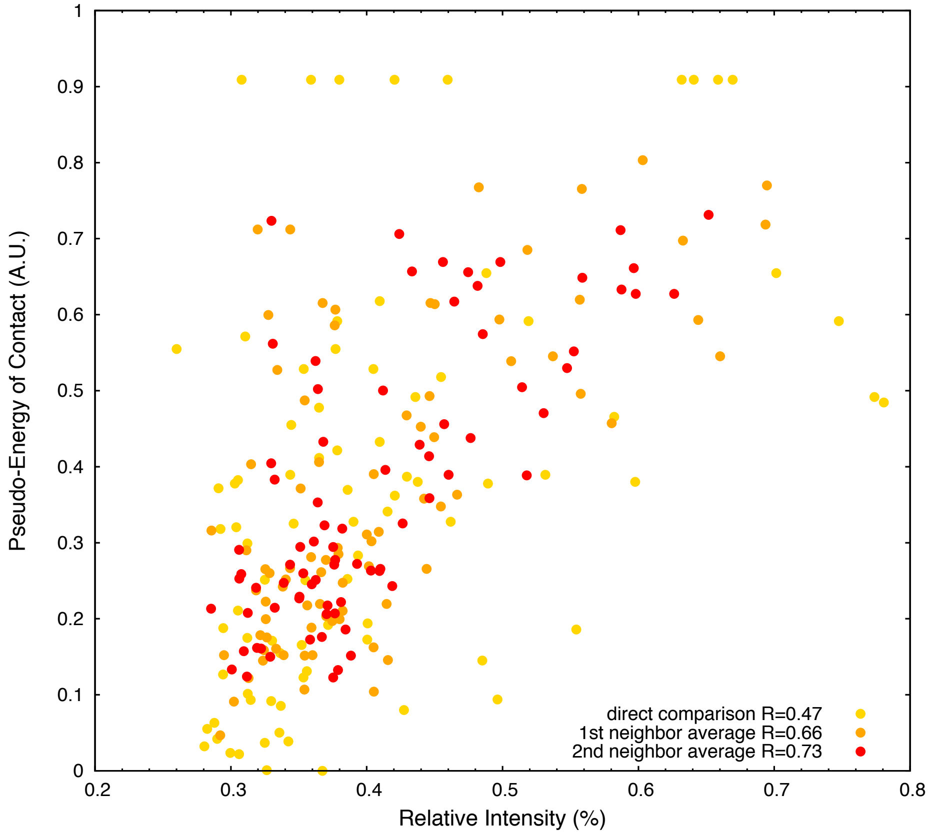

Figure S8. Correlation between the decrease in intensity observed for Spy upon interaction with Im7$_{6\text{-}45}$ and the pseudo-energy of contact, as presented in Figure 2B. Three comparisons are shown: the direct correlation (yellow), the correlation after averaging of the pseudo-energy over the first neighbors (residues *i-1, i, and i+1*) and over the second neighbors (residues *i-2, i-1, i, i+1 and i+2*). The averaging over the neighboring residue is performed to evaluate how the simulations can predict the trends even if missing some details of the interaction. The exchange process observed between Spy and Im7 depends on additional parameters than just the frequency of contact, such as the precise timescales of these interactions, the difference in chemical shift between the bound and the free state, and the relaxation rates within the complex including the potential exchange contribution arising from the multiple conformations of Im7 bound to Spy. For this reason, only a qualitative correlation is expected for the comparison of this data.

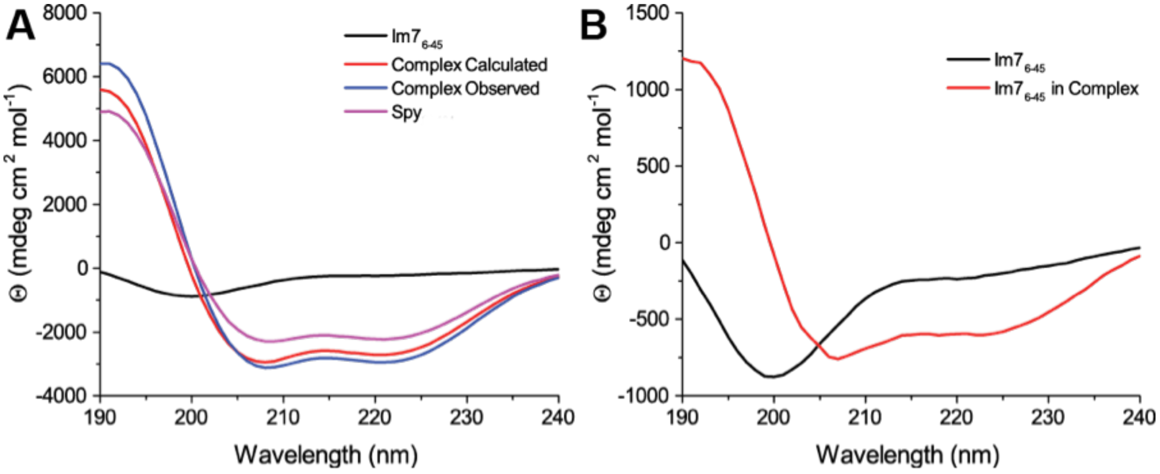

Figure S9. Circular dichroism (CD) spectroscopic analysis of Im7$_{6\text{-}45}$ structural change when bound to Spy. (A) Buffer-subtracted CD spectra of Spy and Im7$_{6\text{-}45}$ individually and in complex. Complex Calculated corresponds to the sum of the spectrum of Spy and Im7$_{6\text{-}45}$ taken individually. (B) CD spectrum of Im7$_{6\text{-}45}$ alone versus in complex generated by subtracting the Spy spectrum from the Spy-Im7$_{6\text{-}45}$ complex spectrum.[3] VARSLC estimates of helicity of Im7$_{6\text{-}45}$ increase from 11% in the unbound state to 40% in the bound state.[4]

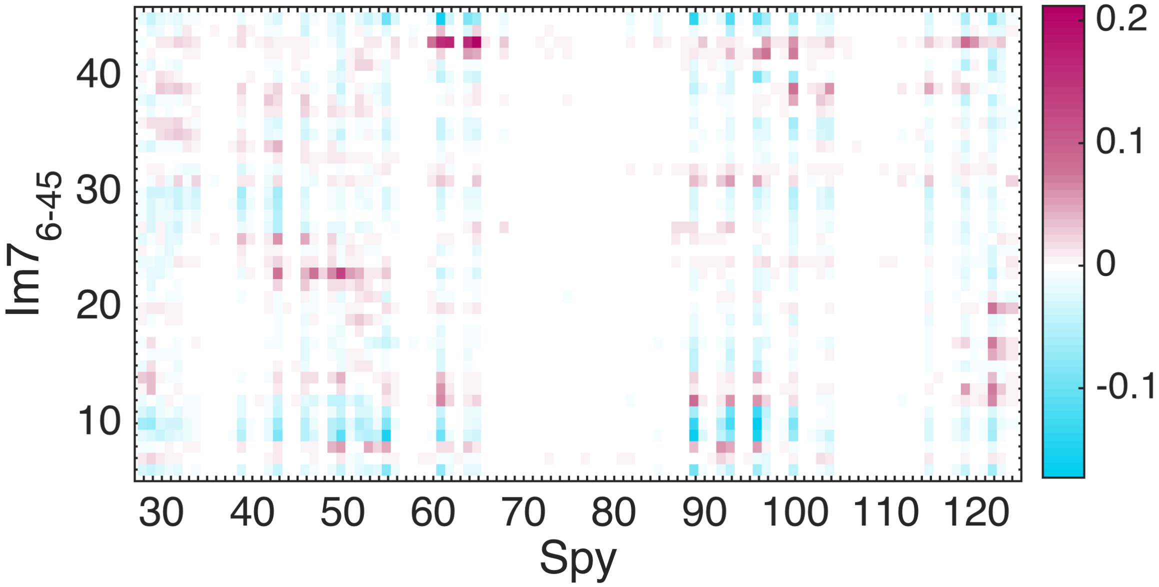

Figure S10. Propensity of Spy residues to interact with $Im7_{6\text{-}45}$ in helical or unfolded conformations. The score is computed based upon the conditional probability of a Spy-$Im7_{6\text{-}45}$ residue-residue contact to occur while the corresponding $Im7_{6\text{-}45}$ residue is in a helix or in a coil (Supporting Methods). Positive (maroon pixels) and negative (light blue) values indicate a higher tendency for the Spy residue to interact with helical and unfolded residues in $Im7_{6\text{-}45}$, respectively. This score is projected onto the surface of the Spy crystal structure in Figure 3B.

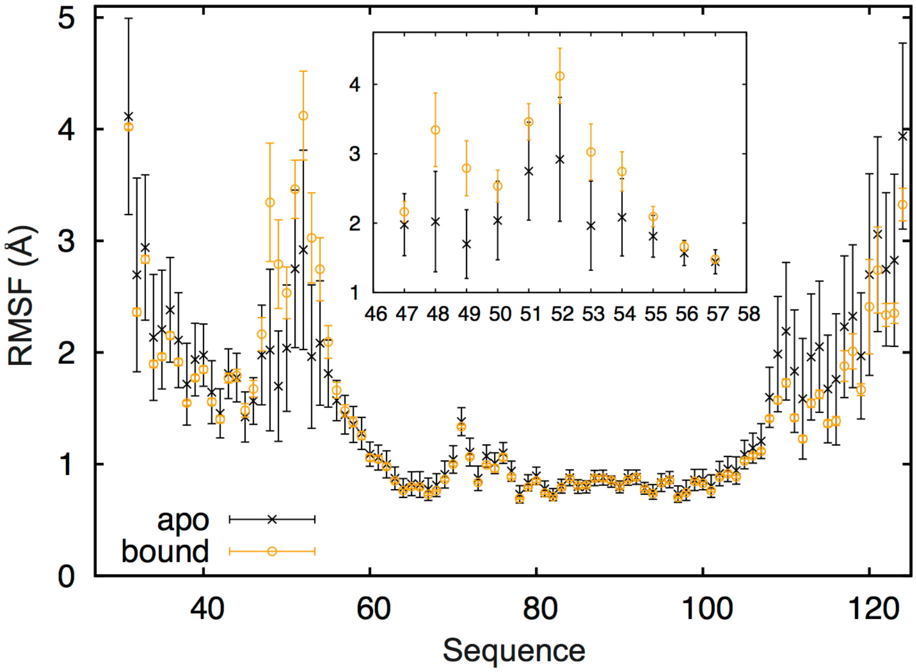

Figure S11. The root-mean-square fluctuations (RMSFs) of residue Cα positions during coarse-grained simulations of apo Spy (black) and bound Spy (orange). Averaged values are identical to Figure 4C. Error bars were estimated by calculating the per-residue standard deviation of average RMSF values across 100 simulations of apo Spy and 10 simulations of bound Spy.

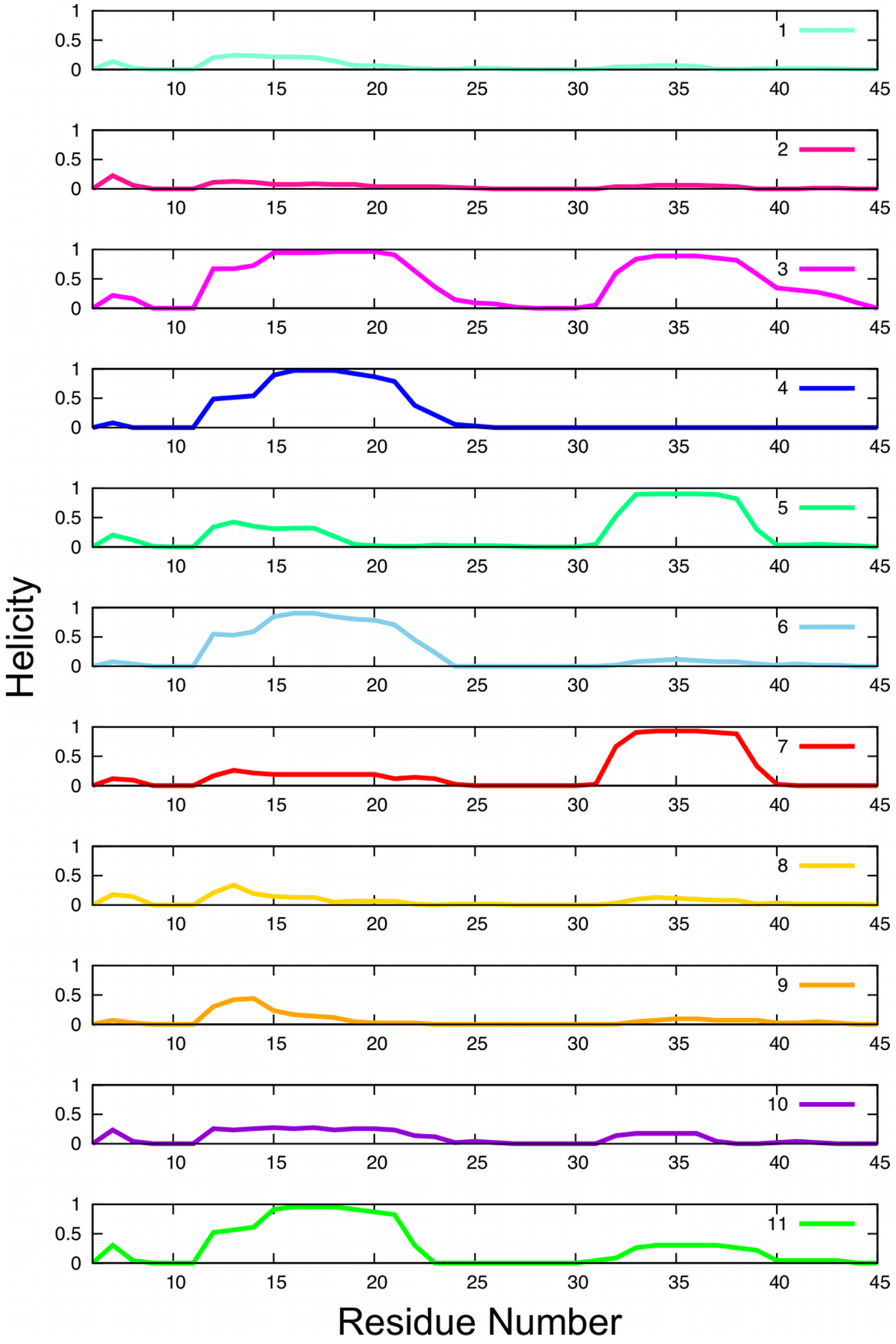

Figure S12. Helical propensity of the different sub-states of Im7$_{6\text{-}45}$ identified in the network analysis. Color coding is the same as in Figure 5 of the main text.

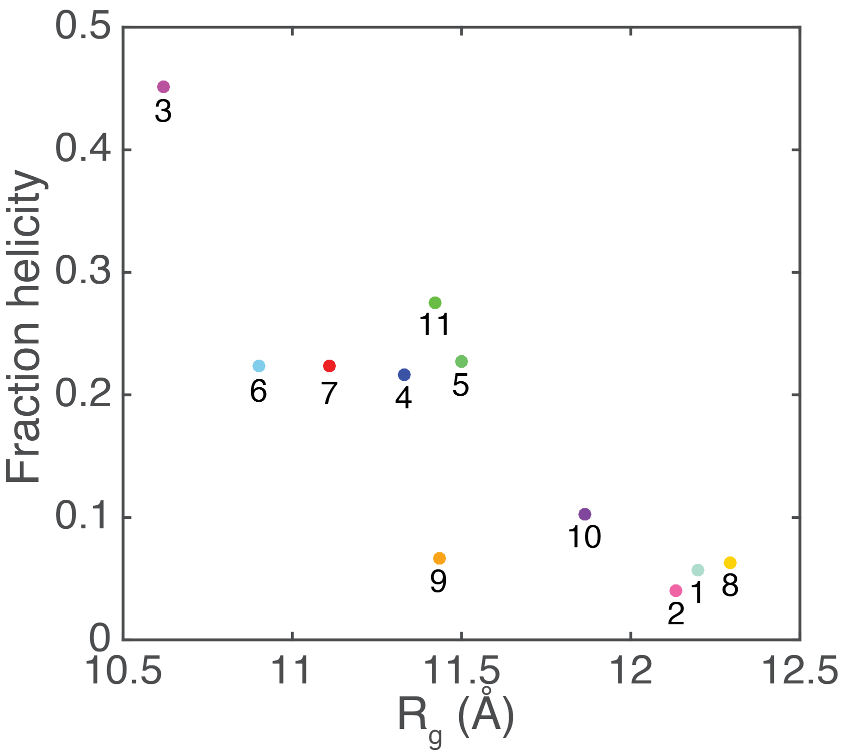

Figure S13. Correlation of the radius of gyration (R_g) and the helicity level of the different sub-states of Im7$_{6\text{-}45}$ identified in the network analysis. Color coding is the same as in Figure 5 of the main text.

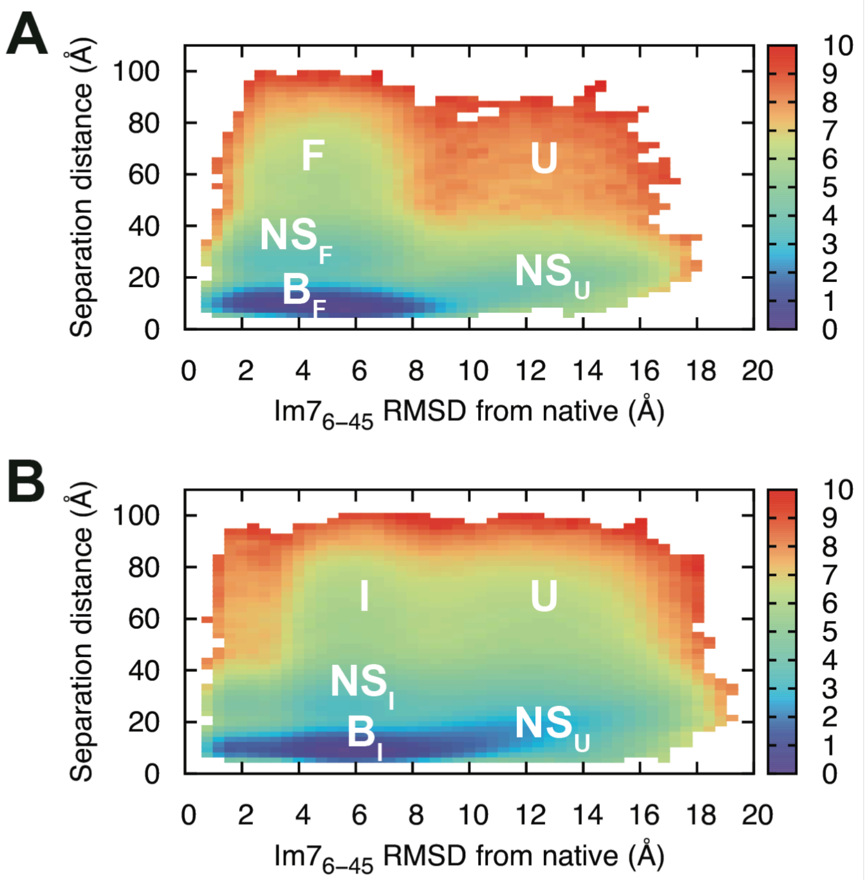

Figure S14. Chaperone-substrate binding mechanism. Free energy landscapes constructed from the (A) optimized and (B) un-optimized Spy-Im7$_{6\text{-}45}$ binding simulations. (A) is identical to Figure 6 in the main text. The surfaces show the Spy-Im7$_{6\text{-}45}$ separation distance and the Cα-RMSD of Im7$_{6\text{-}45}$ to the equivalent segment in the crystal structure of full-length Im7. The separation distance is defined as the distance between the centers of mass of the two partners. The free energy is normalized by the volume of a spherical shell with a radius corresponding to a given distance. The labels signify different conformational states of the Im7 substrate in the presence of the Spy chaperone: B_F (bound and folded), B_I (bound and partially folded intermediate), NS_F (non-specific complex and folded), NS_I (non-specific complex and partially folded intermediate), NS_U (non-specific complex and unfolded), F (unbound and folded), I (unbound and partially folded intermediate), and U (unbound and unfolded Im7$_{6\text{-}45}$). The surfaces are computed at the simulation temperature of 300 K and the colorbars denote the free energy in units of kcal/mol.

Table S1. Agreement of helicity profile at each iteration of the optimization procedure

Iteration	RMSD
0	0.156
1	0.119
2	0.111
3	0.104
4	0.091
5	0.101
6	0.065
α	0.126

The root-mean-square deviation (RMSD) is compared against the experimental low salt SSP profile for iterations 1-6, and compared to the high salt experimental data in the case of α.

Table S2. Rotational diffusion tensor of Spy predicted using the ^{15}N relaxation rates at 800 and 600 MHz

	800 MHz	600 MHz
α (°)	–82 ± 27	–65 ± 26
β (°)	–20 ± 6	–12 ± 6
γ (°)	59 ± 15	77 ± 28
Dx (10^6 s^{-1})	6.9 ± 0.3	7.3 ± 0.2
Dy (10^6 s^{-1})	8.0 ± 0.5	10.7 ± 0.3
Dz (10^6 s^{-1})	11.4 ± 0.5	12.9 ± 0.3

Euler angles follow the zxz convention and were computed in the frame of the apo Spy crystal structure (PDB ID: 3O39).[1] For comparison, the values predicted by HydroNMR[6] are Dz = 12.0, Dy = 7.77, and Dx = 7.53 (10^6 s^{-1}).

Table S3. Order parameters (S^2), timescales of motions (τ_e), exchanges contributions (k_{ex}), slow timescale motions (S^2_{slow}), and their associated errors (err) obtained for Spy during analysis of ^{15}N spin relaxation (600 MHz)

Residue	S^2	S^2 err	τ_e	τ_e err	k_{ex}	k_{ex} err	S^2_s	S^2_s err
31	1.00	0.02						
33	0.92	0.10						
34	0.91	0.02						
35	0.85	0.04	1.3	0.2			0.84	0.06
36	0.92	0.11	0.1	0.7				
39	1.01	0.03	16	7			0.96	0.03
41	1.04	0.02	16	7			0.95	0.03
42	0.97	0.01						
46	0.93	0.04						
47	0.92	0.04	16	6			0.94	0.04
51	0.86	0.07	0.9	0.4				
53	0.82	0.08	0.1	0.3				
55	0.50	0.06	1.6	0.1	16	3		
58	0.81	0.08	1.5	0.5				
60	0.78	0.06	1.4	0.3			0.89	0.06
61	0.80	0.08	1.0	0.4				
62	0.94	0.04						
65	0.89	0.02						
66	0.96	0.03						
67	0.82	0.03	0.05	0.01	5	2		
68	0.87	0.05						
69	0.79	0.02						
72	0.87	0.03	0.04	0.02				
73	0.79	0.13						
75	0.90	0.04						
76	0.80	0.02	0.01	0.01	6	2		
78	0.93	0.02						
81	0.95	0.04						
83	0.96	0.03						
84	0.90	0.04						
89	0.93	0.09						
90	0.94	0.02						
91	0.96	0.02						
92	0.91	0.03						
93	0.88	0.09						
95	0.90	0.02						
97	0.75	0.00			12	0		
100	0.87	0.03						
101	0.96	0.03						
102	0.91	0.02						
104	0.92	0.05						
105	0.86	0.02						
106	0.86	0.02						
107	0.94	0.04	0.1	0.2				
108	0.38	0.09	16	5	16	4		
111	0.94	0.02						
115	0.94	0.04						
116	0.95	0.03						
118	0.95	0.04						
119	0.93	0.03	0.8	0.5				
120	1.00	0.04						
121	0.84	0.03	1.9	0.4			0.92	0.03
122	0.65	0.06	1.4	0.3				
124	0.35	0.01	1.19	0.03	9.2	0.6		

Table S4. **Order parameters (S^2), timescales of motions (τ_e), exchanges contributions (k_{ex}), slow timescale motions (S^2_{slow}), and their associated errors (err) obtained for Spy during the analysis of ^{15}N spin relaxation (800 MHz)**

Residue	S^2	S^2 err	τ_e	τ_e err	k_{ex}	k_{ex} err	S^2_s	S^2_s err
31	0.77	0.03			8	2		
33	0.91	0.14						
34	0.84	0.05						
36	0.91	0.14						
37	0.84	0.08						
39	0.86	0.05						
40	0.83	0.03						
41	0.81	0.04						
42	0.87	0.04						
45	0.83	0.04						
46	0.83	0.03						
47	0.92	0.04						
51	0.81	0.16						
53	0.68	0.08	0.8	0.9				
55	0.58	0.05	2.0	0.6	19	1		
58	0.86	0.09						
60	0.78	0.04	1.6	0.2			0.79	0.04
62	0.92	0.07						
63	0.89	0.04						
64	0.77	0.03			4	1		
65	0.89	0.06						
66	0.86	0.05						
67	0.82	0.05						
68	0.85	0.08						
69	0.85	0.05						
70	0.75	0.03						
71	0.92	0.07						
72	0.83	0.04						
73	0.81	0.09						
75	0.84	0.08						
76	0.74	0.03			8	2		
77	0.92	0.07						
78	0.88	0.05						
80	0.87	0.05						
81	0.83	0.05						
89	0.91	0.06						
91	0.94	0.03						
92	0.86	0.05						
93	0.99	0.15						
94	0.84	0.06						
95	0.82	0.05						
96	0.87	0.06						
99	0.77	0.10						
100	0.83	0.08						
101	0.95	0.07						
103	0.89	0.06						
104	0.87	0.09						
105	0.82	0.05						
106	0.78	0.08						
107	0.82	0.07						
108	0.40	0.12	19	8	19	6		
110	0.74	0.09	0.02	0.02	5	3		
111	0.86	0.03						
112	0.85	0.05						
115	0.90	0.08						
116	0.94	0.06						
117	0.89	0.04						
118	0.86	0.04	0.01	0.00				
119	0.84	0.06	0.2	0.2				
120	0.96	0.03	0.8	0.7				
121	0.83	0.11	3	8			0.80	0.05
122	0.71	0.11	1	6			0.81	0.06
123	0.66	0.06	1	2			0.79	0.05
124	0.44	0.01	1.5 0.1	10.6	0.1			

Table S5. Accuracy of the chemical shift data reproduction as seen using LARMOR$^{C\alpha}$

Nucleus	Spy MAE (ppm)	LARMOR$^{C\alpha}$ MAE (ppm)
HN	0.36	0.25
Cα	0.90	0.90
CO	0.90	1.03
C β	1.09	1.55
N	1.74	2.45

Mean absolute errors (MAE) correspond to the data reproduction for Spy chemical shifts using experimentally measured chemical shifts and the present coarse-grained trajectories. LARMOR$^{C\alpha}$ MAE corresponds to the error estimated for the predictor when tested on a dataset of 61 proteins for which both high-quality X-ray structures and chemical shifts are available (see ref[5] for details). MAE is defined as $MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\delta_i^{calc} - \delta_i^{exp}\right|$, where δ_i is the chemical shift for a given residue *i* in simulation (*calc*) and experiment (*exp*).

Test of the approach on known complexes

To test the validity and the potential general use of our approach, we selected two model systems for which the necessary input information (structural model and carbon chemical shifts of the individual components and the binding affinity of the complex) is available as well as a detailed description of the corresponding complex.

The pKID-KIX system.

Introduction. The first interaction we considered is between the phosphorylated kinase-inducible domain (pKID) of the cAMP response-element binding protein (CBP) and the KID-binding domain (KIX) of the CREB binding protein. The pKID–KIX system has been extensively studied with NMR by Wright and co-workers[9-11] and by coarse-grained simulations in which the intermolecular pKID-KIX interactions are encoded into a Gō-like model.[12,13] While pKID is mainly unfolded on its own, it folds upon binding with the structured KIX domain. The process of complex formation has been described as going through the formation of a first encounter complex, that progressively evolves towards the fully bound state without dissociating.[11]

Available Data and Simulation Procedure. The chemical shifts of free pKID were obtained from BMRB entry 6788.[9] We then built a Gō-like model of pKID using the coordinates from its structure in complex with KIX (PDB ID: 1KDX).[10] To realize a dynamic ensemble consistent with the partially disordered nature of free pKID,[9] the native intra-molecular interactions in this model were renormalized using a similar optimization procedure as described for the $\mathrm{Im7}_{6\text{-}45}$ substrate in the main text (see Methods) and as highlighted below in **Figure S15**. The KIX domain does not undergo large-scale conformational changes upon complex formation, and thus the starting coordinates for KIX were also taken from the 1KDX complex.[10] The unoptimized model of the well-structured KIX domain yielded an ensemble in excellent agreement with its structure in the NMR ensemble of the complex,[10] and thus no further optimization was pursued for KIX.

For the pKID-KIX binding simulations, we ran an initial simulation that yielded a binding affinity of ~0.2 mM. We then used this ensemble with an efficient reweighting procedure to uniformly scale the strength of the intermolecular interactions.[14] This approach yielded binding affinities of 39 μM and 21 μM for the optimized and unoptimized binding simulations, respectively, in close agreement with the experimental value of 31 μM.[15] To track interface formation during simulation, we defined an interface root-mean-square deviation (RMSD) of Cα coordinates for residues that participate in intermolecular interactions in a Gō-like model[16] built for the pKID-KIX complex.

Results and Discussion. We evaluated the ability of our optimized simulations to delineate the binding mechanism of the IDP-partner recognition. We characterized the binding mechanism by using the degree of native complex formation (interface RMSD from the best NMR model) and IDP folding level (pKID RMSD from its bound form) as progress variables to construct two-dimensional free energy surfaces (**Figure S16A-C**). When using the original Gō-like model to describe pKID in the binding simulations, two bound states are observed: native-like bound and folded complexes (N_B) and non-specific, intermediate complexes (I_B) (**Figure S16A**). However, the unbound state of pKID remains either fully or partially folded (F and I_F, respectively), inconsistent with experiment.[9,10] Thus, a coupled folding-and-binding mechanism is not observed. We then applied a backbone dihedral adjustment to pKID that may better account for the unfolded nature of IDPs.[17] With this model, we observe just two stable conformational basins: unbound and unfolded pKID (U) and non-specific complexes (I_B), while native-like complexes (low interface RMSD and low pKID RMSD) are unstable (**Figure S16B**). Upon optimization, the binding simulations clearly show a three-state coupled folding-and-binding mechanism that progresses from unfolded monomers (U) to a non-specific complex (I_B), and then to native-like complexes (N_B) (**Figure S16C**). This mechanism is consistent with NMR measurements that suggest unfolded pKID first forms a non-specific encounter complex that evolves toward a fully bound and folded state without dissociating from the surface of KIX.[11] Thus, of the different models that were tested, the optimized binding simulations yield the most accurate mechanism of flexible recognition for pKID-KIX.

The dominant interaction region from the optimized binding simulations encompasses the majority of the pKID sequence and the C-terminus of KIX, in good agreement with the NMR ensemble (**Figure S16D-E**). A smaller interaction region in the NMR ensemble that involves the C-terminus of pKID (residues 140-145) with the N-terminus of KIX (residues 599-608) is infrequently sampled during the binding simulations. pKID still significantly interacts with the N-terminus of KIX in simulation, but with residues located more toward the middle of pKID sequence. Interestingly, we also observe a region of non-specific interaction, which is consistent with NMR measurements.[11] The ensemble of native-like poses extracted from the full ensemble shows a pKID-KIX interaction pattern (**Figure S16F**) that is highly similar to the contact map constructed for the NMR ensemble (**Figure S16E**), indicating a convergent description of the native complex.

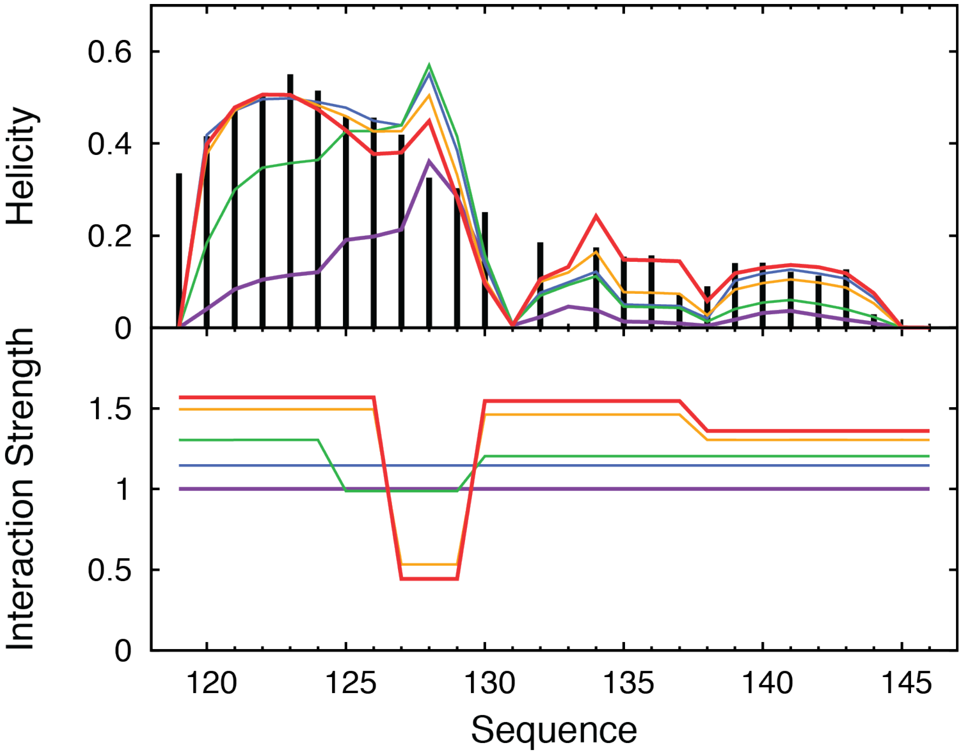

Figure S15. Renormalization of interaction strengths in the Gō-like model for pKID based upon the NMR-determined SSP score. At each iteration, the scaling factors of the interactions in the force field (lower graph) are reweighted to improve reproduction of the experimental SSP profile (upper graph): initial values (violet) and iterations 1 (blue), 2 (green), 3 (orange), 4 (red), and the experimental targeted SSP profile (black bars). Iterative optimization of the interaction strengths results in a force field that reproduces the experimental profile. As the initial trajectories underestimated the helicity in the entire peptide, the groups of residues used to determine the number of parameters in the reweighting procedure were defined progressively during the optimization.

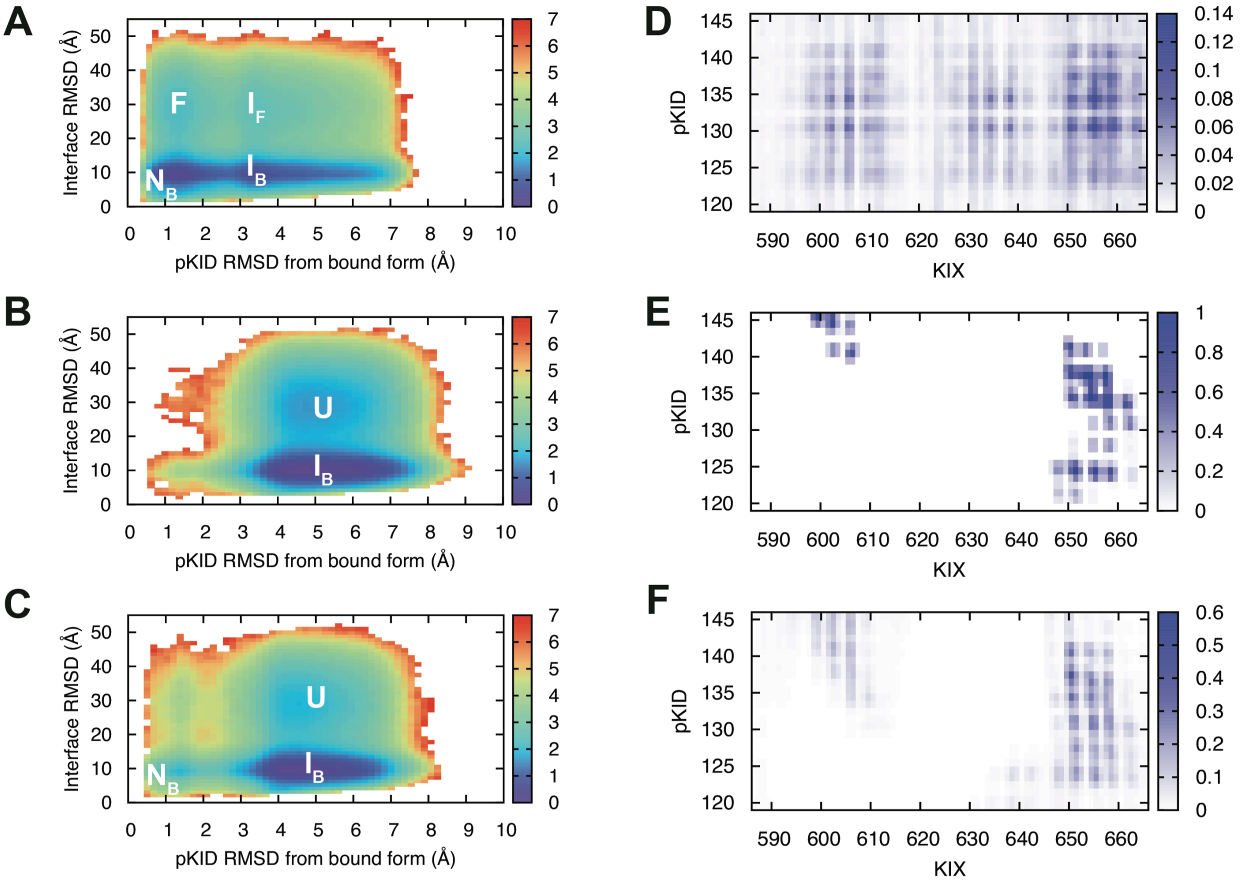

Figure S16. Binding mechanism and interaction mode of the pKID-KIX complex. The left-hand side of the figure shows free energy surfaces that report on the binding mechanism obtained from different pKID-KIX binding simulations: (A) pKID treated with the original Gō-like model (i.e., unoptimized without the backbone dihedral adjustment), (B) unoptimized, and (C) optimized. The surfaces are constructed in the plane of the interface RMSD (i.e., interface formation) and the RMSD of pKID with respect to its bound form (i.e., folding level of pKID). The labels signify different conformational basins: N_B (native-like complexes); I_B (intermediate bound state/non-specific complex); F (unbound and folded pKID); I_F (unbound and partially folded pKID); U (unbound and unfolded pKID). The surfaces are computed at the simulation temperature of 300 K and the colorbars denote the free energy in units of kcal/mol. The right-hand side shows the contact maps obtained from (D) the optimized binding simulations, (E) the pKID-KIX NMR ensemble,[10] and (F) the native-like complexes (interface RMSD $\leq$ 5 Å from the best NMR model) sampled during the optimized binding simulations. The maps are determined as described in the Methods of the main text.

The PX-Ntail system.

Introduction. The second test system we chose to investigate is the complex between the partially folded C-terminal domain of the nucleoprotein of the Sendai virus (Ntail) and the C-terminal subdomain of the phoshoprotein (PX). This system has been extensively studied, especially with NMR by Blackledge and co-workers.[18-20] By itself Ntail undergoes conformational exchange between three helices of different sizes and an unfolded state.[19] PX is well folded and consists of three α-helices.[18] NMR measurements on Ntail-PX complex formation indicate that upon interaction, one of the Ntail helices is stabilized and that the recognition first proceeds through a non-specific and dynamic encounter complex. The encounter complex then locks into the interaction site, mainly mediated through interactions with the third helix of PX.[20] In a joint NMR and X-ray crystallography study of the closely related Measles virus, chemical shift changes indicate that the third helix of PX is the primary interaction site, while a crystal structure of a chimeric construct of the Ntail-PX complex suggests that the interaction between the two partners is mediated by both the second and third helices of PX.[21]

Available Data and Simulation Procedure. For an initial structural model of Ntail, we built an alpha helix using CHARMM[22] for residues 476-492, which encompasses the three helical states observed in the free ensemble.[19] We added five residues to both the N- and C-terminus of this sequence using MODELLER[23] to capture the steric effect of the immediately neighboring disordered residues (i.e., these residues do not participate in any intra- or intermolecular attractive interactions in the model). The chemical shifts of Ntail (BMRB entry number 15123)[24] were then used in the Ntail force-field optimization in order to recapitulate the experimentally determined SSP score of this system (**Figure S17**). A weak harmonic restraint was applied to the change in interaction strengths for this system in order to avoid overly strong rescaling as observed in the primary optimization (data not shown). The structure of PX used to build the Gō-like model was taken from the best model of its NMR ensemble (PDB ID: 1R4G).[18] The interactions in this model were uniformly strengthened by a factor of 1.2, which yielded an SSP profile in good agreement with experiment (data not shown).

For the Ntail-PX binding simulations, we ran an initial simulation that yielded a binding affinity of ~0.5 mM. We reweighted[14] this ensemble to achieve binding affinities of 6.5 μM and 8.5 μM for the optimized and unoptimized binding simulations, respectively, which agree well with the experimental binding affinity determined by isothermal titration calorimetry (8.4 μM).[20] In order to compare the Ntail-PX binding pathway from simulation with the mechanism of complex formation characterized by NMR,[20] we measured two reaction coordinates: 1) the inter-helical angle formed between PX helix 3 and the helical region of Ntail and 2) the distance between the centers of mass of these two helices.

Results and Discussion. To investigate the quality of our initial model for Ntail, we characterized the nature of the helices formed in its free state. Previously, NMR Residual Dipolar Couplings (RDCs)[19] measured for this system showed that Ntail exchanges between four states: three helices of respective positions and populations (H1: 479-484, 36%; H2: 476-488, 28%; H3: 478-492, 11%) and an unfolded state (U) with 25% population. Using the initial force field, we observe the three helical states, but with poorly estimated populations (i.e., H2 is the most populated state, followed by H3 and H1) (**Figure S18**). This incapability in predicting the relative populations of the three helices stems primarily from problems in sampling H1, which translates into an under- and over-estimation of the helicity in the 479-484 region and the 486-490 region, respectively (**Figure S17**). After optimization, the relative populations of the helices show the same trend as the experimental populations, albeit with a slight overestimation of the H1 and H2 populations (**Figure S18**). This minor issue could likely be solved by allowing for more scaling parameters in the force-field optimization. It is worth noting that the nature of the helices (length and position) is not directly encoded into the chemical shifts, which report only on local conformational propensity. Therefore, the improvement observed in the reproduction of the helical populations suggests that the optimization procedure does provide a significantly improved force field.

Upon Ntail-PX binding, it is possible to observe a clear reweighting of the different helical populations. A strong decrease in H1 is balanced by an increase in the H2 and H3 populations, corresponding to an interaction in which PX binding induces an overall folding of Ntail at the interface. A similar trend was observed by NMR relaxation dispersion measurements,[20] where the Ntail H2 helical state was selected during the interaction. Thus, in both simulation and experiment, the folding of Ntail is observed upon interaction with PX.

We also investigated the mechanism of interaction of Ntail and PX using our optimized force field (**Figure S19**). NMR relaxation dispersion measurements[20] revealed the presence of a highly populated, heterogeneous encounter complex en route to specific binding. Fully consistent with this mode of recognition, binding during simulation occurs from the unbound Ntail ensemble (U) through a broad basin corresponding to a non-specific, encounter complex (E) that precedes the formation of a well-defined, specific complex (S). The NMR measurements also suggest that the selection of Ntail helices similar to the H2 state already occurs in the encounter complex.[20] In agreement with this observation, our model shows similar Ntail helical populations in both the encounter and specific complexes (**Figure S18**). Additionally, our model estimates that the encounter and specific complexes comprise 69% and 31% of the total population, in close agreement with the populations obtained from the NMR measurements (74% and 26%). NMR relaxation dispersion experiments suggest that the interaction primarily involves PX-H3 and the groove between PX-H3 and PX-H2.[20,21] We observed a similar pattern of recognition, in which the most significant interaction occurs with PX-H3, albeit with less extensive contacts involving PX-H2. We note that our model cannot distinguish whether the Ntail and PX-H3 helices are parallel or anti-parallel (both are observed in the trajectory with similar populations), but this point also remains experimentally unresolved.

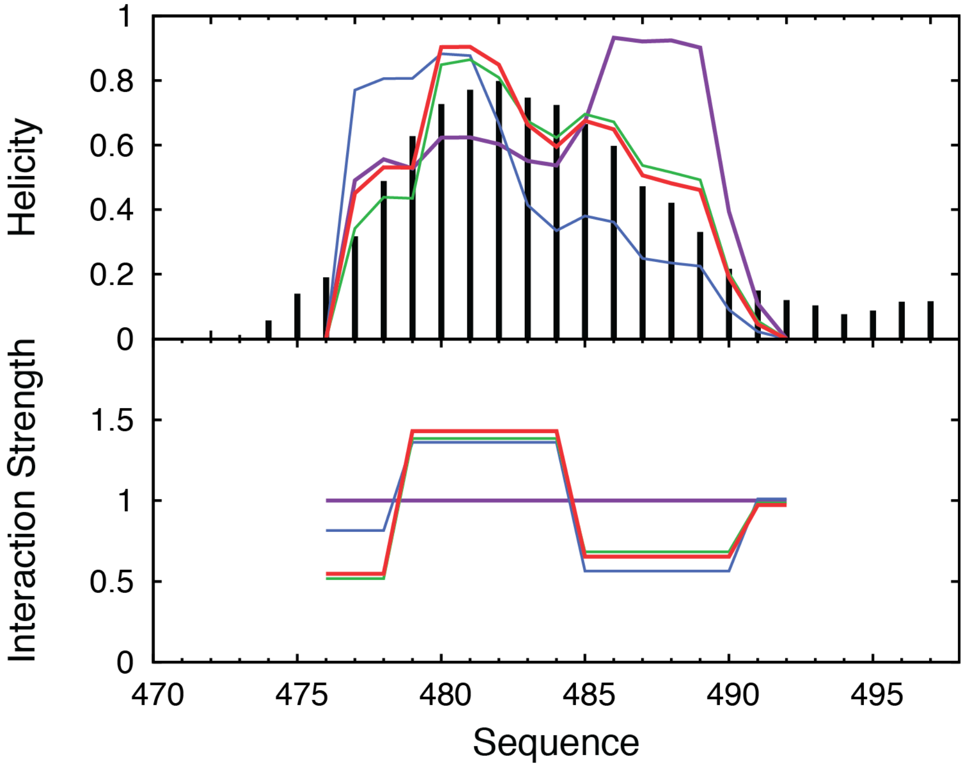

Figure S17. Renormalization of interaction strengths in the Gō-like model for Ntail based upon the NMR-determined SSP score. At each iteration, the scaling factors of the interactions in the force field (lower graph) are reweighted to improve reproduction of the experimental SSP profile (upper graph): initial values (violet) and iterations 1 (blue), 2 (green), 3 (red), and the experimental targeted SSP profile (black bars). Iterative optimization of the interaction strengths results in a force field that closely matches the experimental profile (top panel). Residues 471 to 475 and 493-497 (the first and last set of five residues in the sequence) were not optimized, as they were only included for steric effects.

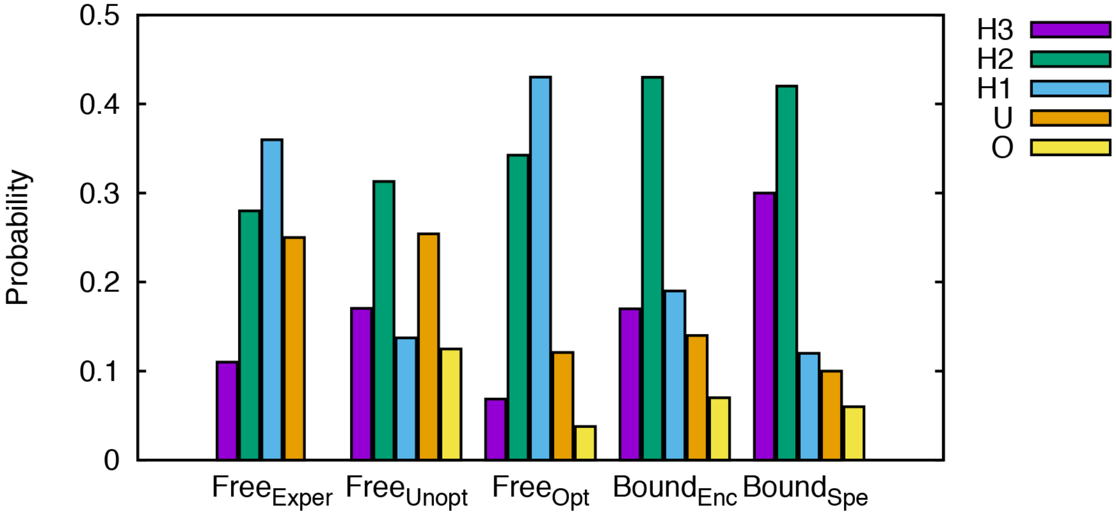

Figure S18. Population of the different helical states of Ntail. Conformers of Ntail are defined as one of the three helical states (H3, H2, or H1)[19] as follows: H3 if they have at least 13 contiguous helical residues in the 478-492 region of the Ntail sequence; H2 if they contain a contiguous helix of at least 9 residues in the 476-488 region and coil elsewhere; and H1 if at least one helical turn is present in residues 479-484 and the conformer is largely coil elsewhere. The remaining snapshots are defined as unfolded (U; five or less helical residues over the entire sequence) or some other state (O) that exhibits significant residual helical structure but does not correspond to H3, H2, or H1. From left to right, the sets of bars correspond to the populations derived from the following data: experiments of free Ntail (Free$_{Exper}$);[19] the un-optimized (Free$_{Unopt}$) and optimized (Free$_{Opt}$) simulations of free Ntail; and the encounter (Bound$_{Enc}$) and specific (Bound$_{Spe}$) ensembles from the optimized Ntail-PX binding simulations.

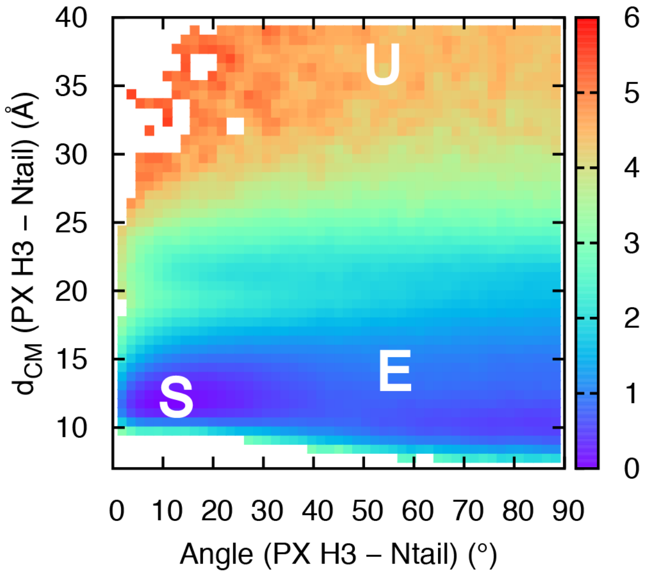

Figure S19. Ntail-PX binding pathway. The two coordinates used for constructing this free energy surface are the distance between the centers of mass (d_{CM}) of the third helix of PX (PX-H3) and the Ntail helical region, and the inter-helical angle between best-fit vectors defined through these two moieties. This angle is restricted between 0° and 90° such that it reports on whether the two partners are oriented parallel or perpendicular with respect to one another, but does not distinguish between parallel and anti-parallel. The surface shows three distinct conformational states: the specific complex (S; $d_{CM} \leq 15$ Å and angle $\leq 30°$), the encounter complex (E, $d_{CM} \leq 25$ Å and excluding the specific region), and the unbound ensemble (U; $d_{CM} >$ 25 Å). The surface is computed at 300 K and the colorbar denotes the free energy in units of kcal/mol.

References

(1) Quan, S.; Koldewey, P.; Tapley, T.; Kirsch, N.; Ruane, K. M.; Pfizenmaier, J.; Shi, R.; Hofmann, S.; Foit, L.; Ren, G.; Jakob, U.; Xu, Z.; Cygler, M.; Bardwell, J. C. *Nat. Struct. Mol. Biol.* **2011**, *18*, 262.

(2) Quan, S.; Wang, L.; Petrotchenko, E. V.; Makepeace, K. A.; Horowitz, S.; Yang, J.; Zhang, Y.; Borchers, C. H.; Bardwell, J. C. *eLife* **2014**, *3*, e01584.

(3) Manavalan, P.; Johnson, W. C., Jr. *Anal. Biochem.* **1987**, *167*, 76.

(4) Whitmore, L.; Wallace, B. A. *Biopolymers* **2008**, *89*, 392.

(5) Frank, A. T.; Law, S. M.; Ahlstrom, L. S.; Brooks, C. L., III *J. Chem. Theory Comput.* **2015**, *11*, 325.

(6) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson.* **2000**, *147*, 138.

(7) Dosset, P.; Hus, J. C.; Blackledge, M.; Marion, D. *J. Biomol. NMR* **2000**, *16*, 23.

(8) Horowitz, S.; Salmon, L.; Koldewey, P.; Ahlstrom, L. S.; Martin, R.; Quan, S.; Afonine, P. V.; van den Bedem, H.; Wang, L.; Xu, Q.; Trievel, R. C.; Brooks, C. L., III; Bardwell, J. C. B. *Nat. Struct. Mol. Biol.* **2016**, doi:10.1038/nsmb.3237.

(9) Radhakrishnan, I.; Perez-Alvarado, G. C.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* **1998**, *430*, 317.

(10) Radhakrishnan, I.; Perez-Alvarado, G. C.; Parker, D.; Dyson, H. J.; Montminy, M. R.; Wright, P. E. *Cell* **1997**, *91*, 741.

(11) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021.

(12) Ganguly, D.; Chen, J. *Proteins* **2011**, *79*, 1251.

(13) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. *PLoS Comp. Biol.* **2008**, *4*, e1000060.

(14) Law, S. M.; Ahlstrom, L. S.; Panahi, A.; Brooks, C. L., III *J. Phys. Chem. Lett.* **2014**, *5*, 3441.

(15) Parker, D.; Jhala, U. S.; Radhakrishnan, I.; Yaffe, M. B.; Reyes, C.; Shulman, A. I.; Cantley, L. C.; Wright, P. E.; Montminy, M. *Mol. Cell* **1998**, *2*, 353.

(16) Karanicolas, J.; Brooks, C. L., III *Protein Sci.* **2002**, *11*, 2351.

(17) De Sancho, D.; Best, R. B. *Mol. Biosyst.* **2012**, *8*, 256.

(18) Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W.; Marion, D. *Virology* **2004**, *319*, 201.

(19) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055.

(20) Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D. F.; Ruigrok, R. W.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2015**, *137*, 1220.

(21) Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. *Proc Natl Acad Sci U S A* **2004**, *101*, 8301.

(22) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

(23) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.

(24) Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W.; Blanchard, L. *J. Virol.* **2007**, *81*, 6807.