Supporting Information:

Structure Determination of Challenging Protein-Peptide Complexes Combining NMR Chemical Shift Data and Molecular Dynamics Simulations

Arup Mondal[1], G.V.T. Swapna[2, 3], Maria M. Lopez[3], Laura Klang[3], Jingzhou Hao[3], LiChung Ma[2], Monica J. Roth[2]*, Gaetano T. Montelione[3]*, Alberto Perez[1]*.

[1]The Quantum Theory Project, Department of Chemistry, University of Florida, Gainesville, FL 32611, USA.

[2]Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.

[3]Department of Chemistry and Chemical Biology, Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

*Corresponding authors: Alberto Perez, Gaetano T. Montelione, Monica J. Roth.

Email: Monica Roth: roth@rwjms.rutgers.edu;

Gaetano Montelione: monteg3@rpi.edu;

Alberto Perez: perez@chem.ufl.edu

Supplementary Results:

Transferring ET-TP CSP Data to ET-NSD3: For NSD3 peptide simulations using the CSP+TALOS data guided the system to the native state (IRMSD=2.0, ILRMSD=2.9, f_{nat} (fraction of native contacts) =72.9%). Repeating the simulation using CSP experimental data from the ET-TP system yielded predictions of similar accuracy (IRMSD=2.5 Å, ILRMSD=2.9Å, and f_{nat}=79%). In both simulations the top cluster had a population of 49%.

Modeling CSP Data: We have taken two approaches to modeling the data. In the first one, we create the combinatorics for all peptide residues with *active* protein residues. For the second one we only create combinatorics between hydrophobic pairs in the two sets. Simulations yielded results of similar quality for all peptides. In the case of BRD3-ET:BRG1 only the hydrophobic pairing selections successfully predicted the experimental structure as the top cluster.

Relative Binding Free Energy of TP/NSD3 to BRD3 from MELD Competitive Binding: We performed triplicates of MELD competitive binding simulations. To calculate relative binding free energy, we used relative population of TP and NSD3 bound states in the lowest temperature MELD trajectory using the following equation.

$$\Delta\Delta G_{TP/NSD3} = -RTln(^{p_{TP}}/_{p_{NSD3}}) \qquad (1)$$

where R is the universal gas constant, T is simulation temperature, p_{TP} is population of TP bound states, and p_{NSD3} is the population of NSD3 bound states. The population of TP and NSD3 states at the lowest temperature from three different trials are given below:

	Trial 1	**Trial 2**	**Trial 3**
TP:	40447	36361	30377
NSD3:	852	386	533

Using equation (1), the calculated $\Delta\Delta G_{TP/NSD3}$ for three trials are: -2.28 kcal/mol, -2.69 kcal/mol, and -2.40 kcal/mol where T is the lowest temperature replica (300K) from MELD simulation. We report the average value for the relative binding free energy, $\Delta\Delta G_{TP/NSD3}$= -2.45 ±0.17 kcal/mol.

Additional Systems:

We applied our MELD-NMR pipeline to study three additional peptides that have been experimentally characterized: BRD4-ET: LANA, BRD3-ET: CHD4 and BRD3-ET: BRG1 (PDB IDs are 2ND0, 6BGG and 6BGH respectively, see Table 1). These three peptides bind as a small antiparallel single β-strand. For consistency, we predicted BRD3-ET: LANA and BRD3-ET: JMJD6. The predicted structures were very similar to their BRD4 counterpart.

CSP Dataset: We transferred the *CSP* dataset of the BRD3-ET:TP system as described in the methods section to study these three systems. With this information, MELD predicts the native complex structure correctly for only BRD3-ET: CHD4 system (40% population excluding termini, see Figure S5). The case of BRD3-ET: BRG1 is interesting since it contains a symmetric binding motif and MELD simulations could not differentiate between an antiparallel or parallel β-strand orientation – with the top cluster incorrectly favoring the parallel orientation when using all possible pairings between residues selected as *active* and peptide residues. Using only the hydrophobic residues of the peptide and the protein result in native antiparallel binding mode with top cluster population around 31%. For BRD3-ET: LANA, MELD could not find the peptide's experimental binding site (Figure S5).

CSP + TALOS Dataset: Like JMJD6, the LANA peptide is a weak binder (with binding affinity 635 μM), and indeed TALOS predictions based on chemical shift shows backbones of this peptide to be dynamics, offering no guiding power for the simulations (see Table S5). The chemical shift lists for BRD3-ET:CHD4 (BMRB code 30367) and BRD3-ET:BRG1 (BMRB code 30368) systems lacked information about C, N atoms which are essential for TALOS to predict backbone dihedral data. Thus, for these three systems we used simulated backbone phi and psi data back calculated from the experimental structure. We enforced phi/psi ranges for the 3 complexes as described in the methods section with standard deviation of 40 degrees for each dihedral (see Table S4, S5, and S6). CHD4 and BRG1 continue to be in good agreement with experiments, with an increased population of top cluster. The MELD prediction for BRD3-ET:LANA complex is partially correct, with the C-terminal floppy region folding back to form a hairpin like structure (Figure S5).

CSP + TALOS + NOE Dataset: Our third type of simulations add the three strongest NOEs for each system. We left out BRD3-ET:CHD4 and BRD3-ET:BRG1 as they were already successful and the BMRB entry contained no NOE peak lists. The BMRB entry for BRD4-ET:LANA (BMRB code 26042) did not contain the NOE peak list either, hence we took the three shortest backbone H-H distances back-calculated from the experimental structures (44H - 75H, 42H - 77H, and 44H - 76H). We enforced them in the MELD simulations as described in the methods section. This resulted in higher accuracy predictions, with the top cluster (with population of 48%) now representing the experimental binding mode (Figure S5).

We complemented our study with AlphaFold predictions (see Methods), finding good agreement with experiments for all three systems (Figure S5). Upon comparing these predictions with MELD successful predictions, we realized the IRMSD values are similar for both MELD and AlphaFold2 (AF2) predicted structures, whereas the f_{nat} values are significantly lower for AF2 predicted structures compared to MELD predictions. Upon further inspection of these systems, the registry between the residue pairing in the receptor and peptide strand is shifted by two residues with respect to the experimental structures (Figure S9).

Supplementary Figures:

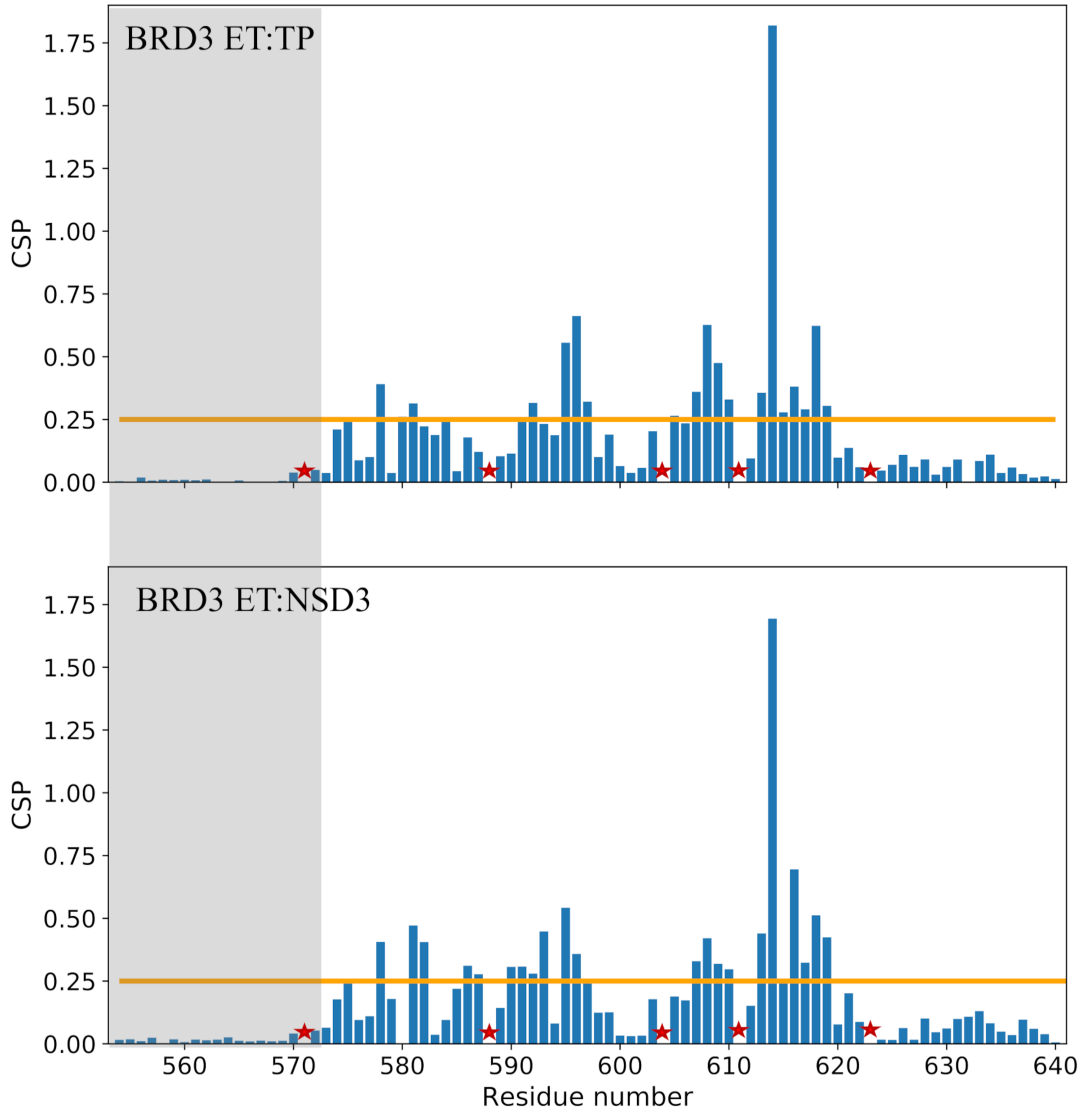

Figure S1. CSP values for each residue along the ET domain represented as bar plots. The orange line shows the threshold used to define *active residues* for the binding process. The shaded region corresponds to the floppy tail in the receptor which was excluded from simulations. The star symbols denote residues for which CSPs could not be measured (e.g. proline). CSP measurement for both BRD3-ET:TP and BRD3-ET:NSD3 were performed at pH=7.0 with a peptide concentration 0.5 μM and a 1:1 ratio of protein to peptide.

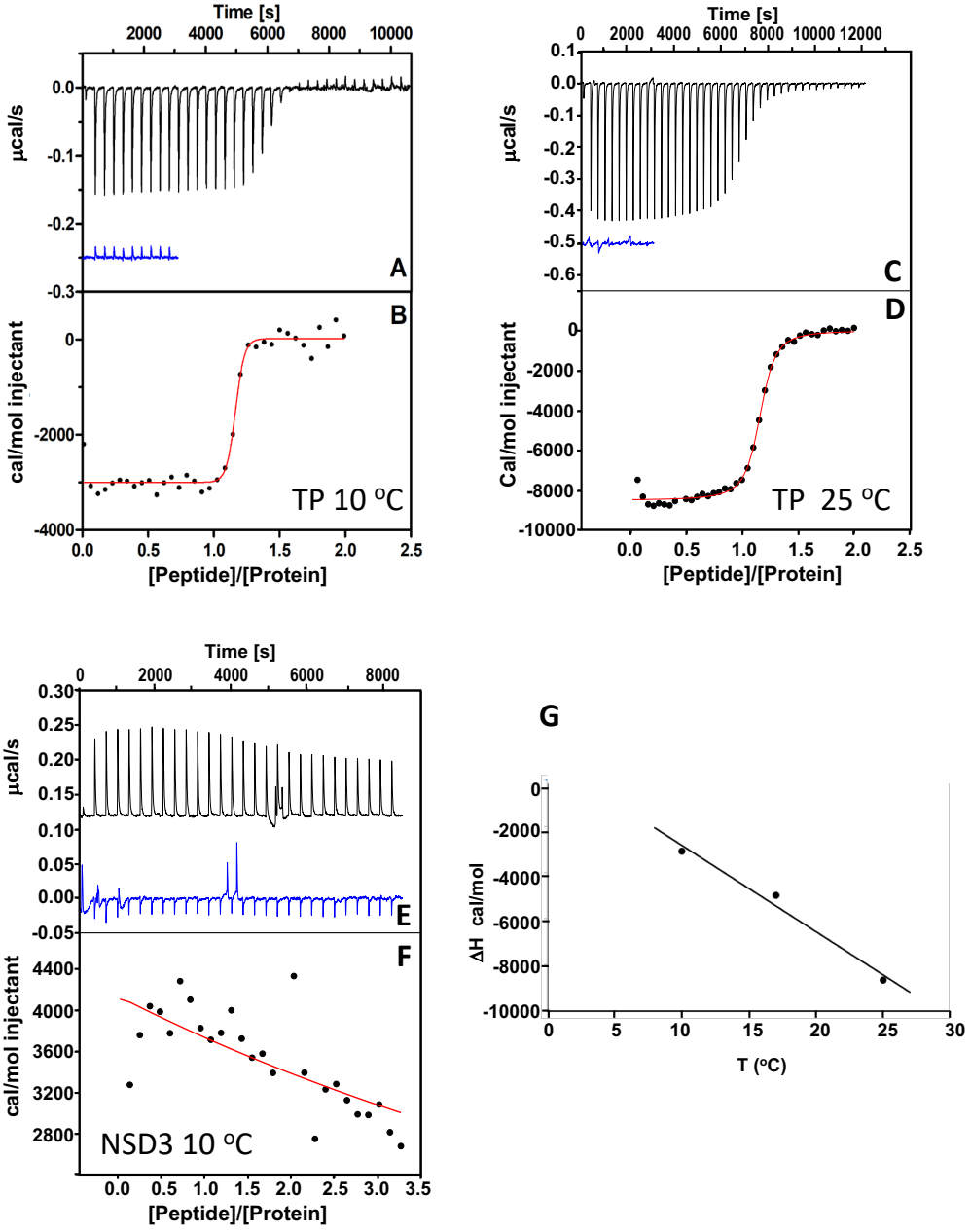

Figure S2. Study of the interaction of the BRD3-ET protein with two peptides by Isothermal Titration Calorimeter. Panels A and B: TP at 10 ºC. The TP (0.277 mM) injected into the BRD3-ET protein (21 µm, black line), and TP (0.277 mM) injected into buffer (blue line) at 10 ºC. The first injection was 1 µL, and all subsequent injections were 6 µL. The protein and peptide solutions were dialyzed in the same buffer (20 mM Tris, 150 mM NaCl, 1mM TCEP, pH 7.5, 24 hours, 2 changes), over a period of 24 hours. Final protein and peptide concentrations were measured by absorbance spectroscopy (λ=280 nm) using extinction coefficients ε_{280} = 4470 and 5500 $M^{-1}cm^{-1}$ for BRD3 ET and IN TP, respectively. The Origin-ITC software was used to calculate the area under the peaks and to subtract the blank injections from the protein titration.

Panel B plots the heat of interaction (cal/mol injectant) as a function of the molar ratio [Peptide]/[Protein]. The data points were fitted to a single set of identical binding sites model according to the equation[1]:

$$\boldsymbol{Q = \frac{nM_t\Delta HV_o}{2}\left[1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{(1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t})^2 - \frac{4X_t}{nM_t}}\right]} \quad (2)[1]$$

Where Q is the total heat effect, V_o is the cell volume, n is the stoichiometry of the interaction, ΔH is the enthalpy of binding, X_t is the total peptide concentration in the cell, K is the association constant, and M_t is the total protein concentration in the cell. The results of the fit are collected in the table below. **Panels C and D:** TP at 25 ºC. Provide the same analysis for BRD3-ET:TP binding at 25 ºC, with injections of IN TP peptide (0.24 mM) into BRD3 ET protein (21 μM, black line) or buffer (blue line). **Panels E and F:** NSD3 at 10 ºC. Injections of the NSD3 Peptide (0.2 mM) into BRD3-ET protein (6 μM) (black line) or buffer (blue line) at 10 ºC. All injections were 5 μL but the first one that was 1 μL. The NSD3 concentration was calculated using its extinction coefficient at 205 nm (ε_{205} = 126,480 M^{-1} cm^{-1}) using literature[2]. The blank injections were subtracted from the titration and the heat of the interaction as a function of the molar ratio is plotted in Panel F. The NSD3 Peptide aggregated at 25 ºC, precluding reliable ITC measurements at that temperature. **Panel G.** The relationship between ΔH and temperature for TP binding ET at three different temperatures (10, 17, and 25 ºC, see Table S10) measured by ITC, indicates a ΔC_p = -375 ± 55 cal/mol-K. For TP, the binding is exothermic ($\Delta H_{binding} < 0$), while for NSD3 it is endothermic ($\Delta H_{binding} > 0$).

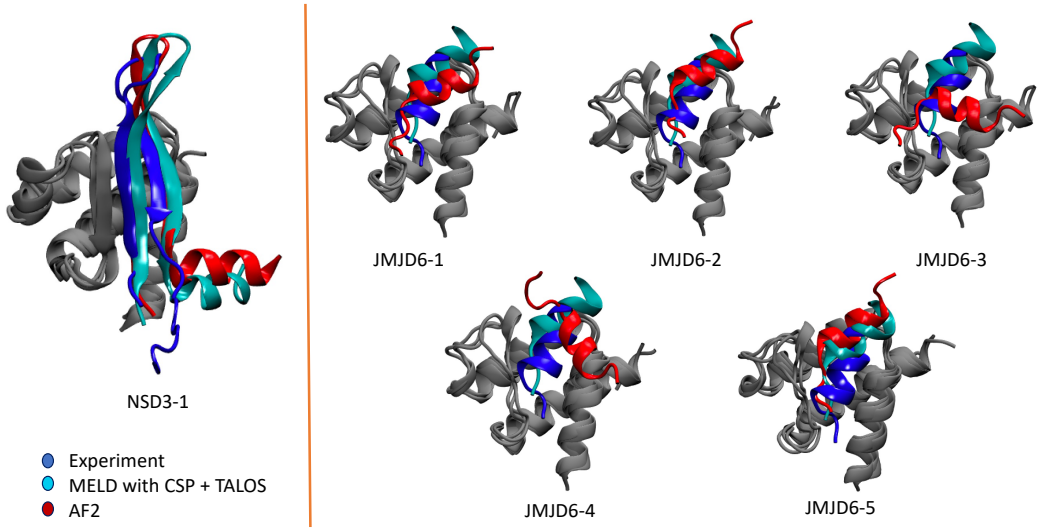

Figure S3. Superposition of the experimental structure, MELD prediction and AlphaFold's best prediction for the BRD3-ET:NSD3 system (left panel). Superposition of the experimental structure, MELD prediction and top five AlphaFold's prediction for the BRD3-ET:JMJD6 system (right panel).

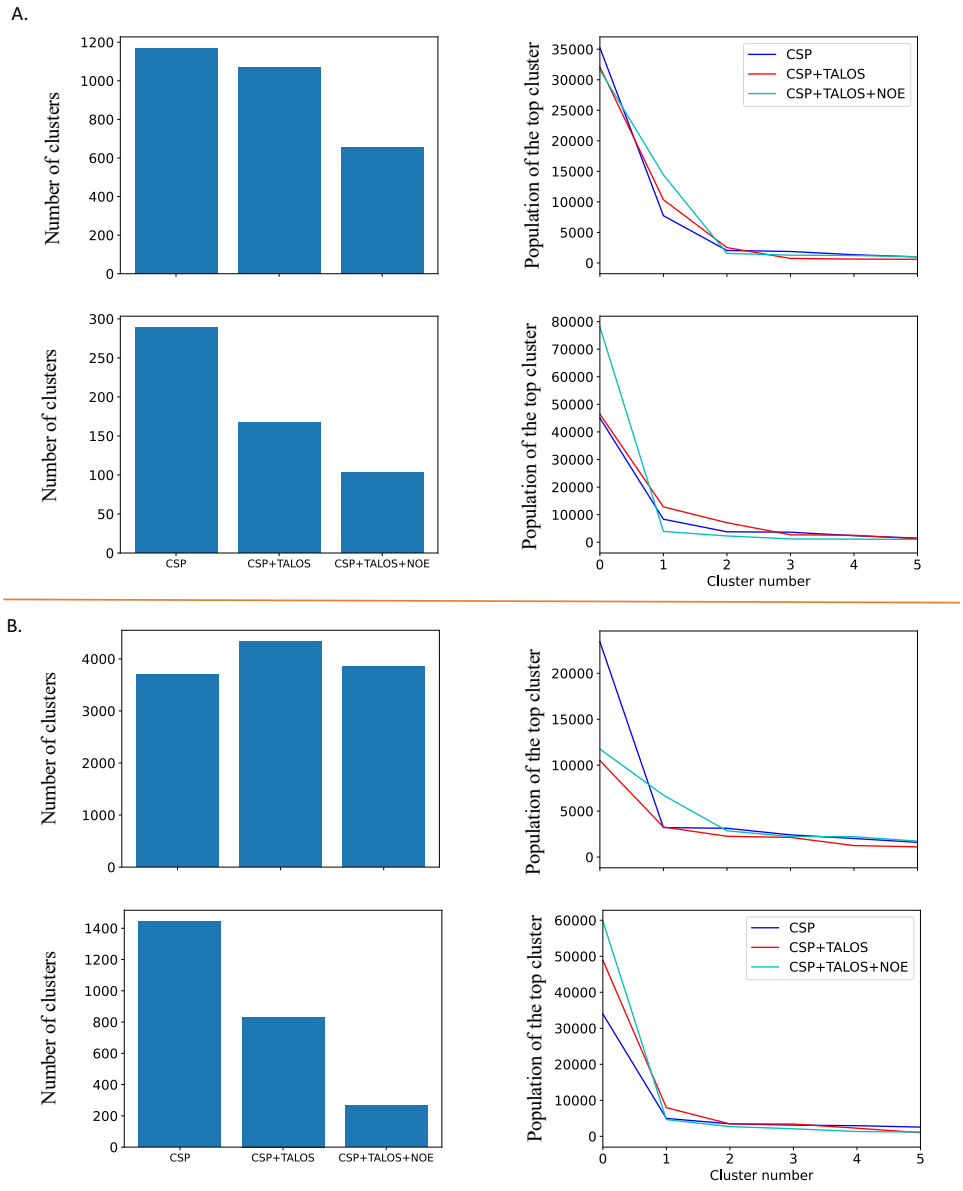

Figure S4. The amount of information used dictates the ability to sample multiple binding modes (left) and identify the native-like one as the highest population cluster (right). (A.) BRD3-ET:TP and (B.) BRD3-ET:NSD3(B.). The top rows of panel A and panel B correspond to clustering over the whole peptide, while the bottom rows exclude the floppy termini (3 residues for TP and 11 residues for NSD3).

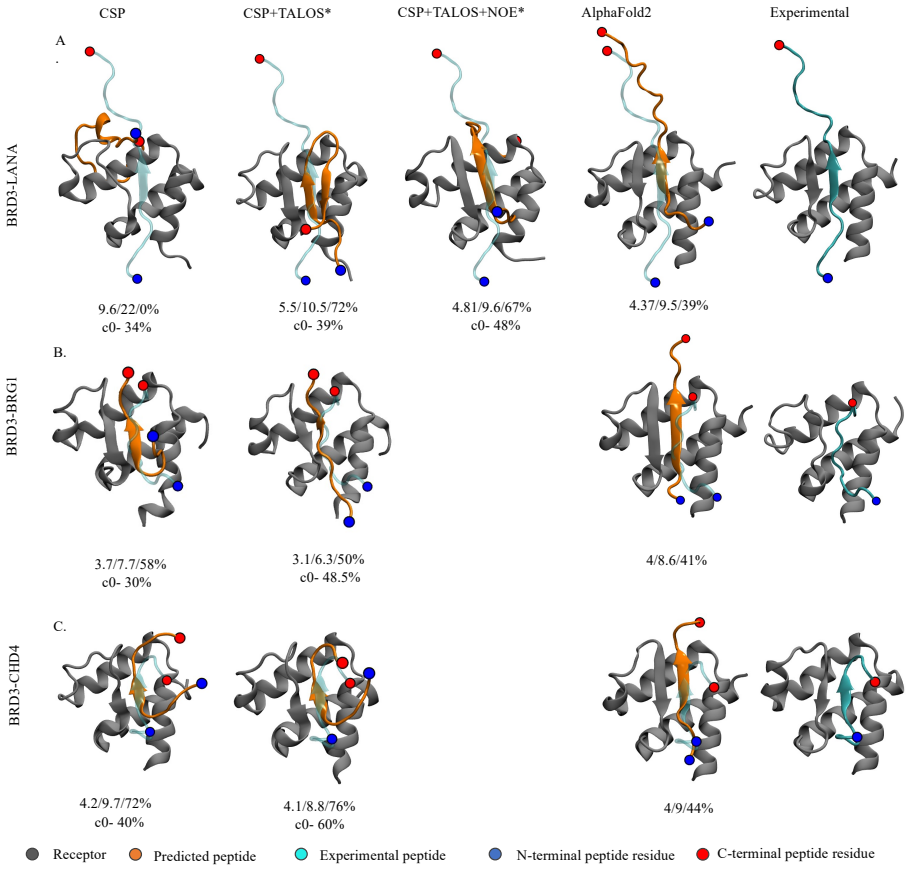

Figure S5. MELD prediction using different dataset (first three columns) for the extended study along with AlphaFold predictions (4th column) and experimental structure (5th column): (A.) BRD3-ET: LANA, (B.) BRD3-ET:BRG1, and (C.) BRD3-ET:CHD4. The numbers stand for IRMSD/ILRMSD/f_{nat} in the first row and the population of the top clusters in the second row.

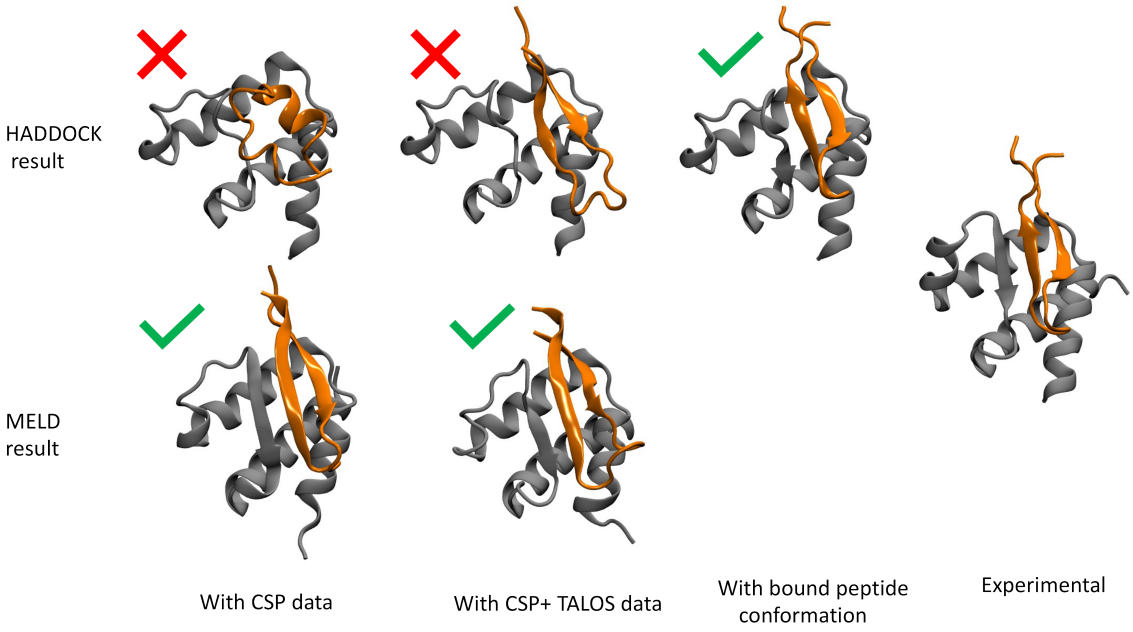

Figure S6. HADDOCK (first row) and MELD (second row) predictions with increased amount of experimental information for the BRD3-ET:TP peptide system. The peptide conformations for HADDOCK originated from either free-peptide or TALOS+free-peptide MD ensembles (1000 structures), or the experimental structure (third column). Using CSP or CSP+TALOS data to guide binding, HADDOCK fails to predict the native structure of the complex. HADDOCK successfully binds the peptide when the bound conformation of the peptide and CSP data are given. MELD correctly predicts experimental conformations in the low information regime, using both CSP (first column), and CSP + TALOS (second column) data.

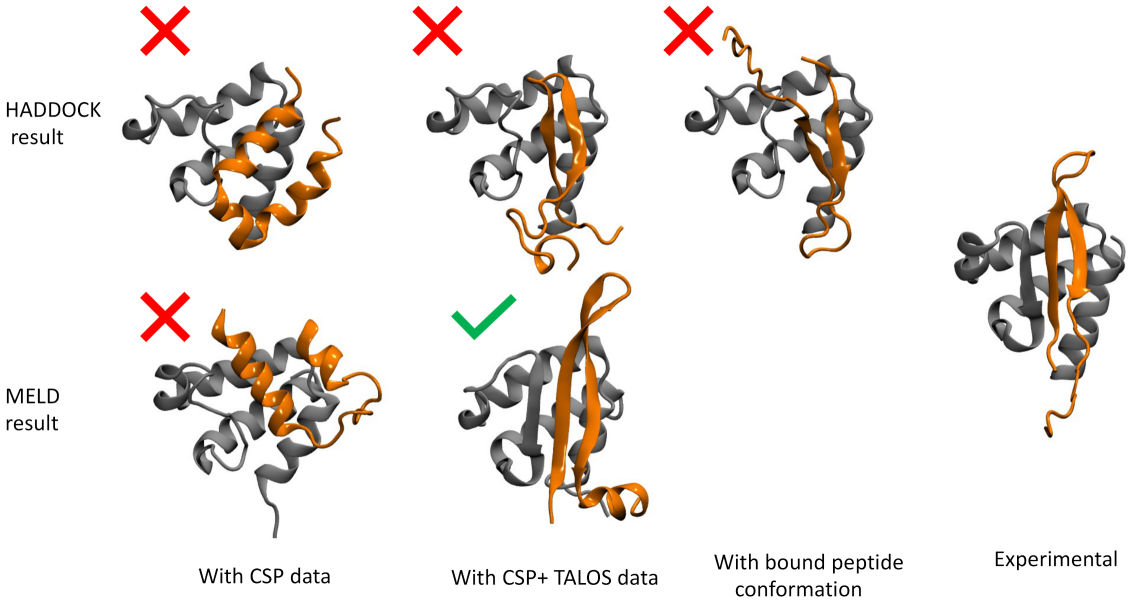

Figure S7. HADDOCK (first row) and MELD (second row) predictions with increased amount of experimental information for the BRD3-ET:NSD3 peptide system. The peptide conformations for HADDOCK originated from either free-peptide or TALOS + free-peptide MD ensembles (1000 structures), or the experimental structure (third column). Using CSP or CSP+TALOS data to guide binding, HADDOCK fails to predict the native structure of the complex. Even with the correct conformation of the peptide (third column), HADDOCK fails to predict the structure of the complex. For this weaker binder, MELD needs CSP + TALOS (second column) data for accurate predictions.

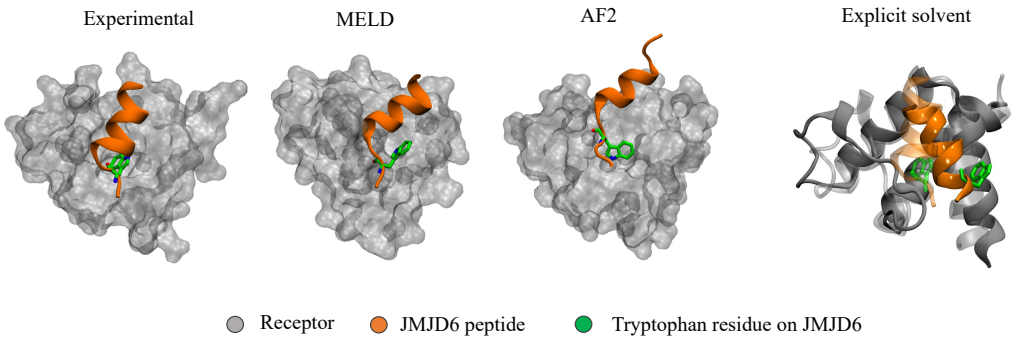

Figure S8. Orientation of tryptophan residues in experiment and simulations. Both implicit and explicit simulations as well as AlphaFold favor conformations in which the tryptophan is not buried in the ET cleft, favoring a change in the binding mode with respect to experiments.

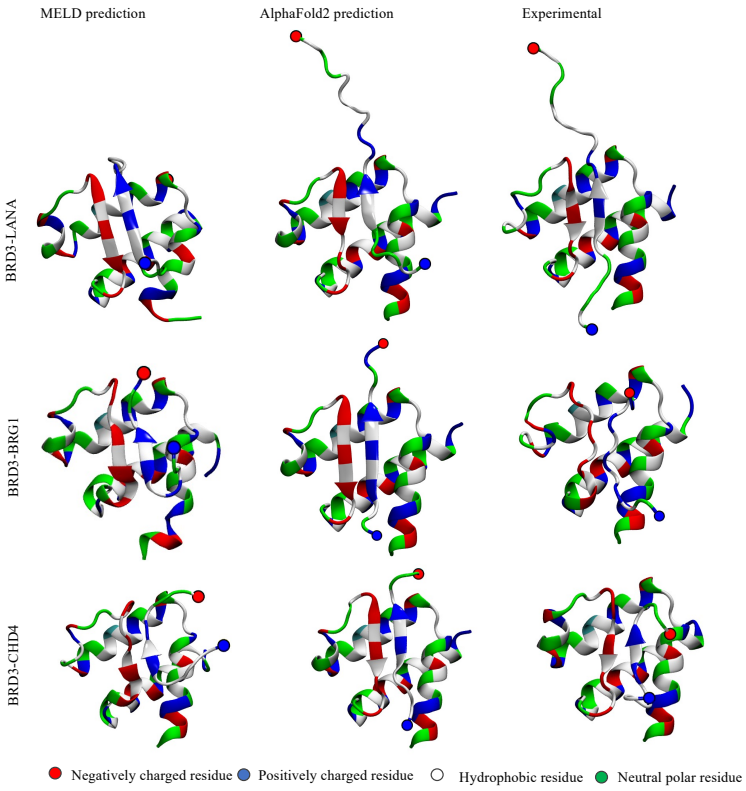

Figure S9. Residue type representation for the top MELD prediction (left column), top AlpahFold prediction (middle column) and the experimental structure (right column).

Supplementary Tables:

Table S1. List of backbone dihedral angle restraint ranges used in MELD for the TP peptide. These dihedral angle restraints were calculated from peptide chemical shift using TALOS+. Peptide residue numbering follow the numbering in MELD simulations (see Table 1). We only used dihedral angle restraint ranges that are categorized as *strong* in the TALOS+ output file.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
72	T	-167.8	-64.6	109.7	173.8
73	W	-151.1	-73.6	93.6	194.8
74	R	-155.6	-103.6	133.5	181.6
75	V	-151.0	-64.6	105.6	145.6
76	Q	-162.0	-54.6	111.2	162.2
77	R	-119.1	-51.4	119.5	159.5
78	S	-147.9	-59.8	96.7	199.0
80	N	-156.2	-79.4	35.6	175.6
83	K	-178.8	-44.3	123.6	178.4
84	I	-152.3	-112.3	131.0	171.0
85	R	-164.2	-98.2	105.1	165.8
86	L	-134.1	-94.1	109.6	164.6
87	T	-139.5	-97.9	111.8	160.4
88	R	-126.2	-79.0	98.8	148.2
89	E	-138.1	-55.4	92.5	157.0
90	A	-140.3	-5.0	119.0	169.2

Table S2. List of backbone dihedral angle restraint ranges used in MELD simulations for NSD3 peptide. These dihedral angle restraints were calculated from peptide chemical shift using TALOS+. Peptide residues number correspond to the residue numbering in MELD simulations (Table 1). We only used dihedral angle restraint ranges that are categorized as *strong* in the TALOS+ output file.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
71	I	-137.9	-57.9	79.4	179.4
72	K	-175.0	-95.0	109.3	209.3
73	L	-176.0	-96.0	90.6	190.6
74	K	-154.4	-74.4	74.8	174.8
75	I	-148.9	-68.9	75.7	175.7
76	T	-159.0	-79.0	70.4	170.4
77	K	-150.9	-70.9	71.6	171.6
78	T	-163.5	-83.5	85.6	185.6
79	I	-150.8	-70.8	73.4	173.4
80	Q	-152.7	-72.7	90.7	190.7
84	E	-131.5	-51.5	76.0	176.0
85	L	-151.9	-71.9	84.5	184.5
86	F	-164.8	-84.8	104.4	204.4
87	E	-183.9	-103.9	101.0	201.0
88	S	-167.0	-87.0	84.6	184.6
89	S	-161.6	-81.6	100.8	200.8

Table S3. List of backbone dihedral angle restraint ranges used in MELD simulations for JMJD6 peptide. Dihedral angle restraints for each residue in the peptide are calculated using based on the experimentally solved NMR ensemble (6BNH) using MDTraj. In MELD simulations, we used dihedral angle restraint ranges only for the residues that a have narrow distribution of dihedral values for 20 NMR structures.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
70	W	-181.1	-101.1	-4.0	76.0
71	T	-118.4	-38.4	-207.4	-127.4
72	L	-82.5	-2.5	-77.5	2.5
73	E	-95.7	-15.7	-110.4	-30.4
74	R	-104.9	-24.9	-97.1	-17.1
75	L	-104.8	-24.8	-103.4	-23.4
76	K	-102.2	-22.2	-77.1	2.9
77	R	-122.0	-42.0	-53.1	26.9
78	K	-136.7	-96.7	-54.9	25.1
79	Y	-129.3	-49.3	-97.1	-17.1
80	R	-116.1	-36.1	-63.0	17.0

Table S4. List of backbone dihedral angle restraint ranges used in MELD simulations for LANA peptide. Dihedral angles restraints for each residue in the peptide are calculated using based on the experimentally solved NMR ensemble (2ND0) using MDTraj. In MELD simulations, we used dihedral angle restraint ranges only for the residues that have narrow distribution of dihedrals for 20 NMR structures.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
74	I	-116.8	-36.8	80.0	160.0
75	V	-140.8	-60.8	109.8	189.8
76	K	-174.1	-94.1	87.8	167.8
77	F	-180.6	-100.6	111.6	191.6
78	K	-137.5	-57.5	91.5	171.5

Table S5. List of backbone dihedral angle restraint ranges used in MELD simulations for BRG1 peptide. Dihedral angle restraints for each residue in the peptide are calculated using based on the experimentally solved NMR ensemble (6BGH) using MDTraj. In MELD simulations, we used dihedral angle restraint ranges only for those residues that have narrow distribution of dihedral values for 20 NMR structures.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
71	V	-136.6	-56.6	-	-
72	K	-	-	76.7	156.7
73	V	-124.3	-44.3	-	-
74	K	-174.2	-94.2	42.1	122.1
75	I	-120.0	-40.0	-	-
76	K	-	-	48.8	128.8
77	L	-117.5	-27.5	-	-

Table S6. List of backbone dihedral angle restraint ranges used in MELD simulations for CHD4 peptide. Dihedral angle restraints for each residue in the peptide are calculated using based on the experimentally solved NMR ensemble (6BGG) using MDTraj. In MELD simulations, we used dihedral angle restraint ranges only for those residues that have narrow distribution of dihedral values for 20 NMR structures.

Residue number	Residue name	PHI-min	PHI-max	PSI-min	PSI-max
70	V	-171.0	-91.0	97.6	177.6
72	P	-	-	98.6	218.6
73	L	-111.4	-31.4	55.2	135.2
74	K	-180.0	-100.0	-	-
75	I	-194.5	-114.5	95.5	175.5
79	G	-134.0	-54.0	-	-
80	F	-168.6	-88.6	-	-

Table S7. List of predicted backbone dihedral angles from TALOS-N for bound JMJD6 based on backbone chemical shift data.

RESID	RESNAME	PHI	PSI	DPHI	DPSI	DIST	S2	COUNT	CS_COUNT	CLASS
69	K	-	-	0	0	0	0	0	6	None
70	W	-67.0	136.59	12.87	6.83	1.403	0.27	6	10	Dyn
71	T	-104.2	145.97	31.86	18.85	0.942	0.37	25	11	Dyn
72	L	-64.76	-30.81	5.85	7.20	0.838	0.5	25	12	Dyn
73	E	-68.50	-21.25	6.82	7.79	0.810	0.635	25	12	Strong
74	R	-74.68	-20.76	7.98	8.75	0.719	0.615	25	12	Strong
75	L	-70.98	139.46	8.27	7.07	0.923	0.558	9	12	Dyn
76	K	-75.95	-20.43	14.67	20.61	2.724	0.486	1	12	Dyn
77	R	-69.14	-33.55	17.70	16.91	0.000	0.431	25	12	Dyn
78	K	-76.62	-30.43	17.55	18.25	-0.90	0.388	7	12	Dyn
79	Y	-97.77	114.36	24.77	41.84	0.000	0.318	10	12	Dyn
80	R	-75.4	-26.96	18.85	19.40	0.000	0.258	10	12	Dyn
81	N	-	-	0	0	0	0	0	8	None

Table S8. List of predicted backbone dihedral angles from TALOS-N for bound LANA based on backbone chemical shift data.

RESID	RESNAME	PHI	PSI	DPHI	DPSI	DIST	S2	COUNT	CS_COUNT	CLASS
69	N	-	-	0	0	0	0	0	9	None
70	L	-69.82	137.21	9.93	9.77	0.507	0.39	25	14	Dyn
71	Q	-68.42	136.06	6.11	8.72	0.442	0.60	25	15	Warn
72	S	-72.85	144.04	9.67	9.36	0.385	0.61	10	15	Warn
73	S	-70.22	143.27	6.99	13.73	0.387	0.54	9	15	Dyn
74	I	-78.00	135.29	15.37	14.47	0.446	0.39	10	15	Dyn
75	V	-77.80	133.68	15.11	14.96	0.438	0.39	25	15	Dyn
76	K	-77.92	126.97	7.19	16.86	0.475	0.45	25	15	Dyn
77	F	-68.58	-25.88	12.28	14.11	0.639	0.67	8	14	Warn
78	K	-64.61	-26.75	6.19	8.42	0.665	0.75	25	14	Strong
79	K	-69.35	134.23	9.82	10.76	3.211	0.73	9	12	Warn
80	P	-68.01	146.20	6.09	12.90	1.167	0.59	10	13	Dyn
81	L	-74.09	134.85	6.85	11.54	0.439	0.50	25	11	Dyn
82	P	-67.13	144.88	6.70	8.39	0.339	0.44	25	13	Dyn
83	L	-73.4	132.37	11.88	12.06	0.310	0.46	10	13	Dyn
84	T	-78.1	144.66	12.5	14.50	0.356	0.45	10	15	Dyn
85	Q	-69.83	125.78	7.11	11.39	0.443	0.43	25	13	Dyn
86	P	-63.84	148.78	8.6	8.9	0.580	0.41	25	12	Dyn
87	G	-	-	0	0	0	0	0	7	None

Table S9. Experimental NOEs for the ET-JMJD6 system, chosen from NOE peak list in four different categories: protein backbone / peptide backbone, protein methyl sidechain/ peptide backbone, protein backbone / peptide methyl sidechain, and protein methyl sidechain / peptide methyl sidechain. In the second row, we further included peaks coming from single Trp70 residue in peptide. Residue numbering corresponds to the residue range in MELD (Table 1).

	Backbone-Backbone	Sidechain-Backbone	Backbone - Sidechain	Methyl - Methyl
Experimental NOE with ILVA methyl	Not present	20 HD21-75HA	13 HA - 75 HD11 9 HA - 75 HD11	12 HD21 - 75 HG21 20 HD11 - 75 HD21 23 HD11 - 75 HG21
Experimental NOEs with ILV methyl and Trp70	Not present	20 HD21-75HA	13 HA - 75 HD11 9 HA - 75 HD11 6 HA - 70 HH2	12 HD21 - 75 HG21 20 HD11 - 75 HD21 44 HD12 - 70 HH2

Table S10. Thermodynamic properties for TP and NSD3 binding to BRD3-ET, calculated from ITC experiment shown in Figure S2 at pH 7.5.

Peptide	Temp (°C)	n	K_d (nM)	ΔH (kcal/mol)
TP	25	1.1	90 ± 10	- 8.64 ± 0.03
TP	17	1.1	37 ± 10	- 4.98 ± 0.06
TP	10	1.1	13 ± 10	- 3.06 ± 0.03
NSD3	10	1.0	250,000 ± 150,000	Endothermic[a]

[a]For such weak binding, the value of ΔH cannot be estimated accurately by ITC.

References:

(1) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Anal Biochem* **1989**, *179*, 131–137. https://doi.org/10.1016/0003-2697(89)90213-3.

(2) Anthis, N. J.; Clore, G. M. Sequence-specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci* **2013**, *22*, 851–858. https://doi.org/10.1002/pro.2253.