Supporting information for "Computed binding of peptides to proteins with MELD-accelerated molecular dynamics"

Joseph A. Morrone,[1,2] Alberto Perez,[1] Justin MacCallum[3] and Ken A. Dill[1,4,5]

1. *Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794 USA*
2. *Present Address: IBM T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA*
3. *Department of Chemistry, University of Calgary, Calgary, Alberta T2N 1N4, Canada.*
4. *Department of Physics & Astronomy, Stony Brook University, Stony Brook, NY 11794 USA*
5. *Department of Chemistry, Stony Brook University, Stony Brook, NY 11794 USA*

Peptide-Protein Binding simulations in MELD

Each simulation performed in this study contains three different molecules: the protein (either MDM2 or MDMX) and two peptides (P1 and P2). In MELD, we use a replica exchange ladder, with 30 replicas to simulate the binding of P1 and P2 to the protein. States along the replica ladder correspond to different temperatures and Hamiltonians to facilitate the binding and unbinding of the peptide. In the replica ladder we map replica number which is discreet (1 to 30) to a continuous variable α that goes from 0 to 1. Based on the value of α, we specify the state (temperature and Hamiltonian) for that replica. For example, to enforce binding of the peptide we will enforce at the lower replicas a restraint that keeps one peptide bound while the other is at a reference state (see Figure 1). However, the strength of this restraint diminishes at higher replicas, becoming zero for $\alpha \geq 0.6$. This enables the two peptides to randomly search at higher replicas, and therefore have the opportunity to exchange which peptide gets selected to bind to the protein. This behavior is seen in Figure 1, were at low replica index the distribution is bimodal – one peptide remains close to the protein while the other one is in a reference state far away from the protein. At low replica index, both peptides are free to sample conformational space around the protein.

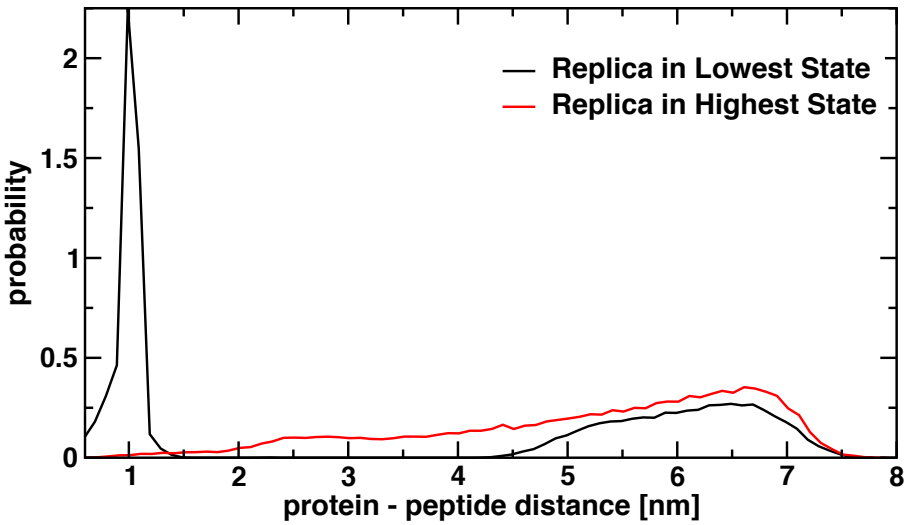

Figure S1. Histogram of distributions of center of mass distance of peptides to the protein. At the lowest replica one of the two peptides remains close to the protein while the other one remains far away in a reference state. At high temperature, there are no distance constraints active, so the peptides have a random distribution of distances around the protein.

Changes in the Temperature

The temperature is scaled geometrically from 300K to 500K along the 30 replicas.

Changes in the Hamiltonian

The basic Hamiltonian (H^o) corresponds to the force field used (ff12SB). H^o is modified by adding restraint potentials that act either on individual molecules or between molecules as described below.

Intramolecular protein restraints:

These are only applied to the protein. They are harmonic position restraints that limit the amount of movement of each amino acid with respect to the original position. These are imposed on the Cartesian coordinates of every Cα atom in the protein. When the Cα position is closer than 3.5 Å to the original (crystal) position, there is no energy penalty. Beyond the 3.5Å offset, the energy increases quadratically with a force constant of 2.50 kJ*mol^{-1}*Å^{-2} per equation 1.

$$E = 0.5k(\Delta x - offset)^2 \quad (1)$$

This restraint acts on all the replicas. There are no intramolecular restraints acting on the peptides.

Intermolecular restraints:

a. Peptide-peptide exclusion: To avoid aggregation, the two peptides are restrained through a flat bottom potential restraint of the form shown in Fig S2. The energy increases quadratically when the distance between the central Cα in each peptide is

below 30 Å, thus forcing them apart. There is no force or energy between 30 Å and 1000 Å. If the peptides are further away, a quadratic energy term is used to bring them closer (e.g. to avoid peptides being at very large separations). In terms of Figure S2, R1 = 0 Å, R2 = 30 Å, R3 = 1000 Å, R4 = 1010 Å. The force constant is 2.50 kJ mol^{-1} $Å^{-2}$. This restraint is not scaled through the replica exchange ladder.

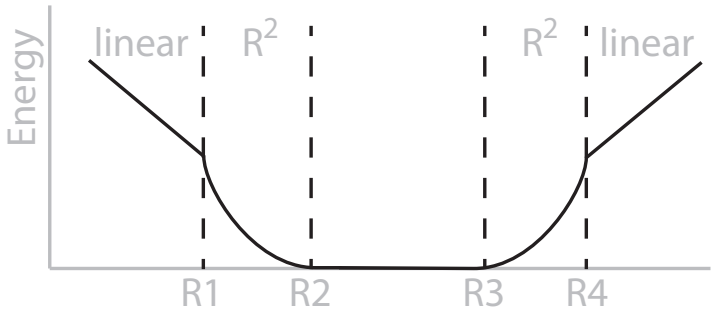

Figure S2. Distance restraint energy term.

b. Overall confinement: There is an additional restraint that is constantly applied at all replicas to keep the peptides from straying too far from the protein (R1 = 0 Å, R2 = 0 Å, R3 = 70 Å, R4 = 80 Å, k = 2.5 kJ mol^{-1} $Å^{-2}$.
c. Protein-peptide in reference state: This restraint is a MELD collection made from two groups, one for each peptide. Inside each group is a restraint between the center of each peptide and the center of the protein. It is of the form shown in Figure S2 with: R1 = 40 Å, R2 = 50 Å, R3 = 70 Å, R4 = 80 Å, and k = 2.50 kJ mol^{-1} $Å^{-2}$. Only one of the two groups (the one with lowest energy) is enforced at any given time. The force constant for this restraint is scaled non-linearly (see Figure S3) between full strength (2.50 kJ mol^{-1} Å2) at the lowest replica and 0 when $\alpha \geq 0.6$. This keeps one of the two peptides confined to the reference state at low α.
d. Hydrophobic binding restraint: This restraint enforces a few contacts between hydrophobic residues in the peptide and hydrophobic residues in the protein (see main text). The number of contacts enforced is the number of hydrophobic residues selected in the protein or the peptide (N), whichever is lower. In MELD terminology, all the possible pairings of hydrophobes are set as restraints belonging to a group, with the N restraints with lowest Energy being enforced. There are two such groups, one for each peptide, together they make up a collection. Only one of the groups in this collection needs to be enforced. Hence, only one of the peptides is drawn close to the protein, whichever has the lowest group energy for that conformation. In terms of the individual restraints, they are defined as distances between Cβ. Their parameters per Figure S2 are as follows: R1 = 0 Å, R2 = 0 Å, R3 = 9Å, R4 = 11 Å, k = 2.5 kJ mol^{-1} $Å^{-2}$. These restraints are also scaled during the replica exchange ladder so that at low replica indexes one of the two peptides will be close to the protein. In combination with restraint *(d)* previously described this keeps one peptide close to the protein at low replica index while the other is in the reference state. The list of

hydrophobic restraints in the (MDM2) protein is [53, 54, 55, 57, 61, 62, 66, 74, 75, 81, 82, 88, 89, 91, 93].

Restraints are switched off in a non-linear fashion

As selected constraints are turned off for higher states in the replica ladder, the force constant is scaled as a function of α (see Table S1 and Figure S3). The effective restraint force constant at a given replica is k'=K·s(α).

Table S1. Scale factor as a function of α.

α	s(α)	K (kJ mol^{-1} Å^{-2})	k' (kJ mol^{-1} Å^{-2})
0.00	1.00	2.50	2.50
0.10	0.50	2.50	1.26
0.20	0.25	2.50	0.62
0.30	0.12	2.50	0.30
0.40	0.05	2.50	0.13
0.50	0.02	2.50	0.04
0.60	0.00	2.50	0.00
0.70	0.00	2.50	0.00
0.80	0.00	2.50	0.00
0.90	0.00	2.50	0.00
1.00	0.00	2.50	0.00

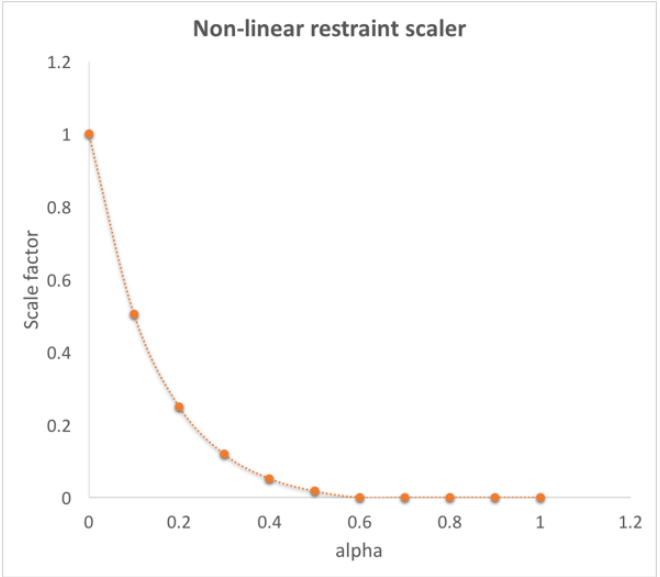

Figure S3. Scale factor as a function of alpha. The non-linearity allows for a smooth transition between no restraints active and restraints fully active. Sharper transitions often result in high restraint forces with associated errors in molecular dynamics runs.

Convergence and sampling of replica exchange simulations

To assess the consistency and convergence of our replica exchange simulation, we first show the RMSD with respect to the crystal pose of a set of the lowest replica states. States are considered up to a cutoff of index 11 as for conformations corresponding to state 11, there is the appearance of (small) violations of the binding constraints (see Figure S4). The peptide RMSD distribution for bound conformations in 6 replica states for 6 different systems are plotted in Figure S5. While there is variation in peak height over the lowest 11 replicas, they are consistently sampling the same peptide structures (as indicated by the value of the RMSD). This is indicative of good mixing amongst the (bottom 11) states and that many replicas attain consistent binding modes. Apart from one system (MDM2-P6W-P53 – an outlier discussed in the main text) conformations within 2 Å of the crystal structure are well represented.

The analysis of bound structures presented above does not guarantee good mixing of replicas amongst the all rungs of the ladder. To assess this, we consider the distribution of the fraction of time that a replica spends in the "binding tier" (that is the lowest 11 replicas) as it moves up and down the ladder. Ideally this distribution should be sharply peaked at 11/30 or 0.367. If mixing is poor, one would find a bimodal distribution with peaks centered near zero and one. In Figure S6, we plot the distribution for the same 6 systems considered in Figure S5. Overall the mixing is quite good for MDM2 binders, and somewhat less so for MDMX binders, where the distributions are considerable broader. Even in the case of MDMX however, it can be seen that there is considerable mixing between replicas.

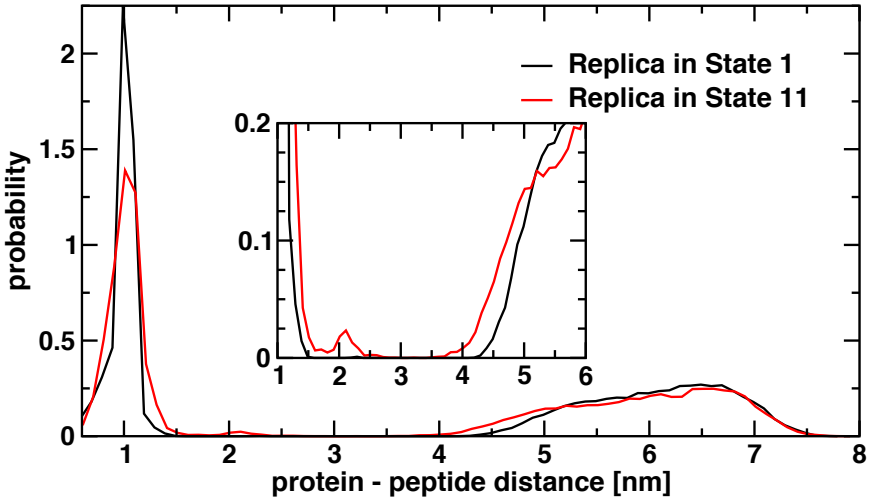

Figure S4 The peptide-protein center of mass distance for replicas in the lowest state 1 (black line) is compared with those that correspond to state 11 (red line). From the region zoomed in by the inset, violations of the assignment of peptides to bound and reference peaks begin to take hold.

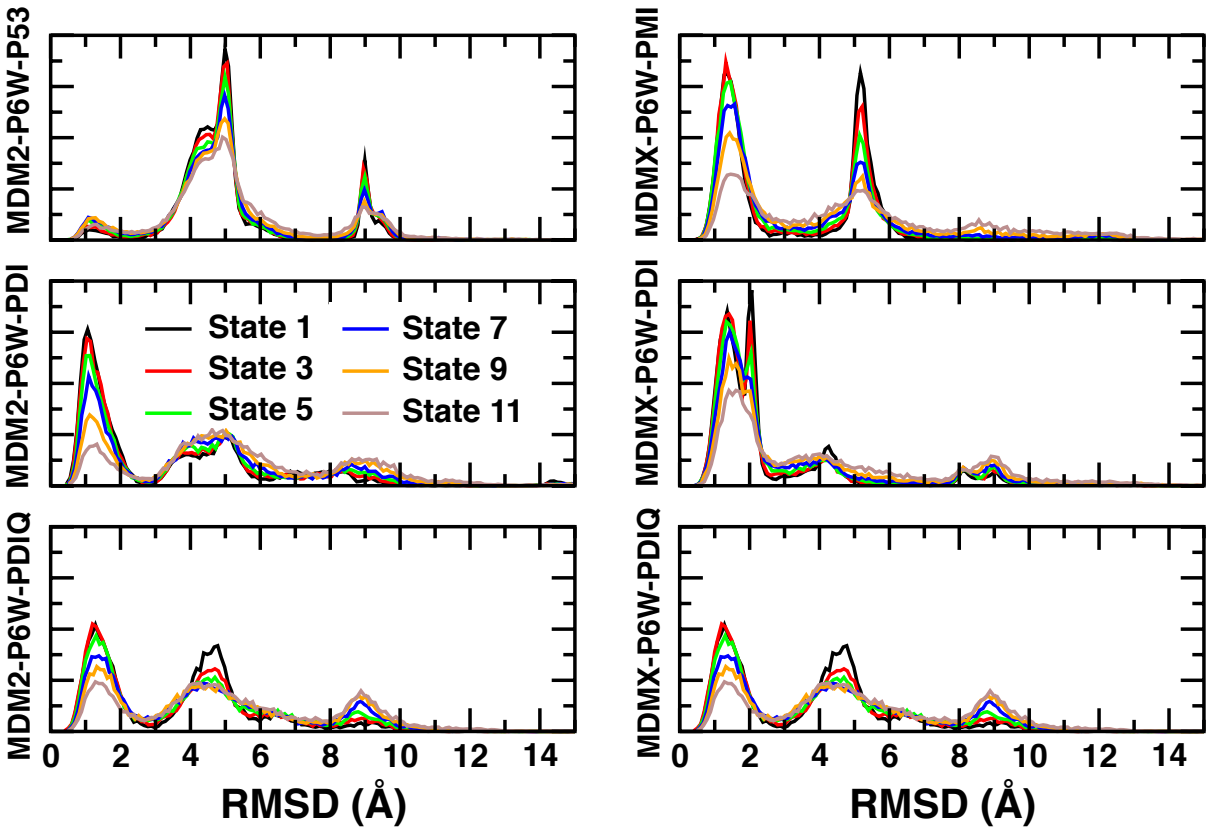

Figure S5: The distribution of peptide RMSD with respect to a crystal reference for 6 selected systems and 6 selected replica states. RMSD is defined as in the main text. Conformations for both peptides present in the simulation are combined for the sake of this analysis.

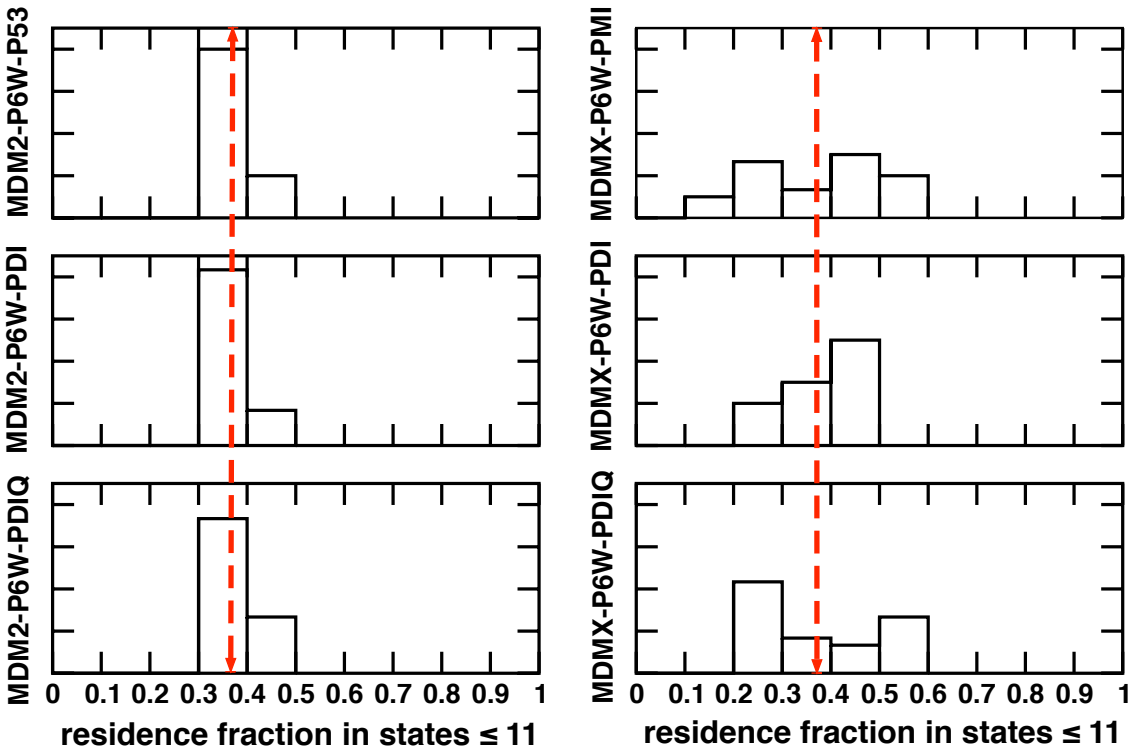

Figure S6 Distribution of the fraction of time that replicas spend in the lowest 11 states for 6 separated replica exchange simulations we have run. Ideally this distribution should be sharply peaked about 0.367 (red, dashed line), and the closeness to this ideal indicates the quality of mixing amongst replicas.

Availability of MELD code and starting configuration

We have made a sample starting configuration (for MDM2-P6W-PDIQ) available for download. The MELD code is available for download from https://github.com/maccallumlab/meld