Supplementary Material

1 How do we choose the 20 systems

We are interested in finding small dimers for which the structure is known experimentally and that are stable under our simulation conditions. In order to do so, we searched the Protein Data Bank archive[37] (PDB) using MDTraj[53] and Prody[54]. We selected only dimers in which one of the two chains has a length between 50 and 100 residues, and the other one between 45 and 155 residues. Additionally, we required 60% of each monomer to have a secondary structure. We also filtered out any structure that contains non-standard residues, DNA, RNA, or ligands. The 96 structures left were tested for stability in our experimental conditions: each PDB structure was used as the starting point for a 100 ns simulation at 300 K with Amber[36]. The monomer and the dimer structures were required to be stable through this simulation. This didn't prove that the crystal pose is the lowest free energy pose according to our setup, but we were able to screen out poses and monomers that were not stable. In principle, it is possible to use different force fields and/or solvent models or even to develop new ones to modify and expand the list of stable dimers, but this is beyond the scope of this paper. We assessed a dimer as stable if it met two conditions during the 100 ns simulation: (i) the average RMSD-L and RMSD-R were below 5 Å (*i.e.* the two sub units do not tend to unfold/misfold) (ii) the average LRMSD is below 10 Å (*i.e.* the docking pose is stable). This screening left us with 49 dimers in total.

Running MELD x MD simulations on all systems would have been prohibitively computationally expensive. Furthermore, we aimed to show how our method is complementary to others. We used CP[11,26–29] as a benchmark of the state of the field. For each of the 49 systems, we fed to the (vanilla) CP server the structure of the two monomers in the crystallographyc holo structure of the dimer. CP is able to correctly predict the structures of 34 dimers (CP TOP1 is the same as CP TOP15, and the CAPRI score is at least *acceptable*). For 14 of the remaining dimers there is a structure at least *acceptable* in the first 15 CP prediction (see Figure SI 1). Only for one structure (3T0E), vanilla CP is not able to provide a valuable answer in the first 15 guesses. For this system ONLY we also includeed a set of native contacts. We tested MELD x MD on these 15 more challenging systems for CP, and on 5 additional systems for which CP is successful in identifying the native state.

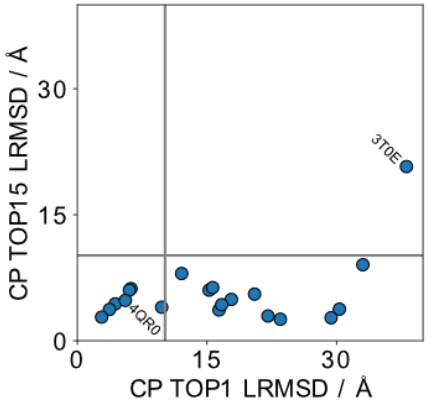

Figure SI 1: **Rigid docking poses are a good source of information for the MELD protocol.** Each point shows the comparison between the LRMSD of the TOP1 and TOP15 CP predictions for the 20 dimers. For only 7 cases does the TOP1 prediction have a LRMSD<10 Å, but in the TOP15 prediction there is an accurate structure for 19 of the 20 cases. Information from the CP TOP15 poses is therefore a good starting point for our MELD x MD protocol. 4QR0 is not regarded as a CP success in the main body of this paper since the TOP1 CP prediction is not the best prediction. No CP TOP15 prediction is accurate enough for 3T0E, for this system only we provide also native contact information.

2 Understanding the metrics we use in the context of flexible docking

Since we are dealing with flexible structures, it is important to develop an intuition about how flexibility is going to affect RMSD values. Dealing with flexible protein monomers introduces "noise" when LRMSD are computed. Part of the noise comes from the flexible nature of the second monomer, and part of it comes from the difficulties in aligning the first monomer. Figure SI 2 shows how these contributions can be parsed. On the left is shown the behavior system where all the atoms of the interface residues are frozen (iRMSD = 0 Å and fTP = 100%. Observed LRMSD distributions are quite spread due to a combination of the monomers internal the flexibility and the difficulty of aligning the receptor with its reference. In the right column is shown the behavior of a system where the interface residues of the receptor and all the atom of the ligand are frozen. In this case all the observed spread of the LRMSD distributions are due to the challenges in aligning the receptor to its structure in the native PDB file.
We note this additional source of uncertainty when evaluating the predicted poses and conformations of flexible monomers. Nonetheless the metrics used to assess rigid docking seem to do a good job in describing the quality of our predictions, so we use them throughout this paper.

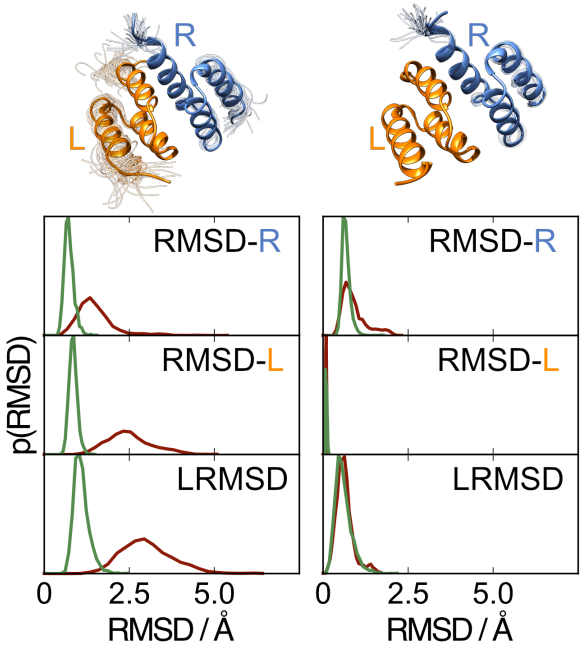

Figure SI 2: **Some ambiguities in defining LRMSD when monomers are flexible.** Here we show the RMSD distributions computed for 100 ns MD simulations of two dimers (2B87 in red, and 2LYJ in green) with different parts of the system frozen. In the panels are shown the RMSD distributions of the receptor backbone (RMSD-R), of the ligand backbone (RMSD-L), and the LRMSD for the two dimers. The left column show those quantities for systems where all interface atoms have been frozen in their native conformation. On the right are shown the same quantities for a system in which the interface atoms of the receptor and all the atoms of the ligand are frozen. On the top of the graphs are shown depictions of the motions of the receptor (blue) and the ligand (orange) monomers of 2B87 for both simulations setups.

3 Convergence of the simulations

We present the RMSDs of Figure 5 of the main text for the 20 systems studied in this paper

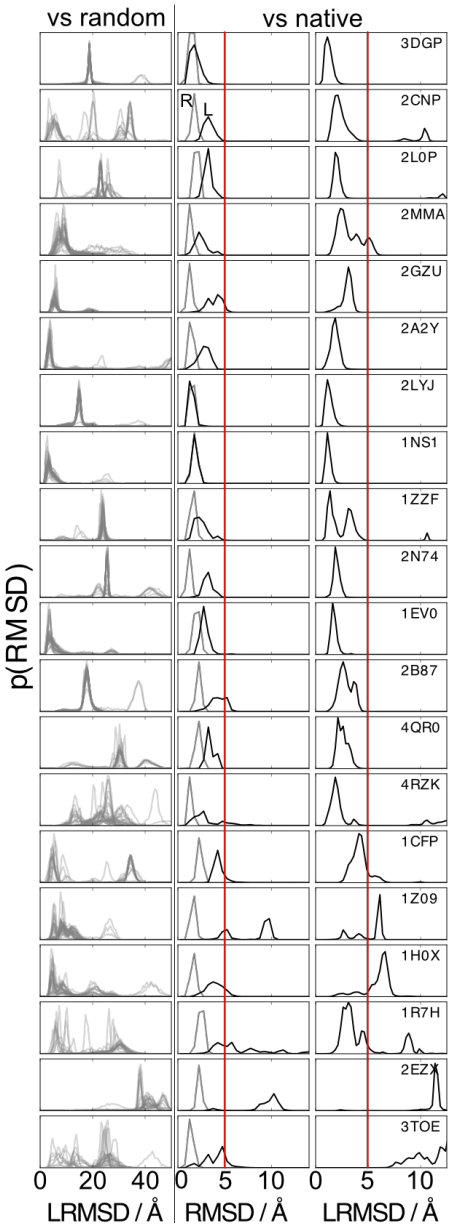

Figure SI 3: **RE convergence and ligand stability are key for good predictions** The RE protocol is converged for all the first 16 systems (*i.e.* the ones we predict correctly). 2CNP and 4RZK are less converged that the other 16, but they are closer to convergence compared the failed prediction of 3T0E and 1R7H. Another key aspect for obtaining good results is the stability of the ligands: all the successful predictions, except 1Z09, have a RMSD-L below 5 Å. In 1Z09, one of the alpha helices of the ligand rearranges, but since this rearrangement happens away from the interface, the docking prediction is still considered of *medium* quality. All the 16 successes (except 1Z09) present clear peaks of the LRMSD below 5 Å of native. 1R7H also belongs to this group, table 1 of the main text shows that there is a conformation with relatively low values of LRMSD and yet with a totally wrong interface.

4 Quality of predictions vs clustering approach

Figure SI 4 shows the effect that different clustering approaches have on our prediction.

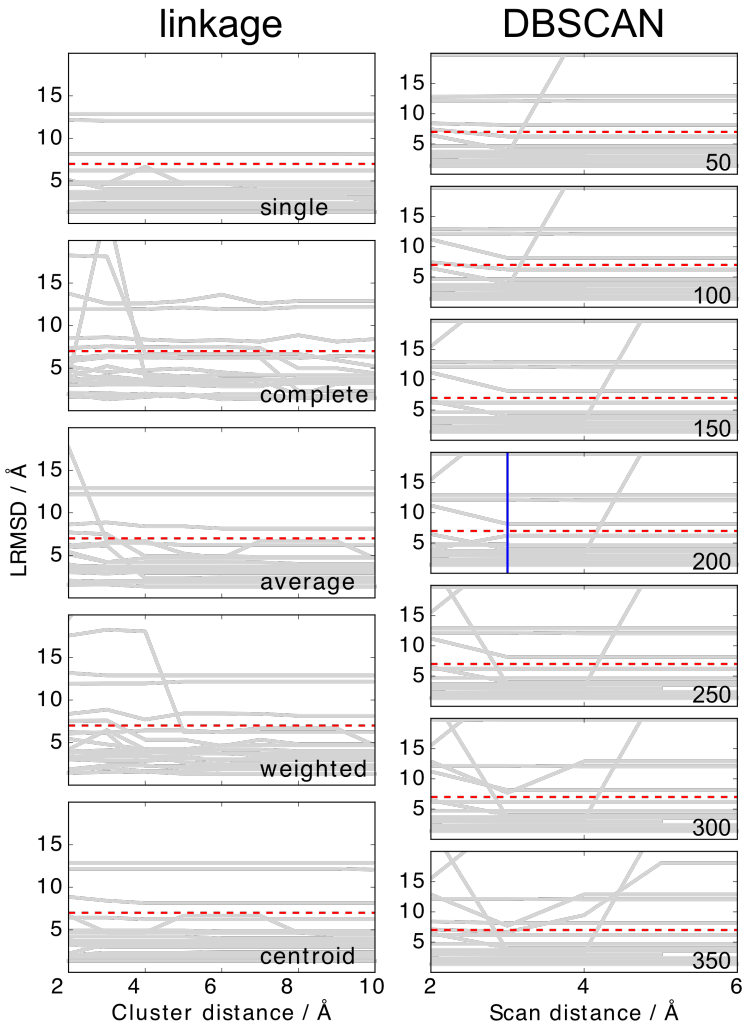

Figure SI 4: **Predictions are relatively sturdy with respect to clustering protocols and parameters.** The gray lines represent the LRMSD of the centroid of the most populated cluster (MELD TOP1) of the twenty dimer simulations according to different clustering method and parameters. On the left are shown results using the linkage clustering method. The different graphs represent different ways to compute the cluster distance, and the x-axes represents different clustering distances. Results on the right were obtained using the DBSCAN clustering approach. The different graphs represent different choices of the number of neighbours that a point must have to be considered part of a cluster (usually denoted as ϵ), and the abscissa represent the different point distances that are considered. The vertical blue line identifies the method and parameters used in the main part of this paper. The red dotted lines highlights a LRMSD of 7 Å. The quality of most of the predictions is not strongly influenced by the choice of different parameters: most of the lines are basically horizontal. What changes with different parameters is the relative population of different structures, but the MELD TOP1 structure is identified consistently.

5 How we draw the protein map

Understanding the quality of a dimer structure prediction is challenging. Computing a series of numerical quantities (*e.g.* LRMSD, iRMSD) is invaluable to rank predictions, but sometimes it reduces too much the dimensionality of the system to get to the heart of what happens. Alternatively, it is possible to compare sets of 3D structures, but this is also challenging because it is difficult by visual inspection to asses the quality of similar models (for example see the different 3D structures in figure 4 in the main text). We propose here a way to built a map of the intermolecular contacts that simplifies the problem of comparing 3D structures and at the same time loses less information than by compressing structures to single value.

The map represents the positions of the contact residues of the receptor. The receptor is approximated as a sphere centred on the receptor center of mass. The *xyz* coordinates of the receptor contact residues are converted to spherical coordinates, and only the angular components (θ and ψ) are kept. Then these coordinates are plotted on a map using Mollweide projection[41], in a totally similar fashion to creating a map of the earth. This displays the full sphere surface in 2D. This means that periodic conditions exist along the x-axis (*i.e.* east-west) and reflective condition exist when poles are reached moving along the y axis (*i.e.* north-south). Furthermore, contacts can be identified as true positive and false positive when compared to a reference structure, and then can be colored in to distinguish them (green and red in our case). Considering the colors of the contacts and their distribution on the map compared to a native structure distribution, we can have an idea of the quality of a structure.

This simplified representation of the system is unlikely to be the final point of the analysis of the results, but we hope it will help clarify what happens when numbers alone do not tell the whole story, and analyzing a plethora of 3D structures would be too tedious.

6 We can use the symmetry of homo-dimers to recover some of the failures

It is known that homo-dimers in nature have a high degree of symmetry[42,43]. In principle we could use this knowledge at different points of our protocol to improve our predictions. Here we briefly show the possibility of using this knowledge to filter our predictions. Table SI 1 good predictions and symmetry of the monomers go hand in hand for most of the systems we investigated. In two cases (3T0E and 1H0X) the physics model favor a different pose. While not affecting the symmetry of the monomers, this is likely to indicate some bias in the model we use to describe the interactions or in the inability of our protocol to escape some local minimum. We also note that restrain we place on the alpha-carbon positions of the receptor includes some information about the native structure of the monomer in the search. This approach seems promising, but further investigations free of the biases we enforced are necessary to unequivocally prove the efficacy of this screening procedure.

Table SI 1: **We can screen out wrong homo-dimer prediction considering the necessary monomer symmetry** Monomers symmetry can be evaluated on the basis of the similarity of the monomers structure. To this end we report the RMSDs of the ligand structure computed against the receptor structure for all the MELD x MD homodimer predictions. The second column reports the RMSDs of the monomers computed on the MELD TOP1 predictions, the third one highlights the ranking and the RMSD of the structures with the most similar monomers for each system. For the first fifteen dimers, the most populated structure is also the one with the most similar monomers (for 2CNP, structure #1 has a slightly better monomers RMSD than #0, but the difference is small enough to not constitute a problem). In three cases (1Z09, 1R7H, and 2EZX) looking for the structure with the lowest RMSD between the monomers allowed us to identify the best dimer docking prediction (see table 2 in the main text). We can not recover the prediction for the last two dimers, which retain a good monomer structure, but prefer a wrong docking pose.

	TOP1 RMSD / Å	TOP10 RMSD / Å	#
2CNP	3.6	2.9	1
2L0P	4.5	4.5	0
2GZU	6.5	6.5	0
2A2Y	2.9	2.9	0
2LYJ	1.1	1.1	0
1NS1	2.2	2.2	0
1ZZF	1.2	1.2	0
2N74	1.3	1.3	0
1EV0	4.7	4.7	0
2B87	3.7	3.7	0
4RZK	4.2	4.2	0
1CFP	3.5	3.5	0
2EZX	8.4	3.4	2
1R7H	12.8	4.0	8
1Z09	8.8	3.5	1
3TOE	3.1	3.1	0
1H0X	2.0	2.0	0