The ability to evaluate free energies is key in answering many biologically relevant questions. The native structure of a protein is lowest free energy one. The best ligand is the one with the lowest binding free energy.

Physics based simulations, like Molecular Dynamics (MD), are in principle a prime method to compute free energies. In practice, the phase space of most biological problem is too vast for plain MD to converge to an answer on reasonable timescale. To tackle this limitation, we recently introduce MELD (Modeling Employing Limited Data). MELD is a MD accelerator, that uses external information to focus the computational effort of MD only to regions of the space that agree with the information.

In the poster left panel we introduce the basic concepts of the MELD method. In green we show a cartoon of a MD simulation potential energy surface. This surface has many minima. Transitioning between them is not hard, but in order to identify the global minimum we need the converge the populations of most of them. The most populated one is the global minimum. MELD uses information to reduce the search space of the problem. It does that by introducing an energy penalty only to structures that do not agree with the external information (blue line). The orange line shows the sum of the two previous potential. MD will naturally avoid high energy structures, and therefore all the computational effort will focus on the few deep minima of the orange potential. Hamiltonian and Temperature Replica Exchange MD (H/T-REMD) protocol is used to jump over the high energy barriers that now separate the minima. MELD is based on a Bayesian framework, that offers an integrative approach where different sets of data contribute to carve the phase-space. Only regions compatible with all data sets are explored by MD. MELD can leverage sparse, ambiguous, and uncertain data. This means data that aren't sufficient alone to predict a structure -*i.e.* crosslink data-, data with multiple possible interpretation -*i.e.* unassigned NMR peaks-, and uncertain data -*i.e.* experimental data with an uncertainty associated with the measure-.

In the center panel we show how MELD can be used to predict globular proteins native structures. Finding the native structure of a protein means finding the lowest free energy configuration. We run H/T-REMD simulations that incorporate different sources of information. We have a set of "heuristic" information, that come from general knowledge about the structure of small globular proteins. For example, *the protein needs to have a hydrophobic core*. We can complement this with contact predicted using machine learning, homology, or metagenomic based methods. Using this protocol, we were able to fold small proteins in an international competition called CASP. Some of our prediction were the best in CASP. We can also leverage experimental data. For example, we can use unassigned NMR peak to fold proteins up to 210 residues long. In CASP 13 (2018) we were the most effective group in leveraging NMR data.

In the right panel we show predictions beyond tertiary structure. Using MELD we can re-rank and refine protein quaternary structure predictions coming from rigid docking method. In this case we use rigid-docking algorithms as inter-protein contact predictors. In our MELD simulation the docking of the monomer uses the predicted contact to guide the process. We can predict amyloid fibril structures using general information about beta-sheet distances and arrangements or leveraging NMR data. We can predict the

binding pose and binding affinities of peptides (and small organic molecules) to proteins. Here we show an example of a CAPRI target, where the closest homolog had the wrong beta-sheet orientation. Only physics-based method like MELD and ROSETTA were able to predict the correct binding pose for this target. We also show the comparison between computed and experimental relative binding affinities for a series of P53 mutants to MDM2/X. Our calculations fall inside a kT error.