Bridging Molecular Timescales with MELD and Blue Waters

Alberto Perez
We need to know protein structures to make new drugs.
We need to know protein structures to make new drugs
Key challenge: develop computational tools to predict protein structures from sequence.

Growth of Protein Databases

- **PDB Structures**
- **UniProt Sequences**
A grand challenge in structural biology is predicting the 3D structure of a protein given the sequence.

6 months of continuous sampling is not enough for even a simple protein.

Molecular Modeling is a computational grand challenge.
Computational brute force will not solve these grand challenges.

When will we be able to fold larger proteins?
We developed MELD to scale to larger systems.
MD is the basis of our method

\[ p(x) \sim \exp[-\beta E_{\text{force}}(x)] \]

+ data

- Sparse
- Ambiguous
- Noisy
I lost my keys in the beach
MELD uses a Bayesian inference approach to incorporate data into simulations

\[
p(x|D) = \frac{p(D|x)p(x)}{p(D)} \sim p(D|x)p(x)
\]

We use Hamiltonian Replica Exchange to enhance sampling

- High Temperature / Vanishing Restraints
- Low Temperature / Strong Restraints
MELD performed high accuracy blind predictions of 3D structure

Top cluster

- **2.8Å**
- 95 residues

- **1.4Å**
- 67 residues

- **1.6Å**
- 68 residues

Blue Waters is key for CASP — the structure prediction competition

- 3 months — daily new targets
- 200 competing groups and methods
- Hundreds of proteins
- Strict deadlines (some as short as 5-7 days)
- We are the only physics-based methodology in CASP
BW’s team help indispensable during CASP

- 30 GPU nodes per protein
- Sparse communication between nodes
- Helping with compilation of the OpenMM/MELD plugin

The aggregate GPU utilization (efficiency) varies significantly by application, with MELD achieving over 90% utilization of the GPU and GROMACS, NAMD, and MILC averaging less than 30% GPU utilization. However, for each of the applications, the
Beyond folding — binding and pathways

Don’t miss these posters for more details!

Protein Folding of Nonthreadables
James Robertson, Alberto Perez, Ken Dill

Protein Structure Prediction Remains Important and Challenging

- Threading methods predict 86% of human protein structures, but many proteins are nonthreadable
- MELD folds proteins fast, is physics-based, and not limited like threading
- Can MELD fold nonthreadable proteins?

MELD is an Accelerator for Molecular Simulations

- MELD uses temperature and Hamiltonian replica exchange molecular dynamics (MD) to enhance conformational sampling and give free energies
- MELD simulations run on GPU-accelerated supercomputers like NCSA Blue Waters

MELD Folds Nonthreadable Proteins

- MELD populations are predictor of folding
- MELD is limited by force field deficiencies

Emiliano Brini

James Robertson
Thanks!

alberto.perez17@gmail.com