NAMD - The Engine for Large-Scale Classical MD Simulations of Biomolecular Systems Based on a Polarizable Force Field

Yun (Lyna) Luo, Wei Jiang, James Phillips
PI: Benoit Roux

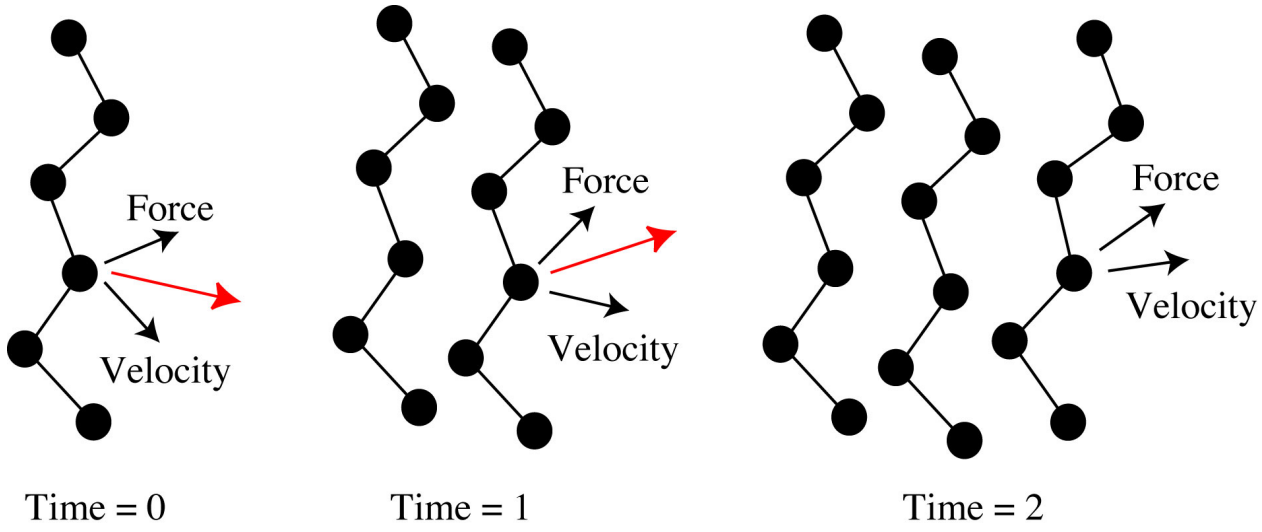
May 16, 2013
Early Science Program Investigators Meeting

- **Benoit Roux (NAMD_esp PI)**
- **Wei Jiang, Ray Loy (ALCF, ANL)**
- **Klaus Schulten, James Phillips (NAMD group, UIUC)**
- **Sameer Kumar (charm++ optimization on BGQ, IBM)**
- **Mikolai Fajer, Lei Huang, Albert Lau, Janamejaya Chowdary (U Chicago)**

- **Alex MacKerell, Pedro Lopes, Jihyun Shim, Jin Huang (CHARMM Force Field group)**

Molecular dynamics (MD) simulation:

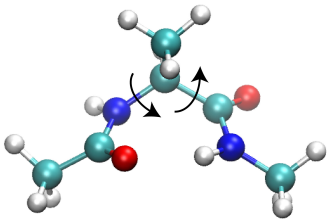
Computational approach to study conformational sampling, estimate free energies and more



$$-\frac{dV}{dx_i} = m_i \frac{d^2 x_i}{dt^2} = Force_i \qquad x(t+\Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2 x(t)}{dt^2}\Delta t^2 + ...$$

CHARMM additive potential energy function

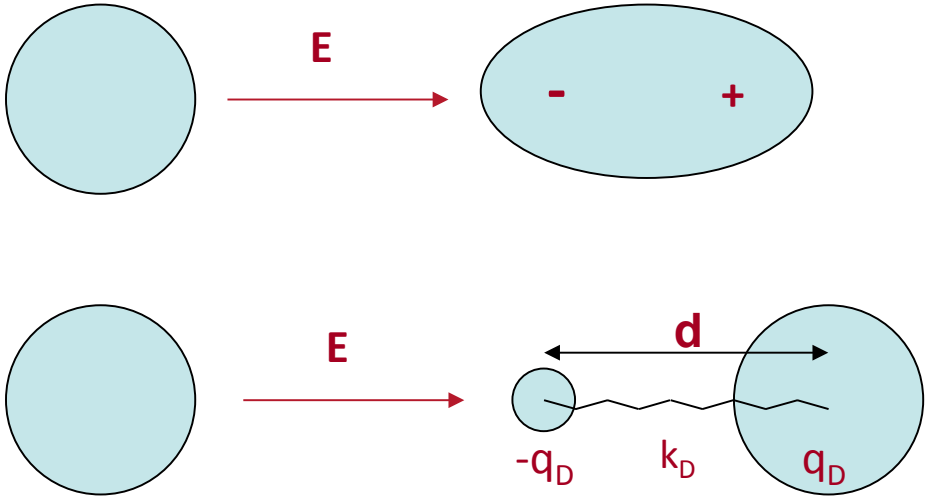
$$V(\vec{R}) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\phi\left[1+\cos(n\phi-\delta)\right]$$
$$+ \sum_{1,3\,pairs} K_{ub}(S-S_0)^2 + \sum_{improper} K_w(w-w_0)^2$$
$$+ \sum_{\mathrm{CMAP}} \left(\sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij} \left(\frac{\phi-\phi_L}{\Delta_\phi} \right)^{i-1} \left(\frac{\psi-\psi_L}{\Delta_\psi} \right)^{j-1} \right)$$
$$+ \sum_{nonbonded} \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi D r_{ij}}$$



Foundation for the polarizable force field

Classical Drude oscillator based polarizable force field

- **Classical Drude oscillator model takes into account the electronic polarizability in molecular systems.**

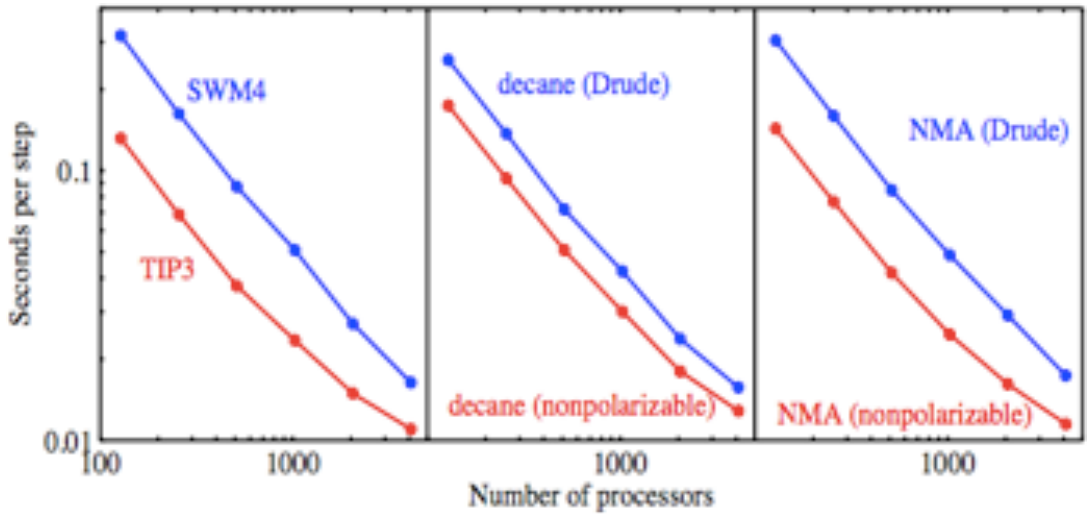


$$q(A) = q_c(A) + q_D(A)$$

$$\alpha(A) = q_D^2(A) / k_D$$

$$U_{Drude} = \sum_{A<B}^{N,N_D} \frac{q_D(A)\cdot q_c(B)}{|\mathbf{r}_D(A) - \mathbf{r}(B)|} + \sum_{A<B}^{N_D} \frac{q_D(A)\cdot q_D(B)}{|\mathbf{r}_D(A) - \mathbf{r}_D(B)|} + \frac{1}{2}\sum_{A}^{N_D} k_D |\mathbf{r}_D(A) - \mathbf{r}(A)|^2$$

Drude model implemented in NAMD: Tests on Blue Gene

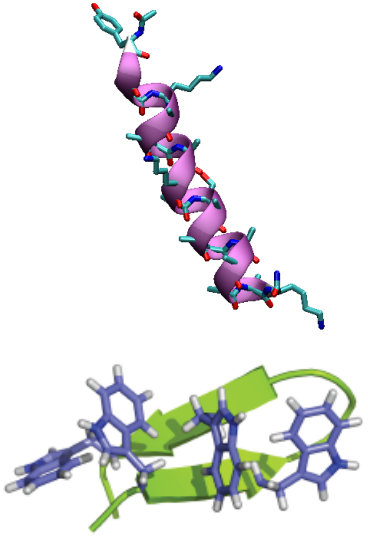


A) Cubic box of 72000 water molecules (ratio 1:2)
B) Cubic box of 8576 decane molecules (ratio 1:1.6)
C) Cubic box of 18944 NMA molecules (ratio 1:1.8)

Jiang, W., Hardy, D., Phillips, J., MacKerell, Jr., A.D., Schulten, K., Roux, B. *J. Phys. Chem. Lett*., 2, 2011.

Towards Macromolecules

- **Nucleic acids: DNA and RNA**
- **Lipids: DPPC monolayer, bilayer**
- **Proteins: Alanine dipeptide, $(Ala)_5$, $(ALA)_7$, $(AAQAA)_3$, Baldwin peptides, TripZip peptide**



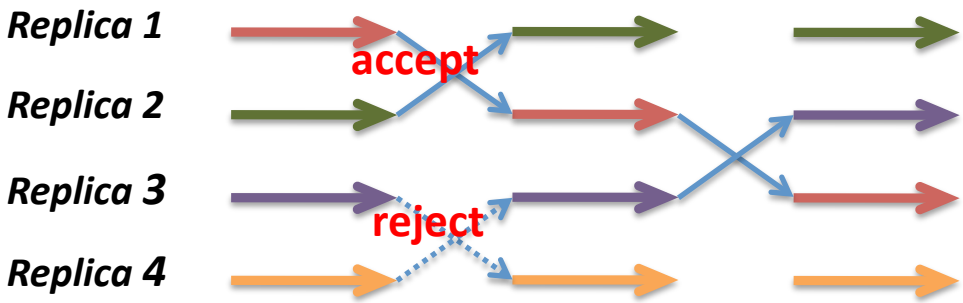
10 sets of new parameters
1 model systems
298k to 398k
100 ns per replica

10*1*100*100 ns =100 µs !

Validating new force field requires supercomputer and efficient sampling algorithm!

Multiple Copy Algorithm (MCA)

Single copy run has scaling limit (100 atoms/core)
Ensemble run that launches large number of copies concurrently



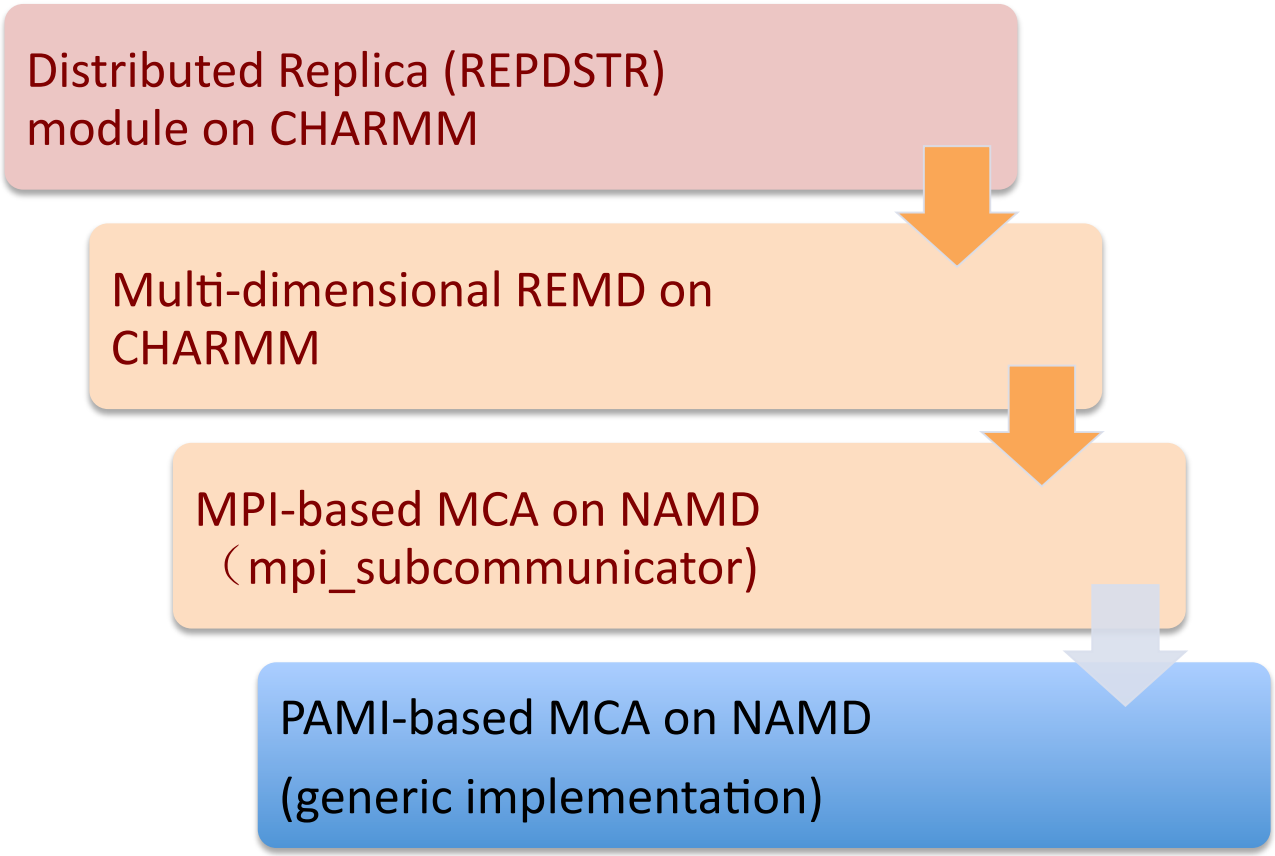
Replica Exchange Molecular Dynamics (REMD):

- **Different Temperature or Hamiltonian (order parameters)**
- **Metropolis criterion to guarantee Boltzmann distribution**
- **Enhance sampling and free energy convergence**

Go beyond single copy simulation scaling limit!

Okamoto and Sugita 1999

Implementations



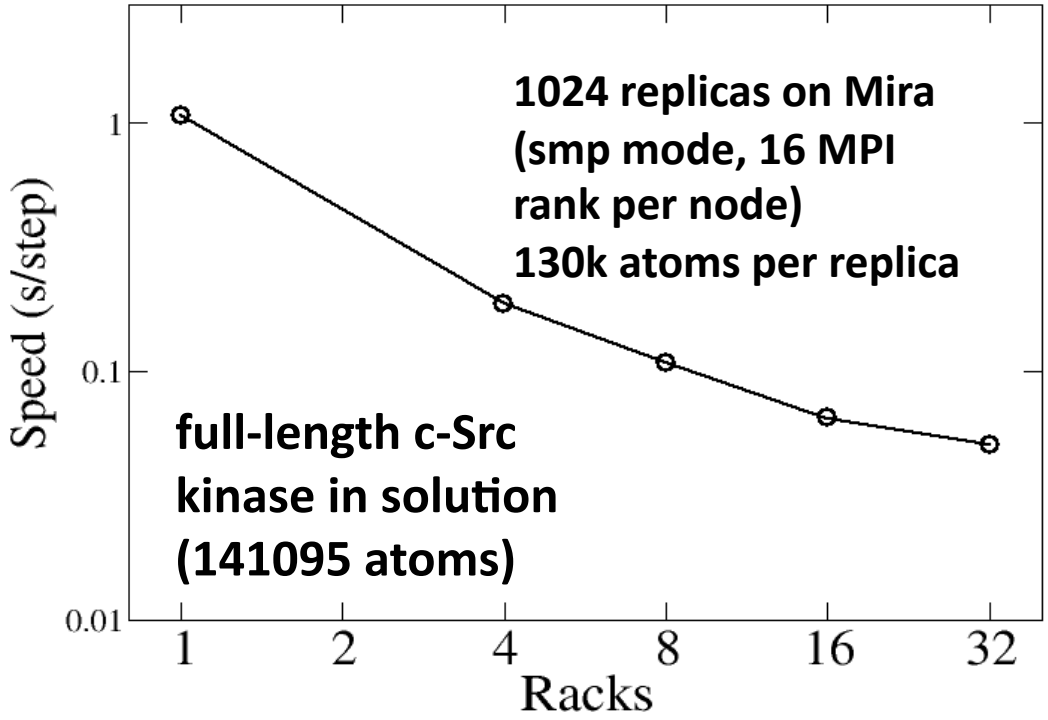
Jiang W, Hodoscek M, Roux*, J. Chem. Theory Comput*, 2009.
Jiang W and Roux B. *J Chem Theory Comput*, 2010.
Jiang, W., Luo, Y. *et al., J. Chem. Theory Comput.* 2012.

Newly Implemented REMD in NAMD2.9

Previous implementation	New implementation
Driven by external job script (small cluster)	MPI based (supercomputer)
Small number of replicas (<100)	Thousands of replicas
Low frequent exchange (<1ps^{-1})	High frequent exchange (>0.1 ps^{-1})
One dimensional exchange	Multi dimensional exchange

Scalability on Mira

- Communication overhead is minimized by swapping order parameters instead of coordinates
- Pthread is implemented in charm++ to take advantage of the multiple hardware threads feature



Each rack contains 1024 nodes, each node contains sixteen 1.6 GHz core processors.

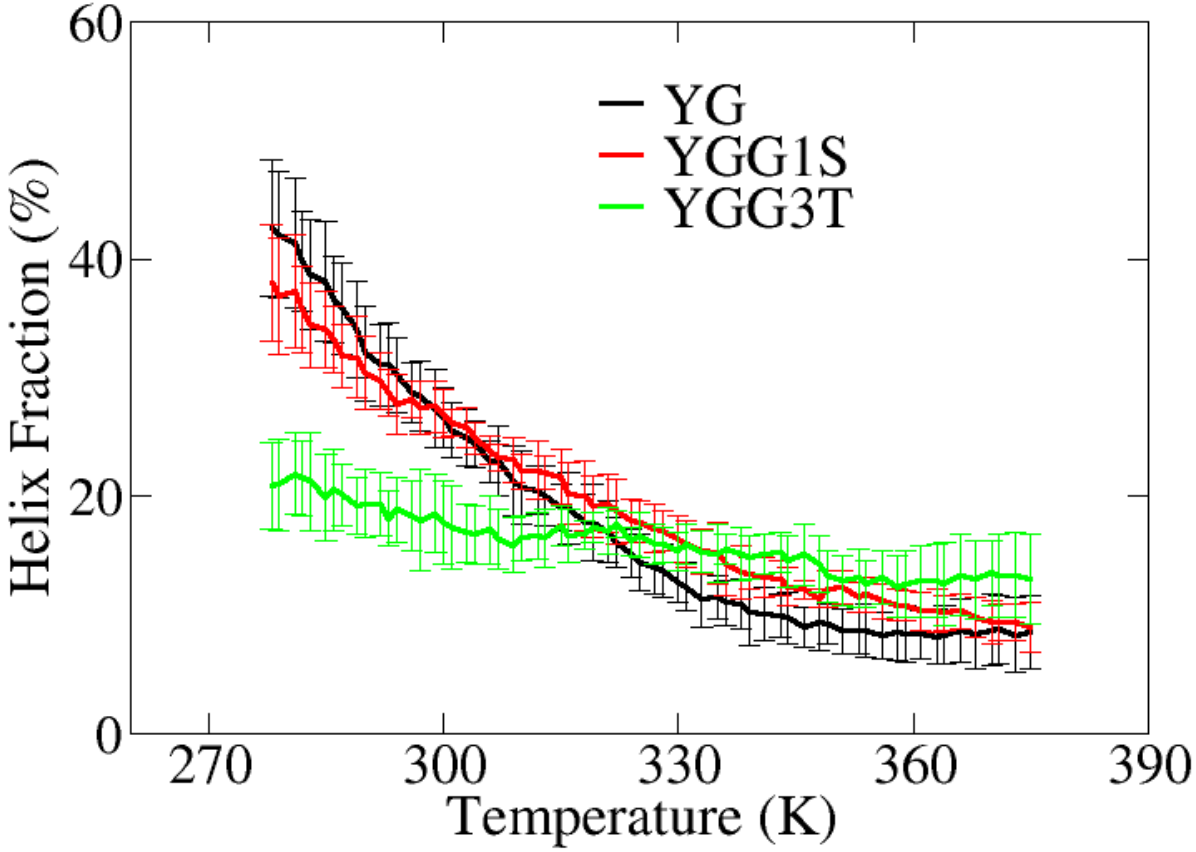
Generality and Flexibility

- **Can be used with variant of MCA algorithms without changing the source code**
  - **T-REMD: temperature replica exchange**
  - **FEP/REMD: free energy perturbation replica exchange**
  - **US/REMD : umbrella sampling replica exchange**
  - **REXAMD: accelerated molecular dynamics replica exchange (Mikolai)**
  - **String Method with Swarms-of-trajectories (Mikolai)**
  - **Others…**

- **Tcl script allows user to define exchange parameters (temperature, lambda or biasing potential in *ColVars module*)**
- **User-defined exchange topology: *replica_neighbors* Tcl**
- **Easy post processing: history file, sortreplicas**

Melting Curve Data of 3 Baldwin peptides

TREMD using CHARMM c36 ff



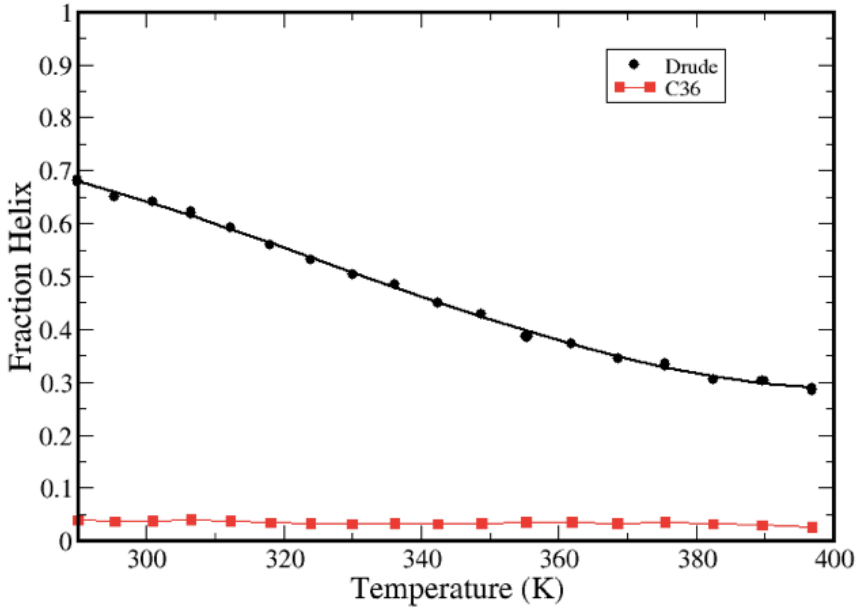
Fraction of α-helix is determined by >=3 residues in the α region of the Ramachandran map (-100<phi<-30 and -67<psi<-7).
Analysis is done using the last 14 ns trajectories (STD: last seven 2 ns blocks).

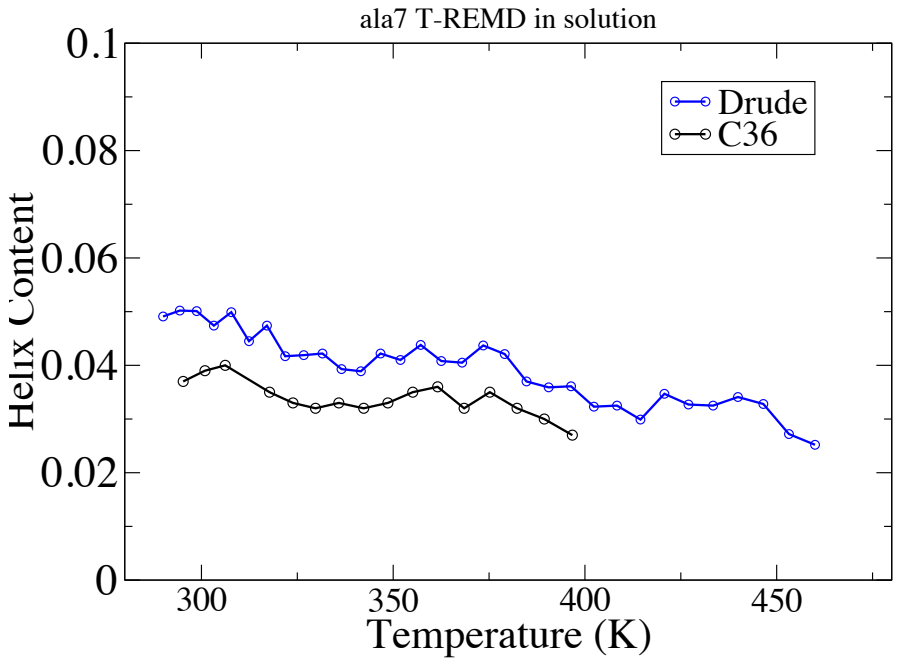
ALA7 TREMD in solution

Bad CMAP

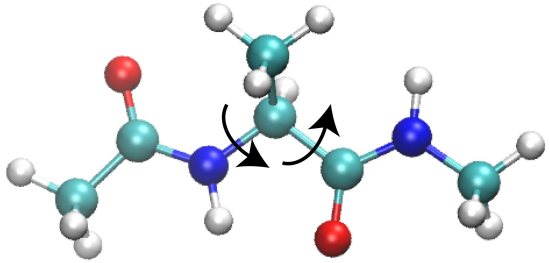
vs.

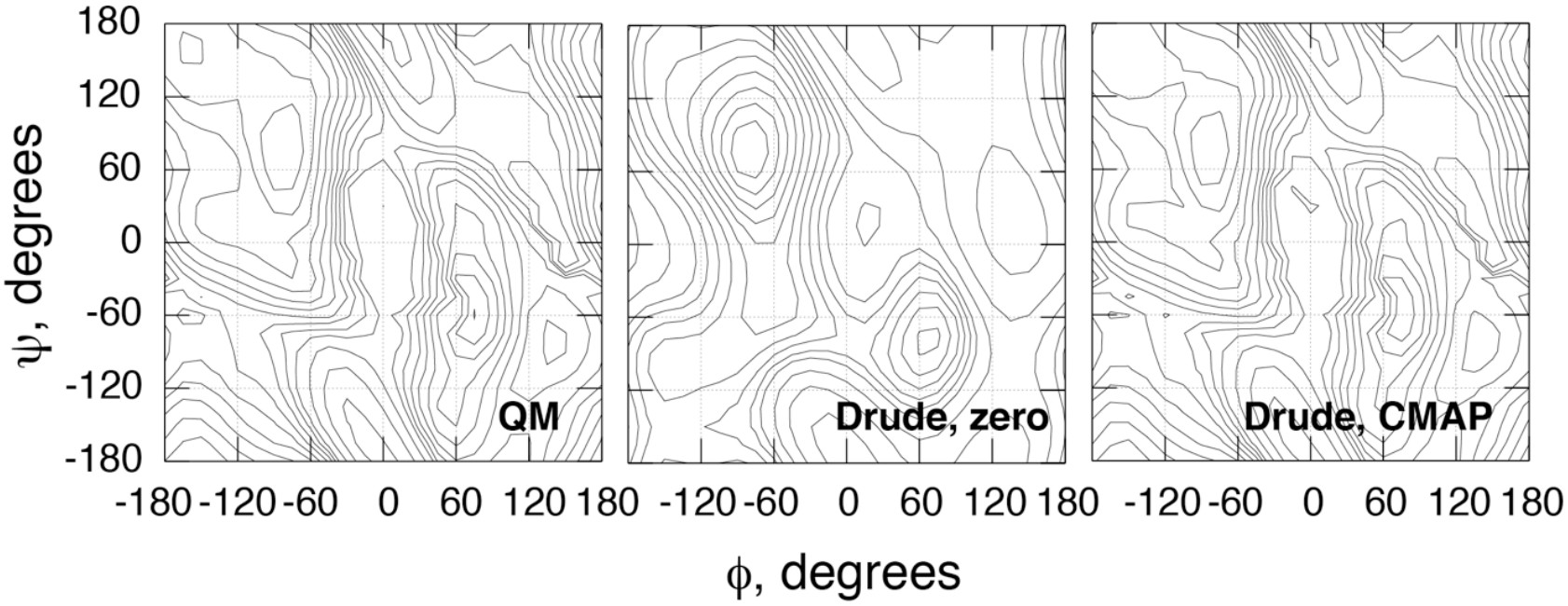
Good CMAP



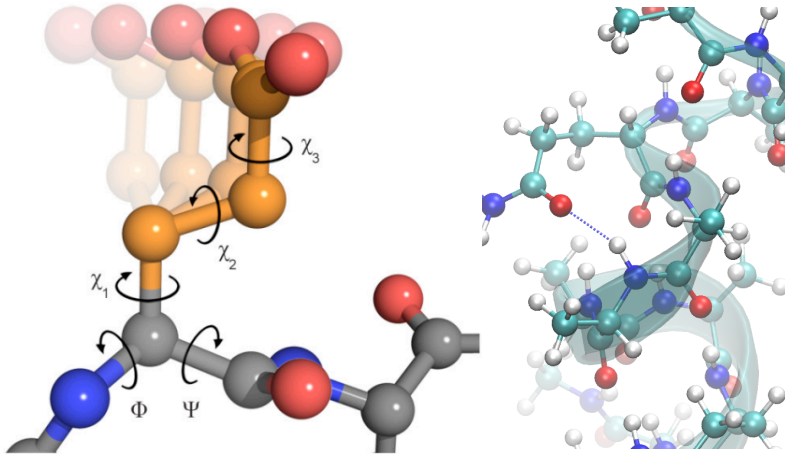


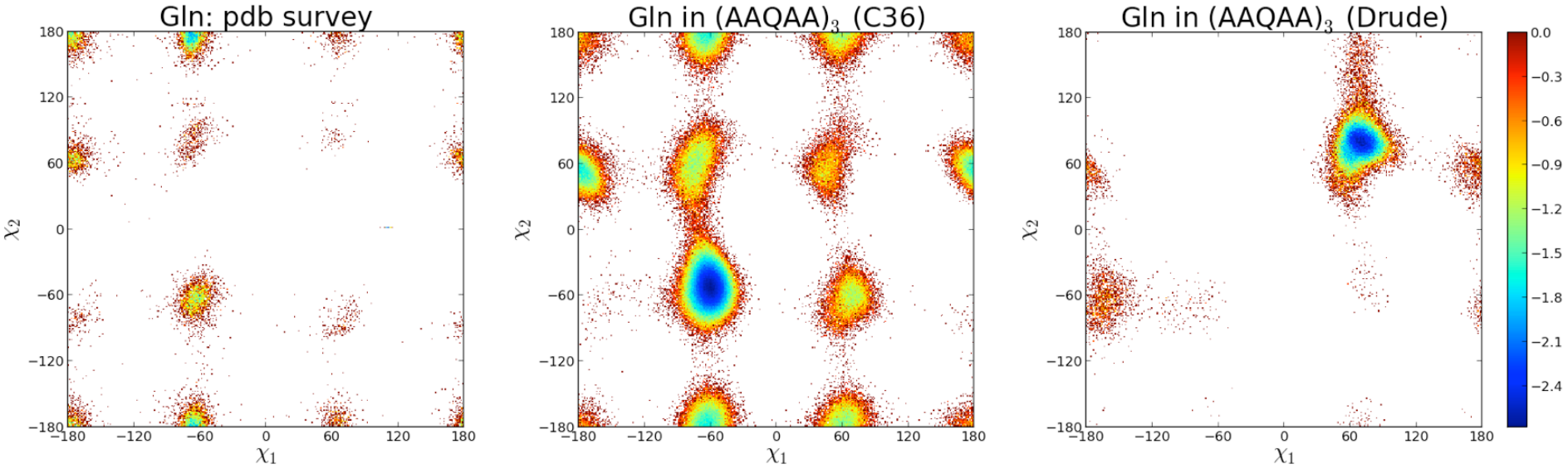
Towards proteins: Alanine dipeptide ϕ,ψ surfaces



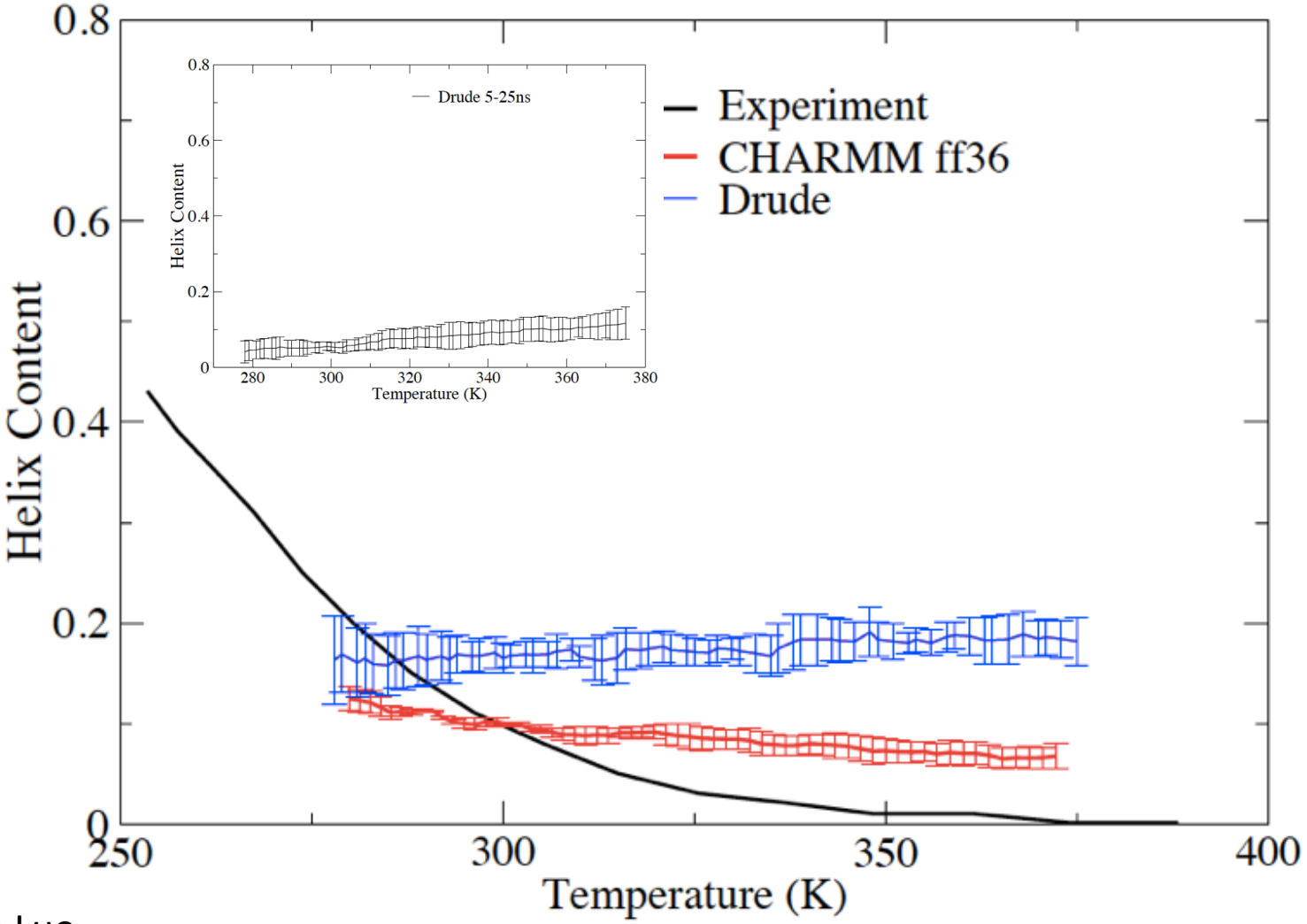


Glutamine side chain conformational sampling





Current status: $(AAQAA)_3$ TREX simulations, 25-50 ns sampling



Yun Luo

Conclusion

NAMD/Charm++ with MPI-level multiple copy algorithms (MCAs)

- Extremely scalable
- Generality and Flexibility
- Easy Post Processing

Drude oscillator polarizable force field

- Take advantage of MPI implementation
- Getting towards the macromolecules

Getting both algorithms and new generation force field ready for Exascale computing!