Accurate and efficient QM/MM molecular dynamics on 86,016 cores of SuperMUC Phase 2

M. Schwörer*, K. Lorenzen*, M. Allalen§, F. Jamitzky§, H. Satzger§, and G. Mathias*§
*Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, LMU München; §Leibniz Supercomputing Centre (LRZ), Munich

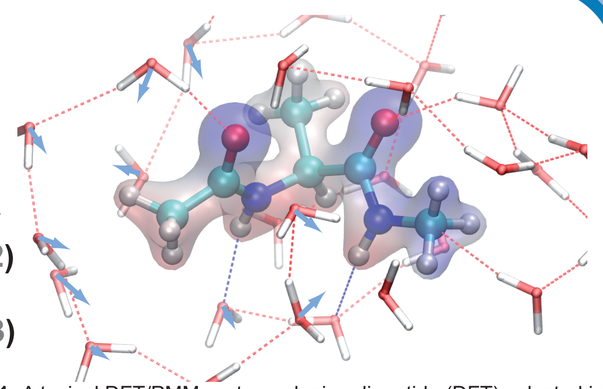
P1 Motivation & Achievements

Hybrid molecular dynamics (MD) simulations [1], in which the forces acting on the atoms are calculated by highly accurate **grid-based density functional theory (DFT)** for a solute molecule and by a **polarizable molecular mechanics (PMM) force field** for a large solvent environment, pose a challenge regarding both the theoretical modeling and the computational realization.

The **Hamiltonian DFT/PMM approach** presented here guarantees energy conservation, excludes artificial distortions of the electron density at the interface between the DFT and PMM fragments, and treats the long-range electrostatic interactions within the hybrid simulation system in a linearly scaling fashion using hierarchically nested **fast multipole expansions** [2,3,4,10].

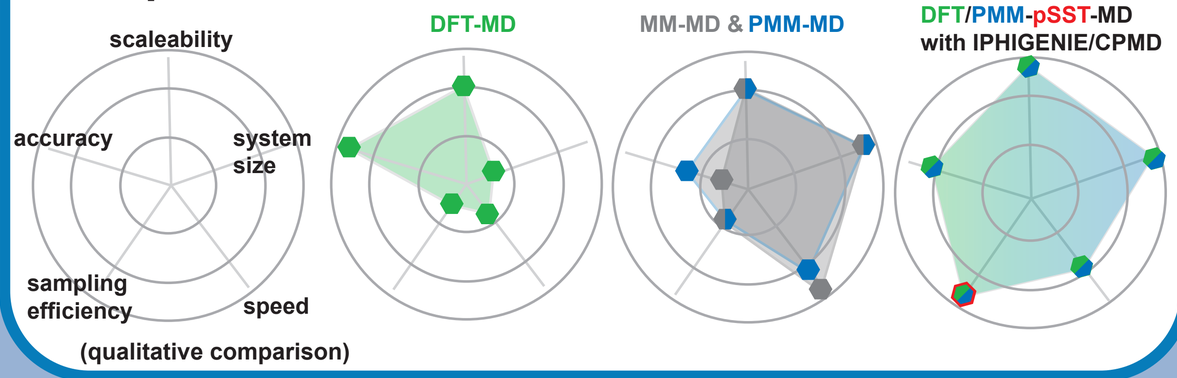
Highlights

- state-of-the-art model (**P1**): **DFT + polarizable MM force field**
- fast multipole QM/MM interactions: 10x performance gain through new **voxel-based implementation** (**P2**)
- generalized pSST ensemble for **enhanced sampling** efficiency (**P3**)
- highly scaleable (**P4**)
- application ready (**P5**)



■ **Fig.1:** A typical DFT/PMM system: alanine dipeptide (DFT) solvated in PMM water. Selected induced PMM dipoles are sketched by blue arrows.

Comparison of methods



P2 Hamiltonian DFT/PMM Coupling

■ **We combine standard MM force fields extended by Gaussian inducible dipoles with grid-based DFT:**

$$H = H_{\mathrm{MM}} + H_{\mathrm{PMM}} + H_{\mathrm{DFT}} + H_{\mathrm{DFT/(P)MM}}$$

with $H_{\mathrm{PMM}} = \frac{1}{2}\sum_{i,j\neq i} q_i \Phi(\mathbf{r}_i|\mathbf{p}_j,\mathbf{r}_j,\tilde{\sigma}_j) - \frac{1}{2}\sum_i \mathbf{p}_i \langle \mathbf{E}^{q,p}(\mathbf{r}_i)\rangle_{\tilde{\sigma}_i} + \frac{1}{2}\sum_i \mathbf{p}_i^2/\alpha_i$ [1] and $H_{\mathrm{DFT/(P)MM}} = H^{\mathrm{vdW}}_{\mathrm{DFT/MM}} + H^{\mathrm{bonded}}_{\mathrm{DFT/(P)MM}} + H^{\mathrm{elec}}_{\mathrm{DFT/(P)MM}}$

■ **Task:** calculate DFT/PMM interaction forces as efficiently as possible from $H^{\mathrm{elec}}_{\mathrm{DFT/(P)MM}} = \int d\mathbf{r}\rho(\mathbf{r})\Phi_{\mathrm{ext}}(\mathbf{r})$

■ **Strategy:** extend the **fast multipole scheme SAMM** [3] to symmetrically compute on the DFT grid the external electrostatic potential Φ_{ext} generated by the PMM atoms and in the remaining simulation system the reverse action of the electronic density on the PMM atoms.

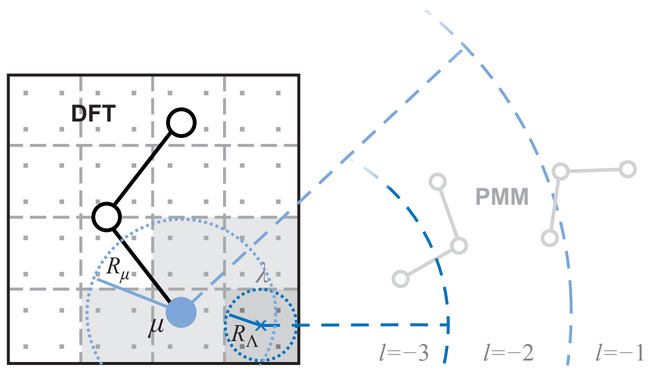


Fig. 2: Decomposition of the simulation system into a hierarchy of nested clusters and interaction levels depending on the sizes R of the depicted voxel Λ and cluster μ.

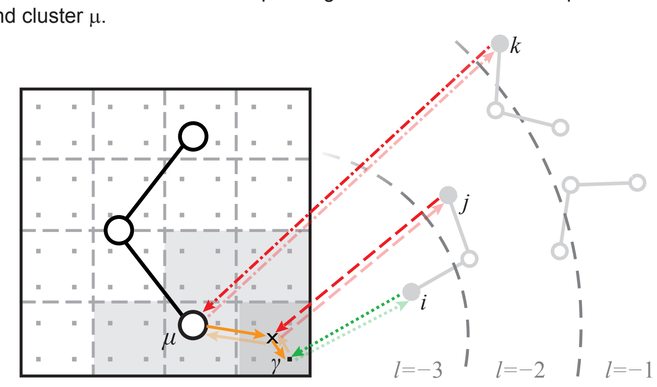


Fig. 3: Evaluation of the external potential Φ_{ext} at the grid point *γ* with SAMM (strong arrows) via direct interaction (green) and multipole and Taylor expansions (red, orange), and, fully symmetrical, calculation of forces on PMM atoms (faint arrows). As in Fig. 2, only three levels of the nested hierarchy are shown.

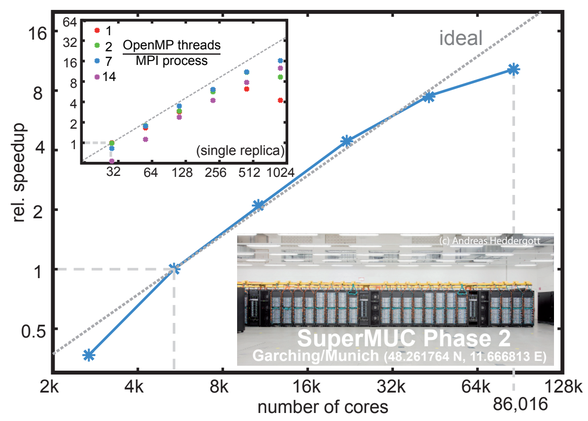
■ Our implementation combines the PMM-MD code IPHIGENIE and the DFT-code CPMD [5] in a highly scalable fashion and makes studies of **large DFT molecules solvated in accurately modeled condensed phase** feasible.

■ **SAMM$_4$:** „structure adapted fast multipole method“, enables the calculation of electrostatic forces **exactly obeying Newton's reaction principle** with **minimal computational effort** [3].

■ The external potential Φ_{ext} is computed on the DFT grid **from the PMM partial charges and induced dipoles** differently for each interaction shell using **direct interaction**, **Taylor expansions** and **multipole expansions**.

⇒ **an efficient and accurate Hamiltonian DFT/PMM multiscale hybrid method!**

P4 Scaling on SuperMUC Phase 2



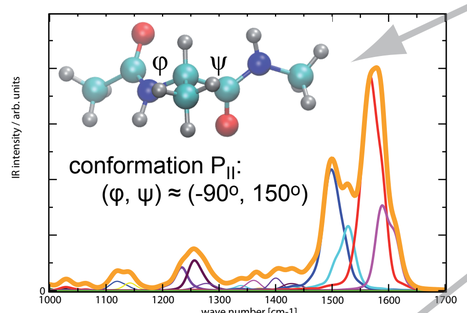
■ **Fig. 5:** The scaling behaviour of a single replica (top inset) vs. the scaling of the pSST generalized ensemble. The single replica setup is used to determine the optimal parallel setup (MPI+OpenMP).

■ **SuperMUC Phase 2 specs:**
- 86,016 Intel Haswell cores, 28 per node
- Infiniband FDR14 interconnect, 3.58 PFlop/s peak performance
- **energy-efficient warm-water cooling**

■ **scale-out setup:**
- 22 DFT atoms of the solute molecule alanine dipeptide and 26,922 PMM atoms in the aqueous solvent
- 128 replicas, up to 96 MPI processes per replica, 7 OpenMP threads per MPI process (optimal performance on the 2x14 core Intel Haswell node architec ture). All 86,016 cores of SuperMUC Phase 2 were used.

P5 Scientific Results: Vibrational Spectra and Free Energy Surface

■ Alanine dipeptide is a minimal model for studying the mechanisms of **protein folding**. Its secondary structure is described by the two dihedral angles φ and ψ.



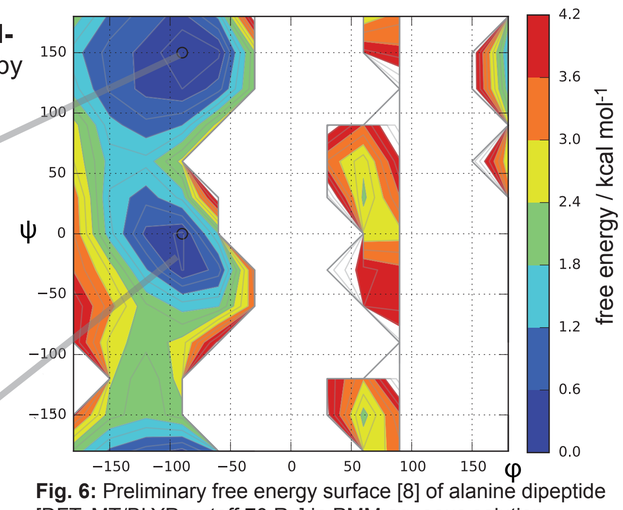


Fig. 6: Preliminary free energy surface [8] of alanine dipeptide [DFT, MT/BLYP, cutoff 70 Ry] in PMM aqueous solution.

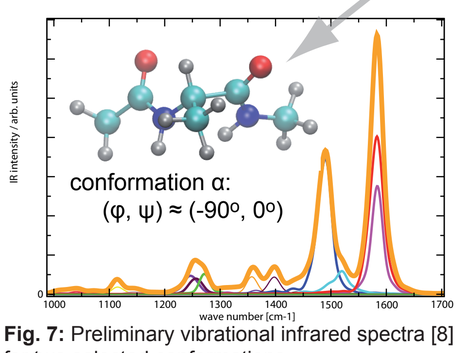


Fig. 7: Preliminary vibrational infrared spectra [8] for two selected conformations.

■ The DFT/PMM-pSST ensemble method allows to access the conformational surface of solvated polypeptides at the DFT level by MD simulations.

■ The high accuracy of DFT enables us to decode the infrared spectrum of the solute molecule in terms of its conformations [7].

■ IPHIGENIE/CPMD on SuperMUC enables computing of both of these observables at a previously unachieved accuracy [8].

P3 Parallel Simulated Solute Tempering (pSST)

Problem:

■ At ambient temperature, the time to visit all possible conformations of a biomolecule is much larger than the possible simulation time
⇒ plain MD has poor conformational sampling efficiency.

Solution:
pSST generalized ensemble

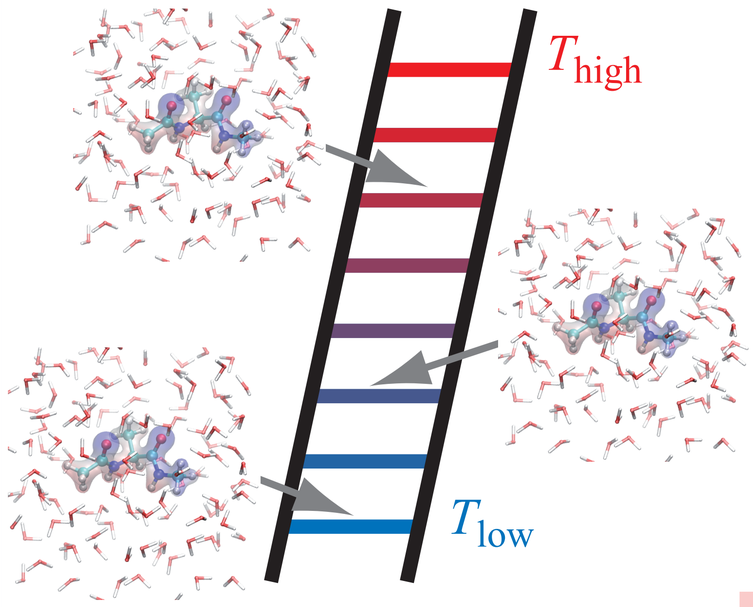


Fig. 4: Sketch of the “temperature ladder“ and three replicas.

■ At higher temperatures, transitions between conformational states occur much more often.

■ Multiple replicas of the simulation system are simulated in parallel at different temperatures $T \in [T_{\mathrm{low}}, T_{\mathrm{high}}]$. Periodically, the replicas may jump to another temperature. A Metropolis criterion ensures that the ensemble is statistically valid. [6,9]

■ All replicas contribute to the SST weight parameters that ensure uniform temperature distribution of the replicas.

⇒ **greatly enhanced sampling efficiency and high scalability**

[1] A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976)
[2] M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias, *J. Chem. Phys.* **138**, 244103 (2013)
[3] K. Lorenzen, C. Wichmann and P. Tavan, *J. Chem. Theory. Comput.* **10**, 3244-3259 (2014)
[4] M. Schwörer, K. Lorenzen, G. Mathias, and P. Tavan, *J. Chem. Phys.* **142**, 104108 (2015)
[5] J. Hutter et al., CPMD: Car-Parinello Molecular Dynamics, v4.1, (c) IBM Corp 1990-2015 and MPI für Festkörperforschung Stuttgart 1997-2001 (*www.cpmd.org*)
[6] R. Denschlag, M. Lingenheil, P. Tavan, G. Mathias *J. Chem. Theory Comput.* **5**, 28474 (2009)
[7] G. Mathias and M. D. Baer, *J. Chem. Theory Comput.* **7**, 2028 (2011)
[8] M. Schwörer and G. Mathias (*in preparation*)
[9] M. Schwörer, C. Wichmann, E. Gawehn and G. Mathias (*submitted to JCTC*)
[10] M. Schwörer et al., Poster @SC13 (2013)

Funding: SFB 749/C4, KONWIHR-III

IPHIGENIE is developed under GPL and can be downloaded at https://sourceforge.net/projects/iphigenie
For more information about SuperMUC visit https://www.lrz.de/services/compute/supermuc