Molecular Dynamics and Monte Carlo methods and their extensions for coarse-grained simulations of polypeptide folding.

Abstract.

The first part of my project was to test Hybrid Monte Carlo algorithm (and its two extensions: Shadow Hybrid Monte Carlo and Separable Shadow Hybrid Monte Carlo) on a simple Lennard-Jones fluid model and compare it with Molecular Dynamics with two different thermostating algorithms: Berendsen thermostat and Nose-Hoover thermostat. I have performed analysis of the temperature histograms for all methods, ergodicity test to evaluate the effectiveness of the methods and the acceptance ratio of the Metropolis test dependence on the number of MD steps between HMC test and on the timestep value itself for the hybrid methods. Results proved that HMC and S2HMC are worthy of further studies.

The main part of the project was the implementation of the HMC algorithm to the UNRES force field, and merging it with already present in the program Multiplexed Replica Exchange Molecular Dynamics. I have named the newly implemented algorithm as Multiplexed Replica Exchange Hybrid Monte Carlo. The newly implemented MREHMC algorithm to UNRES force field was thoroughly tested and compared with the 'classic' MREMD algorithm. The computational cluster allowed for a simulation of about 20 replicas per simulation while performing the computation in reasonable time. The new and the old method was tested with a set of 15 small proteins (12 -97 residues) all with different sequence and tertiary structure. The MREHMC algorithm was tested in several different conditions to evaluate the optimal values of the HMC specific parameters. I have performed tests of the acceptance ratio of the Metropolis test introduced by the hybrid algorithm, defined the best starting structure (extended or native) for the simulations, evaluated the optimal L value and the length of simulations allowing trajectory to converge and finally compared the new method with the 'classic' MREMD. I have performed detailed comparison between two methods, all described in my work. Overall the MREHMC performance was slightly better than the older algorithm with regarding the C^{α} RMSD to the native structure of the best clusters from simulation. Additionally I have proven the superiority of the method

in case of the 1CLB protein for which the MREMD simulations could not find the global structural minima, while the hybrid method was able to do it.

Furthermore I have performed several simulations of the method in much larger scale with the use of newer computational cluster. Simulations performed on the TRYTON supercomputer at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdansk, Poland, allowed for performing about 200 replicas per protein per simulation in a reasonable time (24-48h). Thanks to such approach it was possible to assess the differences between MREMD and MREHMC in case of much larger and much higher in computational cost simulations. Not only the number of replicas (and temperatures) could be expanded but also the total simulation time was doubled. Additional reason for the complementary tests was to study the newly developed ‘MAXLIK’ parameterization of the UNRES force field, and compare the results with the usual ‘1L2Y’ parameterization. The simulations of the methods on the new supercomputer were tested with a set of 10 small proteins (12 – 95 residues) all with different sequence and tertiary structure. Finally the two different simulation methods (MREMD and MREHMC) in two different parameterizations of the force field were compared (thus giving four distinctive types of simulation). The performance of the new ‘MAXLIK’ parameterization was far superior than for the old parameterization, while the difference between MREMD and MREHMC methods were smoothened (because of fact that for such big simulations the MREMD algorithm was able to more effectively search the conformational space).

The second part of my project was to retrieve kinetic information from the MREMD/MREHMC UNRES simulations. For this purpose I have used the original g_kinetics algorithm from Gromacs package.. The output of UNRES simulation had to be specially converted for the needs of the program in order to get two files needed by the g_kinetics program. For the test of the g_kinetics algorithm performance for UNRES results, the two different parameterizations were chosen: ‘1L2Y’ and ‘MAXLIK’, and the three types of performing simulations algorithms: MREMD, MREHMC and classic MD. The performance with regards to time of protein folding of the three methods therefore was tested. The classic MD was used to compare the time of folding and to visualize the improvement in the Replica Exchange methods. Finally the method allowed to get the information for which it was designed, and I could compare the simulation methods regarding the folding times of the proteins. The obtained results allowed to study the

dependence of the folding time on the protein size. It also allowed to compare the simulation folding time with experimental protein folding time values and asses the approximate difference between UNRES protein folding simulation times and the real time of protein folding.

Finally I have performed application of protein folding simulations of Bacteriocins. In this part of the project I was trying to answer the question of one particular bacterial strain susceptible to its own produced bacteriocin. The idea was to obtain the tertiary structures of the proteins that do not have experimental native structures, and look for differences in their structure. Additionally I have performed genome comparison of several closely related strains in search for hints in their genome regarding the self-lethality.