# Development of a TIGER2h-Based Replica Exchange Algorithm with Adaptive Shell Sizes

***M. Kulke, N. Geist, L. Schulig, M. Delcea**, Institut für Biochemie, Universität Greifswald*

## In Short

- Development of a new replica exchange protein folding algorithm
- Determination of a solvent correction-term energy for the usage of adaptive shell sizes
- Folding of insulin alymoid fibrils with the new algorithm

Since the initial development of the replica exchange molecular dynamics (REMD) method in 1999[1], the algorithm has been applied in the field of protein folding, free energy calculations, umbrella sampling and generally conformational ensemble generation. However, with the growing need in larger system sizes, the algorithm is not practicable, because of its poor scaling regarding the number of freedoms of the system. Therefore, multiple improved variants of the algorithm were developed in recent years, including hybrid REMD[2], Hamiltonian REMD, replica exchange with solute tempering (REST) and temperature intervals with global exchanges (TIGER2)[3]. Last year, we presented a sophisticated algorithm TIGER2hs[4] that combines the hybrid REMD and TIGER2 approaches to achieve a fast and accurate structure folding method. It was successfully applied for the folding of fibronectins C-terminal cross linking region comprising 100 amino acids[5]. This would not have been realistically feasible with the original REMD implementation, because of the tremendously high needed computational resource investment.

In the TIGER2hs approach, like in any other REMD variant, multiple parallel MD simulations (replicas) are utilized at different temperatures that allow the exploration of a molecular system's entire phase space. Periodically, the replicas are cooled down to the temperature of interest and the conformational quality is rated based on the potential energy. Subsequently, the Metropolis sampling criterion (MSC) is applied between a replica on a high temperature and its corresponding baseline replica. With a given probability their temperatures are exchanged. This evaluation is performed in the presence of the hybrid solvent that contains approximately the first two water shells around the solute explicitly and a continuum model for water molecules further away. After this MSC step the replicas are heated up in explicit solvent to their respective new temperatures based on the exchange ranking. This cycle is continued until the Boltzmann ensemble converges.

The accuracy of this resulting Boltzmann ensemble is dependent on the correct number of explicit water molecules to represent the hydration shells of the solute for the MSC. To clarify this aspect, let us assume the two extreme cases, only implicit solvent and only explicit solvent. In implicit solvent the water mainly represents bulk properties and the hydration shell of the solute is not considered. This leads to a destabilization of turn, coil and sheet motifs, while helix motifs are overrated and a wrong Boltzmann ensemble is obtained. If the system is solvated in pure explicit solvent, the MSC potential energy is not driven by energy differences in the protein conformation, but by differences in the solvent states. This does not necessarily produce a wrong Boltzmann ensemble but increases the convergence time for the TIGERhs simulation drastically and often leads to unaffordable resource and time requirements. Therefore, choosing the correct number of water molecules is not a trivial task. In our previous study, we analyzed the energy differences in bulk and solvent shell water for the Van-der-Waals and Coulomb energies and provided a rule-of-thumb in taking the water number that completely represents the first two water layers around a solute (Figure 1, left). With this rule most of the energy differences between bulk and hydration shell water molecules are considered and it provided a good Boltzmann ensemble for multiple investigated peptides. At the moment, the optimal number of water molecules is determined from only one conformation or a small ensemble of conformations. Therefore, the method assumes that the shell size and thus number of water molecules is identical for every solute conformation, which is not true (Figure 1, right). This leads to an overestimation of compact conformations as these have generally less water in the hydration shell and are found primarily in the trajectory, where the optimal water number is determined from.
In this project, we will investigate this issue and develop an improved method to overcome the fixed size water shell.

Generally, there are two different approaches to tackle the problem. The first (I) is that for every conformation the optimal number of water molecules is determined by a fixed cut-off from the solute. The

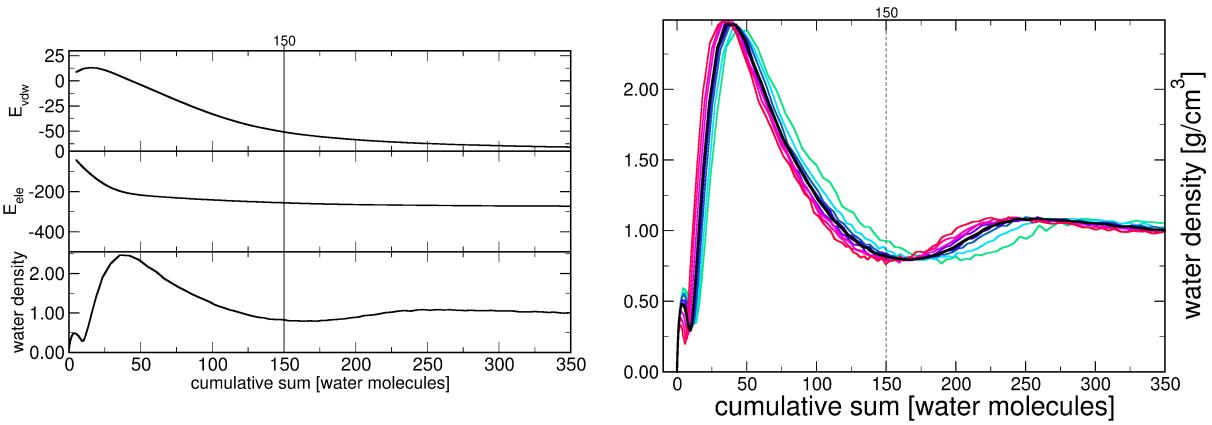


***Figure 1:*** *Analysis of the solute-solvent interface for the peptide $(AAQAA)_3$ from a molecular dynamics simulation trajectory. (Left) The number of molecules in the water shells of the native protein conformation are plotted against the density, the Coulomb energy Eele and the Van-der-Waals energy Evdw. (Right) The number of molecules in the water shells against the density for different protein conformations showing that the first two hydration shells contain between 150 and 200 water molecules, which varies between conformations. The TIGER2hs replica exchange simulation was performed with 150 water molecules.*

main challenge is the comparison of the potential energies in the MSC, because the topologies are different and every explicit water molecule has an inner potential energy. The potential energy differences of two systems are mainly driven by the different number of water molecules then. This results in a preference for extended conformations in the Boltzmann ensemble and the native conformation is often not even found. To make the two protein conformations comparable, an error-correction term in the MSC has to be applied to approximate the inner energy of explicit water molecules. Since this inner potential energy is dependent on simulation parameters like cut-off, temperature, pressure and PME, it must be easily determinable prior to a replica exchange simulation or robust enough to minimize errors in the Boltzmann ensemble if the simulation parameters are changed. We aim to obtain this inner potential energy in two different ways:
(Ia) By simulating water boxes with varying numbers of molecules, the potential energy can be averaged and the inner potential energy of a water molecule is obtained as the slope of a linear regression..
(Ib) Based on initial tests, we narrowed the range for the correct inner potential energy of a water molecule to 44-50 kJ/mol. The dPCA allows the assessment, of inappropriate values compared to the correct inner potential energy by visual inspection. Therefore, a trial and error approach is feasible, where different values are simply tested and compared. With this, the range is narrowed further, until we reach a certainty of at least 0.1 kJ/mol.

The second approach (II) focuses on dynamic estimation of the optimal water molecule number during the simulation based on the so far obtained Boltzmann ensemble.

## WWW

https://biochemie.uni-greifswald.de/forschung/forschung-in-den-arbeitskreisen/ordner-aks-lehrstuehle/biophysikalische-chemie-april-2019/

## More Information


[1] Sugita, Y.; Okamoto, Y., *Chem. Phys. Lett.* **1999**, *314* (1-2), 141-151.

[2] Okur, A.; Wickstrom, L.; Layton, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C., *J. Chem. Theory Comput.* **2006**, *2* (2), 420-433.

[3] Li, X.; Snyder, J.A.; Stuart, S.J.; Latour, R.A., *J. Chem. Phys.* **2015**, *143* (14), 133105.

[4] Geist, N.; Kulke, M.; Schulig, L.; Link, A.; Langel, W., *J. Phys. Chem. B* **2019**, *123* (28), 5995-6006.

[5] Kulke, M.; Uhrhan, M.; Geist, N.; Ohler, B.; Brüggemann, D.; Langel, W.; Köppen, S., *submitted* **2019**.