Electronic Supplementary Information

A Technique of Removing Solvent

We applied a cutoff-removing technique to move the background density of solvent in the emd-5122 system. A cutoff value μ^c is defined as, if $\forall \mu^s(r) < \mu^c$, $\mu^s(r) = 0$. **Figure S1** shows that how the number of virtual particles N changes with μ^c as it increases from 0 to 1 in the process of coarse-graining the emd-5122 system. With μ^c increasing, N reduces to 0 as $\mu^c = 1$. The decreasing of N indicates that the background density of solvent is gradually being removed. When μ^c is greater than 0.35, the solvent in the background is almost completely removed.

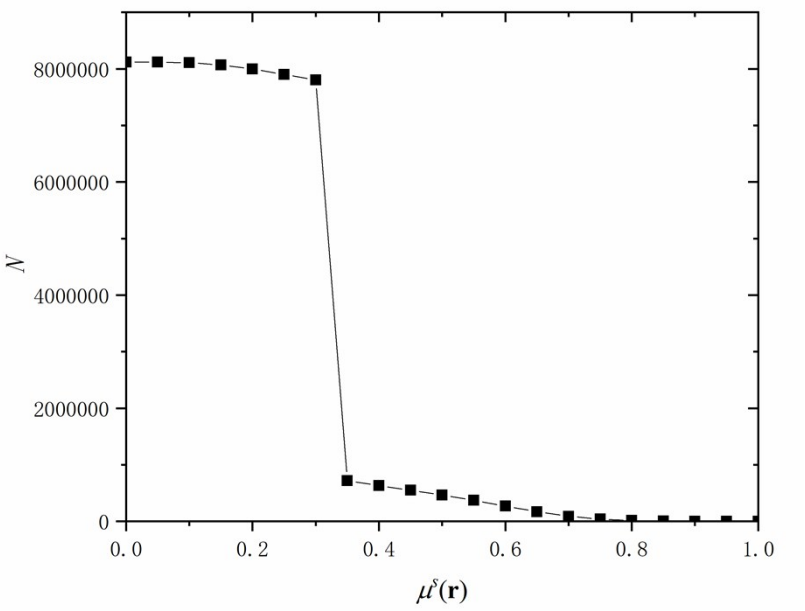


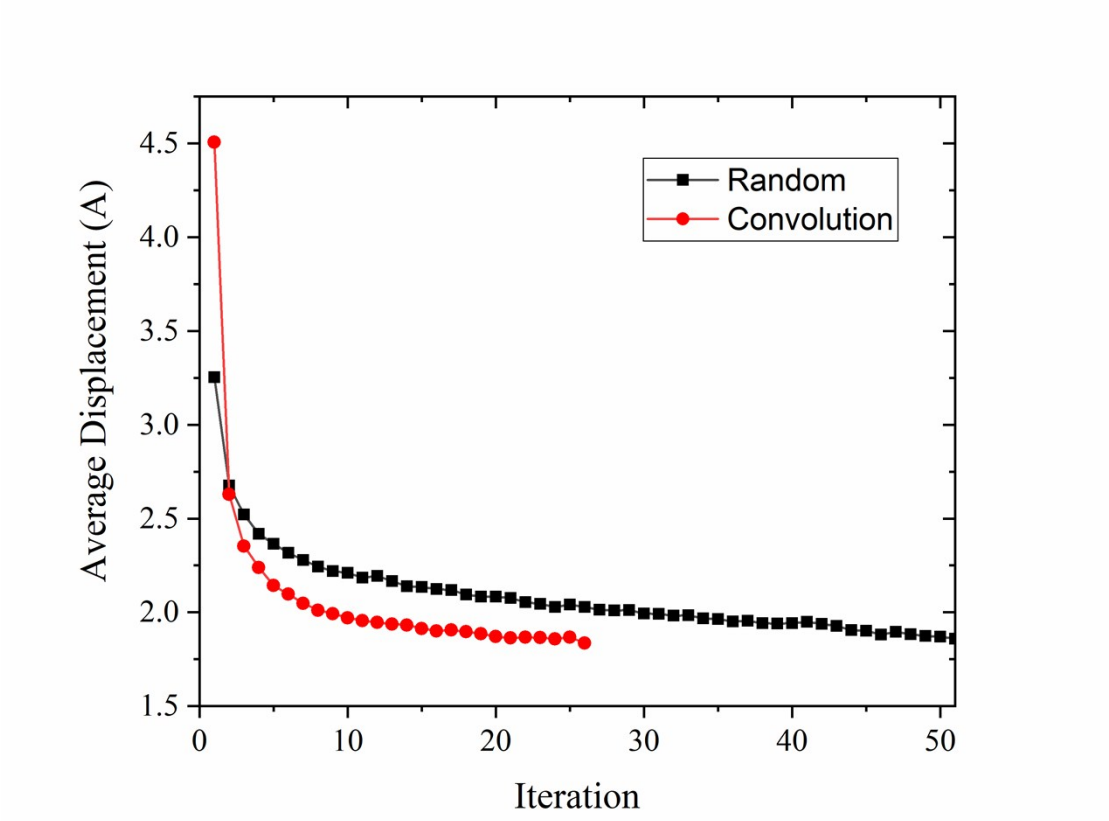
Figure S1. The number of virtual particles N changes with $\mu^s(r)$ from 0.0 to 1.0.

Details about convolution

In this study, the size of the kernel is determined automatically according to the size of EM maps and number of CG beads, which is empirically defined as $int(\sqrt[3]{N \cdot \frac{N_{CG\ per\ kernel}}{N_{CG}}})$. $int(x)$ is a round-off function, N is the number of voxels with non-zero densities, N_{CG} is the target number of CG seeds and $N_{CG\ per\ kernel}$ is the average of CG seeds per kernel. In CK-CG, $N_{CG\ per\ kernel}$ can be an integer larger than 1, which means maxpooling will identify more than one CG seeds. For example, in this work $N_{CG\ per\ kernel}$ is set to 3, that means the first, second and third largest score would be identified respectively. If the number of CG seeds guessed by convolution process is smaller than N_{CG}, more CG beads will be randomly added to kernels having large density.

Comparison of CK-CG and original K-means algorithms

The coarse-graining of emd-5122 system is tested by using the CK-CG and original K-means algorithms, where CK-CG uses an initial guess operated by convolution for while K-means utilized a random initial guess. Both of them are divided into 9860 CG beads, with the convergence criterion of average displacement being set to be smaller than 1.856. The time cost of original K-means is twice slower than that of CK-CG.



Pulling Dynamics Details

To set up pulling dynamics, we first defined the fixed group and pulling group according to the length with the first atom on the left side l, as shown in Figure S2. If $l < l_{fixed}$, those atoms are included in the fixed group. Otherwise, if $l > l_{pulling}$, they are included in the pulling group, where l_{fixed} and $l_{pulling}$ correspond to two empirical parameters. The atoms in the fixed group are restrained by using the command of "fix setforce 0 0 0" in the MD package LAMMPS[1]. Meanwhile, a constant force with the value of $\frac{5000pN}{N_{pulling}}$ along the pulling direction is added to each atom in the pulling group by using "fix addforce", where $N_{pulling}$ represents the number of atoms in the pulling group. The Langevin dynamics[2] simulation was carried out for 20 ns at 300 K.

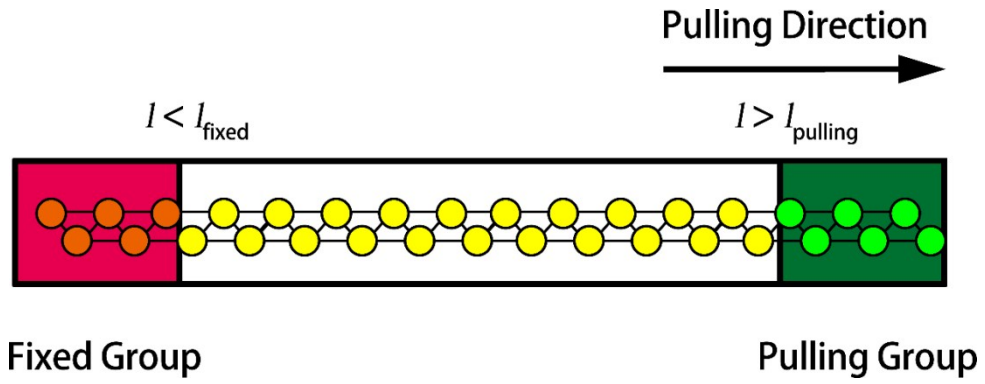


Figure S2 A schematic illustration on pulling dynamics simulation, where the molecule is represented with the colorful spheres connected by sticks. The fixed group is colored in red and the pulling group is colored in green.

References:

1. S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1-19.
2. T. Schneider and E. Stoll, *Phys. Rev. B*, 1978, **17**, 1302-1322.