

# Inverse Boltzmann iterative multi-scale molecular dynamics study between carbon nanotubes and amino acids

Wanying Huang<sup>1</sup>, Xinwen Ou<sup>3</sup> and JunYan Luo<sup>2,\*</sup>

<sup>1</sup> T-Life Research Center, State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai 200433, China; [huangwanying1128@163.com](mailto:huangwanying1128@163.com)

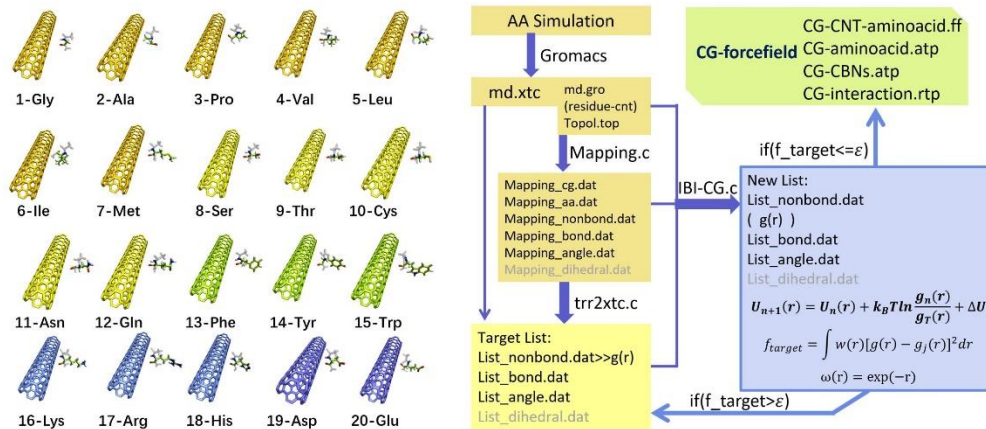
<sup>2</sup> Department of Physics, Zhejiang University of Science and Technology, Hangzhou 310023, China; [jyluo@zust.edu.cn](mailto:jyluo@zust.edu.cn)

<sup>3</sup> Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong 999077, China; [xwou@ust.hk](mailto:xwou@ust.hk)

\* Correspondence: [jyluo@zust.edu.cn](mailto:jyluo@zust.edu.cn); Tel.: (Zhejiang University of Science and Technology, Hangzhou 310023, China)

## Abstract

Our work uses Iterative Boltzmann Inversion (IBI) to study the coarse-grained interaction between 20 amino acids and the representative carbon nanotube CNT55L3. IBI is a multi-scale simulation method that has attracted the attention of many researchers in recent years. It can effectively modify the coarse-grained model derived from the Potential of Mean Force (PMF). IBI is based on the distribution result obtained by All-Atom molecular dynamics simulation, that is, the target distribution function, the PMF potential energy is extracted, and then the initial potential energy extracted by the PMF is used to perform simulation iterations using IBI. Our research results have gone through more than 100 iterations, and finally, the distribution obtained by coarse-grained molecular simulation (CGMD) can effectively overlap with the results of all-atom molecular dynamics simulation (AAMD). In addition, our work lays the foundation for the study of force fields for the simulation of the coarse-graining of super-large proteins and other important nanoparticles.



Keywords:

Multi-scale molecular dynamics, IBI, CBNs

# Introduction

In recent years, molecular dynamics simulation has become an important bridge for the combination of experimental research and theoretical research<sup>[1-3]</sup>. In particular, the use of molecular dynamics to study the molecular dynamics interaction between proteins and nanoparticles<sup>[4]</sup> has attracted close attention including biopharmaceuticals<sup>[5-7]</sup>. At present, in the process of developing new drugs, the use of molecular dynamics for early-stage drug targeted screening has become an important part of new drug development<sup>[8]</sup>. However, with the continuous improvement of computer computing power, especially the assistance of high-performance graphics GPUs for simulation calculations, much research work in molecular dynamics still only calculates the relationship between local globulin and related ligands. The interaction relationship between. In recent years, coarse-grained molecular dynamics simulation, as a pioneer in exploring the simulation of super-large and complex systems, has been playing its unique role<sup>[3; 9; 10]</sup>.

The development between coarse-grained simulation (CG) and all-atom simulation has always been complementary and spiral development <sup>[11-13]</sup>. Among them, the coarse-grained simulated force field represented by Martin (The MARTINI coarse-grained force field) provides a very effective coarse-grained structure for molecular dynamics transmembrane research. In recent years, the multi-scale coarse-grained model has been an important supplement to the traditional coarse-grained simulation, especially several multi-scale simulation methods based on molecular dynamics

simulation to extract the force field, IBI, Energy Decomposition, Fluctuation Matching method, etc. They are derived from the conformational distribution and energy or force derived from the all-atom simulation. Among them, the IBI multi-scale simulation method, as a modification of the classical mean-field coarse-grained force field, it is a very effective coarse-grained force field extraction scheme.

IBI is a multi-scale simulation method that has attracted the attention of many researchers in recent years<sup>[14-18]</sup>. It can effectively modify the coarse-grained model derived from the mean-field. IBI is based on the distribution result  $g_{Target}(r)$  obtained by All-Atom (AA) molecular dynamics simulation. The PMF potential energy is extracted<sup>[19; 20]</sup>, and then the initial potential energy extracted by the average field is used for IBI. Simulation iteration. IBI initially had a very good performance in fluid coarse-grained simulation. Then it plays a key role in the research and simulation system of Polymer. At present, IBI has been in the field of super-large Polymer system research<sup>[21-23]</sup>, represented by the German scholar Muller<sup>[24]</sup>, and much excellent research work has emerged. In addition, some researchers use IBI to conduct researches such as transmembrane research and coarse-grained simulation research of DNA and super large proteins. The interaction between amino acids and nanoparticles has been receiving the attention of many researchers<sup>[25-29]</sup>. It is the preliminary work of the adsorption of super large proteins and nanoparticles<sup>[30]</sup>. Therefore, in our research work, it is very appropriate to choose IBI as the research method for the extraction of carbon nanotubes and 20 kinds of amino acid force fields.

Our work investigated the coarse-grained interaction between 20 amino acids and the representative carbon nanotube CNT55L3. More than 100 simulation iterations were performed using IBI, and finally, the distribution obtained by the coarse-grained simulation (CGMD) can effectively overlap with the result of the all-atomic molecular dynamics simulation (AAMD). In addition, our work lays the foundation for force field research for the simulation of coarse-grained super-large proteins and other important nanoparticles.

## 1 Simulation system and research method

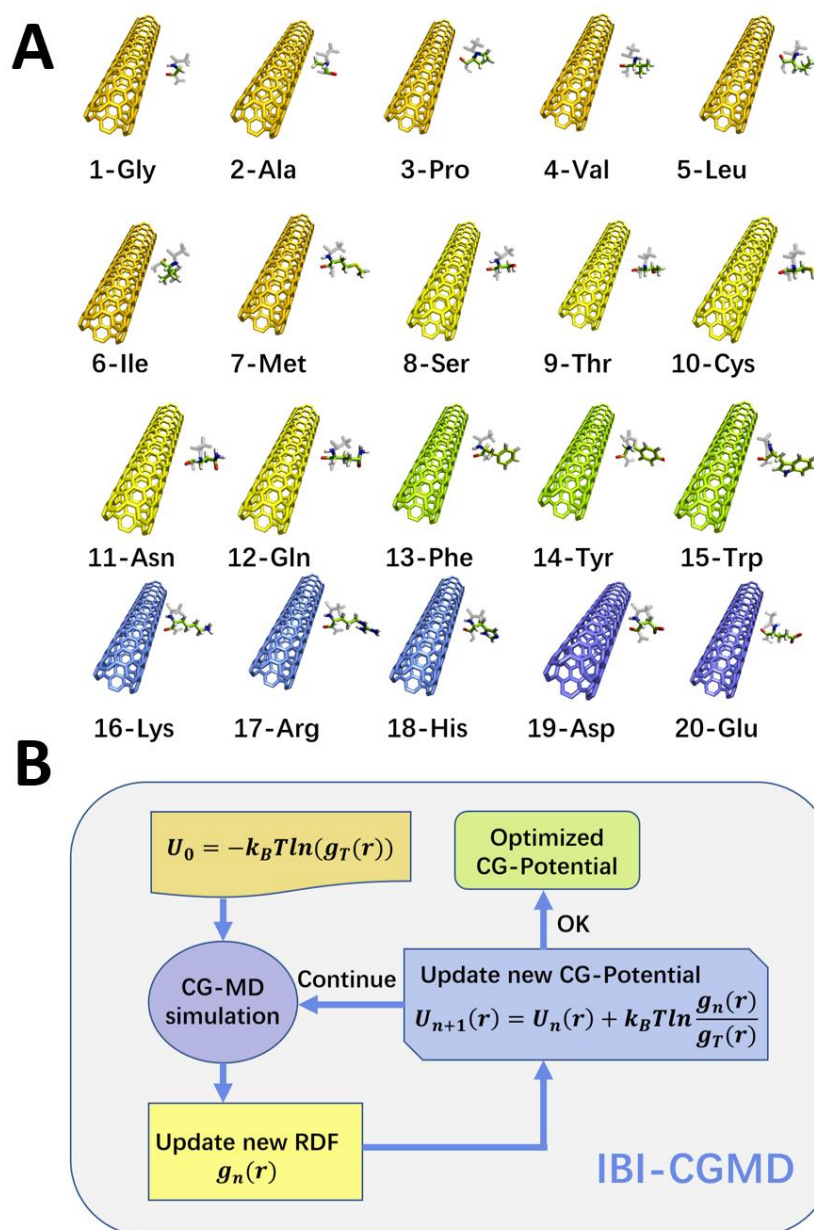


Figure1 (A) Schematic diagram of the 20Aminoacids-CNT55L3 simulation system.

(B) Simple schematic diagram of IBI process, IBI reverse Boltzmann iteration process to extract coarse-grained force field from all-atom trajectories

## 1.1 Simulation system

Our research has constructed 20 molecular systems, namely periodic infinite length CNT55L3×20Aminoacids. We studied the interaction between carbon nanotube CNT55L3 and 20 amino acids in an aqueous solution (Figure1A), and each research system did 3 repeated simulation studies, a total of more than 60 simulation research trajectories, each trajectory 200ns, We hope to conduct objective statistical research

from the atomic level to lay the foundation for the next force field extraction. Our research, based on previous studies of molecular dynamics between carbon-based nanoparticles and proteins by other researchers, has grouped 20 amino acids. As shown in Figure1A, they are (1) Nonpolar amino acid group; (2) Polar uncharged amino acid group; (3) Aromatic amino acid group; (4) Positively charged amino acid group; (5) Negatively charged amino acid group. Such a combination arrangement allows us to conduct research more systematically.

The molecular system of CNT55L3×20Aminoacids in an aqueous solution. We have selected the most commonly used SPC216 model for water molecules. We use Gromacs software throughout our molecular dynamics research. Our research used the leapfrog algorithm, and the simulation step is 2fs. In order to ensure that there is sufficient adsorption time between the amino acid and CNT55L3, the durations of each simulation is 200ns. In the initial stage, the distance between the amino acid and CNT55L3 is about 1.5 nm. Our simulation systems are all NVT system. The simulated temperature of the system is 330K, and we use the Berendsen temperature adjustment method. The Box parameters of the simulation system are:  $L_x = 4.16nm$ ,  $L_y = 4.16nm$ ,  $L_z = 2.93nm$ . We use periodic boundary conditions to make our carbon nanotubes infinitely periodic. In addition, the dynamics of the CNT55L3×20Aminoacids simulation system are studied in detail in the third chapter of this paper, and there is no repeated description. The focus of this research is to extract effective coarse-grained force field parameters from the all-atom simulation system of CNT55L3×20Aminoacids.

## 1.2 Coarse-grained Mapping Method

There are two classic coarse-grained methods: (1) Center-of-Mass (COM) based on the coarse-grained research object; (2) Based on the structure center of the coarse-grained research. This method is suitable for the geometric structure center and the center of mass of the research object to coincide exactly. In addition, in recent years, the use of clustering to coarse-grain some fluid molecular clusters is also a new method. Our research adopts the method of coarse-grained extraction of the COM coordinates, as shown in Figure2A, the calculation method of the COM coordinates is:

$$R_I = \sum_i^n \frac{r_i m_i}{\sum_i^n m_i}$$

Among them,  $R_l$  refers to the position coordinate of each coarse-grained particle (CG-Bead), and  $r_i$  is the coordinate of the all-atomic particle (AA-Bead) included in each coarse-grained particle. And the mass of coarse-grained particles is  $M_l = \sum_i^n m_i$ . Our coarse-grained method can be effectively compared with the AAMD process in the CGMD simulation process, to extract a more appropriate coarse-grained force field.

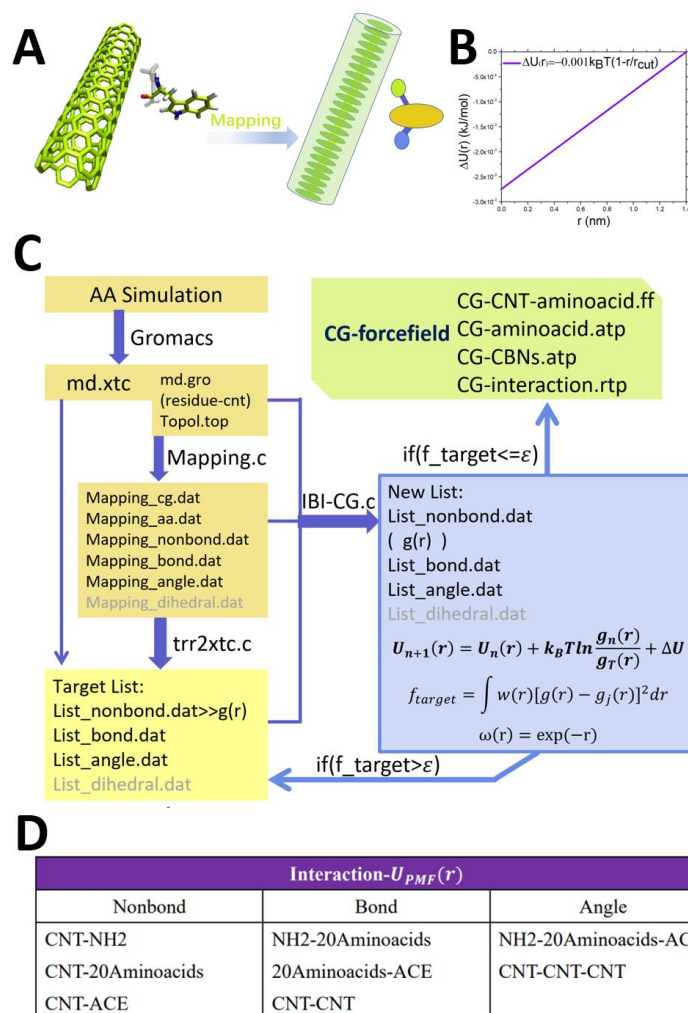


Figure2 (A) schematic diagram of the coarse-grained method of the simulation system Mapping. (B) linear addition function. (C) detailed flow chart of IBI-20Aminoacids-CNT55L3, including the interaction process of each file. (D) table: list of the average field potential energy of each initial interaction relationship.

### 1.3 Iterative Boltzmann Inversion Method

IBI (Iterative Boltzmann Inversion) is a multi-scale simulation method that has attracted the attention of many researchers in recent years. It can effectively modify the coarse-grained model derived from the mean field<sup>[31-35]</sup>. IBI is based on the distribution result  $g_{Target}(r)$  obtained by all-atom molecular dynamics simulation, and extracts the PMF potential energy



$$U_{PMF} = -k_B T \ln(g_{Target}(r)),$$

Then use the initial potential energy extracted by the mean field, using the iterative equation

$$U_{n+1}(r) = U_n(r) + k_B T \ln \frac{g_n(r)}{g_T(r)}$$

Perform simulation iterations<sup>[36-38]</sup>.

The system parameters required by the system we studied are: non-bond interaction, bond interaction, and angle interaction <sup>[14; 32; 33; 38-40]</sup>:

$$U_{Total}(r) = U_{Nonbond}(r) + U_{Bond}(l) + U_{Angle}(\theta)$$

$$P(r, l, \theta) = P_{Nonbond}(r) P_{Bond}(l) P_{Angle}(\theta)$$

According to R. L. Henderson's uniqueness theorem <sup>[41]</sup>, we can conclude that the particle distribution of the system is positively correlated with its potential energy function. Henderson's theorem is the theoretical premise of the potential energy function-like coarse-grained method based on the radial distribution function. The various interactions (Figure2D Table ) are as follows:

$$P_{Nonbond}(r) = g_{Target}(r) \propto \exp \left( -\frac{U_{Nonbond}(r)}{k_B T} \right)$$

$$P_{Bond}(l) \propto \exp \left( -\frac{U_{Bond}(l)}{k_B T} \right)$$

$$P_{Angle}(\theta) \propto \exp \left( -\frac{U_{Angle}(\theta)}{k_B T} \right)$$

$U_{Bond}(l)$  and  $U_{Angle}(\theta)$  are the non-iterative part of the potential energy extracted by PMF, and the  $U_{Nonbond}(r)$  also needs subsequent IBI optimization:

$$U_{Nonbond}(r) = -k_B T \ln(g_{Target}(r))$$

$$U_{Bond}(l) = -k_B T \ln(P_{Bond}(l))$$

$$U_{Angle}(\theta) = -k_B T \ln(P_{Angle}(\theta))$$

Among them,  $U_{Nonbond}(r)$  can be used as the initial potential energy for subsequent IBI iterations, namely

$$U_{Nonbond}^0(r) = U_{Nonbond}(r) = -k_B T \ln(g_{Target}(r))$$

As shown in Figure 1B, the IBI iteration (n\_Iterative) method of non-bond interaction is as follows:

$$U_{Nonbond}^{n+1}(r) = U_{Nonbond}^n(r) + k_B T \ln \left( \frac{g^n(r)}{g_{Target}(r)} \right)$$

Where n is the number of IBI iterations, and  $g_{Target}(r)$  is the target radial distribution function of the system. In the iterative process of IBI, part of the Force calculation

method<sup>[42]</sup> is:

$$\begin{aligned}F_{Nonbond}(r) &= -\frac{1}{\beta} \frac{d}{dr} \ln(g^n(r)) \\F_{Bond}(r) &= -\frac{1}{\beta} \frac{d}{dl} \ln(P_{Bond}(l)) \\F_{Angle}(\theta) &= -\frac{1}{\beta} \frac{d}{d\theta} \ln(P_{Angle}(\theta))\end{aligned}$$

Among them,  $\beta = 1/k_B T$

The general process of the classic IBI is as shown in Figure1B. First, the distribution and interaction function between coarse-grained atoms are extracted by the all-atom simulation, and then the coarse-grained molecular dynamics simulation (CGMD) is performed for a simulation, that is an iteration. Then extract the corresponding new distribution. If the new distribution fits the original target distribution function, the iteration ends; if the error between the new distribution and the original target distribution function does not fit the corresponding requirements, add the function obtained from the iteration to the next iteration of the potential energy function. And the next iteration will be performed until the fitting accuracy meets the research requirements.

In the specific IBI simulation iteration process, there will be many detailed problems that need to be resolved. For example, how to make the format of the system file understudy recognize each other, and how to keep the simulation unit during the transition from AA to CG is the appropriate choice. As shown in Figure2C, the details of IBI's key documents are shown in the flowchart. Among them, our AA part of the simulation used Gromacs molecular dynamics simulation software. The coarse-grained model uses md.gro and topol.top in Gromacs (topological files related to molecular structure characteristics), from which we can read the mass, atomic number, charge, and so on of each AA atom. In the process of extracting the AA target distribution function, we used the official supporting analysis tool XdrFile1.1 of the Gromacs software, in which trr2xtc.c is the trajectory analysis program that comes with the software package. In addition, in the process of IBI-CGMD, we programmed a set of iterative procedures ourselves. Each group of amino acids has undergone more than 100 iterations. The integration step length of each iteration is 0.001 ps, and each iteration runs at least 100,000 steps so that we can get more appropriate statistics. The convergence decision function in the iterative process is:

$$f_{Target} = \int w(r)[g(r) - g_{Target}(r)]^2 dr$$



Among them,  $w(r) = \exp(-r)$  is the system weight function<sup>[21]</sup>. In addition, in order to make the system potential energy update more smoothly, we use interpolation to smooth the potential energy curve, and use a semi-empirical linear function to superimpose the potential energy update process, as shown in Figure2B:

$$U_{Nonbond}^{n+1}(r) = U_{Nonbond}^n(r) + k_B T \ln\left(\frac{g^n(r)}{g_{Target}(r)}\right) + \Delta U(r)$$

$$\Delta U(r) = \alpha \left(1 - \frac{r}{r_{cut}}\right)$$

$$\alpha = -0.001 * k_B T$$

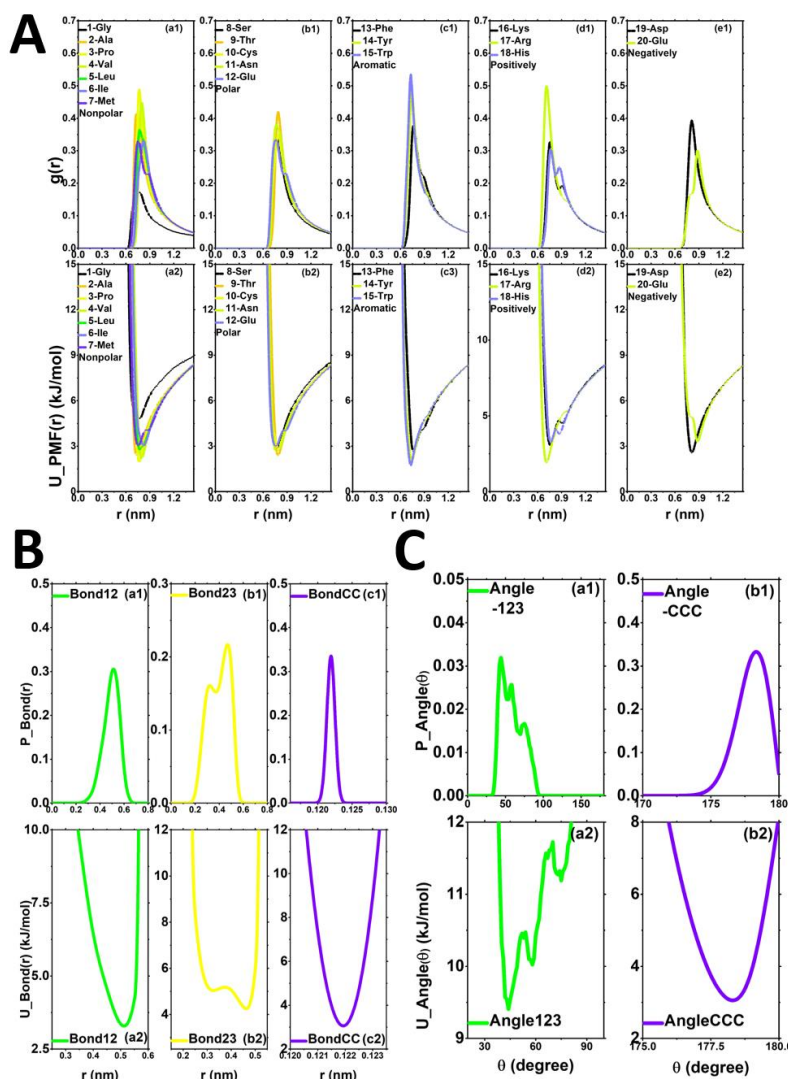
$$r_{cut} = Value(r_{cutoff})$$

Among them,  $\Delta U(r) = \alpha \left(1 - \frac{r}{r_{cut}}\right)$  is a linear function [35], when  $r=r_{cut}$ , the value of the function is 0; when  $r=0$ , the value of the function is  $\alpha$ , that is  $\alpha=-0.001*k_B T$ . Above we have introduced in detail the specific simulation iteration process of IBI and the details of various parameters.

## 2 Results and discussion

### 2.1 Initial PMF of CNT55L3-20aminoacids

Next, we will introduce the results of the average field potential energy extraction of the system from the aspects of Nonbond, Bond, and Angle. Where U-PMF is the average field potential energy derived from the target distribution.



**Figure3(A)** The radial distribution functions between different types of amino acids and carbon nanotubes and their potential energy functions extracted by full-atom simulation: Nonpolar amino acid group RDF (a1) And its U\_PMF(r) (a2); the RDF (b1) of the polar uncharged amino acid group and its U\_PMF(r) (b2); Aromatic amino acid group RDF (c1) and its U\_PMF(r) (c2); The positively charged amino acid group RDF (d1) and its U\_PMF(r) (d2); The negatively charged amino acid group RDF (e1) and its U\_PMF(r) (e2). **(B)** The bond distribution functions extracted by the all-atom simulation and its potential energy function (take CNT55L3-15-Trp as an example): NH<sub>2</sub>-Aminoacids, (a1) P\_Bond(r), (a2) U\_Bond(r); Aminoacids-ACE, (b1) P\_Bond(r), (b2) U\_Bond(r); CNT-CNT, (c1) P\_Bond(r), (c2) U\_Bond(r). **(C)** The distribution of the angle distribution function extracted by the all-atom simulation and its potential energy function (take CNT55L3-15-Trp as an example): NH<sub>2</sub>-Aminoacids-ACE, (a1) P\_Angle( $\theta$ ), (a2) U\_Angle( $\theta$ ); CNT-CNT-CNT, (b1) P\_Angle( $\theta$ ), (b2) U\_Angle( $\theta$ ).

**(1)The initial distribution and PMF of the nonbond group.** Figure3A is a schematic diagram of the nonbond radial distribution function  $g(r)$  between 20 amino acids and CNT55L3 and the potential energy function of each group. We use the coarse-grained distribution of centroids to well reflect the adsorption characteristics of each amino acid and CNT55L3. Among them, the RDF of aromatic amino acids (Figure3A (c1) and (c2)) has a higher First-Peak. There are also differences between the different amino acids within. And in Figure3A (d1) and (d2), we can observe that the potential energy depth of 17-Arg is comparable to that of aromatic amino acids. This is very necessary because this part of the potential energy is as the initial potential energy curve for the next iteration of IBI.

**(2)The initial distribution and PMF of the bond group.** Figure3B describes the distribution of the bond group extracted by the AA system and the potential energy curve. Among them, Figure3B (a1) (a2) is the bond distribution function and potential energy curve of NH2-Trp; Figure3B (b1) (b2) is the bond distribution function and potential energy curve of Trp-ACE; Figure3B (c1) (c2) shows the bond distribution function and potential energy curve of CNT-CNT. The extraction of this part of the potential energy curve is the force field parameter that needs to be used in the next iteration of IBI.

**(3)The initial distribution and PMF of the angle group.** Figure3C describes the distribution function of the angle part and the potential energy curve extracted by the AA system. Among them, Figure3C (a1)(a2) is the distribution of the angle part of NH2-Trp-ACE and the potential energy curve; Figure3C (b1)(b2) is the distribution of the angle part of CNT-CNT-CNT And the potential energy curve. Generally speaking, this part of the average field potential can either choose to iterate or not to iterate. Since the focus of our research is on the non-bonded interaction energy of CNT-Aminoacids, we will not do iterative optimization for this part of the content.

## 2.2 The iteration process of CNT55L3-20aminoacids

A successful IBI iterative process is based on whether the radial distribution functions generated by AAMD and CGMD in the iterative process can overlap well. The higher the accuracy of the fitting function  $f_{Target}$ , the closer the dynamics of our IBI iterative system and AAMD. As shown in Table3.3, the accuracy of the fitting function  $f_{Target}$  can reach  $e^{-4}$ - $e^{-5}$ . Our iterative fitting results can be considered to

have reached a higher accuracy compared with previous IBI related work.

## (1) IBI-Representative Amino Acid Research

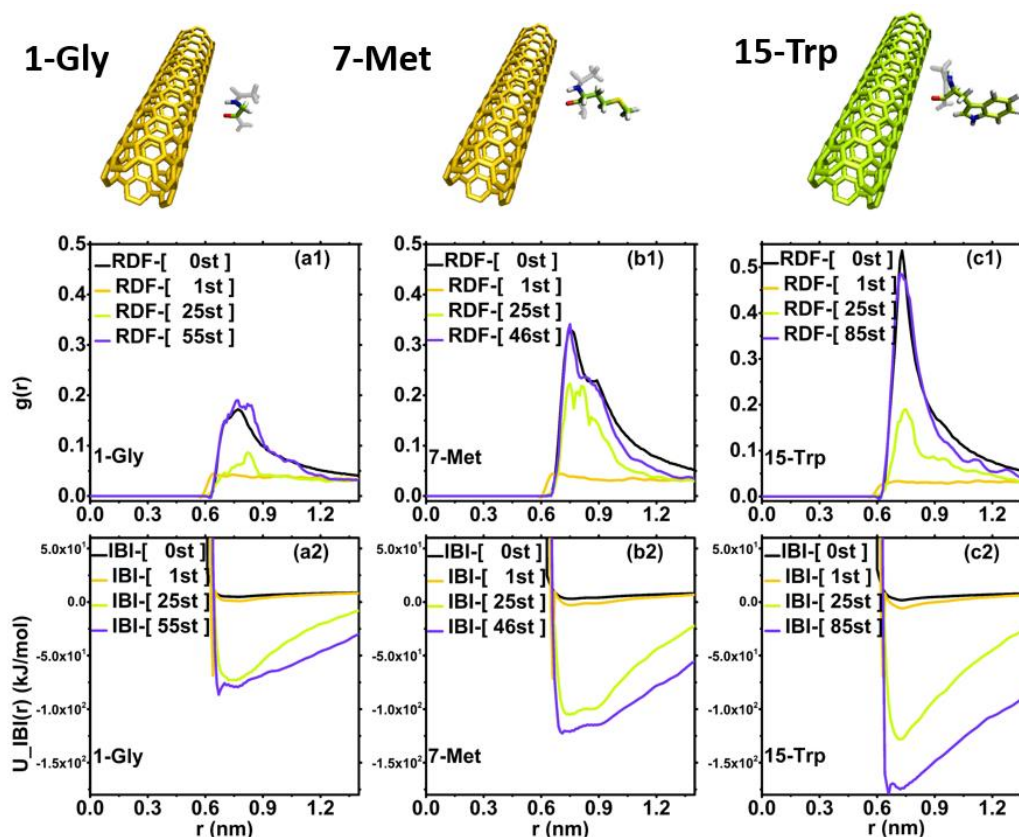


Figure4 Representative amino acid studies: (1) short-chain non-polar amino acid 1-Gly, (a1) is RDF iteration, (a2) is IBI potential energy iteration; (2) long-chain non-polar amino acid 7-Met, (b1) It is RDF iteration, (b2) is IBI potential energy iteration; (3) Aromatic amino acid 15-Trp, (c1) is RDF iteration, (c2) is IBI potential energy iteration

We selected three representative amino acids to show the difference in CNT adsorption strength. These three amino acids are (1) short-chain non-polar amino acid 1-Gly, (2) long-chain non-polar amino acid 7-Met, And (3) 15-Trp, an aromatic amino acid. As shown in Figure4, (a1)(b1)(c1) are their respective RDF iteration process, (a2)(b2)(c2) are their respective IBI potential energy iteration process. Among them, the black line is IBI-[ 0st ], the orange line is IBI-[ 1st ], the green line is IBI-[ 25st ], and the blue line is the number of iterations each reached a certain fitting accuracy. The iteration accuracy of the three amino acids is: (1) the number of optimization iterations for 1-Gly is IBI-[ 55st ], and the fitting accuracy is  $f_{\text{target}} = 1.405139\text{e-}04$ ; (2) the number of optimization iterations for 7-Met is IBI-[ 46st ], the fitting accuracy is

$f_{\text{target}} = 2.522036 \times 10^{-4}$ ; (3) The number of optimization iterations of 15-Trp is IBI-[85st], and the fitting accuracy is  $f_{\text{target}} = 2.881077 \times 10^{-4}$ . We can observe that as the number of iterations increases, the RDF of amino acids is getting closer and closer to  $g_{\text{Target}}(r)$ , and at the same time, its  $U_{\text{IBI}}(r)$  iteration potential becomes deeper.

In addition, we can also clearly observe from Figure4: (1) From the perspective of RDF First-peak:  $\text{RDF-1-Gly} < \text{RDF-7-Met} < \text{RDF-15-Trp}$ ; (2) From the perspective of the potential well depth of the interaction between the three amino acids and CNT, the aromatic amino acid 15-Trp is significantly deeper than other amino acids, while the long-chain non-polar amino acid 7-Met is deeper than the short-chain non-polar Amino acid 1-Gly. This also shows that during the interaction between amino acids and carbon nanotubes, the  $\pi$ - $\pi$  interaction has a greater influence on the adsorption strength than the hydrophobic interaction. This conclusion coincides with the results of AAMD's kinetic analysis, which also shows that our multi-scale IBI research results are trustworthy.

## (2) the Distribution during IBI iteration

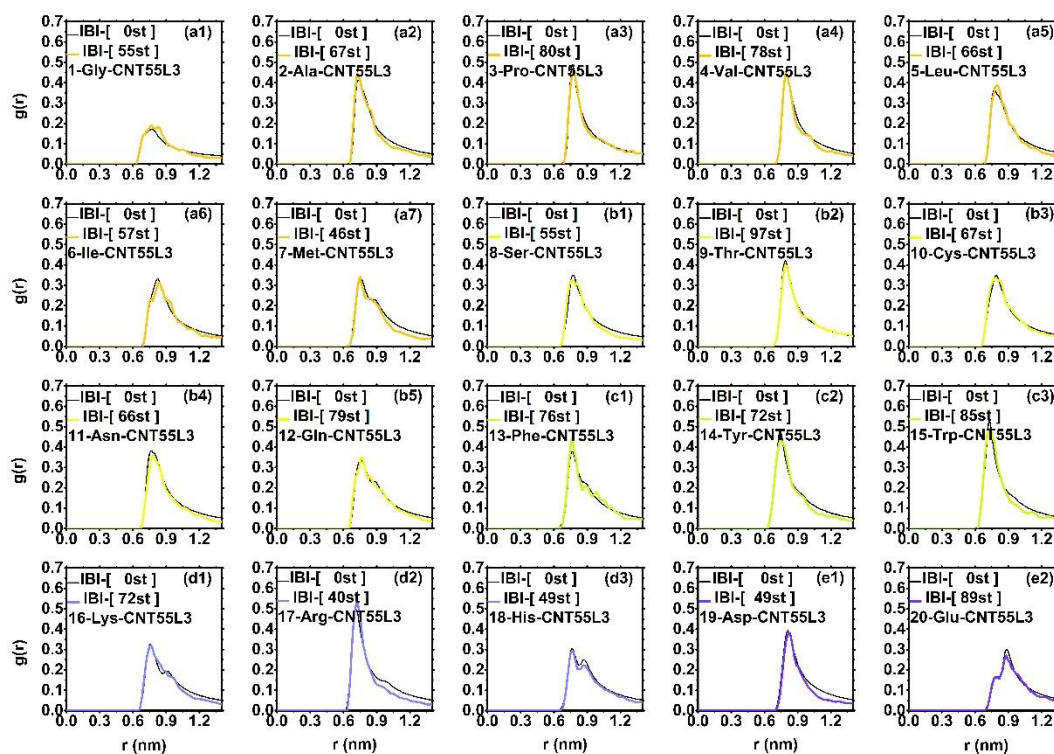


Figure5 The radial distribution function and target radial distribution function between 20 amino acids and CNT55L3 obtained after IBI iterative optimization. The black lines

in each figure are their respective target RDF, and (a1-a7) are non-polar The RDF of the amino acid group, the line color is orange; (b1-b5) is the RDF of the polar uncharged amino acid group, the line color is yellow; (c1-c3) is the RDF of the aromatic amino acid group, the line color is green; (d1-d3) is the RDF of the positively charged amino acid group, the line color is blue; (e1-e2) is the RDF of the negatively charged amino acid group, the line color is purple.

Figure5 shows the iterative process of the radial distribution function of the IBI system between 20 amino acids and CNT55L3. The part of the black line in each data graph is the target distribution function  $g_{Target}(r)$  between each amino acid and CNT55L3. Among them, in order to facilitate the reader to observe the data, we set the amino acid color of the same group to be consistent. We can be very clearly to observe that, within the accuracy range allowed by the error of the fitting function, the 20 amino acids all have a good fitting with the target distribution function  $g_{Target}(r)$ . Slightly flawed, such as 01-Gly-CNT55L3 compared with its target distribution function  $g_{Target}(r)$ , the degree of coincidence is not so good, but its target distribution function  $g_{Target}(r)$  still reaches 1.405139e-04, which is an acceptable error scope. This fully shows that our IBI iterative process is effective. Here, we have completed the fitting of the target distribution function between 20 amino acids and CNT55L3 carbon nanotubes.

### (3) Potential energy in the iterative process of IBI



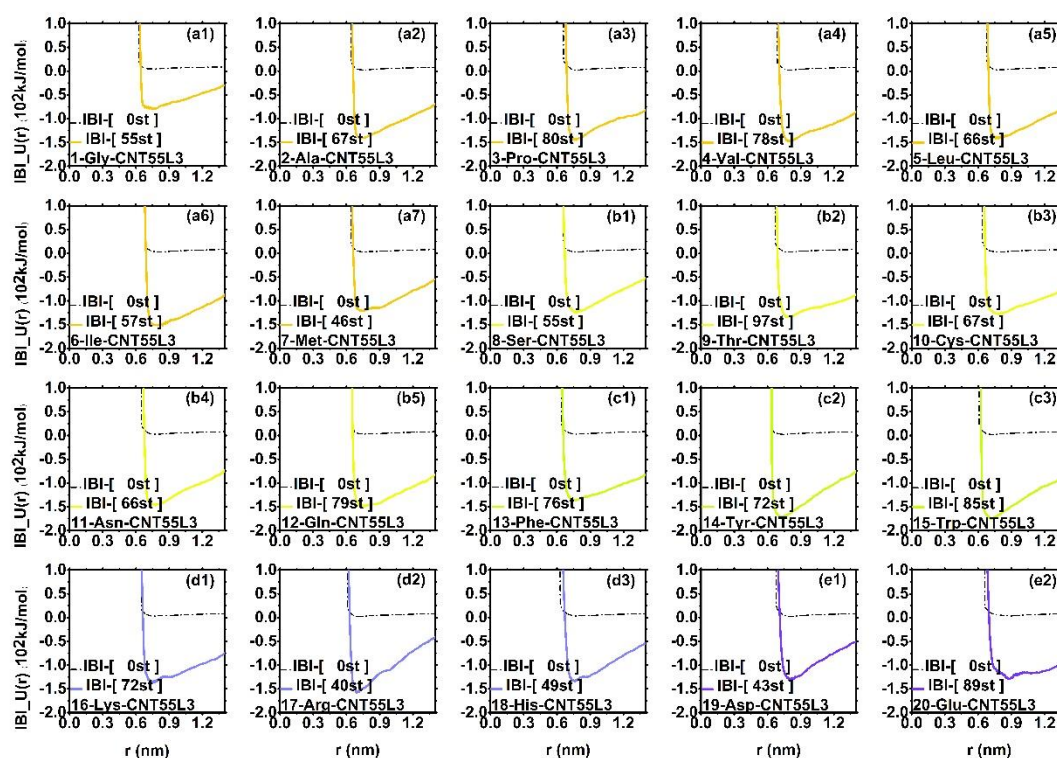


Figure6 The nonbond potential energy function and initial nonbond potential energy function (PMF) between 20 amino acids and CNT55L3 obtained after IBI iterative optimization. The black dotted lines in each figure are their respective initial nonbond PMF, and (a1-a7) are nonpolar PMF of the amino acid group, the line color is orange; (b1-b5) is the non-bonding PMF of the polar uncharged amino acid group, the line color is yellow; (c1-c3) is the nonbond PMF of the aromatic amino acid group, the line color is green; (d1-d3) is the non-bonding PMF of the positively charged amino acid group, and the line color is blue; (e1-e2) is the nonbond PMF of the negatively charged amino acid group, and the line color is purple.

Figure6 shows the change of the non-bond Nonbond potential energy function of the 20 amino acids and CNT55L3 after the iteration of the radial distribution function of the IBI system within the acceptable accuracy range. We can observe that the potential energy of the 20 amino acids has become deeper. That is to say, the potential energy function of the interaction between the 20 amino acids and the CNT55L3 has been well optimized so that the coarse-grained kinetics process can be simulated in all atoms, as shown in Table2.2, a similar effective distribution within the accuracy range.

| Group      | Aminoacids | n_Iterative | f_target     |
|------------|------------|-------------|--------------|
| Nonpolar   | 1-Gy       | 55st        | 1.405139e-04 |
|            | 2-Ala      | 67st        | 2.487304e-04 |
|            | 3-Pro      | 80st        | 1.628479e-04 |
|            | 4-Val      | 78st        | 1.309014e-04 |
|            | 5-Leu      | 66st        | 1.794246e-04 |
|            | 6-Ile      | 57st        | 2.260820e-04 |
|            | 7-Met      | 46st        | 2.522036e-04 |
| Polar      | 8-Ser      | 55st        | 2.312779e-04 |
|            | 9-Thr      | 97st        | 5.229923e-05 |
|            | 10-Cys     | 67st        | 9.155223e-05 |
|            | 11-Asn     | 66st        | 2.181413e-04 |
|            | 12-Gln     | 79st        | 2.359602e-04 |
| Aromatic   | 13-Phe     | 76st        | 3.148886e-04 |
|            | 14-Tyr     | 72st        | 1.940790e-04 |
|            | 15-Trp     | 85st        | 2.881077e-04 |
| Positively | 16-Lys     | 72st        | 1.224145e-04 |
|            | 17-Arg     | 40st        | 4.282584e-04 |
|            | 18-His     | 49st        | 1.075425e-04 |
| Negatively | 19-Asp     | 43st        | 2.234629e-04 |
|            | 20-Glu     | 89st        | 1.136287e-04 |

Table 2.2 The accuracy of the decision function between the initial target radial distribution function of amino acids and CNT55L3 after IBI iterative optimization and the radial distribution function obtained by the best iteration.

The above is the data recording and display of the entire IBI iteration process of 20 amino acids and CNT55L3. As shown in Table2.2, we get that the simulation accuracy can reach e-4 to e-5, which is a relatively high fitting accuracy in related studies using IBI. The results of our research work also show that it is very appropriate to choose IBI as the research method for the force field extraction of carbon nanotubes and 20 standard amino acids, and the coarse-grained force field parameters within a certain accuracy error range are obtained.

### 3 Summary and outlook

Our work investigated the coarse-grained interaction between 20 amino acids and the representative carbon nanotube CNT55L3. Our research process is based on the distribution result  $g_{Target}(r)$  obtained by all-atom molecular dynamics simulation, and then extracts the PMF potential energy, and then uses the initial potential energy extracted by the average field to perform simulation iterations using IBI. Our research results have been carried out more than 100 iterations, and finally, the RDF obtained by the coarse-grained simulation (CGMD) can effectively coincide with the result of the all-atomic molecular dynamics simulation (AAMD). In addition, our work lays the foundation for force field research for the simulation of coarse-grained super-large proteins and other important nanoparticles.

Looking forward to the future, there are three challenges: (1) How to normalize the coarse-grained force field parameters between the 20 standard amino acids and carbon nanotubes that we have obtained, and combine them with some mainstream molecular dynamics simulation software. It is convenient for molecular dynamics simulation researchers to use our force field parameters, and then provide a new way to study the interaction between super large proteins and carbon-based nanoparticles; (2) How to extend this set of IBI force field extraction between carbon-based nanoparticles and amino acids to other nanoparticles, such as gold nanoparticles, silicon-based nanoparticles, etc; (3) We hope to improve our IBI program into a general-purpose small software to facilitate research including fluids and polymers, and other systems with research value such as biological macromolecules.

### Acknowledgement:

Support from the National Natural Science Foundation of China (Grants No. 11647082 and No. 11774311) is gratefully acknowledged. Wangying Huang would like to thank Professor Michael Levitt and Professor Guanghong Zuo for helpful discussions.

### References

[1] N. Luo, J. K. Weber, S. Wang, B. Luan, H. Yue, X. Xi, J. Du, Z. Yang, W. Wei, R. Zhou, and G. Ma, PEGylated graphene oxide elicits strong immunological responses despite surface passivation,

Nature Communications 8, 14537 (2017).

[2] S. Liu, X.-Y. Meng, J. M. Perez-Aguilar, and R. Zhou, An In Silico study of TiO<sub>2</sub> nanoparticles interaction with twenty standard amino acids in aqueous solution, *Scientific Reports* 6, 37761 (2016).

[3] W. Li, J. Wang, J. Zhang, S. Takada, and W. Wang, Overcoming the Bottleneck of the Enzymatic Cycle by Steric Frustration, *Physical Review Letters* 122, 238102 (2019).

[4] G. Zuo, S.-G. Kang, P. Xiu, Y. Zhao, and R. Zhou, Interactions Between Proteins and Carbon-Based Nanoparticles: Exploring the Origin of Nanotoxicity at the Molecular Level, *small*, 9, 1546-1556 (2013).

[5] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu, Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics, *PLoS Comput Biol* 12, e1004619 (2016).

[6] M. G. Saunders and G. A. Voth, Coarse-graining methods for computational biology, *Annu Rev Biophys* 42, 73-93 (2013).

[7] L. Site, What is a Multiscale Problem in Molecular Dynamics? *Entropy* 16, 23-40 (2013).

[8] A. Bianco, K. Kostarelos, and M. Prato, Applications of carbon nanotubes in drug delivery, *Curr Opin Chem Biol* 9, 674-9 (2005).

[9] H. Zorgati, M. Larsson, W. Ren, A. Y. L. Sim, J. Gettemans, J. M. Grimes, W. Li, and R. C. Robinson, The role of gelsolin domain 3 in familial amyloidosis, *PNAS*, 116, 13958-13963 (2019).

[10] J. Kong, J. Li, J. Lu, W. Li, and W. Wang, Role of substrate-product frustration on enzyme functional dynamics, *Physical Review E* 100, 052409 (2019).

[11] Y. Cheng, D. Li, B. Ji, X. Shi, and H. Gao, Structure-based design of carbon nanotubes as HIV-1 protease inhibitors: atomistic and coarse-grained simulations, *J Mol Graph Model* 29, 171-7 (2010).

[12] J. W. Chu, G. S. Ayton, S. Izvekov, and G. A. Voth, Emerging methods for multiscale simulation of biomolecular systems, *Molecular Physics* 105, 167-175 (2007).

[13] J. W. Chu and G. A. Voth, Coarse-grained modeling of the actin filament derived from atomistic-scale simulations, *Biophys*, 90, 1572-82 (2006).

[14] T. D. Potter, J. Tasche, and M. R. Wilson, Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures, *Physical Chemistry Chemical Physics* 21, 1912-1927 (2019).

[15] C. Hu, T. Lu, and H. Guo, Developing a Transferable Coarse-Grained Model for the Prediction of Thermodynamic, Structural, and Mechanical Properties of Polyimides at Different Thermodynamic State Points, *Journal of Chemical Information and Modeling* 59, 2009-2025 (2019).

[16] S. Y. Mashayak, L. Miao, and N. R. Aluru, Integral equation theory based direct and accelerated systematic coarse-graining approaches, *The Journal of Chemical Physics* 148, 214105 (2018).

[17] M. Rebič, F. Mocci, A. Laaksonen, and J. Uličný, Multiscale Simulations of Human Telomeric G-Quadruplex DNA, *The Journal of Physical Chemistry B* 119, 105-113 (2015).

[18] A.-T. Kuo, S. Okazaki, and W. Shinoda, Transferable coarse-grained model for perfluorosulfonic acid polymer membranes [J], *The Journal of Chemical Physics* 147, 094904 (2017).

[19] Li Deng, Yurong Zhao, Hai Xu, and Yanting Wang, Intrinsic defect formation in peptide self-assembly, *Appl. Phys. Lett*, 107, 043701 (2015).

[20] L. Deng and Y. Wang, Multiscale computational prediction of  $\beta$ -sheet peptide self-assembly morphology, *Molecular Simulation*, 1-11 (2020).

- [21] D. Reith, M. Pütz, and F. Müller-Plathe, Deriving effective mesoscale potentials from atomistic simulations [J], *Journal of Computational Chemistry*, 24, 1624-1636 (2003).
- [22] G. Milano and F. Müller-Plathe, Mapping Atomistic Simulations to Mesoscopic Models: A Systematic Coarse-Graining Procedure for Vinyl Polymer Chains, *The Journal of Physical Chemistry B* 109, 18609-18619 (2005).
- [23] K. Kremer and F. Müller-Plathe, Multiscale Problems in Polymer Science: Simulation Approaches, *MRS Bulletin* 26, 205-210 (2001).
- [24] F. Müller-Plathe, Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back, *CHEMPHYSCHEM*, 3, 754-769 (2002).
- [25] Z. He and J. Zhou, Probing carbon nanotube–amino acid interactions in aqueous solution with molecular dynamics simulations, *Carbon* 78, 500-509 (2014).
- [26] L. S. Nechaeva, E. V. Butyrskaya, and S. A. Zapryagaev, Computer simulation of amino acid sorption on carbon nanotubes, *Journal of Structural Chemistry* 58, 217-225 (2017).
- [27] L. Piao, Q. Liu, and Y. Li, Interaction of Amino Acids and Single-Wall Carbon Nanotubes, *The Journal of Physical Chemistry C* 116, 1724-1731 (2012).
- [28] C. Rajesh, C. Majumder, H. Mizuseki, and Y. Kawazoe, A theoretical study on the interaction of aromatic amino acids with graphene and single walled carbon nanotube, *J Chem Phys* 130, 124911 (2009).
- [29] T. Roman, W. A. Diño, H. Nakanishi, and H. Kasai, Amino acid adsorption on single-walled carbon nanotubes, *The European Physical Journal D* 38, 117-120 (2006).
- [30] J. Wang, Carbon-Nanotube Based Electrochemical Biosensors: A Review, *Electroanalysis* 17, 7-14 (2005).
- [31] T. E. De Oliveira, P. A. Netz, K. Kremer, C. Junghans, and D. Mukherji, C-IBI: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids, *The Journal of Chemical Physics* 144, 174106 (2016).
- [32] C. T. Andrews and A. H. Elcock, COFFDROP: A Coarse-Grained Nonbonded Force Field for Proteins Derived from All-Atom Explicit-Solvent Molecular Dynamics Simulations of Amino Acids, *Journal of Chemical Theory and Computation* 10, 5178-5194 (2014).
- [33] Z. Li, X. Bian, X. Yang, and G. E. Karniadakis, A comparative study of coarse-graining methods for polymeric fluids: Mori-Zwanzig vs. iterative Boltzmann inversion vs. stochastic parametric optimization, *The Journal of Chemical Physics* 145, 044102 (2016).
- [34] M. Fukuda, H. Zhang, T. Ishiguro, K. Fukuzawa, and S. Itoh, Structure-based coarse-graining for inhomogeneous liquid polymer systems, *The Journal of Chemical Physics* 139, 054901 (2013).
- [35] C.-C. Fu, P. M. Kulkarni, M. Scott Shell, and L. Gary Leal, A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties, *The Journal of Chemical Physics* 137, 164106 (2012).
- [36] H. A. Karimi-Varzaneh, H.-J. Qian, X. Chen, P. Carbone, and F. Müller-Plathe, IBIsCO: A molecular dynamics simulation package for coarse-grained simulation, *Journal of Computational Chemistry*, 32, 1475-1487 (2011).
- [37] A. Mirzoev and A. P. Lyubartsev, MagiC: Software Package for Multiscale Modeling, *Journal of Chemical Theory and Computation*, 9, 1512-1520 (2013).
- [38] G. Maurel, B. Schnell, F. Goujon, M. Couty, and P. Malfreyt, Multiscale Modeling Approach toward the Prediction of Viscoelastic Properties of Polymers, *Journal of Chemical Theory and Computation*, 8, 4570-4579 (2012).

- [39] B. L. Peters, K. M. Salerno, A. Agrawal, D. Perahia, and G. S. Grest, Coarse-Grained Modeling of Polyethylene Melts: Effect on Dynamics, *Journal of Chemical Theory and Computation* 13, 2890–2896 (2017).
- [40] Y. Yoshimoto, Z. Li, I. Kinefuchi, and G. E. Karniadakis, Construction of non-Markovian coarse-grained models employing the Mori–Zwanzig formalism and iterative Boltzmann inversion, *The Journal of Chemical Physics*, 147, 244110 (2017).
- [41] R. L. Henderson, A UNIQUENESS THEOREM FOR FLUID PAIR CORRELATION FUNCTIONS, *PHYSICS LETTERS*, 49A, (1974).
- [42] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates, *Acta Polymerica*, 49, 61–74 (1998).