



Modeling hydrogen exchange of proteins by a multiscale method

Wentao Zhu(祝文涛), Wenfei Li(李文飞)†, and Wei Wang(王伟)‡

Citation: Chin. Phys. B, 2021, 30 (7): 078701. DOI: 10.1088/1674-1056/abe377

Journal homepage: <http://cpb.iphy.ac.cn>; <http://iopscience.iop.org/cpb>

What follows is a list of articles you may be interested in

Enhancements of the Gaussian network model in describing nucleotide residue fluctuations for RNA

Wen-Jing Wang(王文静) and Ji-Guo Su(苏计国)

Chin. Phys. B, 2021, 30 (5): 058701. DOI: 10.1088/1674-1056/abe1a5

Multi-scale molecular dynamics simulations and applications on mechanosensitive proteins of integrins

Shouqin Lü(吕守芹), Qihan Ding(丁奇寒), Mingkun Zhang(张明焜), and Mian Long(龙勉)

Chin. Phys. B, 2021, 30 (3): 038701. DOI: 10.1088/1674-1056/abc540

Statistical potentials for 3D structure evaluation: From proteins to RNAs

Ya-Lan Tan(谭雅岚), Chen-Jie Feng(封晨洁), Xunxun Wang(王勋勋), Wenbing Zhang(张文炳), and Zhi-Jie Tan(谭志杰)

Chin. Phys. B, 2021, 30 (2): 028705. DOI: 10.1088/1674-1056/abc0d6

Coupling of two-dimensional atomistic and continuum models for dynamic crack

Ren Guo-Wu, Tang Tie-Gang

Chin. Phys. B, 2014, 23 (11): 118704. DOI: 10.1088/1674-1056/23/11/118704

RNA structure prediction: Progress and perspective

Shi Ya-Zhou, Wu Yuan-Yan, Wang Feng-Hua, Tan Zhi-Jie

Chin. Phys. B, 2014, 23 (7): 078701. DOI: 10.1088/1674-1056/23/7/078701

Modeling hydrogen exchange of proteins by a multiscale method*

Wentao Zhu(祝文涛), Wenfei Li(李文飞)[†], and Wei Wang(王炜)[‡]*School of Physics, National Laboratory of Solid State Microstructure, and Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China*

(Received 29 December 2020; revised manuscript received 25 January 2021; accepted manuscript online 5 February 2021)

We proposed a practical way for mapping the results of coarse-grained molecular simulations to the observables in hydrogen change experiments. By combining an atomic-interaction based coarse-grained model with an all-atom structure reconstruction algorithm, we reproduced the experimental hydrogen exchange data with reasonable accuracy using molecular dynamics simulations. We also showed that the coarse-grained model can be further improved by imposing experimental restraints from hydrogen exchange data via an iterative optimization strategy. These results suggest that it is feasible to develop an integrative molecular simulation scheme by incorporating the hydrogen exchange data into the coarse-grained molecular dynamics simulations and therefore help to overcome the accuracy bottleneck of coarse-grained models.

Keywords: coarse-grained model, hydrogen exchange, multiscale method, proteins, integrative molecular simulations

PACS: 87.16.A-, 87.15.ap, 87.14.E-, 87.18.Nq

DOI: 10.1088/1674-1056/abe377

1. Introduction

Molecular dynamics (MD) simulation method is becoming an indispensable technique in the studies of biomolecular structures and dynamics due to its unprecedented spatial and temporal resolutions.^[1–4] However, due to the sampling difficulties, conventional MD simulation at all-atom level can hardly access the timescales involved in most molecular motions of biological relevance. To overcome such efficiency-bottleneck, many kinds of coarse-grained (CG) models have been developed and used in the simulations of biomolecules and prediction of functional sites of proteins.^[5–11] In CG models, a group of atoms are represented by a single bead, so that the degrees of freedom can be largely reduced, which contributes to the increased simulation efficiency.^[12] The main challenge of such CG models is that it is not straightforward to extract the CG force fields with high accuracy, which leads to the so called accuracy-bottleneck. Therefore, improving the accuracy of the CG models in biomolecular simulations attracts much effort in the fields of physics, chemistry, and biology.^[6,7,11–15] Such efforts typically follow two lines: i) multiscale molecular simulations, in which information from all-atom simulations was used to improve the accessible accuracy of the CG models;^[16–20] ii) integrative molecular simulations, in which restraints from available experimental data were used to improve the accuracy of the CG models.^[21,22] With these improvements, CG simulations have shown great success in the studies of structure and functional motions for a wide range of biological systems,

including molecular motors,^[23] molecular chaperonin,^[24] enzymes,^[25,26] transcription machines,^[27] nucleosomes,^[28,29] and even chromosome.^[30,31]

In literatures, many different kinds of experimental data have been used to restrain the MD simulations. Typical examples include the small-angle x-ray scattering (SAXS^[32]), double electron–electron resonance (DEER^[33]), Förster resonance energy transfer (FRET^[34]), hydrogen exchange (HX^[35]) data, and so on. Hydrogen exchange (or hydrogen/deuterium exchange) is a powerful method in investigating the structural dynamics of proteins.^[35–38] In HX experiments, the hydrogen atoms of proteins, particularly the protons in the backbone amide group, can exchange with the deuterium in D₂O solvent, and the hydrogen exchange rate of a residue largely depends on its exposure extent to the solvent, which therefore can be used to characterize the structural dynamics of proteins in solution. Compared to the SAXS, DEER, and FRET techniques, the HX data can provide structural dynamics information at residue-level resolution. However, describing the HX data by MD simulations requires an all-atom level model, and previous studies have successfully used the HX data to restrain the all-atom MD simulations.^[39,40] In order to use HX data to restrain the CG MD simulations, we need to develop a corresponding method to describe the HX data on the bases of the CG model.

In this work, we investigate the possibility to use the CG MD simulations to describe the HX data. For the CG simulations, we used an atomic-interaction based CG model

*Project supported by the National Natural Science Foundation of China (Grant Nos. 11974173 and 11934008) and the HPC Center of Nanjing University.

[†]Corresponding author. E-mail: wangwei@nju.edu.cn

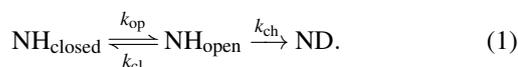
[‡]Corresponding author. E-mail: wfli@nju.edu.cn

(AICG2+),^[41] in which the parameters were optimized based on all-atom MD simulations utilizing a multiscale strategy. Then by mapping the sampled CG structures back to the all-atom structures, we calculated the HX protection factors. The results showed that such a combination of the CG simulations and structure mapping strategy can describe the experimental HX data reasonably well, which therefore builds up a bridge between the experimental HX data and the CG model. Particularly, we showed that the CG model can be further improved by integrating the available HX data via an iterative optimization process. All these results suggest that combining the CG model and HX data is a feasible strategy to achieve high accuracy and high efficiency MD simulations, which can help to extend the application cope of molecular simulations in the studies of biomolecular systems.

2. Method

2.1. Hydrogen exchange rates and local conformational equilibrium

According to previous HX model, hydrogen exchange occurs only when the amide hydrogen in a given residue of protein is exposed to the D₂O, which requires the opening of the local conformation. The scheme can be described as following:^[35–37,42]



Here k_{op} represents the transition rate of the local conformation from the closed state ($\text{NH}_{\text{closed}}$) to the open state (NH_{open}), and k_{cl} is the rate of the reversal process. k_{ch} is the intrinsic exchange rate when the residue is fully exposed to D₂O. The hydrogen exchange rate k_{ex} is then given by^[37,42]

$$k_{\text{ex}} = \frac{k_{\text{op}}k_{\text{ch}}}{k_{\text{op}} + k_{\text{cl}} + k_{\text{ch}}}. \quad (2)$$

Under the EX2 limit ($k_{\text{ch}} \ll k_{\text{cl}}$) and the native-like conditions ($k_{\text{op}} \ll k_{\text{cl}}$), the hydrogen exchange rate can be further written as $k_{\text{ex}} = K_{\text{oc}}k_{\text{ch}}$ with $K_{\text{oc}} = k_{\text{op}}/k_{\text{cl}}$ being the equilibrium constant between open and closed states.^[37,42] In HX experiments, the results were often given by the protection factor (Pf), which is defined by $Pf = k_{\text{ch}}/k_{\text{ex}}$, and it is related to the conformational equilibrium by

$$Pf = 1/K_{\text{oc}}. \quad (3)$$

Therefore, the protection factor Pf can be directly derived from the equilibrated conformational distribution based on MD simulations.

2.2. Coarse-grained MD simulations

The MD simulations were performed by the atomic-interaction based CG model (AICG2+) developed in previous work.^[41] In this model, each residue was represented

by one spherical bead located at the C_{α} position. The AICG2+ energy function was constructed based on the perfect funnel approximation,^[5] with the parameters being optimized based on all-atom MD simulations using a multiscale method.^[41,43] The simulations were conducted by using the software CafeMol^[10] with the temperature being controlled by the Langevin dynamics. The time step of the MD simulation was set as 0.1τ , with τ being the CafeMol time unit. For each protein, we firstly conducted MD simulations at seven temperatures with the length of 5×10^8 MD steps. Then the free energy landscape profiles at different temperatures were constructed by using the weighted histogram analysis method (WHAM^[44]), from which we can determine the folding temperature T_f . Figure 1(b) shows the free energy profiles of the protein CI2 at $T = 1.0T_f$ and $0.92T_f$ projected onto the reaction coordinate Q , which is defined as the fraction of formed native contacts at a given structure. At $T = 1.0T_f$, the free energy difference between the folded and unfolded states is zero according to the definition of the folding temperature. At $T = 0.92T_f$, the free energy difference is $\sim 11.5k_B T$, which is close to the experimental value measured at native condition for the protein CI2.^[42,45] All the conformations sampled in the above simulations at different temperatures were reweighted to the target temperature $T = 0.92T_f$ using WHAM to model the protein dynamics at the native-like condition. We repeated the similar simulations for six different proteins, including CI2, Ubiquitin, SN, S6, HEWL, and I κ B α , for which the HX data are available.^[42,45–56]

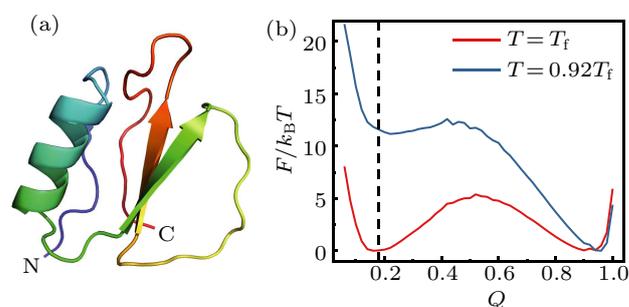


Fig. 1. (a) Cartoon representation of the three-dimensional structure of the protein CI2 (pdb code: 2CI2^[57]). (b) Free energy profile of CI2 along the reaction coordinate Q at the temperatures of $T = T_f$ and $T = 0.92T_f$. The reaction coordinate Q represents the fraction of formed native contacts at a given structure. The dashed line denotes the Q -score of the unfolded state.

2.3. Calculations of HX protection factors

The relation between the experimental data of HX protection factor and the local conformational equilibrium property K_{oc} given by Eq. (3) provides a way to derive the protection factor values from MD simulation data. However, for the CG simulations, the local conformations are often poorly defined as the backbone amide groups contributing to hydrogen exchange are lacking, which makes the calculation of the protection factor a difficult task. In this work, we introduced

the reversal structure mapping strategy to overcome such difficulty. Firstly, we reconstructed the positions for all the backbone atoms starting from the C_α positions by using the BBQ program, which is a backbone building method from quadrilaterals and has been widely tested for the native and near-native protein structures.^[58] Then, the sidechain atoms were further reconstructed by using the software SCWRL based on

the rotamer library, leading to the all-atom conformations for all the sampled CG structures.^[59] Figure 2 shows the illustrative schematic for the reversal structure mapping using the BBQ and SCWRL for the protein CI2. With the obtained all-atom conformations, one can calculate the equilibrium constant of the local conformations for each residue and therefore the protection factor.

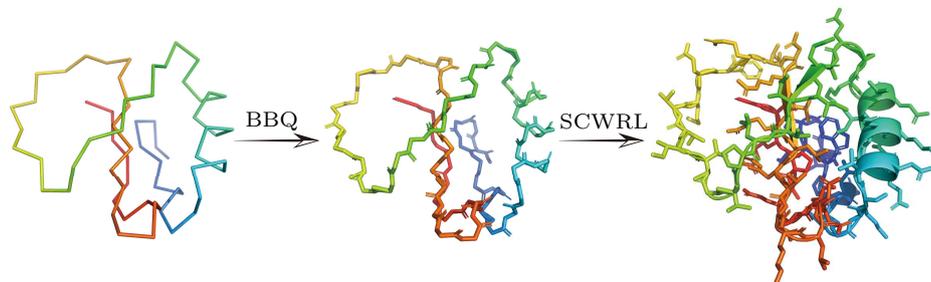


Fig. 2. Illustrative scheme showing the all-atom structure reconstruction from C_α -based coarse-grained model of the protein CI2. At the first step, the positions of the backbone atoms were reconstructed by using the software BBQ.^[58] Then the side-chain atoms were reconstructed by the software SCWRL.^[59]

Several quantities can be used to characterize the exposure extent of the backbone amide groups from given all-atom conformations, including the packing-density, hydrogen-bond, solvent-accessible-surface-area, and so on.^[37–39,60] In this work, we used the packing density and the number of hydrogen bonds. Here the packing density N is given by the number of neighboring (within the distance of 6.5 Å) heavy atoms around the backbone nitrogen atoms for each residue. The hydrogen bonds involving the backbone amide groups were identified based on the distance cutoff of 3.5 Å and angle cutoff of 120°. The equilibrium constant can be approximately given by

$$K_{oc} = \left\langle \exp(-\lambda_1 \cdot N_{HB}) \frac{1}{1 + \exp(\lambda_2 \cdot (N - N_0))} \right\rangle. \quad (4)$$

Here $\langle \dots \rangle$ represents the weighted average among all the sampled conformations. If $K_{oc} \approx 1$, the residue is closer to the open state; otherwise, if $K_{oc} \approx 0$, the residue is closer to the closed state. Note that when $N \gg N_0$, the sigmoid function in Eq. (4) can be reduced to the exponential function, which was used in Refs. [38,39,60]. Based on Eqs. (3) and (4), the HX protection factor values could be extracted from the CG MD simulations. In Eq. (4), the parameters λ_1 , λ_2 , and N_0 were set as 10.0, 0.3, and 20, respectively. For the sake of simplicity, these values were used for all the residues without introducing amino acid type dependence in this work. The use of a common parameter set implies the assumption that the heavy atoms in different amino acids have similar contributions to the screening of the backbone amide group from solvent. In the subsequent section, we will show that even with such common parameter set for all the residues, the HX data can be reproduced reasonably well. It is possible that the performance of the above model can be further improved by introducing amino acid type dependence to these parameters.

3. Results

3.1. Comparison of HX data from MD simulations and experiments

In Ref. [42], the authors calculated the HX data for the proteins CI2, SN, and ubiquitin by using a more sophisticated protein model, in which more detailed structure information was included. Here we calculated the protection factors for the same set of proteins using the method described in Section 2. Figure 3 shows the results of the calculated protection factor (red lines in Figs. 3(a), 3(c), and 3(e)). For comparison, the results from experimental measurements are also shown (black dots and lines). The experimental HX data were taken from Ref. [42], which are the synthesis of the experimental results reported in Refs. [45–53]. One can see that for the above three proteins, the calculated results from the CG MD simulations with AICG2+ model can reproduce the corresponding experimental data very nicely. Both the highly protected regions and the less protected protection regions can be well described, demonstrating the important role of native topology of proteins, as implied in the perfect funnel approximation^[5–7,10] in defining the protein dynamics at native condition, which is consistent with previous discussions.^[42]

To more quantitatively characterize the correlation between the simulation and experimental data, figure 3 also shows the correlation plots and the results of linear fitting (Figs. 3(b), 3(d), and 3(f)). The fitted Pearson's correlation coefficients R are 0.83, 0.73, and 0.68, respectively, for the three proteins. Such a correlation is comparable to the results reported in previous work with a more sophisticated model^[42] (Table 1), although the current model is much more simple and includes only the C_α beads. These results suggest that such a simple multiscale strategy, in which the CG simulations were combined with the structure reconstruction process to produce

the all-atom structure ensemble, can be used to calculate the HX protection factors with reasonable accuracy. In the above calculations, we used the default parameter set given in Section 2. We also calculated the protection factors for the CI2 with the N_0 in Eq. (4) being set as 10, 15, 25, 30, 35, and 40. The resulted correlation coefficients with the experimental values are 0.82, 0.83, 0.83, 0.83, 0.82, and 0.80, respectively, which suggests that the results are insensitive to the parameter N_0 in the calculation of protection factors.

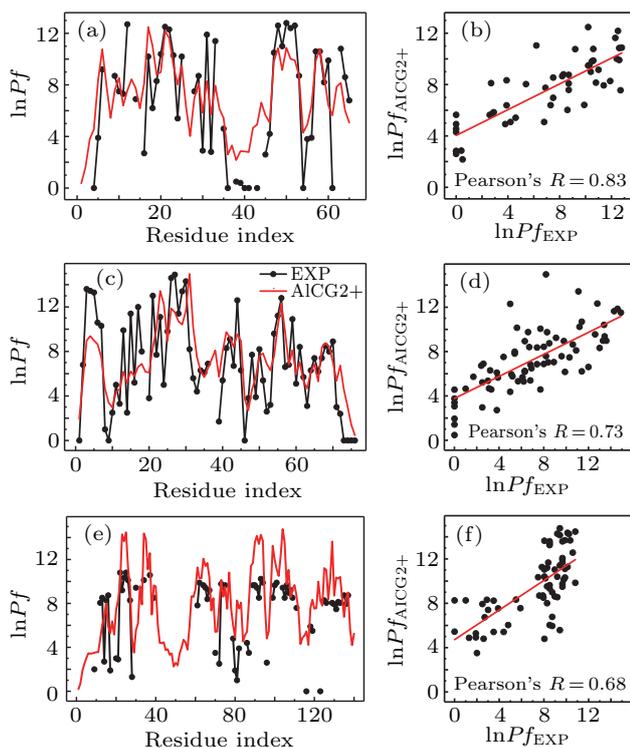


Fig. 3. (a), (c), (e) HX protection factors from the CG MD simulations (red) and experimental data (black) for the residues of the proteins CI2 (a), ubiquitin (c), and SN (e). (b), (d), (f) Correlation plots between the experimental protection factors and the calculated values with AICG2+ force field for the proteins CI2 (b), ubiquitin (d), and SN (f). The red straight lines represent the results of linear fitting. The fitted Pearson's correlation coefficients R were also shown in each panel.

Table 1. The fitted Pearson's R between the protection factors from CG MD simulations with the AICG2+ model and the experimental data for six proteins. The results from previous work^[42] were also shown for comparison.

	AICG2+	Ref. [42]
CI2	0.83	0.70
Ubiquitin	0.73	0.73
SN	0.68	0.72
S6	0.66	–
HEWL	0.62	–
IκBα	0.55	–

We also repeated the above calculations for another three proteins, including the S6, HEWL, and IκBα. The experimental data were taken from Refs. [54–56], respectively. As shown in Table 1, the calculated correlation coefficients are larger than 0.6 for S6 and HEWL, demonstrating reasonable

correlation between the CG simulations and the experimental data for these proteins. In comparison, the obtained correlation coefficient for the IκBα is 0.55, which is significant but much smaller than the values of other proteins discussed above, suggesting the defects in the used CG model. In the next subsection, we will investigate whether it is possible to improve the description of the HX data by optimizing the force field parameters of the above CG model.

3.2. Restraining the CG simulations with HX data

The above results clearly demonstrated the correlation between the simulated HX data and the experimental HX data. However, as shown in Table 1, the correlation coefficients for different proteins can be very different. For the protein CI2, the correlation coefficient is 0.83, which is the highest among the studied proteins in this work. In comparison, the correlation coefficient for the IκBα is much lower (0.55), suggesting the defects of the used CG model in describing the HX data. Meanwhile, the observed correlations encouraged us to investigate whether we can use the HX data to restrain the CG MD simulation and obtain better description of the experimental data. To this end, we constructed an iterative scheme to optimize the parameters of the CG model with the input from experimental data. Specifically, we firstly run MD simulations based on the old CG force field and calculate the protection factor values as described above. Then, we update the parameter ε_i by using the following formula:

$$\varepsilon_i^{(n+1)} = \varepsilon_i^{(n)} \cdot \left(1 + f \cdot \left(P f_i^{\text{EXP}} / P f_i^{(n)} - 1 \right) \right). \quad (5)$$

In the above formula, $\varepsilon_i^{(n)} = \sum_{j \neq i} \varepsilon_{i,j}^{(n)}$ is the total strength of the native contacts related to the i th residue, and $\varepsilon_{i,j}^{(n)}$ is the interaction strength of the residue pair (i, j) in the CG force field. The superscript (n) represents the results obtained at the n th iteration step (thus (0) represents the initial values). The $P f_i^{\text{EXP}}$ and $P f_i^{(n)}$ are the experimental and calculated protection factors. As the correlation coefficients discussed above do not change if we multiply the calculated data for all the residues by a common factor, the $P f_i^{(n)}$ given in the above formula corresponds to the rescaled protection factor. The scaling factor is determined by meeting the condition $\sum P f_i^{\text{EXP}} = \sum P f_i^{(n)}$, so that the global stability of the proteins remains approximately unchanged during the iteration steps. f is the parameter that controls the convergence speed. When $n = 1$, the value $f = 1.0$ was used. In other cases, $f = 0.5$ was used. Because equation (5) updates the values of ε_i , whereas $\varepsilon_{i,j}$ are directly used in the CG MD simulations, we need to further determine the $\varepsilon_{i,j}$ values at each optimization step. Owing to the fact that the $\varepsilon_{i,j}$ is simultaneously related to the ε_i and ε_j values, we calculated the $\varepsilon_{i,j}^{(n+1)}$ in an iterative way, which can be expressed as

$$\begin{cases} \varepsilon_i(k) = \sum_{j \neq i} \varepsilon_{i,j}(k), \\ \varepsilon_{i,j}(k+1) = \varepsilon_{i,j}(k) (\varepsilon_i^{(n+1)} / \varepsilon_i(k)), \text{ if } \exists Pf_i^{\text{EXP}}, \nexists Pf_j^{\text{EXP}}, \\ \varepsilon_{i,j}(k+1) = \varepsilon_{i,j}(k) \sqrt{(\varepsilon_i^{(n+1)} / \varepsilon_i(k)) (\varepsilon_j^{(n+1)} / \varepsilon_j(k))}, \text{ if } \exists Pf_i^{\text{EXP}}, \exists Pf_j^{\text{EXP}}. \end{cases} \quad (6)$$

Here (k) represents the results obtained at the k th iteration step (thus $\varepsilon_{i,j}(0) = \varepsilon_{i,j}^{(n)}$). And the converged $\varepsilon_{i,j}(k)$ is used as $\varepsilon_{i,j}^{(n+1)}$. The optimized parameter $\varepsilon_{i,j}^{(n+1)}$ was used as the interaction strength of the non-bonded native contacting interac-

tion for the residue pair (i, j) in the AICG2+ model.^[41] With the optimization scheme given by Eq. (5), we can gradually improve the CG force field and give better description to the experimental HX data.

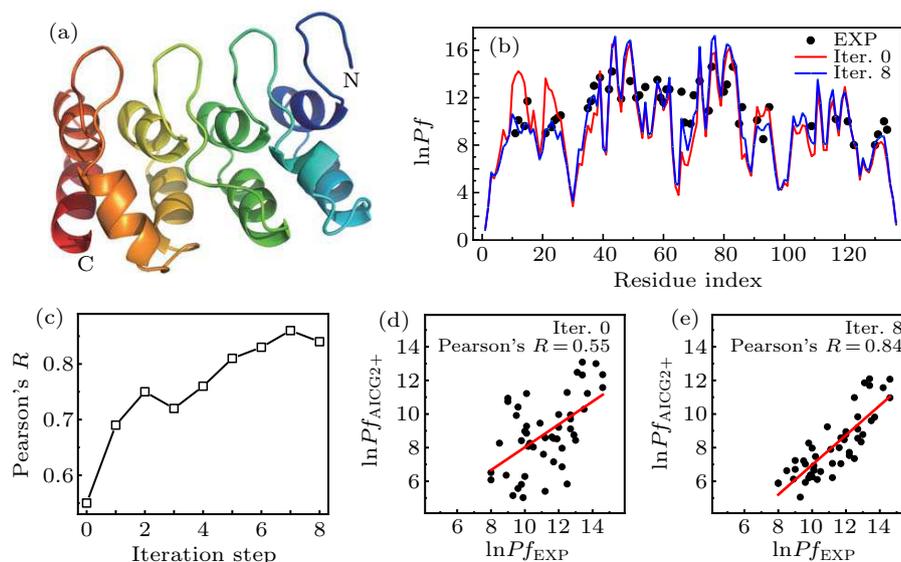


Fig. 4. (a) The cartoon representation of the native structure of $\text{I}\kappa\text{B}\alpha$ (pdb code: 1NFI^[61]). (b) Comparison between the experimental and calculated protection factors (rescaled) before iterative optimization (red) and after eight steps of iterative optimization (blue). (c) The Pearson's correlation coefficient R as a function of iterative optimization steps. (d), (e) The linear correlation between the experimental and predicted protection factors (not rescaled) before iterative optimization (d) and after eight steps of iterative optimization (e). The red straight lines represent the linear fit. The obtained Pearson's correlation coefficients R were also shown.

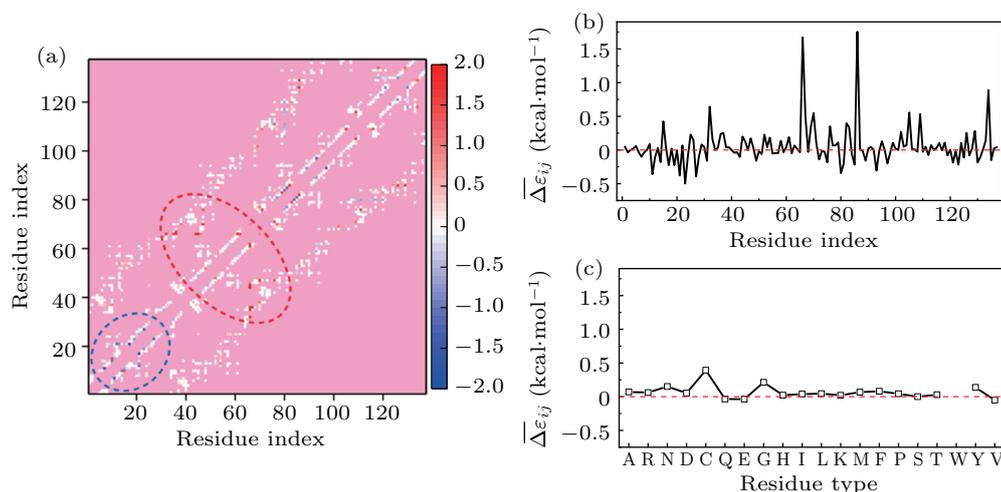


Fig. 5. (a) The differences between the interaction strengths ε_{ij} of the native contacts before iterative optimization and after eight steps of iterative optimization. Red/blue color represents that the interaction strengths become stronger/weaker after optimization. The unit of the color scale is kcal/mol. (b) Same as (a) but with the differences of the ε_{ij} averaged among all the contacting partners of a given residue. (c) Same as (a) but with the differences of the ε_{ij} averaged among all the contacting partners of the residues with a given amino acid type.

Figure 4(c) shows the Pearson's correlation coefficient R for the protein $\text{I}\kappa\text{B}\alpha$ as a function of the optimization step. One can see that the correlation coefficient increases gradually with the progression of optimization. Before optimization, the correlation coefficient is 0.55. After eight steps of

iterative optimizations, the correlation coefficient increases to 0.84 and gets saturated. The final values of the calculated protection factors show great consistency with the experimental data (Figs. 4(b), 4(d), and 4(e)). Notably, the results around the residue 20 were largely improved due to the better descrip-

tion of the relative stabilities for the first and second helices. Figure 5 shows the differences between the non-bonded interaction strengths ϵ_{ij} before iterative optimization and after eight steps of iterative optimization. One can see that the interaction strengths involving the residues with index around 20 become weaker (Fig. 5(a), blue circle line), whereas those involving the residues with index around 70 become stronger (Fig. 5(a), red circle line), which is consistent with the observed changes of the HX data in Fig. 4. Similar results can be obtained from the differences of the ϵ_{ij} averaged among all the contacting partners of each given residue (Fig. 5(b)). However, effect of optimization does not show obvious amino acid type dependence (Fig. 5(c)). Such results demonstrate that with the above iterative optimization scheme, the information from HX data can be effectively integrated into the CG force field to improve the description of the protein structural dynamics by CG MD simulations.

4. Discussion and conclusion

The above results demonstrated that we can improve the CG force fields based on the HX data, although the CG MD simulations were related to the experimental data of the HX by an indirect manner involving an intermediate step of structure reconstruction. Possible reason is that with the C_α positions, the corresponding all-atom structure was relatively well defined as demonstrated in previous works,^[58,59] especially for the structures at the native-like conditions discussed in this work. Apparently, the accuracy of such integrative molecular simulation scheme relies on the performance of the structure reconstruction algorithm. In recent years, developing new structure reconstruction algorithms becomes the focus of several research groups.^[62–64] We believe that with the further improvements of accuracy and efficiency of the structure reconstruction algorithms, the description of the HX data by the CG MD simulations will become more plausible, which can then promote the developments and applications of the integrative molecular simulations with the combinations of the CG model and HX data. Such a combination has obvious advantages. Firstly, the HX data can provide residue-level structural dynamics information, which is superior to other low-resolution experimental data widely used in previous integrative molecular simulations. Secondly, the CG model often has the advantage of high sampling efficiency, but encounters accuracy bottleneck. Because imposing experimental restraints mainly improves the accuracy aspect (instead of the sampling efficiency) of the simulations, such combination can help to overcome the major bottleneck of the CG simulations.

In summary, we used a multiscale scheme to calculate the HX protection factors, in which the structure reconstruction algorithms were used to convert the sampled structures by CG MD simulations into all-atom detailed structures, from

which we can calculate the HX protection factors. By using this method, the HX data can be described with reasonable accuracy by the CG MD simulations. In addition, we showed that the CG force field parameters can be further improved by iterative optimization based on the available experimental data of the HX experiments. Such a combination of the CG MD simulations and HX data can be applied to the studies of other protein dynamics of biological relevance.

Acknowledgment

We thank the insightful discussions with Shoji Takada.

References

- [1] Karplus M and McCammon J A 2002 *Nat. Struct. Biol.* **9** 646
- [2] Hansson T, Oostenbrink C and Gunsteren W V 2002 *Curr. Opin. Struct. Biol.* **12** 190
- [3] Karplus M and Kuriyan J 2005 *Proc. Natl. Acad. Sci. USA* **102** 6679
- [4] Klepeis J L, Lindor-Larsen K, Dror R O and Shaw D E 2009 *Curr. Opin. Struct. Biol.* **19** 120
- [5] Onuchic J N, Luthey-Schulten Z and Wolynes P G 1997 *Annu. Rev. Phys. Chem.* **48** 545
- [6] Noid W G 2013 *J. Chem. Phys.* **139** 090901
- [7] Marrink S J, Risselada H J, Yefimov S, Tieleman D P and de Vries A H 2007 *J. Phys. Chem. B* **111** 7812
- [8] Wang X, Li M, Ye F and Zhou X 2017 *Acta Phys. Sin.* **66** 150201 (in Chinese)
- [9] Zhang P and Shi A 2015 *Chin. Phys. B* **24** 128707
- [10] Kenzaki H, Koga N, Hori N, Kanada R, Li W, Okazaki K, Yao X and Takada S 2011 *J. Chem. Theory Comput.* **7** 1979
- [11] Li W and Takada S 2009 *J. Chem. Phys.* **130** 214108
- [12] Takada S 2012 *Curr. Opin. Struct. Biol.* **22** 130
- [13] Zheng W 2015 *Chin. Phys. B* **24** 128701
- [14] Xu X and Shi J 2020 *Chin. Phys. Lett.* **37** 068701
- [15] Li W, Wang J, Zhang J and Wang W 2015 *Curr. Opin. Struct. Biol.* **30** 25
- [16] Li W, Zhang J, Wang J and Wang W 2015 *Acta Phys. Sin.* **64** 098701 (in Chinese)
- [17] Liu P and Voth G A 2007 *J. Chem. Phys.* **126** 045106
- [18] Lyman E, Ytreberg F M and Zuckerman D M 2006 *Phys. Rev. Lett.* **96** 028105
- [19] Christen M and van Gunsteren W F 2006 *J. Chem. Phys.* **124** 154106
- [20] Li W, Yoshii H, Hori N, Kameda T and Takada S 2010 *Methods* **52** 106
- [21] Larsen A H, Wang Y, Bottaro S, Grudinin S, Arleth L and Lindorff-Larsen K 2020 *PLoS Comput. Biol.* **16** e1007870
- [22] Wen B, Peng J, Zuo X, Gong Q and Zhang Z 2014 *Biophys. J.* **107** 956
- [23] Kubo S, Niina T and Takada S 2020 *Sci Rep* **10** 8225
- [24] Chen J, Zhang Q, Ren W and Li W 2017 *J. Phys. Chem. B* **121** 4987
- [25] Wang Y, Gan L, Wang E and Wang J 2013 *J. Chem. Theory Comput.* **9** 84
- [26] Li W, Peng J, Zuo X, Gong Q and Zhang Z 2019 *Phys. Rev. Lett.* **122** 238102
- [27] Terakawa T and Takada S 2015 *Sci Rep* **5** 17107
- [28] Tan C and Takada S 2020 *Proc. Natl. Acad. Sci. USA* **117** 20586
- [29] Freeman G S, Lequieu J P, Hinckley D M, Whitmer J K and de Pablo J J 2014 *Phys. Rev. Lett.* **113** 168101
- [30] Zhang B and Wolynes P G 2015 *Proc. Natl. Acad. Sci. USA* **112** 6062
- [31] Ashwin S S, Nozaki T, Maeshima K and Sasai M 2019 *Proc. Natl. Acad. Sci. USA* **116** 19939
- [32] Putnam C, Hammel M, Hura G and Tainer J 2007 *Q. Rev. Biophys.* **40** 191
- [33] Larsen R G and Singel D J 1993 *J. Chem. Phys.* **98** 5134
- [34] Ghisaidoobe A B T and Chung S J 2014 *Int. J. Mol. Sci.* **15** 22518
- [35] Hvidt A and Nielsen S O 1966 *Adv. Protein Chem.* **21** 287
- [36] Ferraro D M, Lazo N D and Robertson A D 2004 *Biochemistry* **43** 587
- [37] Petruk A A, Defelipe L A, Rodríguez Limardo R G, Bucci H, Marti M A and Turjanski A G 2013 *J. Chem. Theory Comput.* **9** 658

- [38] Park I, Venable J D, Steckler C, Cellitti S E, Lesley S A, Spraggon G and Brock A 2015 *J. Chem. Inf. Model.* **55** 1914
- [39] Best R B and Vendruscolo M 2006 *Structure* **14** 97
- [40] Skinner J J, Yu W, Gichana E K, Baxa M C, Hinshaw J R, Freed K F and Sosnick T R 2014 *Proc. Natl. Acad. Sci. USA* **111** 15975
- [41] Li W, Wang W and Takada S 2014 *Proc. Natl. Acad. Sci. USA* **111** 10550
- [42] Craig P O, Lätzer J, Weinkam P, Hoffman R M B, Ferreiro D U, Komives E A and Wolynes P G 2011 *J. Am. Chem. Soc.* **133** 17463
- [43] Li W, Wolynes P G and Takada S 2011 *Proc. Natl. Acad. Sci. USA* **108** 3504
- [44] Kumar S, Rosenberg J M, Bouzida D, Swendsen R H and Kollman P A 1992 *J. Comput. Chem.* **13** 1011
- [45] Neira J L, Itzhaki L S, Otzen D E, Davis B and Fershta A R 1997 *J. Mol. Biol.* **270** 99
- [46] Hernández G, Anderson J S and LeMaster D M 2009 *Biochemistry* **48** 6482
- [47] Pan Y and Briggs M S 1992 *Biochemistry* **31** 11405
- [48] Johnson E C, Lazar G A, Desjarlais J R and Handel T M 1999 *Structure* **7** 967
- [49] Bougault C, Feng L, Glushka J, Kupče E and Prestegard J H 2004 *J. Biomol. NMR* **28** 385
- [50] Khorasanizadeh S, Peters I D, Butt T R and Roder H 1993 *Biochemistry* **32** 7054
- [51] Lazar G A, Desjarlais J R and Handel T M 1997 *Protein Sci.* **6** 1167
- [52] Loh S N, Prehoda K E, Wang J and Markley J L 1993 *Biochemistry* **32** 11022
- [53] Mori S, Abeygunawardana C, Berg J M, van Zijl P C M 1997 *J. Am. Chem. Soc.* **119** 6844
- [54] Haglund E, Lind J, Öman T, Öhman A, Måler L and Oliveberg M 2009 *Proc. Natl. Acad. Sci. USA* **106** 21619
- [55] Radford S E, Buck M, Topping K D, Dobson C M and Evans P A 1992 *Proteins* **14** 237
- [56] Sue S C, Cervantes C, Komives E A and Dyson H J 2008 *J. Mol. Biol.* **380** 917
- [57] McPhalen C A and James M N G 1987 *Biochemistry* **26** 261
- [58] Gront D, Kmiecik S and Kolinski A 2007 *J. Comput. Chem.* **28** 1593
- [59] Bower M J, Cohen F E and Dunbrack R L 1997 *J. Mol. Biol.* **267** 1268
- [60] Vendruscolo M, Paci E, Dobson C M and Karplus M 2003 *J. Am. Chem. Soc.* **125** 15686
- [61] Jacobs M D and Harrison S C 1998 *Cell* **95** 749
- [62] Heath A P, Kavraki L E and Clementi C 2007 *Proteins* **68** 646
- [63] Shimizu M and Takada S 2018 *J. Chem. Theory Comput.* **14** 1682
- [64] Badaczewska-Dawid A E, Kolinski A and Kmiecik S 2020 *Comput. Struct. Biotech. J.* **18** 162