Smoothing potential energy surface of proteins by hybrid coarse grained approach*

Yukun Lu(卢禹锟)[1,2], Xin Zhou(周昕)[2,†], and ZhongCan OuYang(欧阳钟灿)[1]

[1] *Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100049, *China*

(Received 10 January 2017; revised manuscript received 13 February 2017; published online 30 March 2017)

Coarse-grained (CG) simulations can more efficiently study large conformational changes of biological polymers but usually lose accuracies in the details. Lots of different hybrid models involving multiple different resolutions have been developed to overcome the difficulty. Here we propose a novel effective hybrid CG (hyCG) approach which mixes the fine-grained interaction and its average in CG space to form a more smoothing potential energy surface. The hyCG approximately reproduces the potential of mean force in the CG space, and multiple mixed potentials can be further combined together to form a single effective force field for achieving both high efficiency and high accuracy. We illustrate the hyCG method in Trp-cage and Villin headpiece proteins to exhibit the folding of proteins. The topology of the folding landscape and thus the folding paths are preserved, while the folding is boosted nearly one order of magnitude faster. It indicates that the hyCG approach could be applied as an efficient force field in proteins.

Keywords: coarse-grained, hybrid force field, enhanced sampling, protein folding

PACS: 02.50.–r, 87.10.Tf, 87.14.E– **DOI:** 10.1088/1674-1056/26/5/050202

1. Introduction

For the past few decades, molecular dynamics (MD) simulation has become one of the most powerful tools for predicting the property of biopolymers like proteins. The development of a high precision empirical all-atom (AA) force field allows investigation of the large structural change in at least small proteins. With the help of the specially designed chip, the accessible timescale of MD can even extend to the millisecond nowadays,[1] which reaches or approaches to the experimental interesting time scales of proteins.

Although there has been success in the fine-grained models, large demand in the study of membrane proteins and genetic materials still requires a coarse-grained (CG) model for further extending both length and time scales of simulations. During the last 20 years, vast CG methods have been proposed for different requirements.[2–15] Although these approaches can currently reach relative accuracy and efficiency in understanding the overall property of polymers, liquids, and biological membrane, still, many unsolved problems remain in transferability and thermodynamic consistency.[16] They have limited ability in describing the elaborate structural change in proteins.[17,18]

Recently, multiscale modelings or hybrid methods have caught a great deal of attention for their ability in preserving the details of protein structure while largely increasing the accessible time scale and length scale of simulations. These methods have been developed into 2 classes: the resolution mixing[19–22] and the resolution scaling.[23–25] The resolution mixing is an actual degree-reducing method, it commonly keeps the high resolution, e.g., the AA, a description for the detailed sensitive region like proteins while the surrounding environment is characterized by low resolution, e.g., the CG. On the contrary, the resolution scaling directly couples the strength of the AA and the CG interaction into a single potential, which is in actuality a fine-grained model but with a modified energy surface. The hybrid interaction in the resolution scaling can be generally written as

$$U_\lambda(r) = U^{\mathrm{cg}}(R) + (1-\lambda)[U^{\mathrm{aa}}(r) - U^{\mathrm{cg}}(R)], \qquad (1)$$

where r represents the coordinates of all atoms and $R = R(r)$ represents the CG coordinates as a function of r. $U^{\mathrm{cg}}(R)$ is the CG potential. The difference between the AA potential and the CG potential, $\Delta U(r) = U^{\mathrm{aa}}(r) - U^{\mathrm{cg}}(R)$, is adjusted by the parameter λ which can switch from $\lambda = 0$ representing the complete AA Hamiltonian to $\lambda = 1$ the pure CG one. The intermediate λ provides a bridge to connect CG with AA, which smooths the roughness of the AA potential relative to the CG potential. In the resolution mixing method, we can also write this kind of hybrid Hamiltonian if thinking λ varies as degrees of freedom.

In the pure CG model ($\lambda = 1$), it is clear that the CG potential should equal to the potential of mean force (PMF)

*Project supported by the National Basic Research Program of China (Grant No. 2013CB932803), the National Natural Science Foundation of China (Grant No. 11574310), and the Joint NSFC-ISF Research Program, jointly funded by the National Natural Science Foundation of China and the Israel Science Foundation (Grant No. 51561145002).

†Corresponding author. E-mail: xzhou@ucas.ac.cn

© 2017 Chinese Physical Society and IOP Publishing Ltd

to reproduce the probability distribution of the AA in the CG space. Many methods, by matching such as the average force, thermodynamic observables, the PMF, or the probability density, are designed to construct the CG potential.[4,14,15] However, in the hybrid CG models ($0 < \lambda < 1$), the remaining AA interaction will also contribute to the PMF, thus the CG potential $U^{\mathrm{cg}}(R)$ alone should be different from that in the pure CG model. So far, to the best of our knowledge, there have been no successful attempts in answering the questions such as which kind of CG potential should be applied in hybrid CG models and how to construct hybrid CG models with higher accuracy than the pure CG model and higher efficiency than the AA model.

In this paper, we intend to propose a novel but economical hybrid approach. This approach can be divided into two stages. The first stage is the same as the normal resolution scaling method but specifies the CG potential as the average of the AA potential. The hybrid potential is much more smoothed so that it is easier to reach equilibrium. It is interesting to note that, the exclusion of entropy in this CG potential is approximately replenished by the existence of the fine-grained interaction in the hybrid model. The second stage is to bridge multiple U_λ by integrating λ from 0 to a λ_{m} (smaller than 1) into a single effective potential. The existence of small λ in the combined potential guarantees the obtained PMF is close to that of the original system, even while the parameterized CG potential could not well match with the average of the AA potential, while the large λ components bring more efficient visiting in the conformational space. Our results show that the hyCG approach could reproduce the folding pathway in small proteins with about one order of magnitude higher efficiency.

The paper is organized as follows. We first show the hybrid protocol in Subsection 2.1. Subsection 2.2 illustrates a construction of an effective potential by combining many λ-hybrid interactions. The simulation details are shown in Section 3. The numerical results and the verification of folding in two small proteins, Villin headpiece (35 residues) and Trp-cage (20 residues), including the efficiency and accuracy of their secondary and tertiary formation, are addressed in Section 4.

2. Theory

In this section, we first construct a hybrid potential to approximately keep the PMF in the CG space, then combine multiple hybrid strengths together to automatically tune the fitness between the CG and AA models.

2.1. λ-hybrid potential

In our approach, first, the λ-hybrid potential U_λ is constructed by linearly combining the AA potential $U^{\mathrm{aa}}(r)$ and CG potential $U^{\mathrm{cg}}(R)$ as in Eq. (1). The λ satisfies $0 \le \lambda < 1$ and acts like a strength tuning parameter to reduce the ruggedness of the potential energy surface.

The free energy landscape of the hybrid system in the CG space, i.e., the PMF, is

$$F_\lambda(R) = -k_{\mathrm{B}}T \ln \int_R \mathrm{d}r\, \mathrm{e}^{-\beta U_\lambda(r)}, \tag{2}$$

where k_{B} is the Boltzmann constant, T is the temperature, and $\beta = \frac{1}{k_{\mathrm{B}}T}$. The integrate is limited at R. It is easy to know

$$F_\lambda(R) = F_0(R) - \lambda \langle \Delta U(r) \rangle_R + \mathrm{o}(\lambda), \tag{3}$$

where

$$\langle \Delta U(r) \rangle_R = \frac{\int_R \Delta U(r)\, \mathrm{e}^{-\beta U^{\mathrm{aa}}(r)}\, \mathrm{d}r}{\int_R \mathrm{e}^{-\beta U^{\mathrm{aa}}(r)}\, \mathrm{d}r} \tag{4}$$

is the average of $\Delta U(r)$ at R under the equilibrium distribution of the $\lambda = 0$ system. Thus, if we choose

$$U^{\mathrm{cg}}(R) = \langle U^{\mathrm{aa}}(r) \rangle_R, \tag{5}$$

the change of PMF is a high-order quantity of small λ,i.e., the PMF is approximately preserved while smoothing the inner roughness of the AA potential. It is equivalent to apply a high temperature to the inner degrees of freedom (DoF) but keep the temperature to the CG DoF. To properly introduce the CG force into the AA potential, the center-of-mass of the CG mapping groups $\boldsymbol{R}_\mu = \left(\sum_{i\in\mu} m_i \boldsymbol{r}_i\right) / \left(\sum_{i\in\mu} m_i\right)$ is usually adopted, with the μ index running over all the mapping sites. In the supporting information, we show that while the inner DoF is tightly bound in CG DoF, the high order term of λ is almost independent on R. It indicates that the average potential is a good selection as the CG potential in the hybrid models to preserve the free energy landscape.

To ensure the proper rigidity of bond and angle connection, the hybridization usually performs on part of AA potentials, such as the non-bonded term only, which includes the Van der Waals (VDW) potential and the Coulomb interaction, thus the total potential in the hybrid model is

$$U_\lambda(r) = U^{\mathrm{cg}}(R) + U_1^{\mathrm{aa}}(r) + (1-\lambda)[U_2^{\mathrm{aa}}(r) - U^{\mathrm{cg}}(R)]. \tag{6}$$

Here $U_1^{\mathrm{aa}}(r)$ represents the bonded potential, $\Delta U_2(r) = U_2^{\mathrm{aa}}(r) - U^{\mathrm{cg}}(R)$, and $U_2^{\mathrm{aa}}(r)$ is the non-bonded part of the AA potential. For implicit solvent simulations, the $U_2^{\mathrm{aa}}(r)$ might contain the generalized Born (GB) potential $U_{\mathrm{gb}}^{\mathrm{aa}}(r)$ and surface tension (SA) potential $U_{\mathrm{sa}}^{\mathrm{aa}}(r)$ to mimic the solvent environment. In this case, $U^{\mathrm{cg}}(R)$ is the average of $U_2^{\mathrm{aa}}(r)$.

2.2. Bridging efficiency and accuracy

A strongly hybrid potential (large λ) can reduce the atomic friction and speed up the simulation, while a lightly hybrid potential ($\lambda \simeq 0$) can provide more accurate atomic detail. A simple way to balance between the two points can communicate multiple λ-hybrid potentials involving $\lambda = 0$. This communication can be implemented through the replica exchange technique.[26] However, the hardware requirement of the replica exchange method is harsher than that of a conventional simulation, which will be n times of the conventional resources. Besides, it could encounter a problem in a low exchange rate. Here we apply another approach, the integrated temperature sampling (ITS),[27] to linearly combine n hybrid potentials to a single effective potential,

$$U_{\text{eff}}(r) = -k_{\text{B}}T \ln \sum_{i=1}^{n} c_i \rho_i(r), \tag{7}$$

where $\rho_i(r) = \frac{1}{Z_i(T)} \mathrm{e}^{-\beta U_{\lambda_i}(r)}$, and λ_i satisfies $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_n = \lambda_{\max} < 1$. The $Z_i(T)$ is the partition function of the λ_i-hybrid potential and c_i refers to the weight of each hybrid level. Thus $\rho_1(r)$ is the AA Boltzmann distribution $\rho_{\text{aa}}(r)$. The effective potential corresponds to the probability distribution $\rho(r) = \sum c_i \rho_i(r)$, which is equivalent to that of the replica exchange simulations at these λ_i. In the ITS approach, the partition function Z_i is iteratively estimated by updating the weight of each hybrid level to the desired c_i during the simulation and then refined by the weight histogram analysis method (WHAM).[28] More details about the iterative estimate of Z_i can be found in the original paper of ITS.[27] Usually, we adopt equal weight for n chosen levels, which is $c_i = 1/n$ for all i. Note that it is not necessary in our model, for example, we can set λ_1 and λ_n relatively larger than other λ's so that the effective potential has more proportion accumulated in the maximum precision and the maximum efficiency. The advantage of the λ-hybrid combination approach is that even though the average potential is not well parameterized, we still have at least $1/n$ proportion correct samples accumulated in the AA model, thus the effective potential still does not deviate much from the original system. The similar shape of PMF for the different λ-hybrid potentials also promises the efficiency of the combined effective potential for the iterative convergence of partition function Z_i.

3. Simulation details

3.1. Parameter of average potential

In this paper, we illustrate the hybrid CG approach in proteins, where the average potential for each CG pair is fitted by the Lennard–Jones (LJ) 12–6 potential and the Coulomb interaction. To mimic the implicit solvent effect with ions, the Coulomb interaction is modified as screened Coulomb function $\frac{1}{\varepsilon_{\text{r}} R} \mathrm{e}^{-R/l_{\text{D}}}$ with the relative dielectric constant $\varepsilon_{\text{r}} = 15$, and the screen length $l_{\text{D}} = 7$ Å. In the current implementation of hyCG, instead of strictly parameterizing the average potential, we directly use the non-bonded parameter of the MARTINI force field[29] as the CG potential. There are several features in MARTINI for supporting this tentative treatment. On the one hand, MARTINI aims at providing a general force field by assigning a phenomenological chemical function for CG beads. The attractive and the repulsive properties are modeled as the LJ potential, which is already fitted to including the VDW and the electric dipole interaction among various chemical environments. On the other hand, the mapping protocol is fine enough to describe the topology structure of the side chain, like the benzene ring in phenylalanine residue. Hence, it can avoid the overlap between different CG groups caused by atomistic diffusion while we largely reduce the atomistic force. We will adopt the latest version 2.2 of MARTINI for our model, which has been improved to correct certain hydrophobic interaction.[30]

However, the closest distance of the LJ potential in MARTINI is a preassigned value, which in our test is found to be more repulsive for polar groups while we integrate it into our hybrid model. To avoid this repulsive force breaking the stability of a native globular protein, any pairs of the CG site within 4 Å in a native structure are excluded from the calculation of the LJ potential. The cut-off 4 Å is chosen to be slightly larger than the distance of two neighbor C_α atoms so that the originally excluded list in MARTINI can be maintained and a very small fraction of non-bonded CG sites will be blind to each other. This treatment allows the energy to become further lower in the native structure.

3.2. Simulation parameters

The solvent effect in our simulation is treated as GB/SA implicit solvent[31,32] with surface tension setting to 0.005 kcal/mol/Å^2 and α-cutoff adopted 10 Å. The ion concentration in GB/SA is set as 0.2 M and all the simulations are carried out at 300 K and adopting charm22-cmap AA force field.[33,34] For the AA force field, the cut-off for both VDW and Coulomb is set as 12 Å and we apply a switched form for VDW. We adopt the same force field treatment for our average potential. The $\lambda_{\max}$ is chosen as 0.5.

4. Results

4.1. Stability of the native state

The stability of protein's native structure is critical for testing the accuracy of a newly developed force field. To understand whether the hyCG can stabilize native structures of proteins, a 35-residue Villin headpiece protein (PDB entry-code 1YRF) is used to test. We run a hyCG simulation and

an AA simulation from the x-ray structure of 1YRF. Both are carried out for 100 ns.

The time evolution of the root of mean squared deviation (RMSD) of the protein backbone related to its native structure, shown in Fig. 1(a), indicates that the AA force field can stabilize the native structure up to 100 ns with the RMSD value below 3 Å. The stability of the hyCG force field can reach nearly 70 ns, and then the protein loses its tertiary structure but maintains a roughly correct secondary structure. The reduced stability of the hyCG indicates a lower potential barrier between the folded and the unfolded states. We select two pairs of distances, the end-to-end distance and the coil-to-coil distance (see Fig. 1(d)), to test the results of hyCG. The distributions of the distances in the hyCG and the AA (Figs. 1(b) and 1(c)) are roughly overlapped with each other. The small deviation in the probability distribution of the hyCG in the CG space can be further refined by considering the difference between the hyCG effective potential in Eq. (7) and the AA potential based on reweighting.[27]

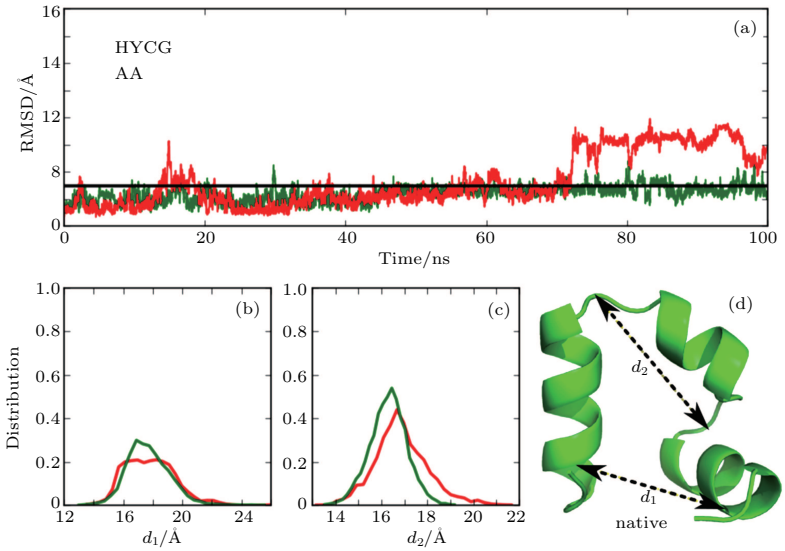


Fig. 1. (color online) Native stability of 1YRF tested by hyCG and AA. (a) Backbone RMSD, in which the red represents hyCG and green represents AA. The color is also adopted by the other figure. The black line indicates a 3 Å limit for the backbone RMSD threshold. Within this threshold, it is commonly considered as a folded state. (b) and (c) Distribution of two selected distances constructed from the first 50 ns trajectory. (d) Schematic representation of 1YRF native structure and the two selected distances.

4.2. Formation rate of secondary structure

To understand the formation of the secondary structure, we study the 1YRF protein again. We perform 250 ns simulation with the hyCG and the AA starting from a complete extended configuration. In order to understand how the non-bonded potential affects the formation efficiency of the secondary structure, we also carry out an energy enhanced sampling simulation on the non-bonded potential as a comparison, which is implemented by selective integrated temperature sampling (SITS).[35]

Two collective variables are chosen to illustrate the formation rate of the helical structure and the correctness of the secondary structure, see Fig. 2. For both collective variables, the hyCG shares a similar growth in the secondary structure formation with the AA but is more efficient and stable. In comparison, although the SITS has a fast formation at the beginning, it will also visit many unfavorable regions for its ability in largely broadening the configuration. The overall effect will slow down the correct formation of the secondary structure in the SITS, which in turn indicates that the rate of the hyCG is most efficient. However, the percentage of the helix in the hyCG shows slightly over helical, yielding the most occupied state differing from the native structure during the short time simulations.

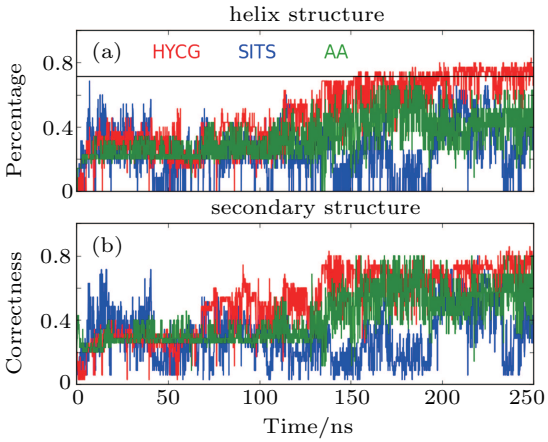


Fig. 2. (color online) Secondary structure formation of 1YRF using hyCG, SITS, and AA. All the secondary structure is determined by STRIDE in VMD.[46] (a) The percentage of helix structure. It is determined by the helical (STRIDE code H,G,I) number divided by the total residue number. The black line indicates the correct percentage of the helix number. (b) The correctness of the secondary structure. It is determined by the percentage of the correct secondary structure compared to the native structure.

4.3. Folding efficiency and folding landscape

To obtain enough samples for constructing the landscape in a short time simulation, we study a smaller protein instead of Villin. Here we study the folding of 20-residues Trp-cage protein (PDB entry-code 1L2Y) to understand the tertiary structure formation efficiency in the hyCG. Trp-cage is a widely studied protein due to its small size. To compare the efficiency of the hyCG and the AA force field, we first run $N = 4$ hyCG and AA simulations at 300 K, respectively, both starting at four different fully extended configurations generated at an equilibrium simulation under 700 K. Each trajectory has $\tau = 250$ ns length. Among these, we find that there are $n = 2$ hyCG trajectories reaching the native structure (with their backbone RMSD smaller than 2 Å) at the time $t_i \approx 70$ ns and 90 ns, respectively, while the other 2 hyCG trajectories do not fold to the native structure within the τ-length simulations. Thus the folding time in the hyCG is estimated by[36,37]

$$t_{\text{fold}} = \frac{1}{n}\left\{\sum_{i=1}^{n} t_i + (N-n)\tau\right\}, \tag{8}$$

which is about 330 ns. The four 250 ns simulations with the AA model all failed to fold, as shown in Fig.3. Considering the known folding time of the protein, about 5 μs,[38] the hyCG has nearly 15 times speed-up. Taking into account the additional calculation costs of the CG force and the effective potential, the hyCG speeds up the folding of the protein about 7.9 times. Because we calculate the CG force and the AA force simultaneously, we can expect a nearly half efficiency drop-down compared to the ideal case due to the additional force calculation.

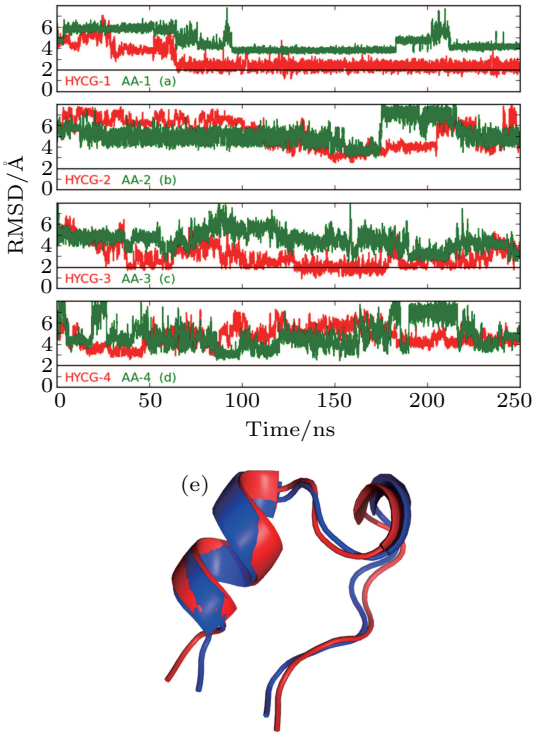


Fig. 3. (color online) Folding trajectory of 1L2Y with hyCG force field and AA force field. (a)–(d) Backbone RMSD of hyCG and AA. The initial configurations are all extended conformation taken from 700 K simulation. Here 2 Å threshold is enough to judge whether the protein is folded. (e) The best folding structure taken from the second trajectory of hyCG with the RMSD equals to 0.88 Å. The red one is the native structure and the blue one is the hyCG's result.

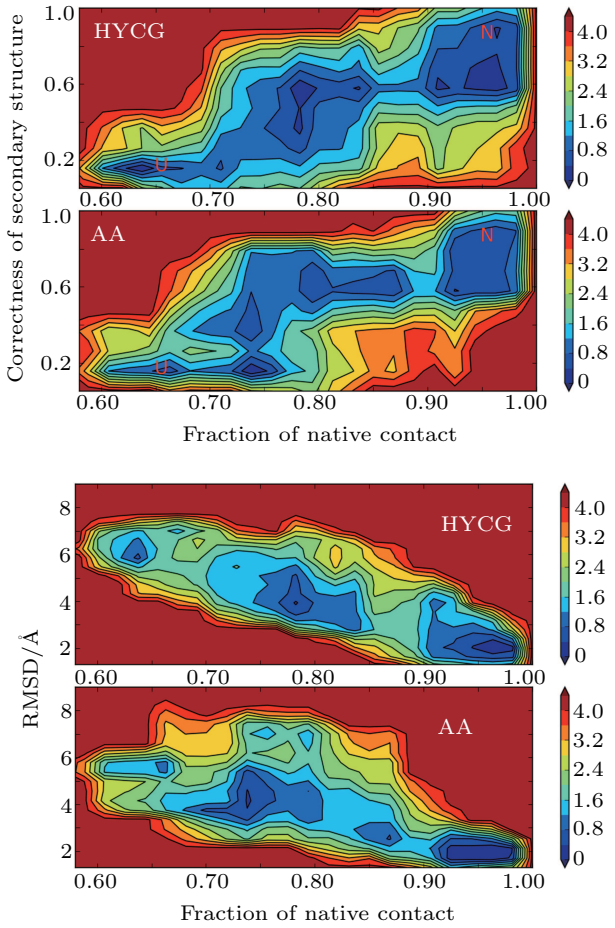


Fig. 4. (color online) Folding landscape of 1L2Y along two given collective variables at 300 K. (a), (c) The fraction of native contact and the correctness of secondary structure. (d), (d) The fraction of native contact and the backbone RMSD (reference is the native structure). The energy is in units of kcal/mol. The free energies of both force fields are constructed by the Markov state model using multiple trajectories. The initial conformations are drawn from a short replica exchange simulation and it contains several nearly folded structures which can enhance the bridging between the folded and the intermediate states in normal simulation.

To investigate whether the folding path is still preserved in the hyCG, we construct the metastable states and their corresponding transitional relationship. A folding path can be described as a transitional network[39] from the unfolded state to the multiple metastable states and to the native state, which can be constructed by the current Markov state model[40–42] or the trajectory mapping method.[43,44] Each metastable state and the main transitional path are visually shown in Fig. 5. We also exhibit two free energy surfaces (FES) to understand the roughness of the folding landscape. Each of the free energy surfaces is constructed by two collective variable sets: the fraction of native contact against the correctness of the sec-

ondary structure and the fraction of native contact against the backbone RMSD (reference is the native structure), respectively, see Fig. 4. The free energy is determined through the Markov state model.

Although some positions of FES local minimums in the hyCG fail to match strictly to that in the AA model, still, the main topology structure is maintained. Besides, the representative structure of important metastable states has a similar secondary and tertiary structure. Further, it verifies the similarity of their folding paths. However, certain metastable states, the V state in Fig. 5 with the backbone RMSD around 7 Å and the fraction of native contact between 0.75 and 0.80 in Fig. 4, are shown to be a branching folding path toward the native structure in the AA and now disappear in the hyCG. These metastable states contain a well-formed helix structure with a very loose coil-like structure, which will slow down the correct contact of the residue. Other simulations starting from this state with the hyCG indicate that the metastable structure in the AA is not stable anymore in the hyCG, see Fig. 6. In turn, it indicates a smoother folding landscape and a more efficient contact rate in the hyCG.

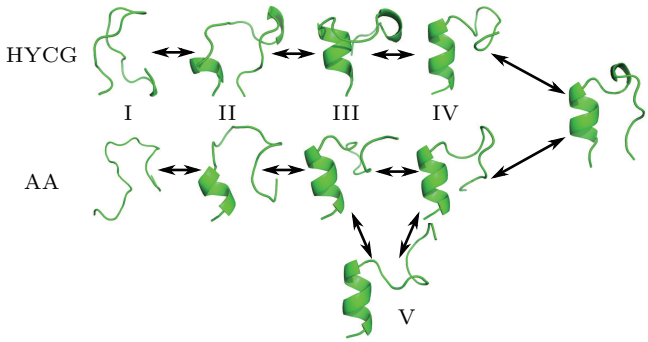


Fig. 5. (color online) Schematically showing the folding path of 1L2Y. Each representative structure is identified by the Markov state model. The black arrows are marked to show the highly frequent transition events. Those states with small transitional rate and very short life are neglected in this diagram for the sake of simplicity.

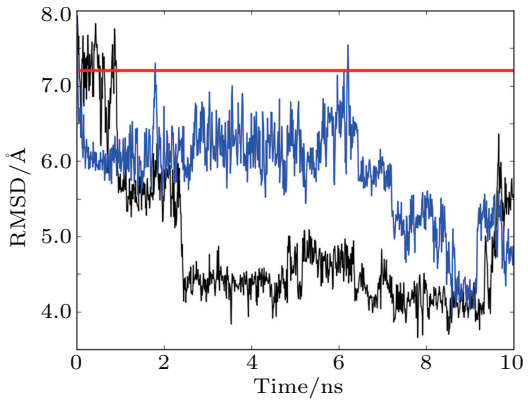


Fig. 6. (color online) Stability of the branching metastable state (V state shown in Fig. 5) in the hyCG force field. The backbone RMSD is determined by comparing to the native structure. The black line and blue line are starting at two randomly selected conformations from the V metastable state. The red line represents the average backbone RMSD of the V state.

5. Conclusion

We develop a novel hybrid approach, named hyCG, and implement it by combining the AA model and the MARTINI model together to efficiently simulate proteins. The hyCG model accelerates the folding dynamics of the small testing proteins about one order of magnitude without changing the folding pathway much. The saving of the time is from the reduced roughness of the potential energy surface in the hyCG. The maintaining of the shape of the potential energy surface in the CG space is responsible for the accuracy of the hyCG in keeping the folding path of proteins. Intuitively, in traditional hybrid CG models, the gaining of efficiency usually brings a sacrifice in the description of the fine interaction, for example, the interaction among two polar groups will gradually switch to a CG LJ potential, and it might change the stability of the hydrogen bond. This, however, is less concerning in the current hyCG where the average potential is applied as the CG part of the potential energy. Our result strongly evidences the practicability of this treatment in comparison with the faster but less accurate CG and the correct but much slower AA model. This hyCG can approximately achieve both stability and accuracy in the native structure of our test proteins, 1YRF and lL2Y. The formation of the secondary structure and the tertiary structure indicates that it has a higher rate than the AA. The folding landscape and transitional network also conclude the similarity in the folding path between the AA and the hyCG. However, we cannot fully expect that the current version of the hyCG can work well in the hydrogen bond abundant system since the non-bonded parameter in the MARTINI force field is originally assigned to describe the hydrophilic and hydrophobic effect.

In comparison with the pure CG scheme, the hybrid CG approach has a shortcoming, the time step in the simulation is required to be the same as that in the AA model, usually femtoseconds in protein, much smaller than that in the pure CG scheme. Therefore, the hybrid approach is actually an AA force field with reducing friction. The multiple time step method[45] may be applied to directly improve the efficiency of the hyCG. In addition, it is also possible to further accelerate the dynamics of the CG DoF in the hyCG then guide the large-scale conformational motion of AA coordinates.

Acknowledgment

The authors thank the support from the massive computation cluster in the Institute of Theoretical Physics of Chinese Academy of Sciences. Most long time simulations were carried out on this platform. Y. L. also appreciates the kind help from Shun Xu, who built a friendly and high-performance

computational cluster that enabled us to quickly verify this idea in a small system and allowed us to frequently test the code without following the queue.

References

[1] Lindorff-Larsen K, Piana S, Dror R O and Shaw D E 2011 *Science* **334** 517
[2] Akkermans R L and Briels W 2001 *J. Chem. Phys.* **114** 1020
[3] Clark A J, McCarty J, Lyubimov I Y and Guenza M G 2012 *Phys. Rev. Lett.* **109** 168301
[4] Izvekov S and Voth G A 2005 *J. Phys. Chem. B* **109** 2469
[5] Liwo A, Lee J, Ripoll D R, Pillardy J and Scheraga H A 1999 *Proc. Natl. Acad. Sci. USA* **96** 5482
[6] Marrink S J, Risselada H J, Yefimov S, Tieleman D P and De Vries A H 2007 *J. Phys. Chem. B* **111** 7812
[7] Shell M S 2008 *J. Chem. Phys.* **129** 144108
[8] Soper A 1996 *Chemical Physics* **202** 295
[9] Tóth G 2007 *J. Phys.: Condens. Matter* **19** 335222
[10] Villa A, van der Vegt N F and Peter C 2009 *Physical Chemistry Chemical Physics* **11** 2068
[11] Villa A, Peter C and van der Vegt N F 2010 *Journal of Chemical Theory and Computation* **6** 2434
[12] Liwo A, Arłukowicz P, Czaplewski C, Ołdziej S, Pillardy J and Scheraga H A 2002 *Proc. Natl. Acad. Sci. USA* **99** 1937
[13] Ołdziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila J, Khalili M, Arnautova Y, Jagielska A and Makowski M 2005 *Proc. Natl. Acad. Sci. USA* **102** 7547
[14] Zhou X, Jiang Y, Rasmussen S and Ziock H 2008 *J. Chem. Phys.* **128** 174107
[15] Lu S and Zhou X 2015 *Communications in Theoretical Physics* **63** 10
[16] Noid W 2013 *J. Chem. Phys.* **139** 090901
[17] Tozzini V 2005 *Current Opinion in Structural Biology* **15** 144
[18] Clementi C 2008 *Current Opinion in Structural Biology* **18** 10
[19] Rzepiela A J, Louhivuori M, Peter C and Marrink S J 2011 *Physical Chemistry Chemical Physics* **13** 10437
[20] Ensing B, Nielsen S O, Moore P B, Klein M L and Parrinello M 2007 *Journal of Chemical Theory and Computation* **3** 1100
[21] Heyden A and Truhlar D G 2008 *Journal of Chemical Theory and Computation* **4** 217
[22] Praprotnik M, Delle Site L and Kremer K 2005 *J. Chem. Phys.* **123** 224106
[23] Christen M and van Gunsteren W F 2006 *J. Chem. Phys.* **124** 154106
[24] Liu P and Voth G A 2007 *J. Chem. Phys.* **126** 045106
[25] Shen L and Hu H 2014 *Journal of Chemical Theory and Computation* **10** 2528
[26] Sugita Y and Okamoto Y 1999 *Chemical Physics Letters* **314** 141
[27] Gao Y Q 2008 *J. Chem. Phys.* **128** 064105
[28] Mori T, Hamers R J, Pedersen J A and Cui Q 2014 *J. Phys. Chem. B* **118** 8210
[29] Monticelli L, Kandasamy S K, Periole X, Larson R G, Tieleman D P and Marrink S J 2008 *Journal of Chemical Theory and Computation* **4** 819
[30] de Jong D H, Singh G, BennettWD, Arnarez C, Wassenaar T A, Schafer L V, Periole X, Tieleman D P and Marrink S J 2013 *Journal of Chemical Theory and Computation* **9** 687
[31] Onufriev A, Bashford D and Case D A 2000 *J. Phys. Chem. B* **104** 3712
[32] Onufriev A, Bashford D and Case D A 2004 *Proteins: Structure, Function, and Bioinformatics* **55** 383
[33] Chen J, Im W and Brooks C L 2006 *Journal of the American Chemical Society* **128** 3728
[34] MacKerell A D, Feig M and Brooks C L 2004 *Journal of Computational Chemistry* **25** 1400
[35] Yang L and Gao Y Q 2009 *J. Chem. Phys.* **131** 214109
[36] Voter A F 1998 *Phys. Rev. B* **57** R13985
[37] Zhou X, Jiang Y, Kremer K, Ziock H and Rasmussen S 2006 *Phys. Rev. E* **74** R035701
[38] Lane T J, Shukla D, Beauchamp K A and Pande V S 2013 *Current Opinion in Structural Biology* **23** 58–65
[39] Rao F and Caflisch A 2004 *Journal of Molecular Biology* **342** 299
[40] Noé F and Fischer S 2008 *Current Opinion in Structural Biology* **18** 154
[41] Noé F, Schütte C, Vanden-Eijnden E, Reich L and Weikl T R 2009 *Proc. Natl. Acad. Sci. USA* **106** 19011
[42] Bowman G R, Huang X and Pande V S 2009 *Methods* **49** 197
[43] Gong L and Zhou X 2010 *J. Phys. Chem. B* **114** 10266
[44] Gong L, Zhou X and OuYang Z 2015 *Plos One* **10** e0125932
[45] Watanabe M and Karplus M 1993 *J. Chem. Phys.* **99** 8063–8074
[46] Humphrey W, Dalke A and Schulten K 1996 *Journal of Molecular Graphics* **14** 33