

University of Groningen

Multiscale Membrane Models

Liu, Yang

DOI:
[10.33612/diss.136221782](https://doi.org/10.33612/diss.136221782)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Liu, Y. (2020). *Multiscale Membrane Models*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.136221782>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 3

Coupling Coarse-Grained to Fine-Grained Models via Hamiltonian Replica Exchange

Yang Liu, Weria Pezeshkian, Jonathan Barnoud, Alex H. de Vries and Siewert J. Marrink

This chapter has been published in *Journal of Chemical Theory and Computation*

Abstract

The energy landscape of biomolecular systems contains many local minima that are separated by high energy barriers. Sampling this landscape in molecular dynamics simulations is a challenging task, and often requires the use of enhanced sampling techniques. Here, we increase the sampling efficiency by coupling the fine-grained (FG) GROMOS force field to the coarse-grained (CG) Martini force field via the Hamiltonian replica exchange method (HREM). We tested the efficiency of this procedure using a lutein/octane system. In traditional simulations, cis-trans transitions of lutein are barely observed due to the high energy barrier separating these states. However, many of these transitions are sampled with our HREM scheme. The proposed method offers new possibilities for enhanced sampling of biomolecular conformations, making use of CG models without compromising the accuracy of the FG model.

Introduction

All-atom (AA) molecular dynamics (MD) simulation has become a powerful technique to investigate a broad range of biomolecular processes at atomistic detail¹¹²⁻¹¹³. MD relies on statistical mechanics and the ergodic hypothesis (the time average is equal to the ensemble average) to connect the microscopic variables to corresponding macroscopic properties that are measured in experimental settings. Therefore, an accurate measurement of the macroscopic properties is only possible if the MD trajectory samples all the important (probable) configurations. Typically, the energy landscape of biomolecular systems contains many local minima that are separated by high energy barriers. Sampling through these minima usually requires simulations much longer than the accessible time to currently available AAMD software¹¹⁴⁻¹¹⁶. On the other hand, coarse-grained (CG) simulation methods, at the expense of molecular details, allow exploring the spatial and temporal evolution of the systems up to 2-3 orders of magnitude higher than AA simulations¹¹⁷⁻

¹¹⁸. Nevertheless, the correct description of many biological processes requires molecular details above the CG resolutions. To resolve this dilemma, one practical solution is to combine AA and CG simulations through a multi-scale approach to benefit from the advantages of both techniques.

There are several different approaches to combine CG and AA simulations, including interface approaches, serial multiscale methods, adaptive resolution, and resolution exchange approaches. The interface approach keeps the details of the AA structure only for a specific part of the system and CG potentials govern the rest of the system^{46, 119-120}. The serial multiscale method—also denoted back mapping, starts with the CG simulations and only switches back to AA resolution if some key states or interesting events occur¹²¹⁻¹²². In the adaptive resolution method, a resolution can be changed or adapted on the fly when atoms or molecules cross the boundary between AA-CG regimes¹²³⁻¹²⁴. In the resolution exchange approach, a set of simulations at different resolutions are performed simultaneously and the configurations between neighboring replicas are exchanged if the detailed balance condition is satisfied. This approach is more generally known as the Hamiltonian replica exchange method (HREM)¹²⁵⁻¹²⁷.

The aim of this article is to combine the CG Martini²⁶ and fine-grained (FG) GROMOS²² force fields via HREM. Previously, Martini and FG force fields have been successfully coupled using the interface approach⁴⁶, in serial multiscaling^{122, 128}, as well as in adaptive resolution simulations^{48, 129}. Christen and van Gunsteren⁵⁴ combined both GROMOS and Martini force fields via HREM through a parameter λ ranging from 0 to 1, where each λ defines a level of resolution or a replica. In the Christen scheme, at each level of resolution, the FG bonded potentials are kept unchanged, whereas the FG non-bonded and the full CG potentials are weighed with a λ -dependent factor. The configurations of the different replicas are exchanged via the HREM framework. This scheme is indeed elegant, however, it is not very effective. Goga et al.¹³⁰ estimated the sampling efficiency, based on the analysis of conformational entropy of systems simulated using this method, and showed that it does not yield a significant speedup. The main reason for the inefficiency of the Christen scheme is the use of unscaled FG bonded potentials. In this work, we take a similar approach but resolve the inefficiency problem by scaling the FG and CG potential of the important targets, which usually have multiple minima in the energy surface, but are separated by high energy barriers. Our approach recovers the FG potentials at the lowest λ value, while the corresponding CG potentials dominate the system at the highest λ .

The rest of this article is structured as follows. First, we introduce the methodological basis of our approach. Then, we validate the new multiscale model based on a pure octane system and on a lutein molecule, an important carotenoid cofactor of the photosynthetic machinery¹³¹. We subsequently evaluate the enhanced sampling of FG/CG HREM on an octane/lutein solution. Lastly, the merits and shortcomings of the method are discussed.

Method

Coupling resolutions

In our HREM framework, we combine the FG GROMOS and CG Martini force fields. A molecule in HREM is represented, simultaneously, at both FG and CG levels of resolution. The CG particles are modeled by virtual sites that are positioned on the center of mass of their corresponding FG particles (Figure 1a, Equation 1a-c), similar to the interfacial coupling approach of Rzepiela et al.⁴⁶.

$$M_k = \sum_{i=1}^{N_k} m_{k_i} \quad (1. a)$$

$$\vec{R}_k = \sum_{i=1}^{N_k} \vec{r}_{k_i} m_{k_i} / M_k \quad (1. b)$$

$$\vec{V}_k = \sum_{i=1}^{N_k} \vec{v}_{k_i} m_{k_i} / M_k \quad (1. c)$$

M_k , \vec{V}_k and \vec{R}_k are the total mass of the particles assigned to VS bead k , the velocity and the position of the VS bead, respectively. The k -th VS bead is constructed from N_k FG atoms with mass m_{k_i} , velocity \vec{v}_{k_i} and position \vec{r}_{k_i} for the i -th atom of the k -th VS bead. Throughout the rest of this article, we refer to the particles of the FG systems as atoms and the particles of the CG systems as beads.

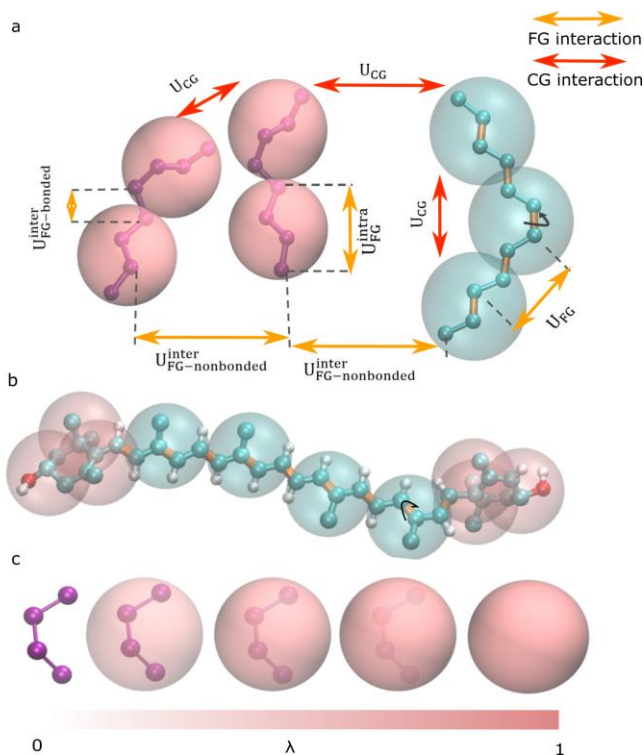


Figure 1. The multiscale system encompasses both FG/CG molecule representations. (a) Multiscale structure of octane (purple) and the backbone part of lutein (cyan). The unrestrained dihedral is indicated by the black arrow (see text for details). Double bonds are shown in orange. (b) Illustration of the dual resolution structure in case of lutein. (c) Schematic illustration of the multiscale structure of butane. λ ranges from 0 to 1 which takes the system from FG to CG descriptions.

The multiscale HREM potential is constructed by combining the underlying FG and CG potentials using a parameter which determines the resolution level of the simulation (Figure 1c). In this work, we propose two different combination schemes (denoted as I and II) that each describe the interactions between different molecules.

Scheme I:

$$U_{\text{multiscale}} = U_{FG}^{\text{intra}}(\lambda_{sc}^{FG}, r) + (1 - \lambda) U_{FG\text{-nonbonded}}^{\text{inter}}(\lambda_{sc}^{FG}, r) + \lambda U_{CG}(\lambda_{sc}^{CG}, R) + U_{FG\text{-bonded}}^{\text{inter}}(\lambda, r) \quad (2)$$

$$U_{FG\text{-bonded}}^{\text{inter}}(\lambda, r) = (1-a\lambda)U_{FG\text{-bond}}^{\text{inter}}(r) + (1-b\lambda)U_{FG\text{-angle}}^{\text{inter}}(r) + (1-c\lambda)U_{FG\text{-dihedral}}^{\text{inter}}(r) \quad (3)$$

$$\lambda = [0 - 1]$$

Scheme II:

$$U_{\text{multiscale}} = (1 - \lambda)U_{FG}(\lambda_{sc}^{FG}, r) + \lambda U_{CG}(\lambda_{sc}^{CG}, R) \quad \lambda = [0 - 1] \quad (4)$$

The specific potential components are illustrated in Figure 1a. The U_{FG} and U_{CG} are the total potential energy (bonded and non-bonded) of FG and CG force fields respectively. U_{FG}^{intra} is the total interaction energy between atoms that are within the same CG bead and U_{FG}^{inter} is the interaction energy between atoms belonging to two different CG beads. a , b and c are constants that are smaller than 1 to keep the weak inter-beads FG bonded potentials at $\lambda=1$. Under the condition of reproducing the ideal CG (VS) bond distribution, these numbers are chosen to keep as much FG bonded potential as possible. The nonbonded potentials are modified to a softer potential¹³² through parameters λ_{sc}^{FG} and λ_{sc}^{CG} to avoid sudden changes in the energy which can produce large forces. λ_{sc}^{FG} and λ_{sc}^{CG} are linear functions of the overall multiscaling parameter λ , ranging in value between 0 and 0.5, and are given in Equations S4-S5. For more information on the use of softcore potentials, see SI (Figure S1).

Scheme I is used to describe the interactions between solvent molecules. Fully FG potentials are maintained at $\lambda=0$. As λ increases, the inter-bead FG interactions decrease and the non-bonded inter-bead FG potentials become softer. Simultaneously, the CG potentials are scaled up by the λ and the softcore CG potential becomes harder. When λ reaches 1, the inter-bead FG interactions go to the lowest value (but are not entirely absent) while the CG interactions are fully switched on. The intra-bead bonded FG potentials are kept constant throughout the replicas. Theoretically, it is not advisable to use the softcore potential on the intra-bead nonbonded FG potential. However, we have to use one tabulated potential for all FG atoms within one replica, so the soft-core potential is also applied to the intra-bead nonbonded FG interactions. This compromise only decreases the computational efficiency (because it only slightly increases the conformational difference between atoms in different replicas), and will not affect the accuracy of the multiscale scheme. Therefore, this scheme produces closely matched structures and potentials at high and low λ , and simultaneously maintains the correct CG structure at $\lambda=1$. In other words, the difference between the potentials at $\lambda=0$ and $\lambda=1$ is small and as a result, the solvent interaction has minimal influence on a decent acceptance ratio. In scheme II, the full FG potentials are also maintained at $\lambda=0$. As λ increases, the FG interactions are scaled down by both the λ and the soft-core potential, while the CG interactions go the opposite way. When λ is close to 1, the hard-core CG potentials are recovered. Note: λ should never reach one. This is important because weak FG bonded potentials are needed to hold the underlying atoms together. At high λ , since the FG potentials are weaker, the FG atoms can easily overcome the energy barriers and visit more

configurations. The configurations can be exchanged through the replicas all the way to $\lambda=0$ and increase the sampled space. Scheme II is usually used for increased conformational sampling of the targets of interest (e.g., a solute). We also treat the interaction between solvent and solute using scheme II. Therefore, the combination of scheme I (solvent) and scheme II (solutes) can increase the sampling efficiency on the target molecules (solutes).

To test our multiscale HREM scheme, we have compared its results and performance against other possible HREM schemes, denoted pure FG HREM, traditional HREM, CG solvent HREM and Christen HREM. In pure FG HREM, the system is only represented by FG interactions. Compared to the multiscale HREM described in Equation 2-4, all the CG interactions are neglected. In traditional HREM (e.g. as used by Fukunishi et al.¹²⁵), the solvent Hamiltonian is kept unchanged and the solute interactions are scaled similarly to the pure FG HREM. In CG solvent HREM, the solvent (octanes in our systems) is treated by unscaled CG potentials and the solute (lutein) interactions are treated by scheme II. In Christen HREM, following the approach of Christen et al.⁵⁴, all the FG and CG potentials are scaled apart from the FG bonded interactions. The λ distribution used is the same for all the systems and it is listed in the SI.

HREM algorithm

As we mentioned earlier, each λ represents a specific resolution (Figure 1c). Exchange between these resolutions is done via HREM¹²⁵. Let's assume $H(S)=H(r^{FG}, p^{FG}, \lambda)$ is the Hamiltonian of a system at a specific state, where r^{FG} and p^{FG} are FG atom positions and momenta, respectively. The probability W_n of configuration S in the n^{th} replica obeys the Boltzmann distribution and is expressed as:

$$W_n(S) = \exp(-\beta H_n(S)) / Z_n \quad (5)$$

$$W_{\text{all}} = \prod_{n=1}^N W_n(S_n) \quad (6)$$

where W_{all} is the probability distribution of the extended system, Z_n is the partition function of the n^{th} replica, N is the number of replicas and $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant and T is the temperature. By satisfying detailed balance, the extended system reaches Boltzmann equilibrium:

$$W_{\text{all}}(S)W(S \rightarrow S') = W_{\text{all}}(S')W(S' \rightarrow S) \quad (7)$$

where $W(S \rightarrow S')$ represents the transition probability from global state S to state S' .

Through Metropolis criterion¹³³, the probability $p(S \rightarrow S')$ of an exchange between replica S to replica S' is expressed as:

$$\begin{aligned} p(S \rightarrow S') &= W_{\text{all}}(S')/W_{\text{all}}(S) = W(S \rightarrow S')/W(S' \rightarrow S) \\ p(S \rightarrow S') &= \exp(-\Delta) \quad \text{when } \Delta > 0 \\ p(S \rightarrow S') &= 1 \quad \text{otherwise} \end{aligned} \quad (8)$$

$$\begin{aligned} \Delta &= \beta(U_{\text{multiscale}}(\mathbf{r}^{\text{FG},S'}, \lambda^S) - U_{\text{multiscale}}(\mathbf{r}^{\text{FG},S}, \lambda^S)) - \beta(U_{\text{multiscale}}(\mathbf{r}^{\text{FG},S'}, \lambda^{S'}) \\ &\quad - U_{\text{multiscale}}(\mathbf{r}^{\text{FG},S}, \lambda^{S'})) \end{aligned} \quad (9)$$

where $U_{\text{multiscale}}(\mathbf{r}^{\text{FG},S}, \lambda^S)$ represents the total potential energy including both FG and CG resolutions of the corresponding system. The coordinates of CG beads ($\mathbf{R}^{\text{CG},S}$) depend on FG atoms ($\mathbf{r}^{\text{FG},S}$) through the appropriate mapping scheme. λ^S represents the λ , $\lambda_{\text{sc}}^{\text{FG}}$ and $\lambda_{\text{sc}}^{\text{CG}}$ defined in Equations 2-4 and S4-S5.

Simulation setup

In this work, we have analyzed three distinct systems *i.e.*, a solvent box containing pure octane, a single lutein in vacuum, and a mixed lutein/octane system. All systems were simulated using standard MD together with different HREM schemes (see Table 1).

Octane solvent: The octane systems consisted of 128 octane molecules (Figure 1a), and was first simulated using the traditional MD technique with combination scheme I at $\lambda=1$ (Table 1 System I). Then we performed HREM simulation on the same system that included 64 replicas, each at different λ which ranges from 0 to 1 (Table 1 System II and see SI for λ values).

Lutein in vacuum: Next we performed HREM on a single lutein molecule in vacuum, considering only a fraction of the lutein backbone (Figure 1a), using scheme II. The simulation included 12 replicas with λ ranging from 0 to 0.8 (Table 1 System III). In addition, two reference simulations, starting either from the trans or cis dihedral structure, were simulated in vacuum using standard MD (Table 1 System IV). The overall translation and rotation were removed at every time step.

Lutein/octane solution: A mixture containing 20 lutein molecules and 128 octane molecules were analyzed by five different simulation techniques, *i.e.*, multiscale HREM, pure FG HREM, traditional HREM, CG octane HREM and Christen HREM. Details are listed in Table 1 System V to IX.

Table 1. Simulation setups in different systems

Systems	Composition	Ensemble	Time	Method	Multiscale scheme	λ
System I	128 octane	NPT	200 (ns)	MD	Scheme I	1
System II (multiscale HREM)	128 octane	NVT	2 (ns) \times 64 replicas	HREM	Scheme I	0-1
System III (multiscale HREM)	1 lutein (backbone only)	NVT	400 (ns) \times 12 replicas	HREM	Scheme II	0-0.8
System IV	1 lutein (backbone only)	NPT	200 (ns)	MD	Scheme II	0
System V (multiscale HREM)	128 octane/ 20 lutein	NVT	3 (ns) \times 64 replicas	HREM	Scheme I/II for octane/lutein	0-0.8
System VI (pure FG HREM)	128 octane/ 20 lutein	NVT	3 (ns) \times 64 replicas	HREM	Scheme I/II for octane/lutein (ignoring all CG interactions)	0-0.8
System VII (traditional HREM)	128 octane/ 20 lutein	NVT	3 (ns) \times 64 replicas	HREM	Pure FG / Scheme II for octane / lutein, while ignoring all CG interactions	0-0.8
System VIII (CG solvent HREM)	128 octane/ 20 lutein	NVT	3 (ns) \times 64 replicas	HREM	Pure CG/Scheme II for octane/lutein	0-0.8
System IX (Christen HREM ⁵⁴)	128 octane/ 20 lutein	NVT	3 (ns) \times 64 replicas	HREM	Both FG and CG potentials are scaled, apart from FG bond interactions	0-0.8

All the simulations were performed using the GROMACS 2016 software³³. The GROMOS 53a6²² and Martini 2.0²⁶ force field were used to describe FG-FG and CG-CG interactions, respectively. Martini and GROMOS are calibrated with two different cut-off lengths and different ways to shift the potentials. Therefore, we exploited a tabulated potential to correctly perform on both force fields. All the GROMOS topologies were generated by the Automatic Topology

Builder¹³⁴. The CG octane model was downloaded from <http://www.cgmartini.nl/> and the CG lutein was developed in ¹³⁵. For details of the multiscale octane topology setups see SI (Table S1, Figure S2). We used an integration time step of 2 fs and updated the neighbor list every 5 steps. The temperature was kept constant at 300 K using Berendsen thermostat³⁴ with $\tau_t = 0.1$ ps. In traditional simulations, the systems were coupled in the isothermal–isobaric ensemble and the pressure was controlled at 1 bar using a Berendsen barostat with $\tau_p = 1$ ps. The compressibility was $4.5 \times 10^{-5} \text{ bar}^{-1}$. In HREM simulations, we used the NVT ensemble for all the replicas and started all the replicas from an identical structure. The exchange trial was performed every 500 steps and the first 20 exchange trials were discarded from the analysis. Note that CG beads cannot be coupled to a thermostat since they are built as virtual sites. However, using the average kinetic energy of the beads, an effective temperature can be measured:

$$T^{\text{CG}} = \sum_i M_i V_i^2 / (3Nk_B) \quad (10)$$

where k_B and N are the Boltzmann constant and number of the CG beads.

The free energy difference between the trans and cis states of lutein were estimated based on the state distribution as

$$\Delta G = -k_B T \ln(p_A/p_B) \quad (11)$$

where p_A and p_B are the probability of visiting the trans and cis states, respectively. A dihedral angle (θ) is defined to be in a cis state if $90 < \theta < 270$ and in a trans state otherwise. The error estimate in the free energy was obtained using block averaging¹⁰¹.

Results

Validation of multiscale solvent system at $\lambda=1$

As it was mentioned earlier, in scheme I, the FG bonded interactions within a bead are unscaled and between beads are not zero. Therefore, even at $\lambda=1$, the force field differs from the standard CG-Martini force field. In this section, we measure the deviation of the scheme I (at $\lambda=1$) from a reference standard CG-Martini simulation. In addition, we compare this result to the multiscale scheme by Christen *et al.*⁵⁴.

To this end, a system of 128 octane molecules was simulated using scheme I at $\lambda=1$ (Table 1 System I). Table 2 compares the CG bead temperature and density of our method with standard Martini and Christen's scheme. Both our multiscale scheme and Christen scheme can reproduce the temperature and density of standard Martini, yet, our scheme is slightly better. Furthermore, in Figure 2 we compare the conformational structure of the octane system, simulated with the two different multiscale schemes and with the standard Martini model, by means of the octane bond length distribution (Figure 2a) and the radial distribution function (RDF) of the octane centers of mass (Figure 2b). The distributions produced by our scheme match very well with the results of the standard Martini model, while the Christen scheme differs substantially. Notably, Christen's scheme produces a shorter bond length (Figure 2a), which results in a clear peak in the RDF around 0.41 nm (Figure 2b). This peak reflects the octane intra-bead distance; if the intra-bead pairs are excluded from the RDF, it is absent. In the standard CG and multiscale HREM, the intra-bead pairs contribute a shoulder to the RDF, reflecting the longer bond at CG level compared to the one resulting from the Christen system. The clear peak at 0.41 nm in the Christen scheme is present because the remaining FG bonded potentials (bond, angle, dihedral) that are fully switched on at $\lambda=1$, dominate the energy landscape relevant to the CG octane intra-bead distance. The shift in position from mapped full FG to the one observed in the Christen scheme is due to the switching off of the 1-4 pair interaction that is used in the FG GROMOS model and contributes to the potential energy landscape of dihedral angles. This can be seen by comparing the purple and red curves in Figure 2a, which reflect a full FG model without and with 1-4 pair interactions. It is to be expected that the use of a dihedral potential in combination with the 1-4 pair interactions would lead to a much better match of the Christen scheme with the full CG and multiscale HREM schemes. Noted that a wider bond distribution is found in the standard CG model compared to the one resulting from mapping the full FG system. The Martini model does not always respect the FG mapping, which was shown for the angles and dihedrals in hexadecane¹³⁶. In our multiscale scheme, the inter-bead FG interactions are weakened enough and the CG bonded potential energy dominant. Therefore, the bond distribution fits better with the standard CG model (Figure 2a).

The shorter CG bond length observed in the Christen scheme explains the somewhat higher mass density. In octane, the effect is mild, but it has been shown that shorter bond lengths can compromise the partitioning behavior of the standard CG model¹³⁷ and should therefore be avoided

also in hybrid schemes. Together, these results clearly show that scheme I, at $\lambda=1$, is able to reproduce pure CG system behavior.

Table 2. Thermodynamic properties of the octane system at $\lambda=1$.*

	Temperature (K)	Density (kg/m^3)
Christen scheme	292.9 ± 12.5	803.0 ± 6.1
Multiscale scheme I	303.6 ± 12.8	795.6 ± 6.1
Standard CG	299.6 ± 8.2	792.3 ± 7.1

*Note, the thermostat does not work on CG virtual sites, the temperature of which depends on the velocity of their underlying FG atoms and is measured using Equation 10. Error margins represent standard deviation after the values reach equilibration (first 10 ns are discarded). Statistical errors $< 0.1\%$.

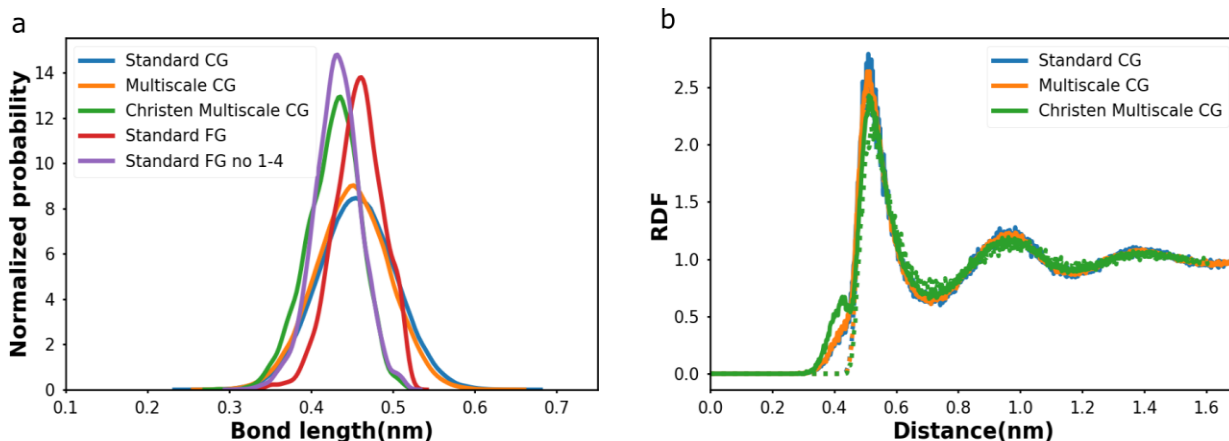


Figure 2. Configurational properties of multiscale octane system at $\lambda=1$. (a) The bond length distribution. A bond is defined as the distance between the two CG beads within an octane molecule. The standard FG no 1-4 represents the system simulated with FG potential excluding the 1-4 interactions. (b) RDF of the centers of mass of octane molecules. The multiscale CG scheme I at $\lambda=1$ is compared to the multiscale scheme of Christen et al.⁵⁴ and the standard CG Martini model. The solid line and dashed line represent the RDF including the intramolecular pairs and excluding the intramolecular pairs, respectively.

HREM on multiscale octane solvent system

To assess the suitability of our multiscaling scheme in HREM, we have performed an HREM simulation with 64 replicas (2 ns for each replica) on a solvent system composed of 128 octane molecules (Table 1 System II). The potential energy of the system was described by scheme I (Equations 2 and 3). The λ space was divided into 64 points, distributed exponentially as shown in Figure 3a. We have observed that the conformations are exchanged between all pairs of

neighboring replicas (Figure 3b and Figure S4a), indicating that the trajectory of the replica at $\lambda=0$ (FG system) is influenced by all other replicas. To investigate the effect of the softcore potential on the exchange rate, we performed a similar HREM simulation but using a hardcore potential. The results show that the exchange rates of the system using a softcore potential are higher than using a hardcore potential at λ close to 0 and 1 (Figure 3b). For a high exchange rate between two replicas, both conformational and potential energy similarities are important (Equations 8-9)¹²⁷. The softcore CG potential at λ close to 0 and 1 increases the topological and conformational similarity between neighboring replicas. Thus, the exchange probability and the sampling efficiency are higher for HREM simulation with softcore potential and therefore incorporated into our Hamiltonian as indicated in Equation 2. At λ close to 1, where the density of replicas is high (Figure 3a), the corresponding exchange rate is still very low (Figure 3b). As the potential energy distributions overlap efficiently in this region of λ space (Figure S3), the poor exchange rate is attributed to a high conformational or topological difference. Therefore, a restricted λ range (from 0 to 0.8) is used in the following sections.

To provide further insight into the factors that govern the exchange probabilities (Figure 3c), we obtained different energy components of the system as a function of λ . As expected, FG (CG) LJ potential decreases (increases) with an increase in the replica index. The total LJ potential energy behaves similarly to the total energy. This means that FG bonded and 1-4 LJ potentials undergo minor changes throughout the replicas because only small parts (namely, inter-bead potentials) are scaled according to the solvent interaction scheme I. The changes in CG bond potential are very small compared to the changes in the LJ potential energy. This indicates that the LJ potential is the dominant factor in the exchange rate.

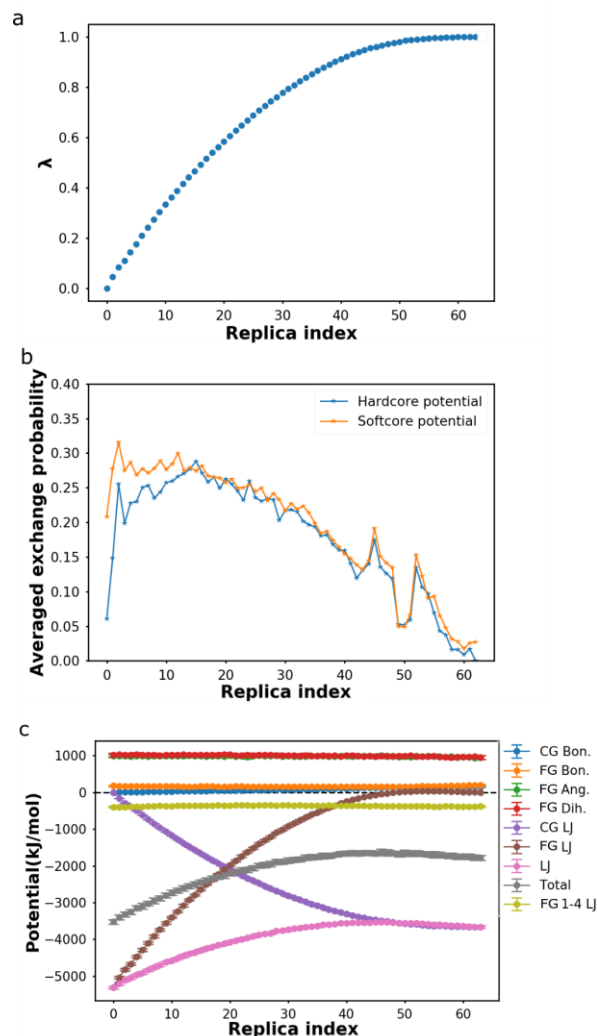


Figure 3. HREM on octane systems. (a) λ distribution of the different replicas. (b) The exchange probability between neighboring replicas (index and index+1). (c) Different energy level components of both FG and CG interactions for all the replicas. The dashed line shows the zero-energy level.

HREM on lutein in vacuum

To characterize the accuracy of scheme II (solute potential) we have performed an HREM simulation of the conjugated backbone of lutein (Figure 1a) in vacuum and compared the results to a standard FG MD simulation (Table 1, System III and IV). The backbone of lutein is composed of conjugated single/double bonds and conformational exchange between the trans/cis configurations cannot be captured by traditional FG simulations because of a high energy barrier separating these states. In order to facilitate the comparison of the HREM results to standard FG

simulations, we only focus on conformational transformations of the central cis/trans double bond (black arrow, Figure 1a). All other dihedrals were restrained. The all-atom reference simulation was either initiated from the cis or trans conformation. In addition, we also performed a temperature REM simulation (temperature range 300-900K) and compared the results to the one obtained by our HREM scheme.

Our results show that, indeed, the HREM simulation samples both trans and cis conformations, while the traditional FG simulations fail to overcome the trans/cis energy barrier on the allocated time scale (Figure 4a). To further check the validity of HREM, we calculated the potential energy distributions of the HREM system (at $\lambda=0$) for the trans and cis conformations separately and compared these to the results from the reference FG simulations (Figure 4b). It is clear that these distributions match very well in both configurations.

The free energy difference between the trans and cis state (ΔG) can be calculated using Equation 11. We found that HREM predicts $\Delta G = -7.4 \pm 0.3$ kJ/mol (trans state has lower energy). In order to validate this value, we also performed temperature REM on the system and found $\Delta G = -7.3 \pm 0.2$ kJ/mol, which is the same as the results from HREM, within the range of the error bar. Thus, we conclude that our HREM approach can sample the trans/cis dihedral distribution of this solute without a bias.

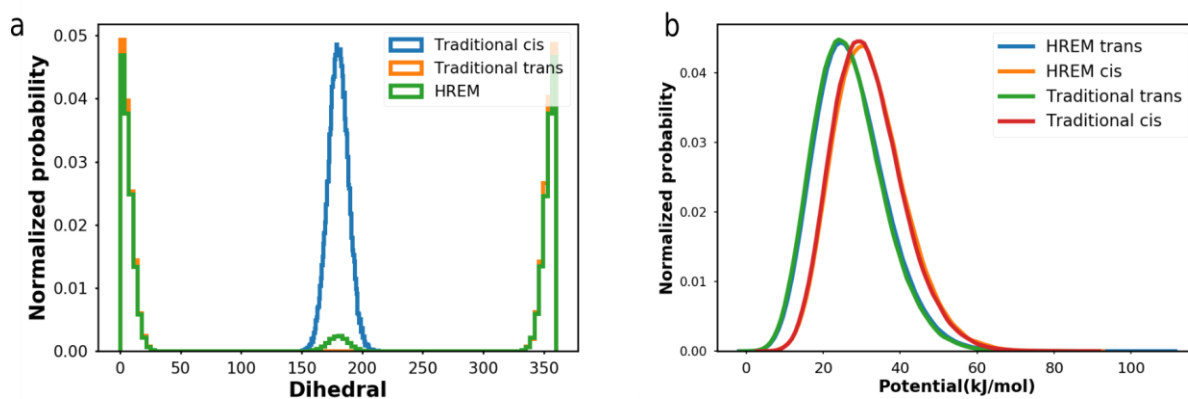


Figure 4. HREM of a solute (lutein tail) in vacuum. Two traditional FG MD simulations, starting either from a trans or cis dihedral structure, are compared to the HREM system at $\lambda=0$. (a) Dihedral distribution of the central dihedrals as marked in Figure 1a. (b) Total potential energy distributions of the different systems samples either trans or cis conformation.

HREM of lutein cofactors in octane solution

In this section, we aim to use both combination schemes (I and II) in a single multiscale simulation. To do this, we performed an HREM simulation on a system composed of 20 lutein (Figure 1b) and 128 octane molecules (Table 1 System V to IX). As shown in the previous section, the backbone of lutein contains several double bonds and conformational exchange between trans/cis conformations cannot be captured by traditional FG simulations. This makes the lutein solution a suitable system to test the power of our HREM approach. Note that here, we do not restrain any of the dihedral angles along the backbone, giving rise to a potential of 512 independent conformations. As we aim to increase lutein conformational sampling, we used scheme II to describe its interactions (both internal and with the surrounding solvent). Solvent-solvent interactions, on the other hand, were described by scheme I.

We compared the results of this hybrid scheme to results obtained from pure FG HREM, traditional HREM, CG solvent HREM or Christen HREM. With these setups, we have observed an appropriate overlap of the energy probability distribution and a decent amount of exchanges between neighboring replicas (for more details see Figure S3 and S4). Acceptance ratios of all HREM simulations are above 10%. Figure 5 shows lutein's dihedral distribution obtained from each HREM method, averaged over all 20 lutein molecules present in the simulation. The analysis is shown for the double bond dihedral that is marked in Figure 1b. Distributions for other dihedrals are shown in Figure S5. It is clear that, apart from Christen HREM, all other HREM schemes capture the transformations between cis/trans structures, in particular at high λ . This is because Christen HREM is the only scheme that does not scale FG bonded interactions throughout the replicas and preserves the high cis-trans barriers. It is interesting to note, however, that the exchange efficiency between neighboring replicas is highest in this scheme and similar throughout the range of λ values (Figure S4), likely due to high conformational similarity. The exchanges between cis/trans conformations, for the same simulation time, is higher in CG solvent HREM and hybrid HREM than pure FG and traditional HREM (Figure 5f). The same trends can be found for the other dihedrals in Figure S5. These results clearly show the advantage of using our HREM scheme over the traditional or pure FG HREM when we are dealing with high energy barriers. Comparing with the pure FG scheme, our hybrid HREM approach is able to sample more cis conformations at $\lambda=0$. Interestingly, fewer cis configurations are sampled at high λ (Figure 5a and b). This suggests that the CG potential at high λ , in our hybrid scheme, limits the sampling of the

FG dihedrals. The mapping scheme and stiffness of the CG bonds, angles, and dihedrals limits the motion of the FG atoms and influences the dihedral distribution at FG level. Thus, more cis configurations are sampled at high λ and exchanged back to $\lambda=0$. In addition, the results show that our hybrid scheme and CG solvent HREM, are equally good at sampling state transition (Figure 5f). Nevertheless, compared with multiscale HREM, since no FG solvent is involved, CG solvent HREM is cheaper, especially when the number of solvent molecules are increased. Even though there is a resolution interface at $\lambda=0$, this virtual site hybrid model can properly reproduce the correct free energy of apolar hybrid compounds in apolar CG solvents⁴⁷. Note that the simulations illustrated in Figure 5 have not fully equilibrated yet. In the lutein backbone system, we need 400 ns in each replica to reach transition equilibrium between trans/cis states. That system only consisted of a lutein backbone in vacuum and was simulated with 12 replicas. However, for the lutein/octane solution simulated with 64 replicas, the HREM simulation is too expensive to reach the equilibrium state for all possible 512 trans/cis states. Here we only want to show that starting with the same structure, the hybrid and CG solvent HREM are able to sample more transitions between trans/cis states than the other schemes.

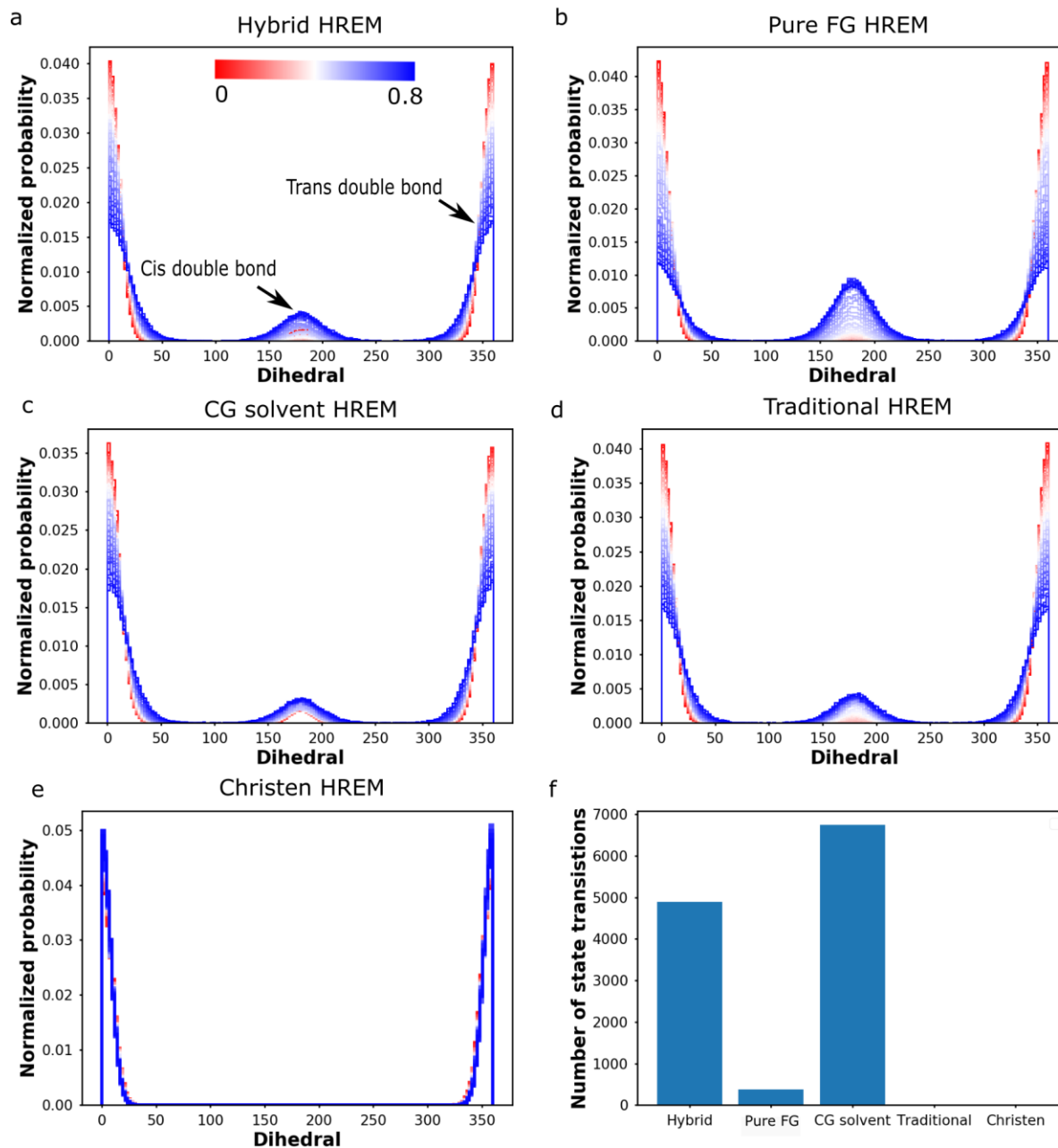


Figure 5. Conformational sampling of a solution of lutein. (a-e) Dihedral distributions for four HREM schemes (our hybrid HREM, pure FG HREM, traditional HREM, CG solvent HREM and Christen HREM). (f) are the number of state (trans/cis) transitions of the five HREM systems at $\lambda=0$. The targeted dihedral is marked in Figure 1b.

Discussion

We describe a novel multiscale approach, based on the HREM algorithm, and we demonstrate it with the efficient sampling of a system of lutein cofactors, containing conjugated bonds with high energy barriers between the possible conformations at atomistic resolution. Our multiscale approach couples the atomistic model to a CG model for which the energy barriers are much lower, allowing efficient sampling of the different conformations. To couple the atomistic and CG resolutions, we introduced two different combination schemes (I, II), one increases sampling efficiency (scheme II), suitable for important targets and the other guarantees accurate CG potential for the solvent (scheme I). We showed that the sampling efficiency of our hybrid HREM scheme is better than other approaches such as pure FG HREM, traditional HREM and Christen HREM.

One of the advantages of using our approach is that it allows having several sampling targets in one set of simulations which can lead to an increase in the sampling. This capacity is absent in other enhanced HREM schemes such as soft-core HREM¹³⁸⁻¹³⁹, which may induce solute overlap leading to unphysical configurations. In our method, the CG potential compensates for the weak FG interactions at high λ (resolution levels toward CG) and prevents unphysical contacts between molecules. Thus, the more physical configurations at high λ can guide and accelerate the sampling for replicas at low λ (resolution levels toward FG) in a more reasonable way. Besides, the CG potential can also compensate for the decreased FG potential at high λ and increase the total potential overlap between replicas (Figure S3).

Note that the CG interactions at high λ cannot accurately predict the FG conformations, especially for scheme II, since CG beads are only the center of mass of their corresponding atoms. Thus, many replicas, which change gradually from one to another, are required to connect these two resolutions. Considering the balance between the cost and efficiency, our approach does not need to go all the way from $\lambda=0$ to $\lambda=1$. This is because the accurate canonical ensemble at $\lambda=0$ is guaranteed by the detailed balance condition and the maximum λ value only affects the sampling efficiency. Therefore, in the current work we use a maximum λ of 0.8. The performance of our HREM scheme can be further boosted in several ways, for instance, it is possible to tune the λ distribution and obtain a more uniform distribution of acceptance ratios between replica pairs which reduces the necessary number of replicas. Another possibility is to use a larger time step at

the coarse-grained level, *e.g.* reversible RESPA¹⁴⁰, since CG-Martini can perform correctly on time steps several-fold larger than 2fs³⁸.

A potential concern is the use of a weak-coupling thermostat in our simulations. For weak-coupling thermostats (*e.g.* Berendsen) that do not produce a proper canonical ensemble, the conformational space distributions are distorted in temperature replica exchange¹⁴¹. However, it is also reported that after increasing the exchange time period and lowering the temperature coupling time in the Berendsen thermostat, the deviation of the potential energy distribution can be counterbalanced¹⁴². In our case, the system is able to sample the correct conformational space (Figure 4a), potential energy distribution (Figure 4b) and thermodynamic property (trans/cis free energy), therefore the choice of thermostat apparently does not lead to noticeable artefacts. It could be that temperature REMD is more sensitive to problems with weak-coupling^{141,142}. Indeed, as far as we know there is no paper discussing the thermostat in HREM, and the Berendsen thermostat is also applied in many other HREM studies¹⁴³⁻¹⁴⁵. To be sure, we also tested other thermostats in our hybrid HREM system: v-rescale¹⁴⁶ and Langevin¹⁴⁷. In the lutein/octane solution, even though the number of trans/cis transitions in the systems coupled with the Langevin thermostat or v-rescale thermostat are lower (Figure S6), the sampling efficiency is still higher than pure FG HREM, traditional HREM and Christen HREM (Figure 5f). This difference in the sampling efficiency is probably caused by the different reactions of thermostats to the solute components. Our hybrid HREM shares many features with the solute tempering replica exchange¹⁴⁸, only the sampled configurations of the important target are increased at high λ , while solvent molecules do not change much. As we used one thermostat group for the whole system, at high λ the effective temperature of the solute (lutein in our case) may increase due to poor coupling with the surrounding solvent. Different thermostats may react differently to this local overheating scenario. A rigorous comparison of different coupling methods is hard since it also depends on how strong the temperature coupling is. Therefore, in hybrid HREM at high λ , the number of trans/cis transitions is different for systems with different thermostats, and thus the sampling efficiency varies.

A shared feature between the Martini and GROMOS force fields is that both are calibrated to reproduce correct free energies (notably partition free energy). Unfortunately, this similarity does not give a great boost in the HREM framework since the exchange criterion is strongly dependent on the structural similarities. Therefore, our scheme will be more efficient if it is

performed on CG and FG force fields, where the configurational properties match better between resolutions, e.g. CG models built through force-matching¹⁴⁹ or iterative Boltzmann inversion approach¹⁵⁰. A related structural incompatibility we encountered in our multiscale simulations, the FG dihedrals between two CG beads (dihedral between C30, C31, C32 and C33 in Figure S5) do not get sampled properly since the stiff CG bond limits the motion of the FG atoms. This problem could be resolved by either neglecting the CG bonded potential in the solute interaction or changing the CG mapping of the solute and placing the sampling target inside a CG bead.

In conclusion, we explored a multiscale HREM scheme based on coupling CG and FG force fields that can provide a substantial speedup of the configurational sampling of a targeted molecule. We demonstrated this method on a system of lutein and octane molecules and obtained the correct distribution of trans/cis conformations of the lutein double bonds. Our scheme opens up a new possibility for enhanced sampling of biomolecules without compromising the accuracy of the FG model.

Supplemental Information

Supporting Methods

Softcore potentials

The nonbonded potentials in both scheme I and II are modified to a softer potential through parameters λ_{sc}^{FG} and λ_{sc}^{CG} for the FG and CG interactions, respectively (Equations 2-4), to avoid sudden changes in the energy which can produce large force. λ_{sc} governs the soft-core interaction by modifying the distance used in the nonbonded potentials, as specified in Equations S1-S3:

$$U_{sc}(\lambda_{sc}, r) = (1 - \lambda_{sc})U(r_A) + \lambda_{sc}U(r_B) \quad (S1)$$

$$r_A = (\alpha\sigma^6\lambda_{sc}^p + r^6)^{1/6} \quad (S2)$$

$$r_B = (\alpha\sigma^6(1 - \lambda_{sc})^p + r^6)^{1/6} \quad (S3)$$

where $U(r)$ is Lennard-Jones (LJ) or electrostatic potentials. σ is the radius of LJ potential. Here we have used $\alpha = 1$, $p = 2$. λ_{sc}^{FG} and λ_{sc}^{CG} are linear functions of the overall multiscaling parameter λ , and are given in Equations S4-S5:

$$\lambda_{sc}^{FG} = 0.5 \times i/(I - 1) \quad (S4)$$

$$\lambda_{sc}^{CG} = 0.5 - 0.5 \times i/(I - 1) \quad (S5)$$

where i is the index of the replica (as defined in Figure 3) and I is the total number of replicas. Figure S1 shows the functional forms of the LJ potential for the range of λ_{sc} used in this work. λ_{sc} ranges from 0 to 0.5. At $\lambda_{sc} = 0$, the hardcore potential is recovered and the potential becomes softer as the λ_{sc} increase.

Octane parameters

For multiscale scheme (scheme I) at $\lambda=1$, the FG interactions within CG beads are retained, while the FG interactions between CG beads are scaled. The FG bond interactions between CG beads are still needed to keep FG atoms together, while the interactions should be weak enough so that they barely affect the CG conformation. Specific parameters are shown in Table S1, octane model is shown in Figure S2.

Table S1. FG octane bonded parameters at $\lambda=1$.

Bond potential	Related atoms	Updated force constant (f_{new})	Original force constant (f_{old})
Bond interaction	C4 and C5	$6000 \times 10^6 \text{ kJ mol}^{-1} \text{ nm}^{-4}$	$715000 \times 10^6 \text{ kJ mol}^{-1} \text{ nm}^{-4}$
Angle interaction	C3, C4 and C5	100 kJ mol^{-1}	530 kJ mol^{-1}
Angle interaction	C4, C5 and C6	100 kJ mol^{-1}	530 kJ mol^{-1}
Dihedral interaction	C3, C4, C5 and C6	3 kJ mol^{-1}	5.92 kJ mol^{-1}

* The rest bond parameters are kept as original GROMOS parameters.

**1-4 nonbonded interaction between C3 and C6 is scaled to 0 at $\lambda=1$, while the rest are kept as original.

*** Corresponding label is found in Figure S2.

**** In Equation 3, a, b and c are computed by $1-f_{\text{new}}/f_{\text{old}}$

***** The topology file is available in the supporting information

Lambda distribution

In HREM simulations of octane the λ is set as: 0.00000 0.04744 0.08445 0.11148 0.14457
0.17725 0.21126 0.24334 0.27433 0.30428 0.33333 0.36155 0.38900 0.41573 0.44179 0.46718
0.49196 0.51610 0.53966 0.56262 0.58499 0.60679 0.62802 0.64867 0.66875 0.68826 0.70720
0.72558 0.74338 0.76061 0.77726 0.79334 0.80884 0.82376 0.83809 0.85183 0.86498 0.87754
0.88950 0.90086 0.91162 0.92177 0.93131 0.94025 0.94857 0.95527 0.96035 0.96582 0.97071
0.97501 0.98033 0.98477 0.98792 0.98990 0.99170 0.99332 0.99474 0.99600 0.99708 0.99800
0.99873 0.99930 0.99970 1.00000

In HREM simulations of fractional lutein and lutein/octane solution the λ is set as: 0.00000
0.01477 0.02954 0.04431 0.05906 0.07381 0.08853 0.10324 0.11792 0.13258 0.14721 0.16181
0.17637 0.19089 0.20537 0.21981 0.23420 0.24853 0.26282 0.27704 0.29121 0.30531 0.31935
0.33331 0.34721 0.36102 0.37476 0.38842 0.40199 0.41547 0.42887 0.44216 0.45537 0.46847
0.48147 0.49437 0.50716 0.51984 0.53240 0.54485 0.55718 0.56938 0.58147 0.59342 0.60525
0.61694 0.62850 0.63993 0.65121 0.66235 0.67335 0.68420 0.69490 0.70545 0.71584 0.72608
0.73616 0.74608 0.75583 0.76542 0.77485 0.78410 0.79319 0.80.

Supporting Figures

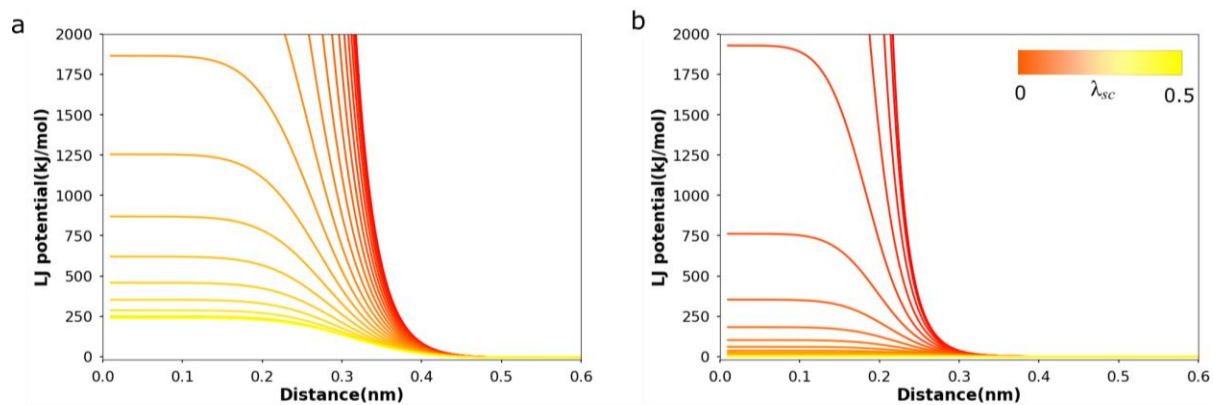


Figure S1. Illustration of softcore Lennard-Jones (LJ) potential. Standard LJ potential is recovered at $\lambda_{sc} = 0$ and the softness increase with the increase of λ_{sc} . (a) and (b) represent CG and FG potential, respectively.

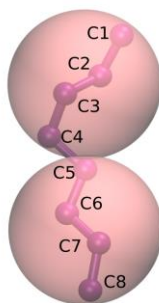


Figure S2. The multiscale model of octane.

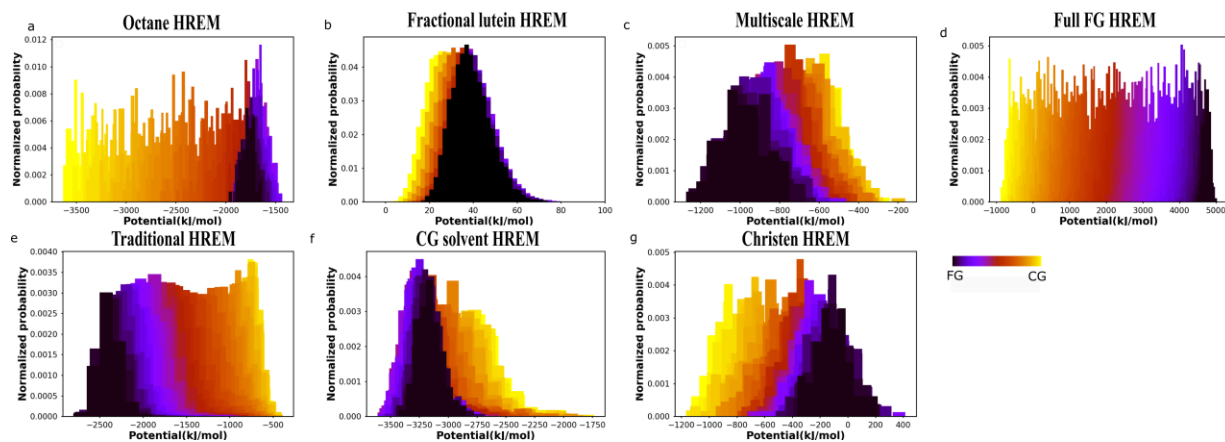


Figure S3. Potential distributions for replica exchange systems. (a) and (b) represent the total potential distributions for octane and lutein fraction system. (c) to (g) illustrate the total potentials of lutein/octane solution applying multiscale HREM, pure FG HREM, traditional HREM, CG solvent HREM and Christen HREM, respectively. The black color represents the replica close to CG side and yellow represents the FG side.

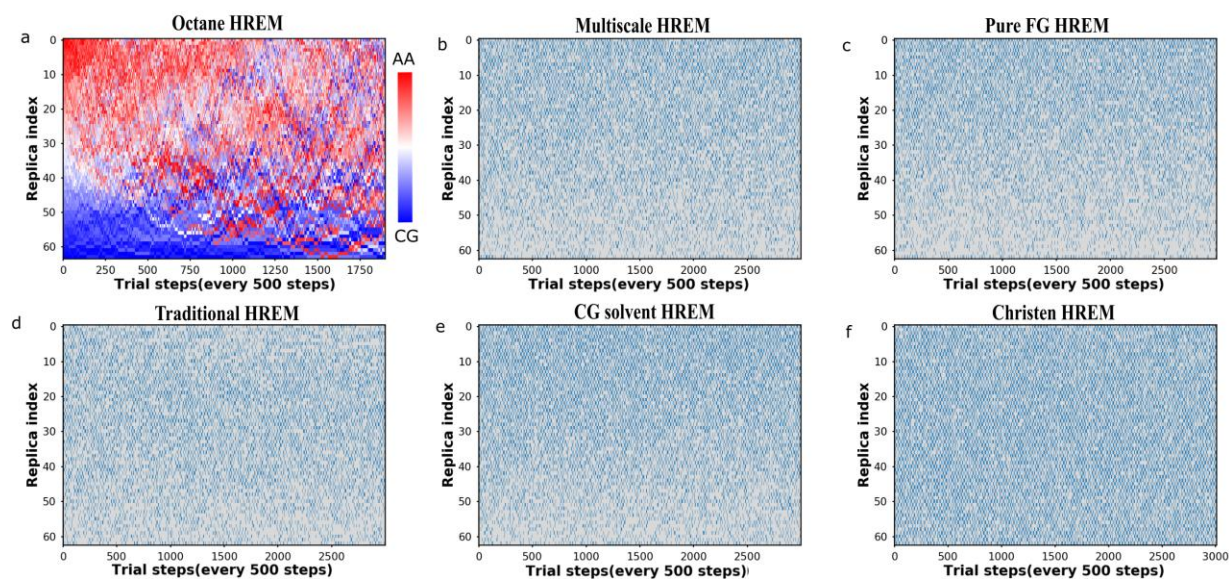


Figure S4. Conformation exchanges in HREM framework. (a) illustrates the octane solution system applying multiscale HREM. The replica index as a function of trial steps are shown. (b) to (f) illustrate the lutein/octane solution systems applying multiscale HREM, pure FG HREM, traditional HREM, CG solvent HREM and Christen HREM, respectively. The conformation exchanges as a function of trial steps are shown. If an exchange is accepted, a blue spot is marked.

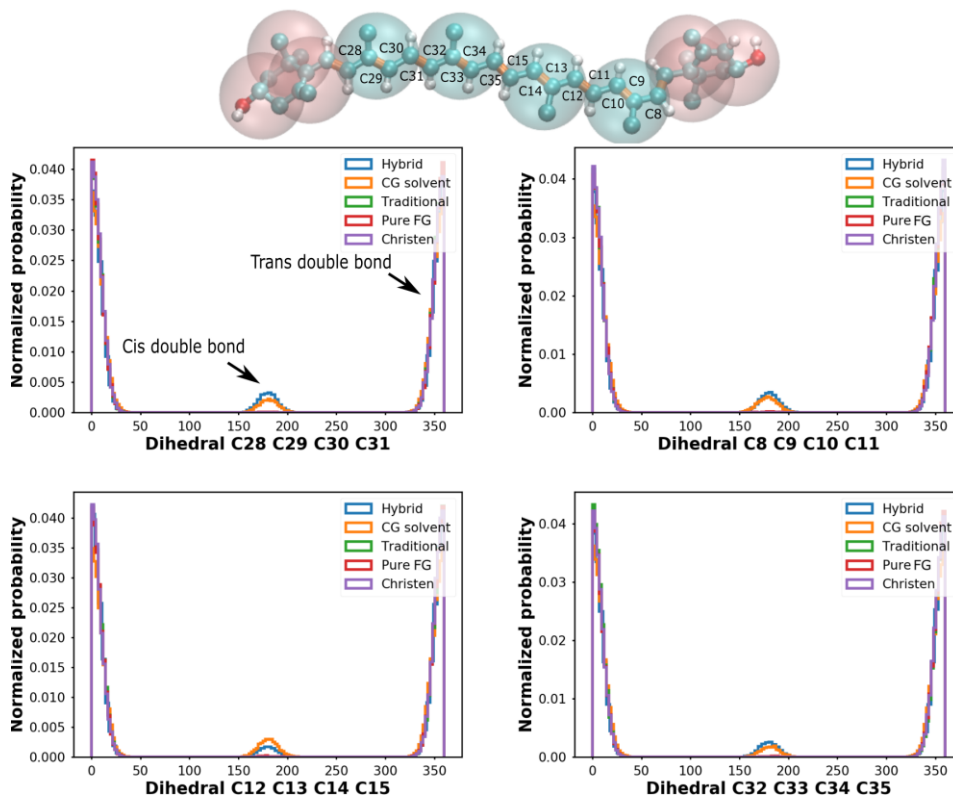


Figure S5. Dihedral distribution for five HREM systems at $\lambda=0$. The hybrid lutein model and corresponding labels are shown in the inset plot and the double bonds are expressed in yellow.

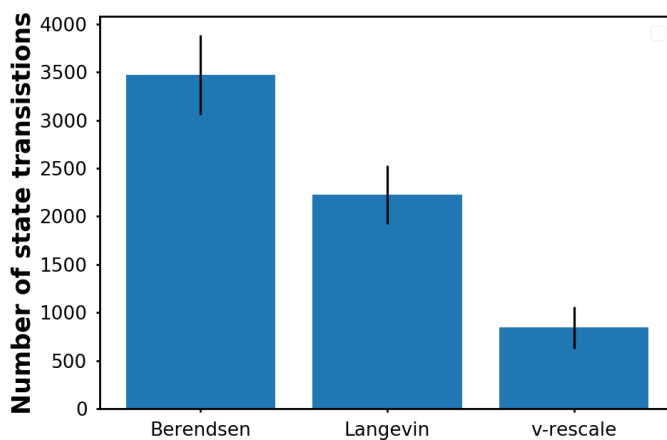


Figure S6. The state transition number of lutein in solution. The number of state (trans/cis) transitions are shown for the multiscale HREM systems at $\lambda=0$ coupled with different thermostats. The targeted dihedrals are the four dihedrals computed in Figure S5. The error bars are the standard errors of the transition numbers of the four dihedrals.