

Supporting Information

Adaptive, geometric networks for efficient coarse-grained *ab initio* molecular dynamics with post-Hartree-Fock accuracy

Timothy C. Ricard, Cody Haycraft and Srinivasan S. Iyengar *

*Department of Chemistry and Department of Physics,
Indiana University, 800 E. Kirkwood Ave, Bloomington, IN-47405*

(Dated: April 13, 2018)

* Email: iyengar@indiana.edu

SI-1. BENCHMARKING $E_{network}$: ABSOLUTE ENERGIES FOR α -HELIX, 3_{10} -HELIX, EXTENDED β -STRAND AND β -STRAND

The extensive nature of the errors in Figures 6 and 8 of the paper are complemented here through per-residue errors in Figures SI-1 and SI-2. Errors in absolute energy are similarly presented in Figures SI-3 and SI-4., we present here the errors for absolute energies per residue. A comparison of using Psi4 for MP2 as opposed to Gaussian09 for MP2 within the coarse-grained approach is presented in Figure SI-5.

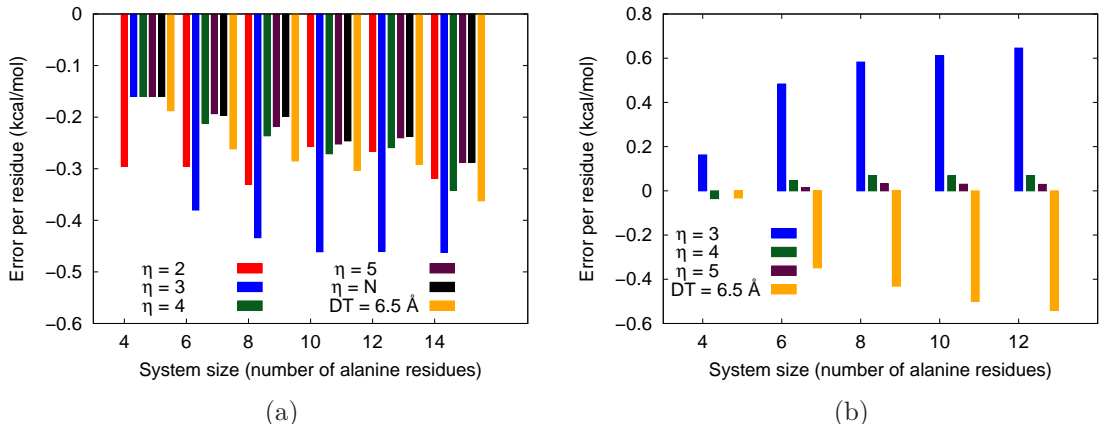


FIG. SI-1: These figures complement Figure 6 in the main paper and present the appropriate per residue errors for stabilization energy and are well within chemical accuracy. B3LYP:PM6 calculations (a) used 6-31++G(d,p) basis and MP2:B3LYP calculations (b) used 6-31+G(d,p) basis. Also note that these CG-calculations are much faster as compared to the respective full system B3LYP and MP2 calculations as seen in Figure 7 of the paper.

SI-2. BENCHMARKING THE COARSE-GRAINED *AB INITIO* MOLECULAR DYNAMICS TRAJECTORIES

Both CG-BOMD and CG-ADMP-pHF trajectories were computed by integrating the equations of motion using velocity Verlet¹ integration with a variety of step sizes with simulation details presented in Tables SI-I, SI-II, SI-III and SI-IV. Four different sets of simulations were performed and these included Ala_3 and Ala_4 with initial structures chosen to be the optimized β -strand and 3_{10} helical conformations. In Table SI-I, we summarize the simulation parameters and energy conservation data for those that use the 3_{10} Ala_3 structure as initial conformation. We further vary η , the simulation time-step and the ini-

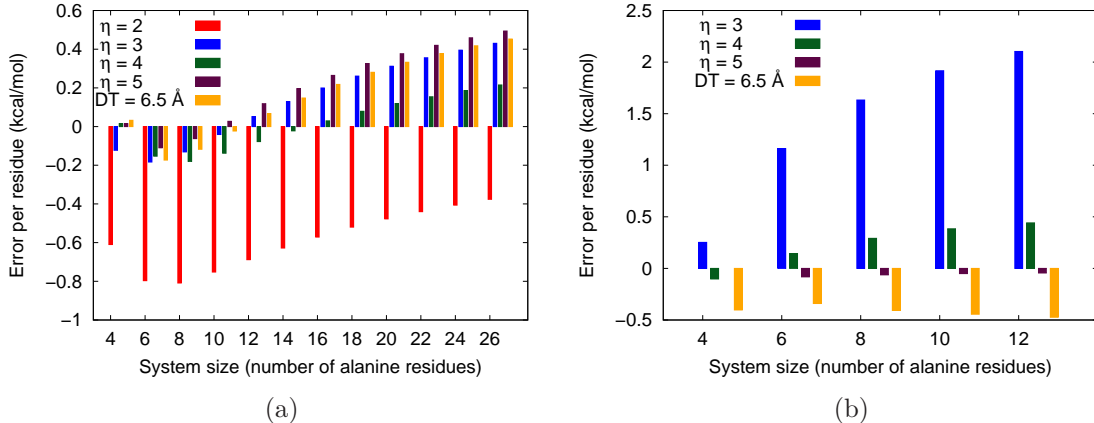


FIG. SI-2: These figures complement Figure 8 in the main paper and present the appropriate per residue values that are well within chemical accuracy. Figure (a): B3LYP:PM6 calculations with 6-31++G(d,p) basis and Figure (b): MP2:B3LYP calculations with 6-31+G(d,p) basis. Also note that these CG-calculations are much faster as compared to the respective full system B3LYP and MP2 calculations as seen in Figure 7 of the paper.

tial nuclear kinetic energy and compare the results from CG-AIMD calculations with those obtained from simulations where the full system is treated at the higher level of theory. As expected, the energy conservation and drift are adversely affected with increased time-step. The analysis of conservation for the simulations that use the Ala_3 - β -strand structure, the Ala_4 - 3_{10} helical structure, and the Ala_4 - β -strand structure as initial geometric configurations are respectively presented in Tables SI-II, SI-III and SI-IV. In these tables the value of η was maintained to include all possible timer configurations to enhance accuracy of the simulations. Furthermore, time-step was also maintained at 0.25fs. In all cases the energy conservation and drift are well within the acceptable range and hence we next gauge the agreement between these trajectories by computing the vibrational density of states as a function of frequency.

The level of hybrid theory used for these calculations include B3LYP/6-31++G(d,p):PM6, and MP2/6-31+G(d,p):B3LYP/6-31+G(d,p). For comparison full system trajectories are computed at the respective higher levels of theory, that is, MP2/6-31+G(d,p) and B3LYP/6-31++G(d,p). These results are used to benchmark the accuracy and efficiency of our CG-AIMD results. Here we present the vibrational density of states for the other trajectories which were not included in the article. Figures SI-6 and SI-7 presents the vibrational density of states for Ala_3 at a higher temperature for B3LYP and MP2 levels of electronic structure respectively. The trajectories which started as β strand have their

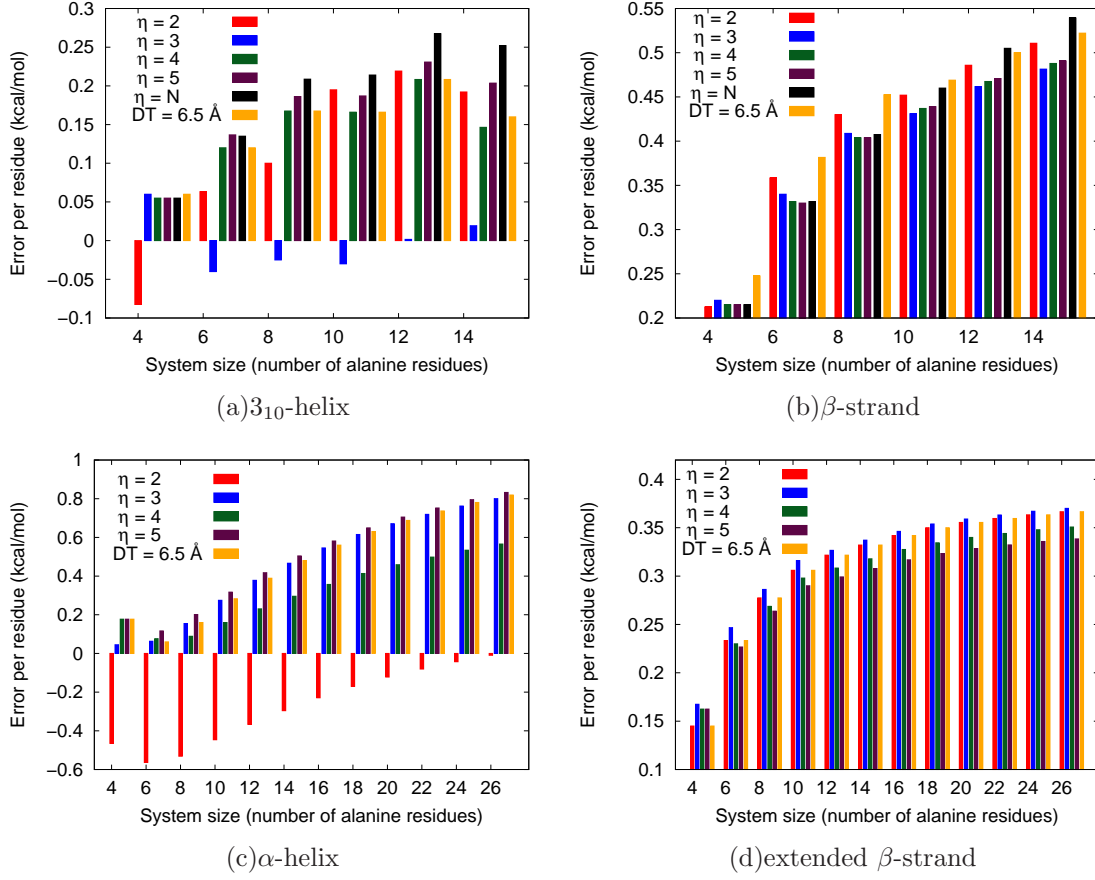


FIG. SI-3: Errors in absolute energies: B3LYP as compared to B3LYP:PM6.

vibrational spectra shown is Figures SI-8 and SI-9. The Pearson correlation coefficient² (ρ) discussed in the main article:

$$\rho = \sqrt{\frac{\mathbf{I}_{V,BOMD} \cdot \mathbf{I}_{V,CG}}{\|\mathbf{I}_{V,BOMD}\| \|\mathbf{I}_{V,CG}\|}} \quad (1)$$

is noted below each CG-AIMD spectral curve.

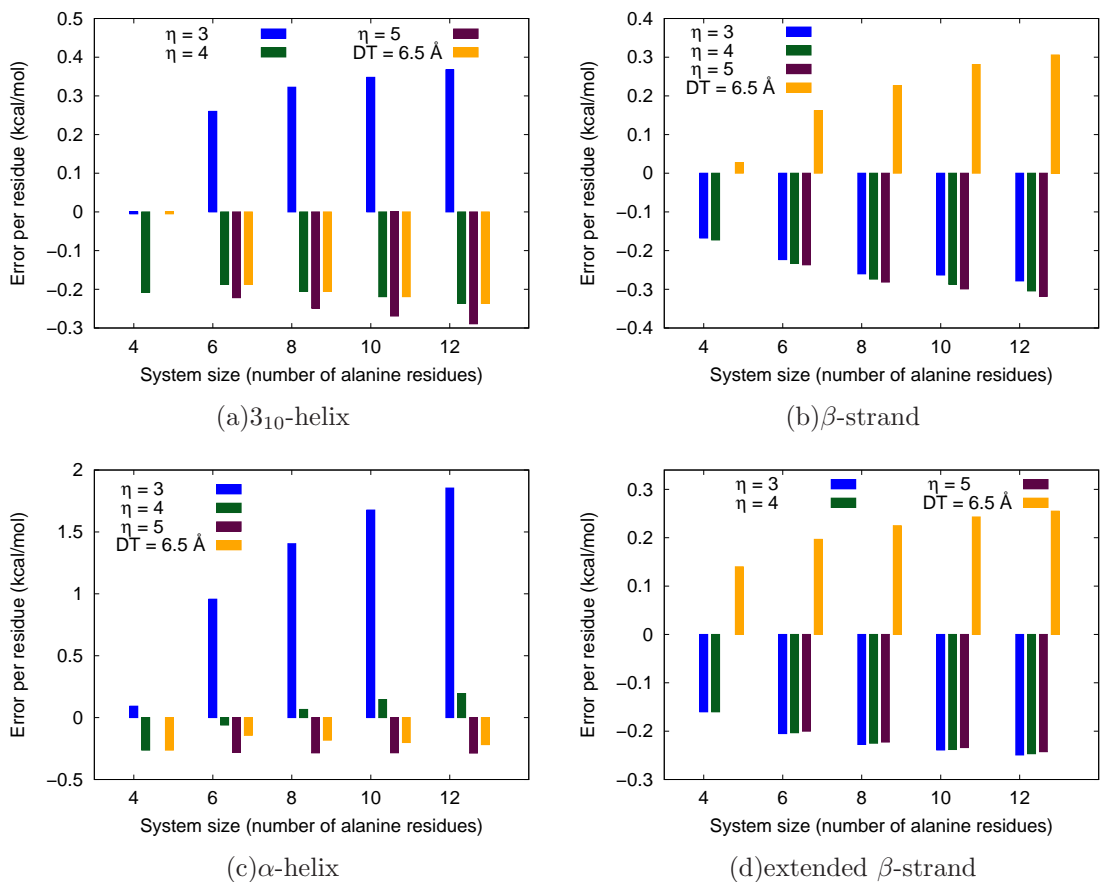


FIG. SI-4: Errors in absolute energies: MP2 as compared to MP2:B3LYP.

SI-3. ACKNOWLEDGMENTS

This research is supported by the National Science Foundation grant NSF CHE-1665336 to SSI.

¹ Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *J. Chem. Phys.* **1982**, 76, 637.

² Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press, New York, 1992.

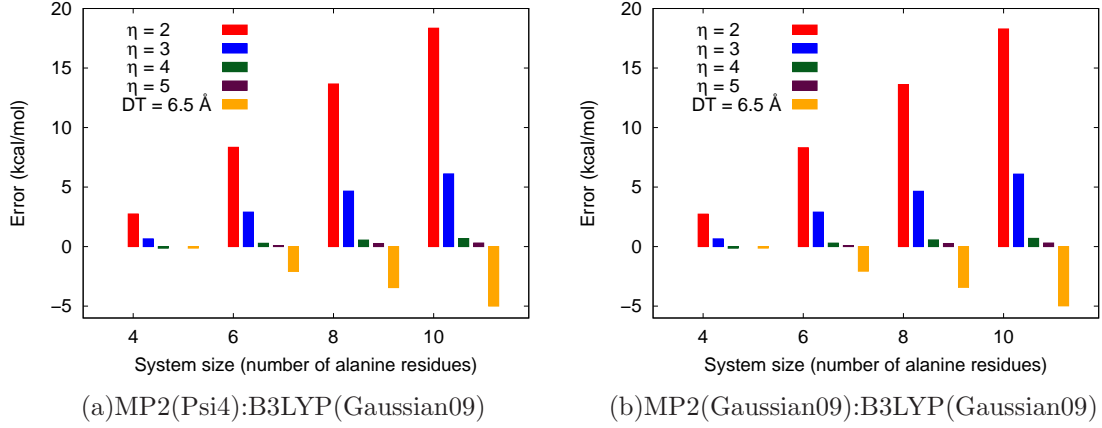
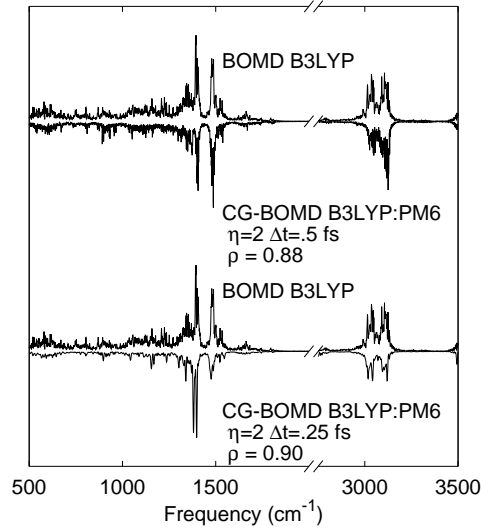


FIG. SI-5: As a complement to Figures 6(b) and 8(b) in the paper, we present here a comparison of errors when the higher level calculation is performed with Psi4 or Gaussian09. While there is no significant difference in accuracy, this does provide us with additional flexibility given the wide range of options available in various individual electronic structure packages.



(a) 3_{10} helical Ala_3

FIG. SI-6: Vibrational density of states calculated from B3LYP level dynamical trajectories for helical initial conformations of tri-alanine (a) at elevated temperatures. We present the full system, with a positive amplitude, for each of the coarse-grained dynamics (CG), with a negative amplitude, for ease of comparison.

TABLE SI-I: Energy conservation properties for dynamical simulations with initial conditions corresponding to the 3_{10} Ala_3 structure. All simulations are micro-canonical.

Theory	η	Total Simulation Time ^a	Δt ^b	$\langle K \rangle$ ^c	$\Delta \mathcal{H}$ ^d	\mathcal{H}_{Drift} ^e
BOMD B3LYP		10.00	.25	340.42K+/37.63K	0.02	0.01
CG-BOMD B3LYP:PM6	3	10.00	.25	335.79K+/36.25K	0.11	0.09
CG-BOMD B3LYP:PM6	3	10.00	.5	327.89K+/35.93K	0.06	-0.06
BOMD B3LYP		15.00	.25	416.62K+/44.38K	0.03	0.08
CG-BOMD B3LYP:PM6	3	27.28	.5	413.91K+/43.66K	0.18	0.14
CG-BOMD B3LYP:PM6	2	5.32	.25	416.22K+/47.78K	0.06	0.18
CG-BOMD B3LYP:PM6	2	10.00	.5	415.88K+/45.97K	0.11	0.22
CG-BOMD B3LYP:PM6 ^f	2	7.96	.25	410.20K+/44.21K	0.09	0.27
BOMD MP2		2.50	.25	345.83K+/39.06K	0.01	-0.00
CG-BOMD [MP2(Gaussian09)]:B3LYP	3	2.45	.25	343.73K+/40.15K	0.07	0.18
CG-BOMD [MP2(Psi4)]:B3LYP	3	2.50	.25	344.02K+/39.52K	0.07	0.15
CG-ADMP-pHF MP2:B3LYP ^g	3	2.60	.25	322.17K+/35.18K	0.07	0.14
BOMD MP2		3.18	.25	429.23K+/51.03K	0.01	0.00
CG-BOMD MP2:B3LYP	2	5.55	.5	423.09K+/47.53K	0.43	0.05
CG-BOMD MP2:B3LYP ^f	2	4.08	.25	418.50K+/48.05K	0.06	-0.11

^atimes reported in picoseconds

^btimes steps reported in femtoseconds

^cAverage nuclear kinetic energy noted in units of Kelvin. This value also determines the extent of conformational sampling. The Kelvin value for the average nuclear kinetic energy is determined assuming equipartition theorem, $\frac{3}{2}(N-1)kT$. The initial kinetic energy, which was randomly distributed, to obtain an average value corresponding to approximately 300K

^dRMS deviation of total energy in kcal/mol. The total energy reflects the total Hamiltonian for the system

^eThe drift in total energy, is obtained as the difference between the average total energies for the first and last 100fs of the dynamics data (in kcal/mol)

^f*CG-node* is defined as the amino acid residue, thus the peptide bond is broken

^gThe fictitious inertia tensor, $\mu_{valence}$, is 180au. The valence orbitals are mass-weight with $\mu_{valence}$ whereas the core orbitals are weighted based on their respective Fock matrix elements.

TABLE SI-II: Energy conservation properties for dynamical simulations with initial conditions corresponding to the β -strand Ala_3 structure. All simulations are micro-canonical

Theory	η	Total Simulation Time ^a Time ^f	Δt^b	$\langle K \rangle^c$	$\Delta \mathcal{H}^d$	\mathcal{H}_{Drift}^e
BOMD B3LYP		15.00	.25	328.05K+/36.14K	0.06	0.12
CG-BOMD B3LYP:PM6	3	10.01	.25	328.66K+/38.73K	0.14	-0.41
BOMD MP2		4.25	.25	335.12K+/38.88K	0.02	0.00
CG-BOMD [MP2(Gaussian09):B3LYP	3	1.37	.25	333.81K+/44.14K	0.03	0.07
CG-BOMD [MP2(Psi4)]:B3LYP	3	2.67	.25	334.25K+/41.94K	0.04	0.11
CG-ADMP-pHF MP2:B3LYP ^g	3	0.92	.25	303.87K+/43.30K	0.04	-0.01

^atimes reported in picoseconds

^btimes steps reported in femtoseconds

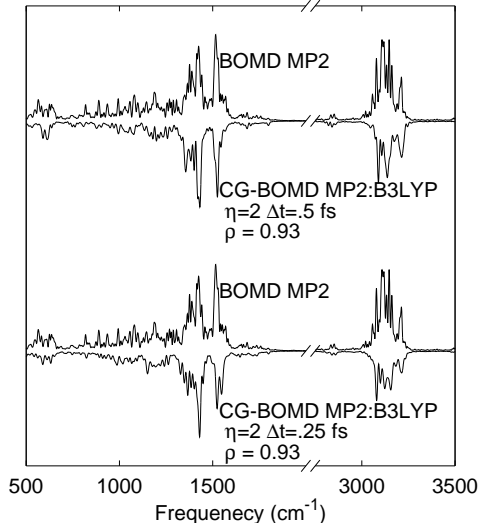
^cAverage nuclear kinetic energy noted in units of Kelvin. This value also determines the extent of conformational sampling. The Kelvin value for the average nuclear kinetic energy is determined assuming equipartition theorem, $\frac{3}{2}(N-1)kT$. The initial kinetic energy, which was randomly distributed, to obtain an average value corresponding to approximately 300K

^dRMS deviation of total energy in kcal/mol. The total energy reflects the total Hamiltonian for the system

^eThe drift in total energy, is obtained as the difference between the average total energies for the first and last 100fs of the dynamics data (in kcal/mol)

^ftimes reported in picoseconds

^gThe fictitious inertia tensor, $\mu_{valence}$, is 180au. The valence orbitals are mass-weight with $\mu_{valence}$ whereas the core orbitals are weighted based on their respective Fock matrix elements.



(a) 3_{10} helical Ala_3

FIG. SI-7: Vibrational density of states calculated from MP2 level dynamical trajectories for helical initial conformations of tri-alanine (a) at elevated temperatures. We present the full system, with a positive amplitude, for each of the coarse-grained dynamics (CG), with a negative amplitude, for ease of comparison.

TABLE SI-III: Energy conservation properties for dynamical simulations with initial conditions corresponding to the 3_{10} Ala_4 structure. All simulations are micro-canonical

Theory	η	Total Simulation Time ^a	Δt ^b	$\langle K \rangle$ ^c	$\Delta \mathcal{H}$ ^d	\mathcal{H}_{Drift} ^e
BOMD B3LYP		11.27	.25	331.66K+/31.68K	0.02	0.00
CG-BOMD B3LYP:PM6	4	15.00	.25	326.98K+/30.16K	0.18	0.34
BOMD MP2		1.32	.25	329.84K+/36.30K	0.02	-0.00
CG-BOMD [MP2(Gaussian09)]:B3LYP	4	1.99	.25	329.43K+/32.97K	0.07	0.03
CG-BOMD [MP2(Psi4)]:B3LYP	4	2.50	.25	330.44K+/32.08K	0.07	0.00
CG-ADMP-pHF MP2:B3LYP ^f	4	1.12	.25	329.84K+/36.30K	0.02	-0.00

^atimes reported in picoseconds

^btimes steps reported in femtoseconds

^cAverage nuclear kinetic energy noted in units of Kelvin. This value also determines the extent of conformational sampling. The Kelvin value for the average nuclear kinetic energy is determined assuming equipartition theorem, $\frac{3}{2}(N-1)kT$. The initial kinetic energy, which was randomly distributed, to obtain an average value corresponding to approximately 300K

^dRMS deviation of total energy in kcal/mol. The total energy reflects the total Hamiltonian for the system

^eThe drift in total energy, is obtained as the difference between the average total energies for the first and last 100fs of the dynamics data (in kcal/mol)

^fThe fictitious inertia tensor, $\mu_{valence}$, is 180au. The valence orbitals are mass-weight with $\mu_{valence}$ whereas the core orbitals are weighted based on their respective Fock matrix elements.

TABLE SI-IV: Energy conservation properties for dynamical simulations with initial conditions corresponding to the β -strand Ala_4 structure. All simulations are micro-canonical

Theory	η	Total Simulation Time ^a	Δt ^b	$\langle K \rangle$ ^c	$\Delta \mathcal{H}$ ^d	\mathcal{H}_{Drift} ^e
BOMD B3LYP		13.64	.25	312.04K+/29.88K	0.02	-0.01
CG-BOMD B3LYP:PM6	4	10.99	.25	311.39K+/30.58K	0.20	0.66
BOMD MP2		1.01	.25	319.87K+/34.97K	0.02	0.00
CG-BOMD [MP2(Gaussian09)]:B3LYP	4	2.39	.25	318.49K+/32.72K	0.07	0.03
CG-BOMD [MP2(Psi4)]:B3LYP	4	0.92	.25	318.36K+/36.42K	0.07	0.14
CG-ADMP-pHF MP2:B3LYP ^f	4	0.98	.25	292.18K+/30.14K	0.13	0.34

^atimes reported in picoseconds

^btimes steps reported in femtoseconds

^cAverage nuclear kinetic energy noted in units of Kelvin. This value also determines the extent of conformational sampling. The Kelvin value for the average nuclear kinetic energy is determined assuming equipartition theorem, $\frac{3}{2}(N-1)kT$. The initial kinetic energy, which was randomly distributed, to obtain an average value corresponding to approximately 300K

^dRMS deviation of total energy in kcal/mol. The total energy reflects the total Hamiltonian for the system

^eThe drift in total energy, is obtained as the difference between the average total energies for the first and last 100fs of the dynamics data (in kcal/mol)

^fThe fictitious inertia tensor, $\mu_{valence}$, is 180au. The valence orbitals are mass-weight with $\mu_{valence}$ whereas the core orbitals are weighted based on their respective Fock matrix elements.

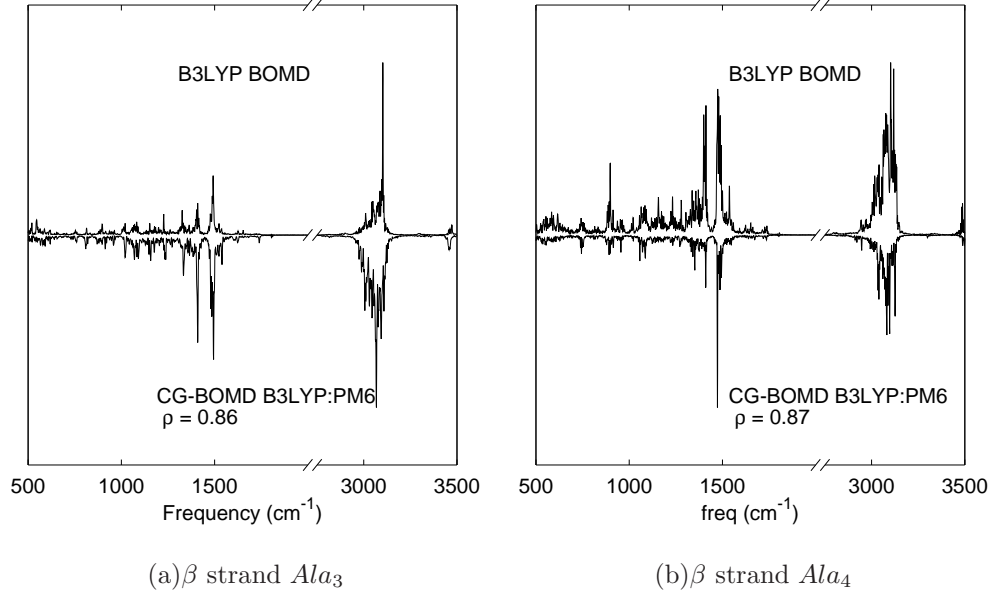


FIG. SI-8: Vibrational density of states calculated from B3LYP level dynamical trajectories for β strand initial conformations of tri-alanine (a) and tetralanine (b). We present the full system, with a positive amplitude, for each of the coarse-grained dynamics (CG), with a negative amplitude, for ease of comparison.

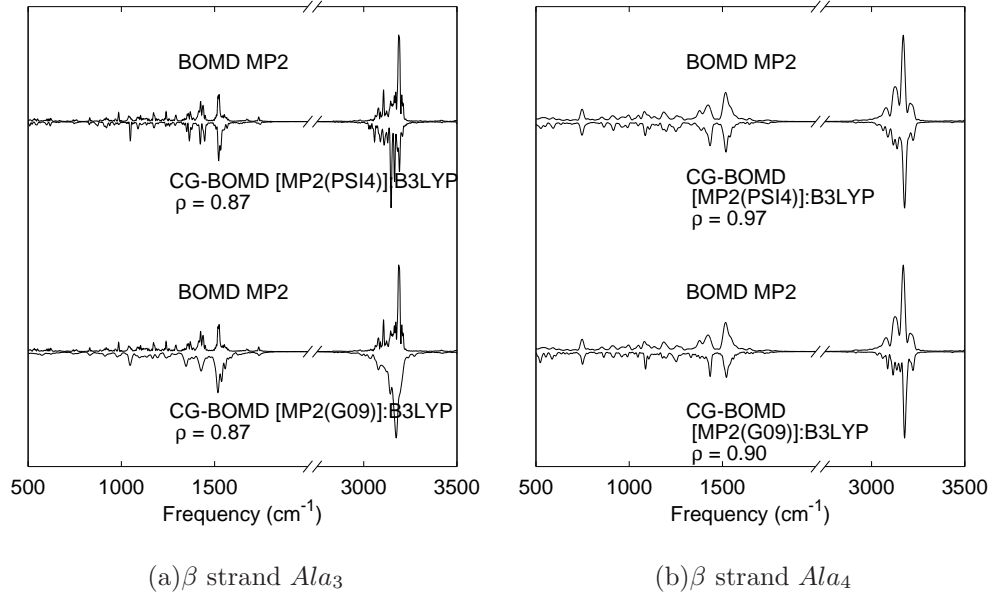


FIG. SI-9: Vibrational density of states calculated from MP2 level dynamical trajectories for β strand initial conformations of tri-alanine (a) and tetralanine (b). We present the full system, with a positive amplitude, for each of the coarse-grained dynamics (CG), with a negative amplitude, for ease of comparison.