"Gold-Standard" Δ-Machine Learned and Transferable Potential for Linear Alkanes

Chen Qu,[*,†] Thomas C. Allison,[‡] Apurba Nandi,[*,¶] Paul L. Houston,[∥] Qi Yu,[⊥] Riccardo Conte,[#] and Joel M. Bowman[*,@]

†Independent Researcher, Toronto, Ontario M9B0E3, Canada

‡National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

¶Department of Physics and Materials Science, University of Luxembourg, L-1511, Luxembourg City, Luxembourg.

§Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy

∥Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, USA and Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

⊥Department of Chemistry, Fudan University, Shanghai, 200438, P. R. China

#Dipartimento di Chimica, Università Degli Studi di Milano, via Golgi 19, 20133 Milano, Italy

@Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, USA.

E-mail: szquchen@gmail.com; apurba.nandi@uni.lu; jmbowma@emory.edu

1

Abstract

The conformational properties of linear alkanes, C_nH_{2n+2}, have been of intense interest for many years. Experiments and corresponding electronic structure calculations were first reported in the mid-2000s and continue to the present time. These focus on the minimum chain length where the transition from the linear minimum to the hairpin minimum occurs. We recently reported a transferable many-body permutationally invariant polynomial (MB-PIP) for linear alkanes using B3LYP electronic energies, which do not account for dispersion. Here we report a Δ-ML approach to elevate this B3LYP-based and new PBE0+MBD MB-PIP potentials using PNO-LCCSD(T)-F12 energies. The new Δ-corrected potentials predict the difference in these minima accurately, compared to benchmark CCSD(T) results, over the range $C_{12}H_{28}$ to $C_{28}H_{58}$. Vibrational power spectra are also reported for $C_{14}H_{30}$ and $C_{30}H_{62}$ using the uncorrected and Δ-ML B3LYP. These new PIP-MB potentials for linear alkanes are the most accurate ones currently available and can be used in studies of properties of linear alkanes.

Hydrocarbons are important in fuels, plastics, and other industrial products. Consequently, there have been many studies, not only of their chemistry, but also of their physical and molecular properties. The conformational properties of linear alkanes, C_nH_{2n+2}, have been of particular interest for many years. Early studies of hydrocarbons have been reviewed by Patterson.[1] Molecular mechanics models have been developed by Allinger,[2] and by Jorgenson,[3–5] among others.

Experiments and corresponding electronic structure calculations were reported in the mid-2000s, with a focus on the minimum chain length where the transition from the linear minimum to the hairpin minimum occurs.[6–8] Theoretical electronic structure work continues to the present. The paper by Liakos and Neese[8] is a recent and comprehensive study of the linear and hairpin minima using high-level *ab initio* theory, namely Domain Based Pair Natural Orbital Coupled Cluster theory (DLPNO-CCSD(T)). Rather than repeat the excellent background given there, we simply refer the interested reader to that paper. A major point of this careful study was the quantitative examination of the difference in geometry of these two minima and the resultant sensitivity of the energy difference. This was examined for alkanes ranging from C_6H_{14} to $C_{18}H_{38}$, with the conclusion that the hairpin minimum drops below the "extended" linear one at $C_{16}H_{34}$ or $C_{17}H_{36}$. This conclusion is in accord with earlier ones using lower-levels of electronic structure theory.[6,7] A final point about these "single-point" studies is that the importance of dispersion was noted by Bartlett and co-workers who wrote "As the length of unbranched alkane chains reaches some critical length, intramolecular dispersion forces cause a self-solvation effect in which the chains assume a folded conformation."[7] More recent, extended computational studies of linear alkane conformational energies by Ehlert et al.[9] created the ACONFL dataset of energies. Subsequent work using this dataset was reported by Santra and Martin[10] and most recently in 2023 by Werner and Hansen.[11] The carefully benchmarked PNO-LCCSD(T)-F12 approach used by Werner and Hansen is used here (see below for details). Figure 1 depicts the extended and hairpin conformers of alkanes, exemplified for $C_{14}H_{30}$.

Linear, Extended

Hairpin

Figure 1: Linear extended and hairpin structures for the alkane $C_{14}H_{30}$

Since 2023, there has been substantial progress in methodology in developing machine-learned potentials; for a recent review see ref. 12. With respect to linear alkanes, we recently reported permutationally invariant polynomial (PIP) machine-learned potentials for $C_{14}H_{30}$ from precise training of roughly 250 000 B3LYP electronic energies. One of those potentials was based on a novel many-body "targeted" approach, denoted MB-PIP, that we showed is transferable to general linear alkanes.[13] It was recently used in a study of the folding landscape.[14]

This MB-PIP transferable potential is the focus of work here. Specifically, we report a Δ-ML approach to elevate the previous B3LYP-based and new PBE0+MBD MB-PIP potentials using 4515 PNO-LCCSD(T)-F12 energies.[15] The new Δ-ML potentials provide energy differences between the hairpin and linear minima that are compared to directly calculated PNO-LCCSD(T)-F12 results for alkanes $C_{12}H_{26}$ to $C_{28}H_{58}$.

The paper is organized as follows. First, we give a brief review of many-body approaches in machine-learned potentials, including those based on PIP regression. Then, details of the approach taken for the linear alkanes are reviewed followed by details of the Δ-ML approach. The results of the approach are given and assessed by comparing the energy difference between the hairpin and linear minima against the benchmark numbers. Power spectra of $C_{14}H_{30}$ are presented using B3LYP and Δ-B3LYP Potentials at two total energies corresponding to temperatures of 10 and 300 K.

Many-body approaches to machine learned potentials of large molecules, clusters, and condensed phases are becoming the standard approach; however, the details vary greatly depending on the method. For a recent review including a list of more than 90 open-source software programs, see reference 12. Among this large group of approaches, permutationally invariant polynomial regression has played an important role, both historically and also very recently.[16] This is reviewed briefly below which concludes with the details of a new Δ-machine learning approach for our MB-PIP approach. That approach is applied to two DFT-based MB-PIP potentials for linear alkanes.[13,17,18]

Many-body representations of potentials are generically given by

$$V(1, \cdots, N) = \sum_{i=1}^{N} V_{1-b}(i) + \sum_{i>j}^{N} V_{2-b}(i,j) + \sum_{i>j>k}^{N} V_{3-b}(i,j,k) + \sum_{i>j>k>l}^{N} V_{4-b}(i,j,k,l) + \cdots, \quad (1)$$

This representation is well-known for non-covalent interactions, e.g., for water, where each "body" is a water monomer.[19-23] In these widely-known potentials each n-body interaction is fit to gold-standard CCSD(T) energies using a basis of PIPs. The most recent of these, q-AQUA,[21] q-AQUA-pol,[22] and MB-pol(2023)[23] fit up to 4-body interactions.

For covalent interactions, i.e., for a single (large) molecule, MB-PIP approaches are just appearing in the literature. In this case, there isn't an obvious partitioning in terms of "bodies". Paesani and co-workers recently suggested a physically based partitioning of linear hydrocarbons and developed a transferable PIP-based model.[24] Shortly after, we proposed and tested an atom-based PIP-approach that is different from – albeit inspired by – a PIP representation of an atom-centered expansion approach exemplified by work of Csanyi and co-workers.[25] In our approach[17,18] the monomers are atoms, and these interactions are enumerated for all atoms in the molecule. For an n-body interaction among N atoms, there are a total of $N!/[n!(N-n)!]$ interactions. To limit the number of interactions, physical range cutoffs are used for each n-body interaction, and details can be found in references 13,17 which report our first application to the 44-atom alkane $C_{14}H_{30}$ and transferability to alkanes as large as $C_{30}H_{62}$, respectively. In this example (and also for other two-element molecules, clusters, materials, etc.,) there are three types of 2-b interactions, i.e., CC, HH, and CH interactions, four types of 3-b, and five types of 4-b interactions. (We do not consider 5-b interactions.) For each interaction beyond 2-b, an appropriate PIP basis is used and there are fewer of these than the types of interactions. To see this, consider the 3-b CCC and HHH interactions. While these are different, the PIP basis is the same, namely the basis for A_3 but the corresponding sets of linear coefficients are different. The full expression, up to V_{4-b}, for this MB-PIP representation has been given previously;[13] here we just illustrate

this for V_{3-b}.

$$
\begin{aligned}
V_{3-b} =& \sum_{C_i,C_j,C_k} V_{3-b}^{(CCC)}(y_{ij}, y_{ik}, y_{jk}) + \sum_{C_i,C_j,H_k} V_{3-b}^{(CCH)}(y_{ij}, y_{ik}, y_{jk}) + \\
& \sum_{C_i,H_j,H_k} V_{3-b}^{(CHH)}(y_{ij}, y_{ik}, y_{jk}) + \sum_{H_i,H_j,H_k} V_{3-b}^{(HHH)}(y_{ij}, y_{ik}, y_{jk}) \\
=& \sum_m \sum_{C_i,C_j,C_k} c_m^{(CCC)} p_m^{(A_3)}(y_{ij}, y_{ik}, y_{jk}) + \sum_m \sum_{C_i,C_j,H_k} c_m^{(CCH)} p_m^{(A_2B)}(y_{ij}, y_{ik}, y_{jk}) + \\
& \sum_m \sum_{C_i,H_j,H_k} c_m^{(CHH)} p_m^{(A_2B)}(y_{jk}, y_{ij}, y_{ik}) + \sum_m \sum_{H_i,H_j,H_k} c_m^{(HHH)} p_m^{(A_3)}(y_{ij}, y_{ik}, y_{jk}),
\end{aligned}
\tag{2}
$$

where now the permutational symmetry of the two 3-b PIPs are A_3 and A_2B.

For the expansion up to 4-b terms, the general permutational symmetry of the PIPs bases are A_4, A_3B, and A_2B_2. Also, we note that the distinct 2-b, 3-b, and 4-b PIP bases can have different range parameters for the corresponding Morse variables, and different maximum polynomial orders. Once these bases are set up, the linear coefficients are determined using a standard over-determined linear least-squares fit to the total energies of this molecule.

As shown previously, this MB-PIP is transferable to linear alkanes and this was demonstrated for alkanes with as many as 30 carbon atoms.[26] We also note that the MB-PIP approach is to be distinguished from fragmentation representations[27] and the use of a fragmented (PIP) basis.[28] This approach was used successfully to represent the PES for $C_{14}H_{30}$.[17]

This MB-PIP potential was fit to a dataset of roughly 250 000 B3LYP energies for $C_{14}H_{30}$; details of the dataset and fitting are given elsewhere.[13,17] We note here and also in a very recent paper that dispersion is not accounted for in these straight B3LYP energies. So the central objective of the work here is to correct this in a single step by using Δ-ML at the CCSD(T) level. In addition, the above expression for the MB-PIP is employed here for a new set of PBE0+MBD energies,[29–31] computed with the "intermediate" basis settings of the FHI-aims electronic structure package,[32,33] for the same dataset of configurations. Since this approach does contain dispersion, Δ-ML applied to this PES is not as heavy a "lift".

The Δ-ML approach is given by the general equation[34,35]

$$V_{LL \to CC} = V_{LL} + \Delta V_{CC-LL}, \tag{3}$$

where $V_{LL \to CC}$ is the corrected PES, V_{LL} is a PES fit to low-level DFT electronic data, and ΔV_{CC-LL} is the correction PES based on high-level (usually coupled cluster) energies. In the present context V_{LL} is the above MB-PIP fit to B3LYP or PBE0+MBD energies and ΔV_{CC-LL} is the analogous fit to the difference between high-level and DFT energies. Here we apply PNO-LCCSD(T)-F12 energies, implemented in Molpro,[36][†] to obtain the correction term ΔV_{CC-LL}.

In general, the expectation is that the 4-b interactions are shorter-ranged than 3-b than 2-b, so we applied smaller Morse range parameter and cutoff distance for 4-b, and larger values for 2-b. The 1-b energy is simply the sum of the energies of isolated carbon and hydrogen atoms computed using B3LYP/cc-pVDZ theory and it is not a trainable parameter. For the 2-b, a single Morse range parameter of 2.5 bohr was used, with a maximum power of 10, so the total number of 2-b coefficients is 30. The cutoff distance for 2-b is 18.0 bohr. For the 3-b, the Morse range parameter is 1.8 bohr, and the maximum polynomial order of the PIPs is 8. The switching function is applied when the maximum distance in a trimer is between 12.3 and 14.2 bohr, and the energy contribution is 0 when the maximum internuclear distance in the trimer is beyond 14.2 bohr. For the 4-b bases, the Morse range parameter is 1.2 bohr, and the maximum polynomial order is 6. The switching range for 4-body is between 8.5 and 10.4 bohr. All the bases are purified,[37,38] that is, any polynomial that does not go to zero when an atom is infinitely far away from the remaining atoms is removed from the bases. Using the polynomial orders mentioned above, there are 32, 78, 78, 32, 40, 115, 174, 115, and 40 PIPs for CCC, CCH, CHH, HHH, CCCC, CCCH, CCHH, CHHH, HHHH, respectively. These sum up to 734 undetermined linear coefficients, including the 30 from

---

[†]Identification of certain commercial equipment, instruments, software, or materials does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

all the 2-b interactions.

These coefficients were determined in a single least-squares fit to *ab initio* energies of $C_{14}H_{30}$. (See ref. 39 for details of the least-squares implementation.) The training (on $C_{14}H_{30}$) gives a root-mean-square error on energies of 115 cm^{-1} (0.33 kcal/mol), and corresponds to an error of 2.6 cm^{-1} (0.007 kcal/mol) per atom. This is very small precision error compared to typical ML values for such a large molecule and dataset. Even though it should be clear, *prima facie*, that these many-body terms, trained on $C_{14}H_{30}$, are transferable to all alkanes. The precision of such transferability needs to be verified of course. This was done by us where excellent precision was shown over the range C_5H_{12} to $C_{30}H_{62}$.[13]

For ΔV_{CC-LL}, a corresponding MB-PIP fit was done to 4514 energy differences between PNO-LCCSD(T)-F12b/AVTZ' energies (where AVTZ' means cc-pVTZ basis for H and aug-cc-pVTZ for C) and B3LYP or PBE0+MBD ones for $C_{14}H_{30}$. The 4514 configurations were chosen as follows: A first set of 1359 configurations comes from molecular dynamics simulations. Specifically, along the minimum energy path between the linear and hairpin structure, 7 additional local minima were located on the low-level PES. Molecular dynamics calculations (NVE) starting from these 9 stationary points (7 plus the linear and hairpin) were carried out and 151 configurations were collected from each trajectory. The second set of 459 configurations was obtained by assigning random displacement of coordinates to the 9 stationary points, 51 configurations each. The last 2544 configurations were randomly picked from the data set used for the low-level PES. For all the 4514 configurations, PNO-LCCSD(T)-F12b energies were obtained. The fitting also used the MB-PIP approach, with the same hyperparameters as the low-level PES, except the polynomial order. The maximum polynomial orders of 2-b, 3-b, and 4-b of the ΔV_{CC-LL} are 6, 7, and 5, respectively, resulting in 383 unknown coefficients. This very small number results in a trivial linear least-squares problem. The fitting errors are 12 cm^{-1} for the Δ-B3LYP data and 7 cm^{-1} for the Δ-PBE0+MBD data; these indicate very precise fitting. These small fitting errors are expected because the range of the energy difference is much smaller, specifically, the energy differences

9

between PNO-LCCSD(T)-F12 and B3LYP span a range of 12 966 cm^{-1}, and the differences between PNO-LCCSD(T)-F12 and PBE0+MBD span only 3904 cm^{-1}. We note that the larger range for the former than the latter is not surprising, because the B3LYP energies are dispersionless whereas the PBE0+MBD ones describe dispersion accurately. That the precision of the two correction PESs, ΔV_{CC-LL}, are about the same is certainly a gratifying outcome of our machine learning approach.

To sum up thus far, we have obtained transferable low-level MB-PIP PESs based on B3LYP (with no dispersion) and PBE0+MBD energies, V_{LL}, and corresponding correction PESs ΔV_{CC-LL}, all trained on $C_{14}H_{30}$. We now examine the performance of the transferability of the corrected PES, $V_{LL \rightarrow CC}$ for the determination of the energy difference between the hairpin and extended conformations over the range $C_{12}H_{26}$ to $C_{28}H_{58}$. For consistency, we use stationary configurations for these alkanes, obtained from direct optimization of the relatively efficient and (as seen below) accurate PBE+MBD method.

To begin we show results for the B3LYP and B3LYP- corrected PES shown graphically in Fig. 2. To be clear the "B3LYP" results are from the transferable MB-PIP B3LYP PES (V_{LL}) and those labeled Δ-B3LYP are from the transferable MB-PIP corrected B3LYP PES ($V_{LL \rightarrow CC}$). As seen, there is a major improvement of the B3LYP PES (with no dispersion) such that the corrected PES is very close to the benchmark CCSD(T) results. The correction is not perfect, as expected, with a gap of roughly 0.5 kcal/mol. This can be traced mainly to the fact that ΔV_{CC-LL} was trained on the difference between direct B3LYP and CCSD(T) energies and not the difference between the B3LYP PES and CCSD(T) energies. Across the range of alkanes shown the mean absolute error of the PES is 0.5 kcal/mol

Next, consider the performance of the PBE0+MBD MB-PIP and corrected PBE0+MBD, MB-PIP PESs for these energy differences. These are shown in Fig. 3. As seen the PBE0+MBD results are in good agreement with the CCSD(T), with differences of roughly 1-1.5 kcal/mol. The corrected PBE0+MBD PES differences are reduced substantially to about 0.5 kcal/mol; essentially the same difference at seen for the corrected B3LYP MB-PIP
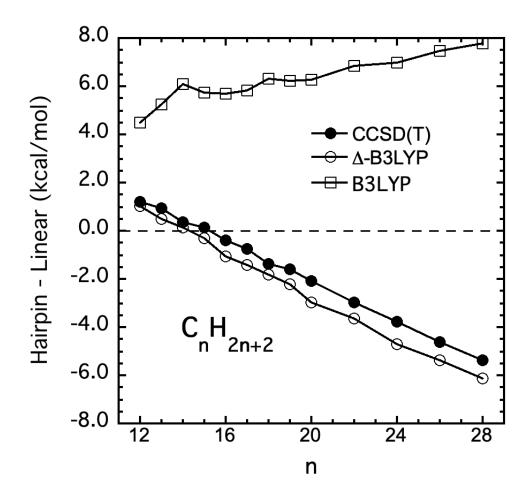
Figure 2: Energy difference between hairpin and linear minima from indicated sources and where "CCSD(T)" refers to PNO-LCCSD(T)-F12

PES. The numerical values shown in Figures 2 and 3 are given in Table. 1. These results indicate that the new transferable corrected MB-PESs based on B3LYP and PBE+MBD have successfully been elevated to the "gold standard" level, especially the previous B3LYP one.
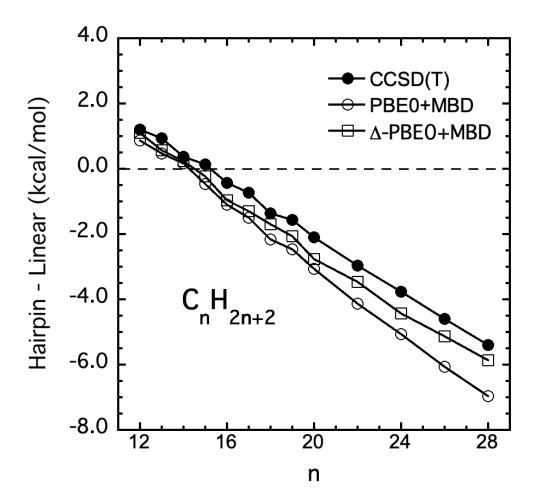

Figure 3: Energy difference between hairpin and linear minima from indicated sources and where "CCSD(T)" refers to PNO-LCCSD(T)-F12

Table 1: Energy difference (kcal/mol) between hairpin and linear minima from indicated calculations, and where "CCSD(T)" refers to direct PNO-LCCSD(T)-F12 calculations, and all other values were obtained using the corresponding MB-PIP PES. These are all evaluated at the PBE0+MBD optimized configurations.

n	CCSD(T)	B3LYP	Δ-B3LYP	PBE0+MBD	Δ-PBE0+MBD
12	1.20	4.50	1.00	0.92	1.10
13	0.92	5.24	0.50	0.46	0.58
14	0.37	6.07	0.14	-0.10	0.22
15	0.12	5.75	-0.33	-0.52	-0.25
16	-0.42	5.69	-1.08	-1.24	-0.98
17	-0.75	5.82	-1.42	-1.67	-1.29
18	-1.36	6.30	-1.83	-2.29	-1.71
19	-1.58	6.23	-2.20	-2.67	-2.06
20	-2.09	6.25	-2.96	-3.34	-2.77
22	-2.98	6.85	-3.65	-4.26	-3.45
24	-3.77	6.96	-4.70	-5.31	-4.43
26	-4.61	7.45	-5.39	-6.07	-5.12
28	-5.39	7.78	-6.15	-6.88	-5.86

There are numerous studies that can be done with these corrected MB PESs. One that we present here are calculations of the vibrational power spectra of $C_{14}H_{30}$ from molecular dynamics simulations, using the B3LYP PES and Δ-B3LYP PES. These were performed as we did recently,[18] by running a classical trajectory initiated from a specific configuration; the final spectra are averages for five trajectories, run for approximately 1 ps, at each temperature and potential. Here we considered both the linear and hairpin minima. The trajectories were run using the NVE protocol and for zero total angular momentum using our in-house software. The total energy is obtained from the simple correspondence between the average energy $<E>$ and temperature T for classical harmonic oscillators. (For $C_{14}H_{30}$ $<E>$ = 126RT.) Temperatures of 10 and 300 K were considered with the goal of examining both the difference in the spectra at these two temperatures and also the differences between the B3LYP PES and Δ-B3LYP PES. The power spectra were obtained by using an open source web platform,[40,41] recently developed at the University of Milan, which is based on the time-averaged Fourier transform of the Cartesian velocity autocorrelation function.[42]

We examined such trajectories in detail previously,[18] and noted that at 300 K many

conformations are visited and so the final spectra are virtually identical whether starting from the hairpin or linear minimum. However, at 10 K this is not the case, and the spectra obtained are more representative of the harmonic normal mode spectra at these minima. With this in mind consider the 10 K results first. There are clear differences between the hairpin and linear spectra in the range 0 to 1000 cm^{-1} and this is seen for both the B3LYP and Δ-B3LYP potentials. And, not surprisingly, we also see significant differences between these potentials in this spectral range. These differences are largely quenched for the 300 K spectra for both potentials. So, one conclusion from these results is that signatures of the effects of dispersion are present at 10 K in the spectral range 0 to 1000 cm^{-1}. This conclusion is consistent with the double harmonic analysis done by Luttschwager and Suhm in 2013.[6]

A second, and again not surprising, conclusion from these spectra is that the high-frequency CH-stretch region is largely insensitive to temperature and potential.
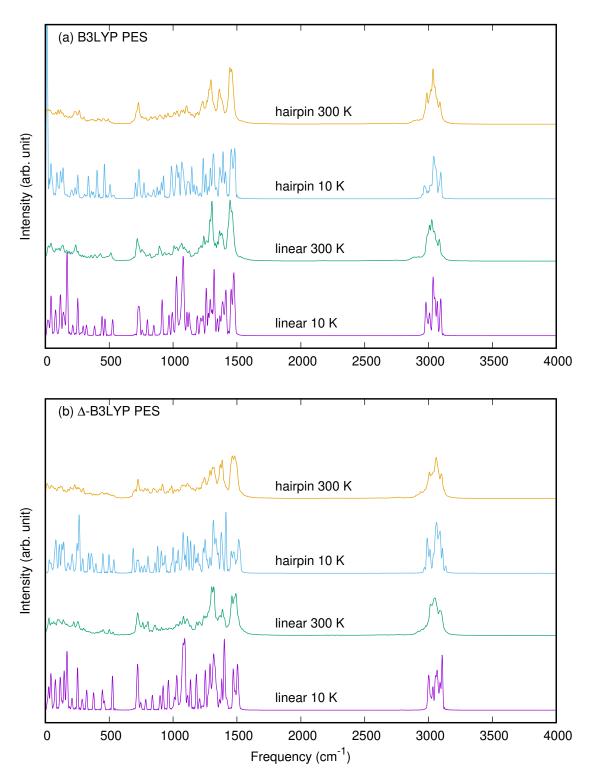
Figure 4: Power Spectra of $C_{14}H_{30}$ using B3LYP and Δ-B3LYP PES at indicated temperatures.

The present results demonstrate that it is straightforward to elevate a low-level, e.g., DFT, MB-PIP potential for a large class of molecules, e.g., linear alkanes, to the CCSD(T) level. This was done here for MB-PIP architecture, where all terms are fit "at once". That the method is successful in elevating a low-level PES based on dispersion-less B3LYP calculations is very encouraging, albeit not necessarily surprising. Note the correction to the low-level PES was done by fitting a difference potential using the "at once" approach. The high-level energies were obtained for $C_{14}H_{30}$ using efficient and certified PNO-LCCSD(T)-F12 methods in Molpro 2023. Also, we note that correction terms could be obtained using a smaller alkane. This is something to be investigated in the future.

The success was explicitly demonstrated for the energy difference between the hairpin and linear conformers of alkanes ranging from 12 to 28 carbon atoms against CCSD(T) energies. As noted above the most recent "single-point" study of alkanes considered datasets of 13, 18, and 22 configurations for alkanes with 12, 16, and 20 carbon atoms.[9] The present work considered 4514 configurations, spanning a large energy range, at which PNO-LCCSD(T)-F12 energies were obtained. These were used to precisely fit a MB-PIP difference potential with precision of less than 0.03 kcal/mol.

Finally, we note a different, "building block", approach was recently introduced by Paesani and co-workers for polyalanine chains.[43] This work uses and builds on the many-body approach based on the selection of monomers, "chemically intuitive building blocks."[24] In the more recent paper on polyalanine, each building block is separately elevated to the CCSD(T) level (using the DLPNO-CCSD(T) method). This is certainly a reasonable approach. However, a statement made in that paper about our MB-PIP approach is contradicted by the present results. Specifically, it was stated that "...the 'many-body strategy' builds PIP terms for interactions involving sets of 2, 3, or 4 atoms, which are then summed to predict the total molecular energy.... because these models require whole-molecule reference data for training, they are typically limited to DFT-level calculations due to the prohibitive cost of higher-level quantum methods." The present work demonstrates that the MB-PIP approach

can extend to higher-level methods using a standard Δ-ML approach and without the need to select molecule-specific building blocks.

To summarize, the present paper reports two significant results. One is the successful elevation of DFT-based, transferable MB-PIP potentials for alkanes to the CCSD(T) level. Second, we demonstrated quantitatively the major and increasing importance of dispersion in determining the folding conformation of linear alkanes.

This new transferable potential can now be used in many applications and tests of other potentials for this important class of molecules. For example, our most work was a study of energies and configurations of tens of thousands of local minima and first order saddle-points of $C_{14}H_{30}$ using the B3LYP MB-PIP PES.[14] It is now possible to conduct that study for a MB-PIP PES that includes dispersion and also for a range of alkanes. Also, the new MB-PIP PES can be used to test the new generation of universal machine-learned ones, such as MACE-OFF,[44] SO3LR,[45] ANI,[46] Allegro[47] and OMNI-P1.[48]

Acknowledgement

Data Availability

The PNO-LCCSD(T)-F12 electronic energies used in this paper are available at https://github.com/jmbowma/QM-22

References

(1) Patterson, G. *Modern Applications of Flory's "Statistical Mechanics of Chain Molecules"*; ACS Symposium Series; American Chemical Society, 2020; Vol. 1356; Chapter 1, pp 1–13.

(2) Allinger, N. I. *Molecular Structure*; Wiley Online Books, 2010; pp 51–91.

(3) Thomas, L. L.; Christakis, T. J.; Jorgensen, W. L. *J. Phys. Chem. B* **2006**, *110*, 21198–21204.

(4) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(5) Jorgensen, W. L.; Ghahremanpour, M. M.; Saar, A.; Tirado-Rives, J. *J. Phys. Chem. B* **2024**, *128*, 250–262.

(6) Lüttschwager, N. O. B.; Wassermann, T. N.; Mata, R. A.; Suhm, M. A. *Angew. Chem. Int. Ed* **2013**, *52*, 463–466.

(7) Byrd, J. N.; Bartlett, R. J.; Montgomery, J., John A. *J. Phys. Chem. A* **2014**, *118*, 1706–1712.

(8) Liakos, D. G.; Neese, F. *J. Chem. Theory Comput.* **2015**, *11*, 2137–2143.

(9) Ehlert, S.; Grimme, S.; Hansen, A. *J. Phys. Chem. A* **2022**, *126*, 3521–3535.

(10) Santra, G.; Martin, J. M. *J. Phys. Chem. A* **2022**, *126*, 9375–9391.

(11) Werner, H.-J.; Hansen, A. *J. Chem. Theory Comput.* **2023**, *19*, 7007–7030.

(12) Xia, J.; Zhang, Y.; Jiang, B. *Chem. Soc. Rev.* **2025**, *54*, 4790–4821.

(13) Qu, C.; Houston, P. L.; Allison, T.; Bowman, J. M. *J. Chem. Theory Comput.* **2025**, *21*, 3552–3562.

(14) Allison, T. C.; Bowman, J. M.; Houston, P. L.; Pillai, Y.; Qu, C.; Wales, D. J. *J. Chem. Phys.* **2025**, *163*, 104307.

(15) Ma, Q.; Werner, H.-J. *J. Chem. Theory Comput.* **2019**, *15*, 1044–1052.

(16) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. *J. Chem. Phys.* **2025**, *162*, 180901.

(17) Qu, C.; Houston, P. L.; Allison, T.; Schneider, B. I.; Bowman, J. M. *J. Chem. Theory Comput.* **2024**, *20*, 9339–9353.

(18) Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. *J. Phys. Chem. A* **2024**, *128*, 10713–10722.

(19) Wang, Y. M.; Huang, X. C.; Shepler, B. C.; Braams, B. J.; Bowman, J. M. *J. Chem. Phys.* **2011**, *134*, 094509.

(20) Babin, V.; Medders, G. R.; Paesani, F. *J. Chem. Theory Comput.* **2014**, *10*, 1599.

(21) Yu, Q.; Qu, C.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. *J. Phys. Chem. Letts.* **2022**, *13*, 5068–5074.

(22) Qu, C.; Yu, Q.; Houston, P. L.; Conte, R.; Nandi, A.; Bowman, J. M. *J. Chem. Theory Comput.* **2023**, *19*, 3446–3459.

(23) Zhu, X.; Riera, M.; Bull-Vulpe, E. F.; Paesani, F. *J. Chem. Theory Comput* **2023**, *19*, 3551–3556.

(24) Bull-Vulpe, E. F.; Riera, M.; Bore, S. L.; Paesani, F. *J. Chem. Theory Comput.* **2023**, *19*, 4494–4509.

(25) Allen, A. E. A.; Dusson, G.; Ortner, C.; Csányi, G. *Mach. Learn.: Sci. Technol.* **2021**, *2*, 025017.

(26) Chen, C.; Braams, B.; Lee, D. Y.; Bowman, J. M.; Houston, P. L.; Stranges, D. *J. Phys. Chem. Lett.* **2010**, *12*, 1875–1880.

(27) Collins, M. A.; Bettens, R. P. A. *Chem. Rev.* **2015**, *115*, 5607–5642.

(28) Nandi, A.; Qu, C.; Bowman, J. M. *J. Chem. Phys.* **2019**, *151*, 084306.

(29) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

(30) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(31) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, 236402.

(32) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.

(33) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. *New J. Phys.* **2012**, *14*, 053020.

(34) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. *J. Chem. Phys.* **2021**, *154*, 051102.

(35) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. *J. Chem. Theory and Comput.* **2023**, *19*, 1–17.

(36) others„ et al. MOLPRO version 2023: a package of ab initio electronic structure programs. https://www.molpro.net/, accessed 2025-09-19.

(37) Paukku, Y.; Yang, K. R.; Varga, Z.; Truhlar, D. G. *J. Chem. Phys.* **2013**, *139*, 044309.

(38) Conte, R.; Houston, P. L.; Bowman, J. M. *J. Chem. Phys.* **2014**, *140*, 151101.

(39) Pandey, P.; Arandhara, M.; Houston, P. L.; Qu, C.; Conte, R.; Bowman, J. M.; Ramesh, S. G. *J. Phys. Chem. A* **2024**, *128*, 3212–3219.

(40) Gandolfi, M.; Moscato, D.; Aieta, C.; Pindaro, M.; Campana, E.; Valtolina, S.; Ceotto, M. A web platform for time averaged Fourier transform of autocorrelated data. 2024; `http://semisoft.unimi.it/`, Accessed on Sep. 20, 2025.

(41) Conte, R.; Gandolfi, M.; Moscato, D.; Aieta, C.; Valtolina, S.; Ceotto, M. *J. Comput. Chem.* **2025**, *46*, e70118.

(42) Conte, R.; Aieta, C.; Botti, G.; Cazzaniga, M.; Gandolfi, M.; Lanzi, C.; Mandelli, G.; Moscato, D.; Ceotto, M. *Theor. Chem. Acc.* **2023**, *142*, 53.

(43) Zhou, R.; Bull-Vulpe, E. F.; Pan, Y.; Paesani, F. *J. Chem. Theory Comput.* **2025**, *21*, 6194–6212.

(44) Kovács, D. P.; Moore, J. H.; Browning, N. J.; Batatia, I.; Horton, J. T.; Pu, Y.; Kapil, V.; Witt, W. C.; Magdău, I.-B.; Cole, D. J.; Csányi, G. *J. Am. Chem. Soc.* **2025**, *147*, 17598–17611.

(45) Kabylda, A.; Frank, J. T.; Suárez-Dou, S.; Khabibrakhmanov, A.; Medrano Sandonas, L.; Unke, O. T.; Chmiela, S.; Müller, K.-R.; Tkatchenko, A. *J. Am. Chem. Soc.* **2025**, *147*, 33723–33734.

(46) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. *Nat. Commun.* **2019**, *10*, 2903.

(47) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. *Nature Comm.* **2023**, *14*, 579.

(48) Chen, Y.; Dral, P. O. *J. Chem. Theory Comput.* **2025**, *xx*, xx.