Full-dimensional, ab initio potential energy surface for glycine with characterization of stationary points and zero-point energy calculations by means of diffusion Monte Carlo and semiclassical dynamics

Riccardo Conte,[1, a)] Paul Houston,[2, b)] Chen Qu,[3] Jeffrey Li,[4] and Joel M. Bowman[4, c)]
[1)]*Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy*
[2)]*Department of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853, U.S.A. and Department of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A*
[3)]*Department of Chemistry & Biochemistry, University of Maryland, College Park, Maryland 20742, U.S.A.*
[4)]*Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, U.S.A.*

(Dated: 2 December 2020)

A full-dimensional, permutationally invariant potential energy surface (PES) for the glycine amino acid is reported. A precise fit to energies and gradients calculated at density functional theory DFT/B3LYP level of electronic-structure theory with Dunning's aug-cc-pVDZ basis set is performed involving 20,000 low-energy points and associated Cartesian gradients plus about 50,000 additional higher-energy points. The fact that the DFT/B3LYP energies for the main stationary points are close to the coupled cluster CCSD(T) values, recently reported in the literature, provides reassurance about the accuracy of the constructed PES. Eight conformers and numerous saddle points are identified and characterized by describing geometries, relative stability, and harmonic frequencies. Stochastic and dynamical approaches are employed to study the vibrational ground state. Diffusion Monte Carlo simulations and approximate quantum dynamics, performed by means of the adiabatic switching semiclassical initial value representation technique, provide zero-point energies in excellent agreement with each other. The PES we report is sufficiently complete to permit spectroscopic and dynamical studies on glycine which may be of interest to the biochemical and astrochemistry communities.

I. INTRODUCTION

Glycine, the smallest amino acid, is a molecule that has drawn the attention of experimentalists and theoreticians for several decades. The main interest in this building block of life lies in the quest for its presence in the interstellar medium, which would represent a milestone in the development of theories about the origin of life. Furthermore, it turns out that the energy landscape of this amino acid is particularly complicated as it involves large amplitude motion of the hydrogen atoms, opening up questions of purely theoretical interest about the possibility of a high-level mathematical and chemical description of its potential energy surface (PES).

The open debate about glycine concerns its presence in the interstellar medium. Millimeter-wave observation of rotational transitions has been widely employed for studying the issue. In 2003 Kuan et al.[1] reported the observation of 27 glycine lines from several interstellar sources, but these observations were not reproduced by Snyder et al.,[2] who offered alternate common molecular candidates for the reported absorptions. Jones et al.[3,4] also searched for two conformers of glycine in the same region of Sagittarius B2 (LMH) where Kuan et al. had reported finding it, but they did not collect any positive observation above a detection limit. They claimed that the sensitivity of the Australia Telescope Compact Array employed should have easily been able to detect glycine if it were at present at the concentration reported by Kuan et al. Subsequently, Lattelais et al.[5] argued that other forms of glycine atomic composition, $C_2H_5O_2N$, were more stable isomers and that these might be more abundant in the interstellar medium. So, it appears that the presence of glycine in the interstellar medium is currently unresolved and that the previous positive identification was likely in error.

The elusiveness of glycine conformers has triggered several experimental as well as theoretical investigations. To the best of our knowledge, the study of gaseous glycine conformers was pioneered by Csaszar in 1992.[6] That was a theoretical investigation that predicted up to 8 conformers but aimed at supporting rotational constant measurements for two conformers only. Moreover, further studies were necessary and indeed later performed to deal with the vibrational problem. For instance, Balabin[7,8] reported the first jet-cooled Raman spectrum of glycine in 2010. His results, coupled with MP2 and density functional theory (DFT) calculations using an aug-cc-pVTZ basis set, confirmed the structures of the two most stable glycine conformers, known from electron diffraction and microwave spectra,[9,10] and provided evidence for a third, which at that time had not been previously reported experimentally. However, a previous study by Stepanian and Adamowicz in 1998,[11] had already identified a different "third" conformer, so that the total number of experimentally known glycine isomers actually increased to four after Balabin's investigation. A theoretical study by

[a)]Electronic mail: riccardo.conte1@unimi.it
[b)]Electronic mail: plh2@cornell.edu
[c)]Electronic mail: jmbowma@emory.edu

Barone's group employed second-order vibrational perturbation theory (VPT2) calculations at MP2/aug-cc-pVQZ level of theory to describe the vibrational spectra associated to these four conformers.[12] Bazsó et al.[13,14] examined the tunneling between two conformers of glycine using near-IR radiation and mid-IR absorption in low temperature rare gas matrices at 12 K. Several vibrational frequencies were reported. Coussan and Tarczay reported infrared laser induced conformational and structural changes of glycine in low-temperature matrices.[15] Finally, in 2020, the Czako group reported a benchmark study based on CCSD(T)-F12b/aug-cc-pVTZ energy and CCSD(T)-F12b/aug-cc-pVDZ harmonic frequency calculations in which a total of eight glycine conformers were characterized,[16] improving on Csaszar's original results.[6,17]

Energy calculations have not been the only target of glycine investigation. Indeed, harmonic and anharmonic vibrational frequencies of glycine have been widely studied too. After Csaszar and Adamowicz's pioneering harmonic frequency calculations,[6,11] more refined approaches have been adopted. For instance, Gerber et al. performed vibrational self-consistent field (VSCF) calculations to describe some of the vibrations of the global minimum conformer.[18] Successful studies based on model Hamiltonians have been undertaken by Fernandez-Clavero and coworkers,[19] while Hobza et al. have determined quantum mechanically for the first time the whole set of 24 fundamental frequencies of vibration by means of VPT2.[20] All these investigations concerned the global minimum conformer or Conformer 1. However, as anticipated, the glycine PES is characterized by several minima and the first comprehensive VPT2 study of vibrational frequencies for the 4 long known conformers was reported by Barone and coworkers.[12] Finally, another step forward was contributed by one of us within the recent semiclassical dynamics investigation of the Ceotto group.[21] In that case the multi-well nature of the PES was taken into account, a fundamental aspect that will be better pointed out in this manuscript.

Very recently there has been quite some interest in the construction of an analytical potential energy surface for glycine. This is in part due to the continuous effort in improving fitting techniques and partly to the importance of glycine as a prototypical system for more complex and sizable biomolecules. A first "demonstration surface" was obtained by Zhu et al. via a neural network (NN) machine learning approach, in which the original Density-Functional Tight Binding (DFTB) energies have been corrected towards DFT values.[22] A further extension has been provided by the Tkatchencko group on the basis of a new NN machine learning model.[23] In these approaches DFTB values are corrected towards DFT ones because for complex chemical systems calculations are commonly thought not to be easily accomplished by means of DFT or higher level ab initio quantum chemistry methods. We actually do that in the present paper using a DFT approach that has been shown to produce energies for conformers of glycine in close accord with the previously anticipated CCSD(T) ones.[16] Also, we note that these previous PESs appear to be limited to one conformer and the focus of these studies was mainly method-based and not on developing a global PES. In general, the recent machine learning approaches are undertaken with the main aim of generating a new kind of general-purpose biological force field based on *ab initio* quantum mechanical electronic calculations. These approaches are expected to outperform in accuracy the currently available force fields, even if they cannot be applied to very large molecular systems and additionally do not achieve spectroscopic accuracy. Conversely, our PES is specialized for glycine and very good accuracy is obtained.

In this paper we present a new full-dimensional, permutationally invariant PES for glycine including complete characterization of conformers and saddle points for use in dynamical simulations. Many times in the past our group has constructed very accurate *ab initio* PESs for dynamics applications, as in the case of quasi-classical trajectory simulations for molecular energy transfer,[24–27] or adiabatic switching simulations for spectroscopy.[28] Here we employ the new glycine PES for spectroscopy applications by running approximate quantum dynamics simulations by means of semiclassical (SC) dynamics. The accuracy of SC results is determined by comparison to diffusion Monte Carlo (DMC) benchmarks. The paper is organized as follows. In Section II we illustrate the computational methods employed. Section III is reserved to presentation of results, and, finally, in Section IV we draw our conclusions.

II. COMPUTATIONAL METHODS

A. Potential Energy Surface

1. The Data Base

Although glycine has only ten atoms, the number of conformations, the range of energies, and the desire for gradients as well as energies at each geometry limited the choice of *ab initio* methods to density functional theory, which we employed with B3LYP and an aug-cc-pVDZ basis set. All electronic structure calculations were determined with Molpro.[29,30] In addition to the global minimum, denoted as Conformer 1, we determined the structures for seven other conformers as well as for 15 of the saddle points between them. The equilbrium structures of these conformers are given in Fig. 1.

At each stationary point we determined the optimum geometry and the harmonic vibrational frequencies. The stationary points are discussed below. In order to provide data for an accurate potential energy surface, an energy and full Cartesian gradients were determined at each of over 70,000 geometries. We next describe in some detail the methods that we use to develop and analyze the data set used for the PES fit.

Geometries for the data set are chosen by an iterative process involving a combination of classical and quantum sampling. The classical sampling is performed either using *Ab Inito* Molecular Dynamics (AIMD) or, for detailed information concerning stationary points, using calculations on a grid of randomly generated geometries centered on the stationary point. Classical sampling has the merits that, in many cases, it can provide gradients as well as energies and that the calculation is fairly fast. Following the generation of points by

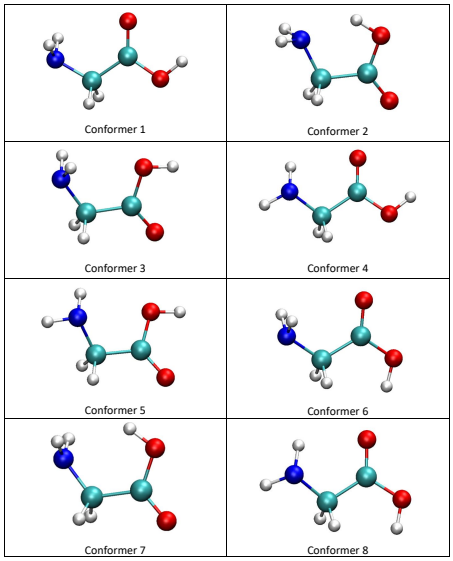

FIG. 1. Eight low-energy conformers of glycine

classical sampling, a fit is performed and evaluated. Quantum sampling is then performed using Diffusion Monte Carlo (DMC) exploration of the potential surface.[31–33] DMC, discussed in detail below, has the advantage that it explores the surface quantum mechanically up to the zero-point energy (ZPE) of the molecule, which is determined by the calculation and is relatively high for most large molecules. While DMC is usually employed simply to determine the ZPE, it also has the benefit of indicating where there are "holes" in the PES, locations where the energy of an improper fit is negative relative to that of the global minimum structure. The holes occur when there is too little data in a region or when the number of fitting coefficients is too large compared to the number of data base geometries. Analysis of these hole geometries finds them to be of two types. Those for which each Morse variable is within the range of the data set, and those for which there are Morse variables outside that range. For the former type we include all of the geometries in a new round of electronic structure calculations, while for the latter set we determine the mean values of those Morse variables that are outside the range and include geometries with these mean values in the new calculations. Adding the new points (perhaps augmented by more points from additional classical sampling), refitting the PES, and rerunning the DMC calculation provides an iterative method for improving the surface. Although we strongly recommend this method, it does have the disadvantage that evaluation of the energy from the PES fit must be fairly fast because the converged DMC calculation (needing roughly 10 independent runs) typically involves about 10^{11} potential calls.

For glycine, we began with AIMD trajectories of 6000–25000 steps starting from the optimum glycine geometry (Conformer 1) and using total energies of 100, 500, 1000, 5000, 10000, 14000, 17500, 25000, 30000, 40000, 60000, 70000, and 80000 cm^{-1}. For most of the AIMD trajectories, every tenth point on the trajectory was included in the data set. We then augmented the data set with AIMD runs starting from Conformers 2, 3, and 4 using similar energies from 14000–60000 cm^{-1}, as well as starting from Saddle points 13, 14, and 23 using 60000 cm^{-1}. Even with these trajectories, it was apparent that we did not have enough data points near the conformers to determine accurately their harmonic vibrational frequencies. Thus, we augmented the above set with grids of points centered on the geometries of conformers 1–4. DMC calculations then located the "holes" in the PES, resulting in new geometries whose energy and gradients were calculated. The iterative process of adding the new geometries, refitting the PES, and rerunning the DMC calculation greatly improved the surface. During this process we also added geometries from new AIMD trajectory calculations to determine information about Conformers 5–8 and the saddle points joining these to Conformers 1–4. We found the mixed DMC-AIMD approach very effective in fixing "holes" on the surface. This is actually a topic which has drawn quite some interest very recently and other approaches are possible. One remarkable example is provided by recent work by Poirier and coworkers who have developed their own fixing protocol and provided the community with dedicated software.[34,35]

2. *Analysis of the Data Base*

We now illustrate a few methods that are helpful in assessing the coverage of the data base. The first is a histogram of the energies, shown for the glycine data set in Fig. 2. The energies for Conformers 1–8 lie in the region of the first peak in the distribution. Beyond that peak, there appears to be a sufficient density of points up to about 35000 cm^{-1}.

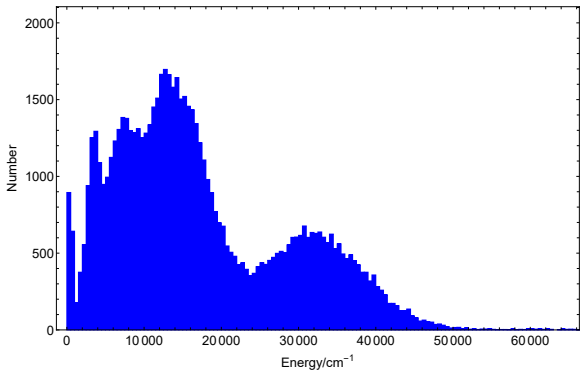

FIG. 2. Histogram of energies for geometries used as the data set for glycine. The bin size for the abscissa is 500 cm^{-1}. The data base has 70,099 energies and $30 \times 70,099$ gradient components.

An alternative method for determining the coverage of geometries in the data base is to make a 2D plot of energies along the ordinate and positions along the abscissa. Since there are $n(n-1)/2$ inter-nuclear distances, where n is the number of atoms (10, for glycine), the question of which distance or what linear combination of distances should be used remains to be discussed. A standard way of choosing the best dispersal is

Principal Component Analysis.[36] Briefly, we construct a matrix whose columns are the 45 (for glycine) inter-nuclear distances, referenced to the mean of each column, and whose rows represent the 70,099 geometries of the data base. We then premultiply the transpose of this matrix by the matrix itself and determine the eigensystem of the square (45×45) resultant. The eigenvalues of the resultant give the degree to which the dispersion is determined by each of the 45 principal components, and the eigenvectors, when multiplied by the transpose of the original matrix, give the projections of the data onto each principal axis. In the case of glycine and our data set, about 41% of the variance is determined by the first principal component. We then plot the projections of the data set onto the first principal axis, which gives the maximum dispersal of the data in one dimension, given the inter-nuclear distances as variables. A plot of energy as a function of projection onto the first principal axis for the glycine data set is shown in Fig. 3.

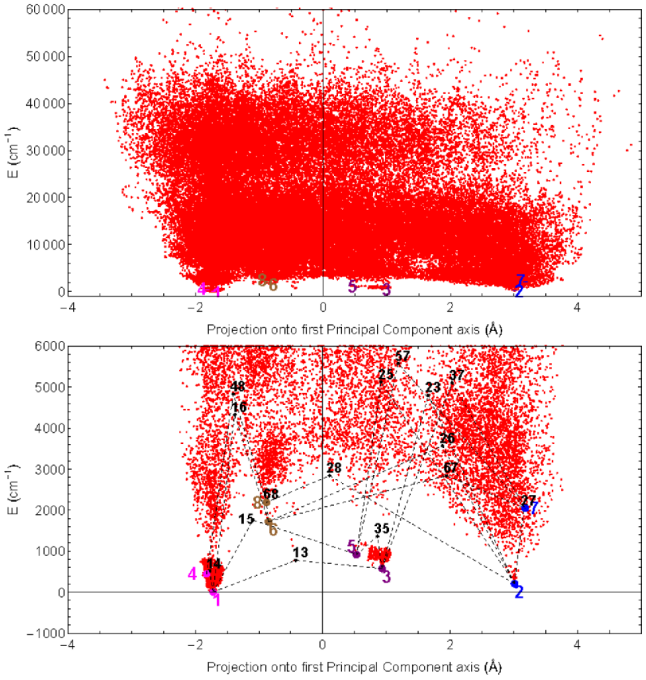

FIG. 3. Points of the glycine data base giving energy as a function of projection of the data onto the first principal component axis for energy ranges 0–60,000 cm^{-1} (top panel) and 0–6,000 cm^{-1} (bottom panel). The colored numbers indicate the identities of the conformers, with conformers that are related by a partial rotation of the NH_2 group around the C-N bond axis represented by the same color. In the bottom panel, the two-digit black numbers identify saddle points, based on the fit to the PES, while the dashed black lines are a guide to the eye in connecting the conformer locations with those of the saddle points between them.

Figure 3 provides a method for identifying regions of the data set that might benefit from more data. These are areas where there are interesting saddle points or conformers but where there is not a high density of points. For example, more points between conformers 6 and 8 might be useful, as well as would more points near saddle points 35, 13, 28, and 15. On the other hand, there appears to be sufficient data near conformers 1, 3, and 4, for example. Note also that one learns from the diagram that the saddle points between pairs 1 and 4, 6 and 8, 3 and 5, and 2 and 7 are quite close to one or both of the conformers in each pair. This is because each pair of similarly colored conformers is related to its partner by a partial rotation of the NH_2 group around the C-N bond axis.

We also investigated inter-conformer distance relationships. In this case we consider the 45 bond lengths for each conformer in increasing order. Then we define a simple metric to distinguish the distance between conformers. This is just the square root of the sum of the squares of bond-length distances between pairs of conformers. The results are shown in Table I. It is easy to determine from this which conformers are near one another. For example, in increasing order of distance Conformer 1 is nearest 4, 6, 8, 5, 3, 2, and then 7. The remaining closeness numbers are 2: 7, 3, 5, 6, 8, 1, 4; 3: 5, 7, 2, 1, 4, 6, 8; 4: 1, 8, 6, 5, 3, 2, 7; 5: 3, 4, 1, 7, 2, 8, 6; 6: 8, 1, 4, 3, 5, 2, 7; 7: 2, 3, 5, 6, 8, 1, 4; and, 8: 6, 4, 1, 5, 3, 2, 7. Note that conformers in the pairs {1,4}, {6,8}, {3,5}, and {2,7} always appear next to one another because they are related by the rotation of the NH_2 group around the C-N bond.

TABLE I. Distance metric between conformers in Å

	1	2	3	4	5	6	7	8
1	0.	4.89778	3.42631	1.60179	3.42081	1.99666	4.95141	2.47169
2	4.89778	0.	3.29411	5.18013	3.69346	4.28879	0.988108	4.52848
3	3.42631	3.29411	0.	3.6268	1.71099	3.95133	3.11758	4.08296
4	1.60179	5.18013	3.6268	0.	3.09869	2.57775	5.25665	2.06025
5	3.42081	3.69346	1.71099	3.09869	0.	3.96457	3.63599	3.74662
6	1.99666	4.28879	3.95133	2.57775	3.96457	0.	4.42053	1.47715
7	4.95141	0.988108	3.11758	5.25665	3.63599	4.42053	0.	4.65539
8	2.47169	4.52848	4.08296	2.06025	3.74662	1.47715	4.65539	0.

A third indicator of the data base coverage is to see which points of the data base are closest to each of the conformers. Using the positions of all atoms and the complete 70,099 geometry data base, a histogram of how many data base points were closest to each isomer is shown in Fig. 4. A more even distribution would be more desirable, but even the current data base provides more than 3500 close geometries for each conformer. As seen, conformer 1 has the largest number. This is the lowest energy conformer and so it is not surprising that geometries close this one are most represented. Another plot, Fig. SI-1, shows the correlation plot of geometric configurations for the data base in terms of closeness to Conformer 1 and to Conformer X, where $X = 2 - 8$ is the next closest conformer.

3. *The PES Fit*

Having described how the energies and gradients were calculated and how one might analyze the coverage of the data base, we now turn to the fit of a potential energy surface to the data. In general, permutationally invariant polynomials are used for the fitting basis set; a brief overview of the method was provided in Section II of ref. 37. Several recent publications have described the use of fragmentation, and the

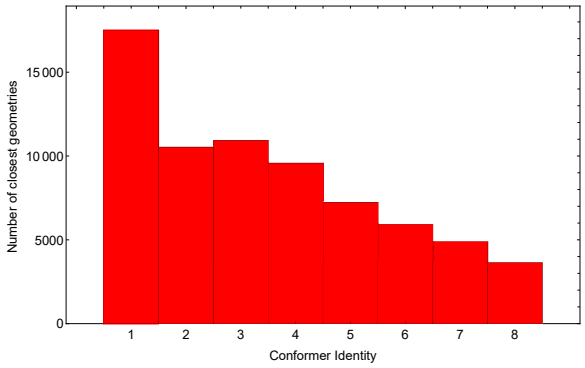

FIG. 4. Histogram plot of the number of data base geometries closest to each of the indicated conformers.

augmentation or pruning of the polynomial basis sets.[37–42] For glycine, we investigate several fits based on the numbering scheme of Fig. 5. We have previously reported fits of the glycine PES using 1, 2, or 3 fragments.[43] The best schemes for the current report were based on a single "fragment", here the parent, and with the symmetry designations either 2221111 with atoms {2,3,5,6,8,9,1,4,7,10} or 322111 with atoms {2,3,10,5,6,8,9,1,4,7}. In the first of these, it is assumed that each of the atoms pairs {2,3}, {5,6} and {8,9} permute, whereas in the second it is assumed that the triplet {2,3,10} and each of the pairs {5,6} and {8,9} permute. Because the second scheme allowed an additional permutation that might be important to one of the conformers, all calculations reported here were performed with the 322111 permutational symmetry. The fit was then to 4^{th} order in this permutationally invariant basis set; the total number of polynomials/coefficients was 22,250. Important practical points are that we use both energies and gradients in the fit, and that we use a mix of geometries, some with both energies and gradients and some with energies only. Of the 70,099 geometries, those with lowest 20,000 energies were used along with their gradients in the fit, while the remaining 50099 geometries were used in the fit with their energies but not with their gradients. The fits were inverse-energy weighted; in addition, the weights of the gradients were taken as 1/3 of the weights of the energy.

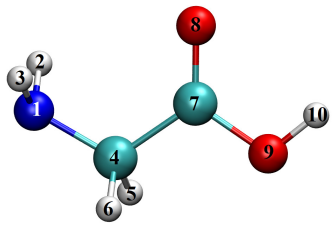

FIG. 5. Numbering scheme used for glycine.

Table II shows fitting parameters and results for three of the final fits. In the first fit, with 2221111 permutational symmetry, the polynomials were augmented to a total of 29,962, which resulted in lower RMSE values (at somewhat higher computational cost). The two fits of 322111 permutational symmetry differed in the total number of geometries fit and in the number of those geometries that included both energy and gradients, as opposed to solely energies. The weighted RMSE values for all three fits are quite acceptable, but the 322111 fit allows for one more permutable atom. All results reported below are based on the fit of the third column, using 322111 permutational symmetry and based on 70099 geometries. The last row of the table indicates the time taken by a DMC calculation of 30,000 walkers and 55,000 steps. The calculation consisted of roughly 1.6×10^9 potential evaluations, with each evaluation taking about 0.068 ms when using a 1.2 GHz Intel Xeon16 processor. To put this result in perspective, the PES fit for acetylacetone recently reported[44] took 27.5 hours for the same number of walkers, steps and potential evaluations, giving an average time for a potential evaluation of 0.062 ms when using the same 1.2 GHz Intel Xeon16 processor.

TABLE II. Glycine 4^{th}-order fitting results for three fits. The number of coefficients is also the number of polynomials used in the fit. Weighted RMS error values (wRMSE) are in cm^{-1} and cm^{-1}/bohr for potentials and gradient components, respectively. The maximum energy for the gradients is in cm^{-1}.

Perm. Sym.	2221111	322111	322111
Num. Coefs.	29,962	22,250	22,250
Tot. Num. Geoms.	52,352	52,325	70,099
Num. (E + Grads.)	15,000	15,000	20,000
Max E for Grads.	8,777	8,777	9,721
wRMSE (pot)	10.5	20.3	28.7
wRMSE (grad)	7.8	13.1	17.1
DMC time (hrs.)*	28.5	29.9	29.9

*time for DMC calculation of 30,000 walkers and 55,000 steps, consisting of roughly 1.6×10^9 potential evaluations.

B. Diffusion Monte Carlo Calculations

Diffusion Monte Carlo (DMC) simulation is an approach to compute the quantum zero-point energy (ZPE) and ground-state wavefunction of a molecule.[31,32] The simple unbiased algorithm,[31–33] briefly described next, was applied. An ensemble of random walkers is used to represent the nuclear wavefunction of the molecule. At each step, a random displacement in each degree of freedom is assigned to each walker, and this walker may remain alive (and may give birth to a new walker) or be killed by comparing its potential energy, E_i, with a reference energy, E_r. For the ground state, the probability of birth or death is given as:

$$P_{\rm birth} = \exp\left[-(E_i - E_r)\Delta\tau\right] - 1 \ (E_i < E_r) \qquad (1)$$

$$P_{\rm death} = 1 - \exp\left[-(E_i - E_r)\Delta\tau\right] \ (E_i > E_r), \qquad (2)$$

where $\Delta\tau$ is the step size in imaginary time.

After removing all dead walkers, the reference energy is

updated using the equation

$$E_r(\tau) = \langle V(\tau)\rangle - \alpha\frac{N(\tau)-N(0)}{N(0)}, \tag{3}$$

where τ is the imaginary time; $\langle V(\tau)\rangle$ is the average potential over all the walkers that are alive; $N(\tau)$ is the number of live walkers at time τ; α is a parameter that can control the fluctuations in the number of walkers and the reference energy. Finally, the average of the reference energy over the imaginary time gives an estimate of ZPE.

C. Semiclassical Calculations

The main feature of semiclassical (SC) dynamics is represented by its ability to add quantum effects to simulations based on classically-evolved trajectories. In this way semiclassical simulations represent an approximate, but hopefully accurate, treatment of quantum dynamics. Indeed one of the main goals of the technique is to extend the study of quantum dynamics to systems not feasible to treat by purely quantum mechanical approaches. For instance, SC calculations are suitable for spectroscopy studies, and the field has been in constant evolution for almost 50 years with periodic resurgences of interest. As a result of these efforts, it is nowadays possible to calculate semiclassical vibrational and vibronic spectra for molecular systems up to tens of atoms with high accuracy.[45,46]

Among the many theoretical issues that have been solved to make SC spectroscopy a viable and practical simulation tool, some are relevant for the present work. The SC formalism can be derived by applying a stationary phase approximation to Feynman's path integral representation of quantum mechanics. This leads to a difficult-to-solve double boundary problem which has been simplified through introduction of Miller's Semiclassical Initial Value Representation (SCIVR).[47] The initial value representation allows one to perform SC calculations simply by generating a set of starting conditions in phase space for trajectories to be evolved classically. Another major issue concerns conservation of unitarity of the SC propagator sufficiently long in time. This aspect has been improved by adoption of the Heller-Herman-Kluk-Kay (HHKK) SC propagator,[48–50] which contains a suitable pre-exponential factor that has also a primary role in the description of quantum effects. This pre-factor is constructed starting from the elements of the classical monodromy or stability matrix associated with the trajectory, and it serves as a probe for assessing the level of deterministic chaos that affects the SC simulation. Deterministic chaos hampers SC calculations, and chaotic trajectories are commonly discarded from those contributing to the SC simulation due to numerical instability. Quite recently, an adiabatically switched version of SC dynamics (AS SCIVR) has been introduced by one of us.[51] This technique allows one to sample the trajectory initial conditions in a way that the percentage of discarded trajectories is reduced, and accuracy and precision of calculations improved. Another major challenge for SC dynamics is the extension of calculations to sizeable systems commonly not treatable with purely quantum mechanical methods. A first step in this direction was made by Kaledin and Miller who worked out a time averaged version of semiclassical dynamics (TA SCIVR).[52] Further and more recent advances have been accomplished by means of Ceotto's multiple coherent semiclassical initial value representation (MC SCIVR),[53] which allows one to get accurate results on the basis of just a single tailored trajectory, and through the divide-and-conquer semiclassical initial value representation (DC SCIVR), which has so far permitted applications up to 60-atom systems.[54]

In this work on glycine we were able to perform AS-SCIVR calculations in full dimensionality (24 degrees of freedom). AS-SCIVR simulations are made of two steps. In the first part, the adiabatic switching procedure is performed starting from an initial harmonic quantization of zero-point-energy. In the second part, the atomic positions and momenta obtained from the adiabatic switching run are employed as starting phase-space conditions for the SC dynamics simulation.[51] SC calculations are based on Kaledin and Miller's working formula

$$I(E) = \frac{1}{(2\pi\hbar)^{N_{vib}}}\int d\boldsymbol{p}_0\int d\boldsymbol{q}_0\frac{1}{2\pi\hbar T}\times \left|\int_0^T dt'\, e^{i[S_{t'}(\boldsymbol{p}_0,\boldsymbol{q}_0)+\phi_{t'}(\boldsymbol{p}_0,\boldsymbol{q}_0)+Et']/\hbar}\langle g_{t'}(\boldsymbol{p}_0,\boldsymbol{q}_0)|\Psi\rangle\right|^2, \tag{4}$$

from which the semiclassically approximate quantum density of vibrational states, $I(E)$, can be calculated as a function of the vibrational energy, E. In principle the exact quantum mechanical evaluation of $I(E)$ may be obtained by Fourier transforming the survival amplitude of an arbitrary reference state $|\Psi\rangle$

$$I(E) = \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} dt\, e^{iEt/\hbar}\,\langle\Psi|e^{-i\hat{H}t/\hbar}|\Psi\rangle. \tag{5}$$

In practice the SC spectral density was obtained from Eq. (4) by means of a Monte Carlo phase-space integration of a function involving classical quantities. Specifically, $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$ are phase space variables providing the initial conditions for glycine trajectories. The trajectories were numerically integrated and instantaneous dynamical values collected to evaluate Eq. (4). These include the action

$$S_{t'} = \int_0^{t'} dt(T-V), \tag{6}$$

where T and V are the kinetic and potential energy respectively, the quantum overlap between the time-evolved coherent state $|g_{t'}(\boldsymbol{p}_0,\boldsymbol{q}_0)\rangle$ and the arbitrary quantum reference state $|\Psi\rangle$, and the phase $\phi_{t'}$ of the Herman-Kluk prefactor (commonly indicated as $C_t(\boldsymbol{p}_0,\boldsymbol{q}_0)$).[52] Coherent states are represented in coordinate space by the following equation

$$\langle\boldsymbol{q}|g_t(\boldsymbol{p}_0,\boldsymbol{q}_0)\rangle = \left(\frac{\Gamma}{\pi}\right)^{1/4}\exp\left[-\frac{\Gamma}{2}(\boldsymbol{q}-\boldsymbol{q}_0)^2+\frac{i}{\hbar}\boldsymbol{p}_0^T(\boldsymbol{q}-\boldsymbol{q}_0)\right]. \tag{7}$$

Γ is a matrix of width parameters which is usually chosen to be diagonal with elements numerically equal to the harmonic

frequencies of vibration. The Herman-Kluk prefactor can be obtained starting from the monodromy matrix elements (i.e. the instantaneous derivatives of $\boldsymbol{p}(t)$ and $\boldsymbol{q}(t)$ with respect to the initial phase-space conditions $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$)

$$C_t(\boldsymbol{p}_0,\boldsymbol{q}_0) = \sqrt{\left|\frac{1}{2}\left(\frac{\partial \boldsymbol{q}(t)}{\partial \boldsymbol{q}_0} + \Gamma^{-1}\frac{\partial \boldsymbol{p}(t)}{\partial \boldsymbol{p}_0}\Gamma - i\hbar\frac{\partial \boldsymbol{q}(t)}{\partial \boldsymbol{p}_0}\Gamma + \frac{i\Gamma^{-1}}{\hbar}\frac{\partial \boldsymbol{p}(t)}{\partial \boldsymbol{q}_0}\right)\right|} \tag{8}$$

Eventually, the SC estimates of vibrational eigenvalues were located in correspondence of peaks in the plot of $I(E)$ vs E. This is a semiclassical power spectrum, which is exact for the harmonic oscillator potential and returns a very accurate approximation to quantum mechanics otherwise. It provides information about the quantum vibrational eigenenergies, so SC spectroscopy is able to estimate not only the zero-point energy but also the quantum frequencies of excited states, including fundamentals and overtones. Recent and still in-progress developments of SC dynamics techniques also include the possibility to calculate eigenfunctions and IR spectra.[55–57]

III. RESULTS AND DISCUSSION

A. Conformers and Saddle Points

Figure 1 shows the structures of the eight conformers studied here, and Table III shows their energies from the PES in comparison to those optimized at the B3LYP/aVDZ and CCSD(T)-F12b/aVTZ levels of theory. The agreement is excellent, even when comparing to the higher-level CCSD(T)-F12 results. Tables SI-1 and SI-2 list the conformer energies and coordinates optimized using B3LYP/aVDZ.

TABLE III. Energies relative to the global minimum (in kcal/mol unless otherwise indicated in the Table) of each conformer from the PES or indicated electronic structure theories.

Conformer	PES (cm^{-1})	PES	B3LYP	CCSD(T)-F12 [a]
Conf 1	0.0	0.00	0.00	0.00
Conf 2	205.1	0.59	0.58	0.68
Conf 3	577.1	1.65	1.64	1.73
Conf 4	450.4	1.29	1.27	1.23
Conf 5	945.9	2.70	2.61	2.62
Conf 6	1719.4	4.92	4.91	4.80
Conf 7	2043.6	5.84	5.84	5.89
Conf 8	2174.3	6.22	6.25	6.06

[a] From ref. 16

The optimized saddle points determined in this work are listed with their energies and coordinates in Tables SI-3–SI-6. Table IV compares the energies of the saddle points for the fitted PES to those calculated for the B3LYP/aVDZ optimized structures. The agreement is quite good, particularly for the lower-energy saddle points. For the 6 saddle points whose energies are greater than 4000 cm^{-1} (SP 16, 23, 25, 37, 57, 48), the agreement is not as good.

TABLE IV. Saddle point energies relative to the global minimum from fitted PES and from B3LYP/aVDZ optimized structures. All energies are in cm^{-1}.

SP	PES	B3LYP
13	777.9	775.2
14	500.0	515.1
15	1733.2	1755.8
16	4013.1	4320.0
23	4548.8	4801.3
25	4781.0	5118.7
26	3543.7	3560.3
27	2055.6	2048.7
28	2831.8	2838.0
35	1367.1	1359.2
37	4864.8	5097.1
57	5282.8	5567.9
48	4516.0	4823.8
67	2843.8	2836.4
68	2185.0	2210.0

Harmonic frequencies for Conformers 1–8 based on the PES fit are shown in Tables SI-6 and SI-7, while similar information for the saddle points is provided in Tables SI-8–SI-12. A comparison of these frequencies to those for the optimized structures (at the B3LYP/aVDZ level) gave a mean absolute error (MAE). These values (in cm^{-1}) are shown in Table V. Similar to the energetics, the errors for the 6 high-energy saddle points (whose energy is larger than 4000 cm^{-1}) are relatively large, while for the other stationary points, the agreement between PES and B3LYP frequencies is very good.

TABLE V. Mean absolute errors (in cm^{-1}) for the harmonic frequencies calculated from the PES fit as compared to those for the optimized structures.

Geom	MAE	Geom	MAE
Conf 1	4.3	SP 23	13.2
Conf 2	3.5	SP 25	12.4
Conf 3	4.7	SP 26	8.5
Conf 4	6.2	SP 27	5.9
Conf 5	6.2	SP 28	7.9
Conf 6	8.0	SP 35	6.3
Conf 7	7.4	SP 37	17.7
Conf 8	6.1	SP 48	15.7
SP 13	4.1	SP 57	17.7
SP 14	5.2	SP 67	6.7
SP 15	7.3	SP 68	5.2
SP 16	17.4		

B. Diffusion Monte Carlo and Semiclassical Results

The vibrational zero-point energies (relative to the electronic energy of conformer 1) of each conformer, calculated using DMC and semiclassical simulations, are presented in Table VI. From a global point of view the molecule has just

one ZPE value, which is, of course, the lowest vibrational energy level allowed. However, if the several conformers are quite isolated from each other, then for each conformer a "local" ZPE value can be calculated.

TABLE VI. Zero-point energies (in cm^{-1}) of each conformer, obtained from DMC and semiclassical calculations. Values are relative to the electronic energy of the global minimum (Conf 1). For Conf 4, Conf 5, Conf 7, and Conf 8 SC calculations provide an excited state energy. Values in parenthesis are referred to the bottom of the corresponding well (see text for details).

Conformer	DMC	Semiclassical
Conf 1	17151 ± 5	17157
Conf 2	17430 ± 3	17439
Conf 3	17717 ± 4	17724
Conf 4	17150 ± 6	17571 (17121)
Conf 5	17720 ± 6	18105 (17736)
Conf 6	18843 ± 6	18831
Conf 7	17430 ± 4	19105 (17267)
Conf 8	18841 ± 4	19222 (18767)

An interesting observation is that some conformers share the same ZPE. In fact there are 4 pairs of conformers, Conf 1 & 4, Conf 2 & 7, Conf 3 & 5, and Conf 6 & 8; the two conformers in each pair have the same ZPEs. This indicates that the ground states of the two conformers in each pair are the same state, which is confirmed by computing the ground-state wavefunctions using the walkers from the DMC calculations. Our results show that the two conformers in each pair have the exact same ground-state wavefunctions, which are reported in Figure 6.

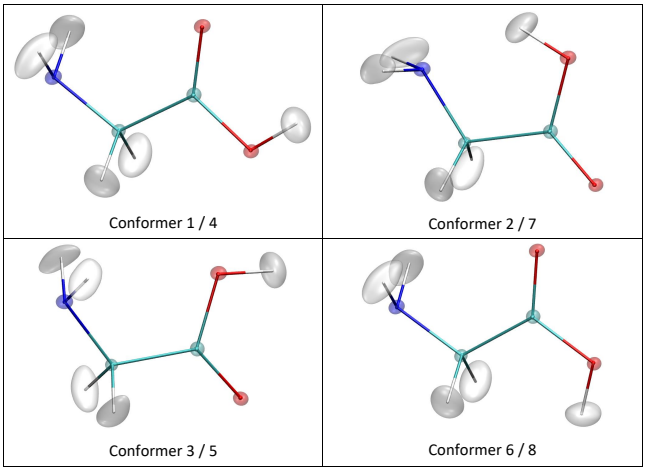

FIG. 6. Nuclear ground-state wavefunctions of all glycine conformers. Note that the two conformers in each pair have the same wavefunction.

The results can be explained by the fact that the barrier that separates the two conformers within the pair is low (for example, SP 27 is only 12 cm^{-1} above Conf 7; the highest barrier is between Conf 3 and 5, which is just 421 cm^{-1}). So the two conformers in each pair form an asymmetric double well and the barrier that separates these two wells is low. On the other hand, the barriers that separate two different double-wells are much higher, for example, SP 23, 25, 37, 57 are all more than 2500 cm^{-1} above Conf 7, and therefore the four double-wells are well separated and we get 4 different ZPEs. (An exception is that SP 13 is only 201 cm^{-1} higher than Conf 3; however, the transition between Conf 1 and 3 involves the motion of two O atoms, and since O is a quite heavy atom compared to hydrogen, this transition is unlikely in spite of the low barrier height.)

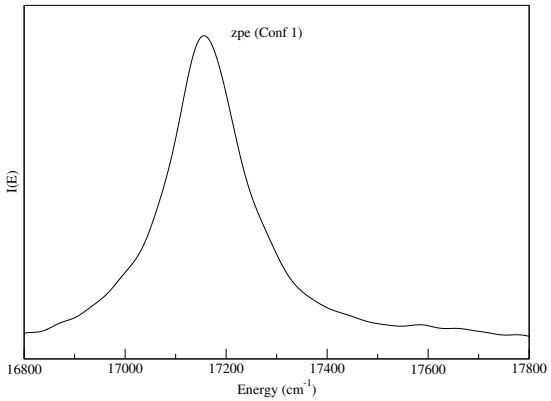

FIG. 7. Zero point energy power spectrum for Conf 1. The density of states I(E) is plotted against the vibrational energy.

Figure 7 reports the zero-point-energy portion of the calculated SC power spectrum for Conf 1. We note that (see Table VI) semiclassical ZPE estimates for Conf 1, Conf 2, Conf 3, and Conf 6 are in excellent agreement with DMC calculations and discrepancies are in the order of just a few wavenumbers. These conformers are the low energy members of the double wells. For the other conformers (Conf 4, Conf 5, Conf 7, and Conf 8) AS-SCIVR returns the energy of an excited state. This is expected because, as demonstrated by De Leon and Heller,[58] SC dynamics is able to yield the exact quantum eigenvalue even by running a single trajectory in case the trajectory is run at that exact eigenenergy. This important feature has led to the development of techniques (MC SCIVR, AS-SCIVR) able to overcome the issue of sampling the whole phase space making computationally prohibitive simulations affordable without loss in accuracy of results. In practice, the quantum energy level is not known a priori and a harmonic energy approximation to the target energy level is commonly employed. AS-SCIVR improves on the energy sampling of initial phase-space conditions with respect to the basic harmonic approximation, but when running AS-SCIVR starting from the higher energy conformer of any asymmetric double well the expected outcome is an excited state energy. As an exercise, for Conf 4, Conf 5, Conf 7, and Conf 8, we report in parenthesis in Table VI the energy values relative to the bottom of the corresponding well. We notice that in some cases (Conf 4 and Conf 5) we still get good estimates of the zero-

point energy, while in other instances (Conf 7 and Conf 8) the anharmonicity of the potential is clearly pointed out.

All AS-SCIVR simulations are based on 10,000 trajectories with a trajectory rejection rate of about 75-80% depending on the specific conformer. Each complete trajectory is 2500-step long and a time step of 10 a.u. is employed for a total evolution time of about 600 fs. A very accurate 4-th order symplectic integrator is adopted to ensure conservation of energy. All calculations are performed in full dimensionality.

IV. SUMMARY AND CONCLUSIONS

We have presented an accurate permutationally invariant PES for glycine suitable for quantum calculations. The potential energy surface of this small amino acid is quite complex and as many as 70,000 energy points and 20,000 Cartesian gradients had to be fitted to get to a reliable analytical surface. Investigation of the energy landscape by means of the common Newton minimization method has permitted us to identify 8 conformers and 15 saddle points connecting them.

For each conformer DMC calculations of zero-point energy and ground state wavefunction have been performed. This has allowed us to determine the presence of 4 asymmetric double-well structures (for a total of 8 conformers) each characterized by a specific ground state energy. In this way DMC results have been shown to be independent of the starting conformer within a specific double-well, while the simulations were clearly able to distinguish between different double wells. Consequently, even if the proper zero-point energy for glycine is certainly that one relative to the global minimum or Conf 1, the fact that the asymmetric double wells are basically isolated systems permitted us to define a local zero-point energy for each of the 4 double wells.

DMC outcomes have been corroborated by semiclassically approximated quantum dynamical simulations which provide zero-point energy estimates in very close agreement. This demonstrates the suitability of the constructed PES for stochastic and dynamical quantum simulations even at high energy. As expected, in absence of a full sampling of the phase space, SC zero-point energy estimates depend on the starting conformer and the trajectory energy. This implies that for an accurate semiclassical determination of the zero-point energy level one would need to identify the global minimum of the system under investigation. This task could be virtually impossible for very complex systems like, for instance, condensed phase or biochemical ones. However, in such systems the several local minima are generally separated by a very small amount of electronic energy (just a few wavenumbers) and zero-point energy estimates may still be obtained with good accuracy, as demonstrated in the case of glycine Conf 4 and Conf 5. Furthermore, one is generally interested in vibrational frequencies which are directly comparable to the experiment. In semiclassical spectroscopy they are calculated by difference with respect to the zero-point energy and so the discrepancies between values collected starting from different low-lying conformers are expected to be even smaller than they are for ZPE. This couple of observations leads to the remarkable consequence that in practice one can perform reliable AS-SCIVR calculations even in very complex environments by starting the dynamics from any of the several low-energy local minima of the PES.

In conclusion we believe that with this work we have provided the community with an accurate potential energy surface for an important amino acid. On the one hand this may allow for new calculations of interest for the astrochemistry community. On the other hand this surface is an example of a specialized PES for a small biomolecule, opening up the possibility to improve accuracy in the study of bioactive molecules with respect to the commonly employed but less specialized force fields.

V. SUPPLEMENTARY MATERIAL

The supplementary material file contains a correlation plot of configurations and the coordinates and harmonic frequencies for the stationary points of the surface. Fortran subroutines for the PES are also available.

VI. ACKNOWLEDGEMENTS

JMB thanks the Army Research Office, DURIP grant (W911NF-14-1-0471), for funding a computer cluster where most of the calculations were performed.

VII. DATA AVAILABILITY

Data supporting the findings of this study are available from the corresponding authors upon reasonable request.

[1] Y.-J. Kuan, S. B. Charnley, H.-C. Huang, W.-L. Tseng, and Z. Kisiel, Astrophys. J. **593**, 848 (2003).
[2] L. E. Snyder, F. J. Lovas, J. M. Hollis, D. N. Friedel, P. R. Jewell, A. Raemijan, V. V. Ilyushin, E. A. Alekseev, and S. F. Dyubko, Astrophys. J. **619**, 914 (2005).
[3] P. A. Jones, M. R. Cunningham, P. D. Godfrey, and D. M. Cragg, Mon. Not. R. Astron. Soc. **374**, 579 (2007).
[4] M. R. Cunningham, P. A. Jones, P. D. Godfrey, D. M. Cragg, I. Bains, M. G. Burton, P. Calisse, N. J. M. Crighton, S. J. Curran, T. M. Davis, J. T. Dempsey, B. Fulton, M. G. Hidas, T. Hill, L. Kedziora-Chudczer, V. Minier, M. B. Pracy, C. Purcell, J. Shobbrook, and T. Travouillon, Mon. Not. R. Astron. Soc. **376**, 1201 (2007).
[5] M. Lattelais, F. Pausat, J. Pilme, Y. Ellinger, and C. Ceccarelli, Astron. Astrophys. **532**, A39:1 (2011).
[6] A. G. Csaszar, J. Am. Chem. Soc. **114**, 9568 (1992).
[7] R. M. Balabin, J. Phys. Chem. Lett. **1**, 20 (2010).
[8] R. M. Balabin, Phys. Chem. Chem. Phys. **14**, 99 (2012).
[9] P. D. Godfrey and R. D. Brown, J. Am. Chem. Soc. **117**, 2019 (1995).
[10] V. Kasalová, W. D. Allen, H. F. Schaefer Iii, E. Czinki, and A. G. Császár, J. Comput. Chem. **28**, 1373 (2007).
[11] S. G. Stepanian, I. D. Reva, E. D. Radchenko, M. T. S. Rosado, M. L. T. S. Duarte, R. Fausto, and L. Adamowicz, J. Phys. Chem. A **102**, 1041 (1998).
[12] V. Barone, M. Biczysko, J. Bloino, and C. Puzzarini, J. Chem. Theory Comput. **9**, 1533 (2013).
[13] G. Bazsó, G. Magyarfalvi, and G. Tarczay, J. Mol. Struct. **1025**, 33 (2012).
[14] G. Bazsó, G. Magyarfalvi, and G. Tarczay, J. Phys. Chem. A **116**, 10539 (2012).
[15] S. Coussan and G. Tarczay, Chem. Phys. Lett. **644**, 189 (2016).

[16]E. M. Orjan, A. B. Nacsa, and G. Czakó, J. Comput. Chem. **41**, 2001 (2020).
[17]A. G. Csaszar and A. Perczel, Prog. Biophys. Mol. Biol. **71**, 243 (1999).
[18]B. Brauer, G. M. Chaban, and R. B. Gerber, Phys. Chem. Chem. Phys. **6**, 2543 (2004).
[19]M. Senent, M. Villa, R. Dominguez-Gomez, and A. Fernandez-Clavero, Int. J. Quantum Chem. **104**, 551 (2020).
[20]O. Bludsky, J. Chocholousova, J. Vacek, F. Huisken, and P. Hobza, J. Chem. Phys. **113**, 4629 (2000).
[21]F. Gabas, R. Conte, and M. Ceotto, J. Chem. Theory Comput. **13**, 2378 (2017).
[22]J. Zhu, V. Q. Vuong, B. G. Sumpter, and S. Irle, MRS Commun. **9**, 867–873 (2019).
[23]M. Stöhr, L. Medrano Sandonas, and A. Tkatchenko, J. Phys. Chem. Lett. **11**, 6835 (2020).
[24]C. Chen, B. Braams, D. Y. Lee, J. M. Bowman, P. L. Houston, and D. Stranges, J. Phys. Chem. Lett. **12**, 1875 (2010).
[25]Y.-C. Han, A. Sharma, and J. M. Bowman, J. Chem. Phys. **136**, 214313 (2012).
[26]R. Conte, P. L. Houston, and J. M. Bowman, J. Phys. Chem. A **117**, 14028 (2013).
[27]R. Conte, P. L. Houston, and J. M. Bowman, J. Phys. Chem. A **119**, 12304 (2015).
[28]C. Qu and J. M. Bowman, J. Phys. Chem. A **120**, 4988 (2016).
[29]H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Scchutz, WIREs Comput. Mol. Sci. **2**, 242 (2012).
[30]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, *et al.*, https://www.molpro.net (2019).
[31]J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
[32]J. B. Anderson, J. Chem. Phys. **65**, 4121 (1976).
[33]I. Kosztin, B. Faber, and K. Schulten, Am. J. Phys. **64**, 633 (1996).
[34]A. Pandey and B. Poirier, J. Chem. Phys. **152**, 214102 (2020).
[35]A. Pandey and B. Poirier, J. Phys. Chem. Lett. **11**, 6468 (2020).
[36]G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R* (Springer, New York, 2013).
[37]P. L. Houston, R. Conte, C. Qu, and J. M. Bowman, J. Chem. Phys. **153**, 024107 (2020).
[38]C. Qu and J. M. Bowman, J. Chem. Phys. **150**, 141101 (2019).
[39]A. Nandi, C. Qu, and J. M. Bowman, J. Chem. Theory Comput. **15** (2019).
[40]R. Conte, P. L. Houston, and J. M. Bowman, J. Chem. Phys. **140**, 151101 (2014).
[41]R. Conte, C. Qu, and J. M. Bowman, J. Chem. Theory Comput. **11**, 1631 (2015).
[42]C. Qu, R. Conte, P. L. Houston, and J. M. Bowman, Phys. Chem. Chem. Phys. **17**, 8172 (2015).
[43]R. Conte, C. Qu, P. L. Houston, and J. M. Bowman, J. Chem. Theory Comput. **16**, 3264 (2020).
[44]C. Qu, R. Conte, P. L. Houston, and J. M. Bowman, Phys. Chem. Chem. Phys., , in press (2020).
[45]F. Gabas, R. Conte, and M. Ceotto, J. Chem. Theory Comput. **16**, 3476 (2020).
[46]M. Wehrle, M. Sulc, and J. Vanicek, J. Chem. Phys. **140**, 244114 (2014).
[47]W. H. Miller and T. F. George, J. Chem. Phys. **56**, 5637 (1972).
[48]E. J. Heller, J. Chem. Phys. **75**, 2923 (1981).
[49]M. F. Herman and E. Kluk, Chem. Phys. **91**, 27 (1984).
[50]K. G. Kay, J. Chem. Phys. **101**, 2250 (1994).
[51]R. Conte, L. Parma, C. Aieta, A. Rognoni, and M. Ceotto, J. Chem. Phys. **151**, 214107 (2019).
[52]A. L. Kaledin and W. H. Miller, J. Chem. Phys. **118**, 7174 (2003).
[53]M. Ceotto, S. Atahan, G. F. Tantardini, and A. Aspuru-Guzik, J. Chem. Phys. **130**, 234113 (2009).
[54]M. Ceotto, G. Di Liberto, and R. Conte, Phys. Rev. Lett. **119**, 010401 (2017).
[55]M. Micciarelli, R. Conte, J. Suarez, and M. Ceotto, J. Chem. Phys. **149**, 064115 (2018).
[56]M. Micciarelli, F. Gabas, R. Conte, and M. Ceotto, J. Chem. Phys. **150**, 184113 (2019).
[57]C. Aieta, M. Micciarelli, G. Bertaina, and M. Ceotto, Nat. Commun. **11**, 4348 (2020).
[58]N. De Leon and E. J. Heller, J. Chem. Phys. **78**, 4005 (1983).