JCTC Journal of Chemical Theory and Computation

Spectroscopy and Excited States

Efficient Generation of Permutationally Invariant Potential Energy Surfaces for Large Molecules

Riccardo Conte, Chen Qu, Paul L. Houston, and Joel M Bowman

J. Chem. Theory Comput., **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.0c00001 • Publication Date (Web): 26 Mar 2020

Efficient Generation of Permutationally Invariant Potential Energy Surfaces for Large Molecules

Riccardo Conte,*,† Chen Qu,‡ Paul L. Houston,*,¶ and Joel M. Bowman*,§

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

‡*Department of Chemistry & Biochemistry, University of Maryland, College Park, Maryland 20742, U.S.A.*

¶*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*

§*Department of Chemistry and Cherry L. Emerson Center for Scientific Computation, Emory University, Atlanta, Georgia 30322, U.S.A.*

E-mail: riccardo.conte1@unimi.it; plh2@cornell.edu; jmbowma@emory.edu

Phone: +1 404 727-6592

Abstract

An efficient method is described for generating a fragmented, permutationally invariant polynomial basis to fit electronic energies and, if available, gradients for large molecules. The method presented rests on the fragmentation of a large molecule into any number of fragments while maintaining the permutational invariance and uniqueness of the polynomials. The new approach improves on a previous one reported by Qu and Bowman by avoiding repetition of polynomials in the fitting basis set and speeding up gradient evaluations while keeping the accuracy of the PES. The method is demonstrated for $CH_3-NH-CO-CH_3$ (N-methyl acetamide) and NH_2-CH_2-COOH (glycine).

1 Introduction

Developing high-dimensional, *ab initio*-based potential energy surfaces (PESs) is a long-term and currently very active area of theoretical and computational research. In the past 15 years, significant progress has been made in the development of non-parametric, machine learning approaches to fit large data set of electronic energies for polyatomic molecules and clusters.[1–15] These approaches include several that have been extensively applied to date. They are permutationally invariant polynomials (PIPs),[1,14] Neural Networks (NN),[3–6,16–19] NN with PIP inputs,[10–13,20,21] Gaussian Process regression (GPR),[8,15,22] and GPR with PIP inputs.[23] There is a major motivation to extend these methods to large molecules of interest in chemistry, biochemistry and materials science. However, there are significant challenges in doing this for the various approaches.

Our group has developed the PIP approach over the last 15 years to represent high dimensional PESs of molecules and molecular clusters with numerous applications.[1,14,24–26] This method makes use of Morse variables, which are transformed internuclear distances. In 2003 the approach was first applied to the CH_5^+ cation to construct a global PES that is invariant with respect to the 120 possible permutations of the five equivalent H atoms.[27]

Generally the data sets consist of 10^4–10^5 scattered electronic energies, typically obtained at the CCSD(T) level of theory. Here "scattered" means non-grid based energies; typically the data are from a number of low-level direct dynamics trajectory calculations run at different total energies and in some instances from very different initial configurations. More details can be found elsewhere.[1] This approach has been applied to obtain PESs for more than 50 molecules, including reactive systems, and molecular clusters.[14] Of particular interest to this paper, there are PESs for 7, 8, 9, and 10 atom systems, e.g., CH_3CHO, with many minima and saddle points,[28] CH_3CHOO,[29] malonaldehyde,[30] and the ten-atom formic acid dimer,[31] respectively.

There are bottlenecks for the PIP methods as the molecular size increases, and these have been discussed previously.[32] To recap these briefly, the inputs are the values of all the Morse variables and the number of variables grows as order N^2. For the PIP-NN approach the input is the minimum number of PIPs to correctly describe the symmetry of the molecule. This number is larger than the number of Morse variables and grows rapidly with the number of atoms. The growth in the size of PIP bases with the number of atoms depends on both the total polynomial order and the order of symmetric group.[1] As noted above, the PIP approach has been applied for molecules with as many as 10 atoms and this value has been cited in the literature as the practical limit for the PIP, while the PIP-NN approach has been applied to the seven atom $OH+CH_4$ reaction.[33]

The "ten-atom limit" using the PIP approach was broken for the 12-atom *trans*-*N*-methyl acetamide (*trans*-NMA).[32] The major point of that paper, which is preliminary to the present one, was to describe a fragmented PIP approach able to extend the PIP method to larger molecules. As an aside, we mention that the 10-atom limit was also just exceeded using PIPs in the construction of an interaction surface for the CH_4-H_2O-H_2O system,[34] and in a calculation of anharmonic rovibrational partition functions including torsional motion.[35]

As for ML approaches, we note that they have been extensively developed and applied mainly in the context of materials chemical physics.[4,7,36] These approaches have in common

the development of either a NN or GPR of the atomic energy of each atom. Thus, these methods have been extended to large numbers (e.g., hundreds) of atoms at the cost of a large number of NN or GP evaluations. A recent paper comparing important aspects of atom-based NN and PIP-NN approaches for several molecules[37] indicates that for "small" molecules the PIP approach is probably the preferred one.

There is, we reasoned,[32,38] a regime for molecules with more than 10 atoms and probably less than hundreds of atoms where the PIP approach could be extended. The basic observation that enables this extension is that Morse variables go asymptotically to zero at large internuclear distances. Thus for large molecules many Morse variables are essentially zero and thus any PIP basis function containing these variables is zero and can be dropped from the total basis. This observation allows one to fragment the larger target molecule into smaller moieties for which PIP basis sets can be generated efficiently.[32,38] The pruning approach is indeed successful and it was used with recent, extended MSA software that includes gradient data for fits.[39] However, several issues with the algorithm used were noted. These included redundant terms in the basis and also the cost of evaluating gradients, as described in detail below.

New software, described in this paper, solves these two problems. To begin, recall that the PIP approach represents the potential as follows, using compact notation:

$$V = \sum_{i=1}^{n_p} c_i p_i, \tag{1}$$

where c_i are coefficients, p_i are permutationally invariant polynomials, denoted as PIPs, (the basis set) and n_p is the total number of polynomials for a given maximum polynomial order. The p_i are typically functions of Morse variables, which themselves are functions of the interatomic distances, $r_{\alpha,\beta}$ (by the usual exponential relationship $\exp(-r_{\alpha,\beta}/\lambda)$, where λ is commonly chosen to be equal to 2 Bohr). Following the practice of the MSA software, Morse variables are in this work denoted by x_l. We further stipulate that the basis functions should be chosen to maintain permutational invariance among identical atoms, or at least

some of them. The linear coefficients are obtained using standard least squares fits to large datasets of electronic energies at "scattered" geometries.

In standard PIP approaches, the computational issue arises when the basis set for the parent molecule is completely unwieldy – too big to be useful in practical terms, either because calculating the proper basis set takes too long or because the number of coefficients is so large that the least squares optimization becomes problematic. The size of the basis varies in a complicated and generally non-linear way with respect to the number of Morse variables, the maximum polynomial order, and the order of the symmetric group.[1] This growth in the size of the PIP basis is the basic consideration in stating the 10-atom limit for the approach.

However, as noted above, the fragmented basis approach is an effective way to break this limit. Clearly, by fragmenting a parent molecule into groups of smaller molecular moieties the basis for each smaller moiety can be calculated rapidly and then combined with those of other fragments to provide a compact and hopefully still precise representation of the potential energy.[32] To be specific, consider a simple example of a 5-atom molecule with atoms labeled as 1-5 and a scheme in which the molecule is fragmented into three fragments, say {1,2,3}, {2,3,4}, {3,4,5}. In this 3-fragment scheme the potential is given compactly by

$$V = \sum_i c_i \, p_i(\boldsymbol{x}_1, \boldsymbol{m}_1) + \sum_j c'_j \, p'_j(\boldsymbol{x}_2, \boldsymbol{m}_2) + \sum_k c''_k \, p''_k(\boldsymbol{x}_3, \boldsymbol{m}_3), \tag{2}$$

where $\{p\}$, $\{p'\}$, and $\{p''\}$ are PIP bases for the n^{th} fragment, n = 1,2,3, $\{c\}$, $\{c'\}$, $\{c''\}$ are the corresponding linear coefficients, $\boldsymbol{x}_n$ represent the set of corresponding Morse variables, and $\boldsymbol{m}_n$ indicate a set of monomials built from the Morse variables. Morse variables between atoms 1 and 4, atoms 1 and 5 and atoms 2 and 5 are assumed to be zero and hence not in the fragmented bases. In this example there are some Morse variables in common among the fragments, and it should be clear that there are indeed some redundant basis functions in this expression in terms of common Morse variables.

These issues were pointed out previously;[32,38] however, they were not "fatal" ones, because the linear least squares method used was able to deal with a modest number of identical basis functions. Nevertheless, there is compelling motivation to eliminate these redundant basis functions and thereby reduce the size of the basis. We do note the redundant-term issue is similar to an issue identified earlier for developing PIP representations of interaction potentials that should rigorously vanish in asymptotic regions where there is no interfragment interaction. In that case the issue concerned basis functions involving Morse variables of fragments that do not go to zero at large internuclear distances where rigorously there is no inter-fragment interaction. An effective "empirical" pruning procedure was then employed to eliminate such basis functions and applied to several systems.[34,40,41]

An "empirical" pruning approach is also developed here as a post-processing task performed on the standard complete PIP bases of the fragments. The entire computational approach is described in detail in the next section followed by illustrations for N-methyl acetamide (NMA) and glycine. These are not large molecules; however, they serve to test the effectiveness of various fragmentation approaches, as full PIP basis fits can be done for these molecules. NMA does contain a low-barrier methyl rotor and so this example should be relevant to many large molecules with methyl rotors. We will make further comments and suggest some guiding principles on how fragmentation of the basis might be applied to challenging molecules that undergo isomerization and/or unimolecular break-up.

2 Computational Methods

Following the work of Xie and Bowman,[42] Xie has provided software[43] based on a monomial symmetrization approach (MSA) that generates the permutationally invariant basis set for many permutational symmetries and polynomial orders. The monomials are functions of the Morse variables and are combined to form permutationally invariant polynomials up to a select order in a clever and efficient system that allows each polynomial to be recursively

determined from simple sums of products of previously calculated monomials or polynomials. In this way, the specific polynomials $p_i = p_i(\mathbf{x}, \mathbf{m})$ in Eqs. (1) and (2) are calculated from the monomials, $\mathbf{m}$, as well as from other polynomials p_j with $j < i$ rather than from the repetitive and more complicated evaluation of functions of the Morse variables, $\mathbf{x}$. For example, the MSA program may determine that p_{128} may be written as $p_{128} = p_{32}p_{61} + m_4$. At the time when p_{128} is evaluated, the components p_{32}, p_{61} and m_4 have already been calculated, making the computation of p_{128} very efficient. A Perl program, `postemsa.pl`, is used to convert the recursion relationships determined by the MSA software to a Fortran program `bemsa*.f90`, where * stands for the permutational symmetry numbers and the desired fit order. In addition to subroutines that efficiently calculate the monomial and polynomial basis function values, the Fortran program also includes a function that uses these to calculate the energy based on a list of coefficients and the inputs of Cartesian coordinates for each atom rigorously entered in the order of the chosen permutational symmetry. The coefficients are determined by using the `bemsa*.f90` program to fit a number of known energies for different geometries. (This software was recently extended to obtain gradient of the energy as well.[44]) As mentioned in the Introduction, it is often quite time consuming to generate the MSA output, particularly for large molecules. For example, using a single Intel xeon16 (1.2 GHz) processor, our calculation of a basis for 10-atom glycine, with maximum total polynomial order of 4 took more than 12 days. This basis contains 46654 polynomials. Fragmentation of the parent compound can speed up this process substantially and provide an efficient basis set that allows for faster energy evaluation without sacrificing accuracy. This has been demonstrated already.[32,38] However, as anticipated, an issue of replicated basis functions was noted.

Owing to the efficient recursive algorithm in the MSA software it is not trivial to identify and eliminate replicated basis functions, and then re-order the basis. We do that here with new Mathematica software.[45] This software starts with the `bemsa*.f90` files created by the MSA software for each of the fragments and generates a list of the unique Morse variables,

monomials, and polynomials (here denoted collectively as "$(\mathbf{x}, \mathbf{m}, \mathbf{p})$") while maintaining permutational invariance. Mathematica was chosen because it has both the tools to do complicated string manipulations and the ability to turn strings of texts into commands that can be immediately evaluated. Thus, the MSA Fortran program can be read as a string, the strings can be converted to Mathematica format, and the Mathematica format can be then be evaluated. Conversely, once a new recursive scheme has been developed combining the fragment basis sets and eliminating duplicate variables, it can then be converted to a text string, translated into Fortran format, and output as the Fortran program (`DuplicatesDeletedbemsa.f90`). Additionally, if desired, the new recursive scheme can efficiently incorporate fitting of gradients if these are available. In summary, there are two main goals of the Mathematica program: 1) to increase efficiency by eliminating duplicated Morse variables, monomials and polynomials in the basis set of the ensemble of fragments, and 2) to provide an efficient method for calculating derivatives of the energy as a function of the Cartesian coordinates. We outline the methods for achieving these goals next.

As for point 1), we developed two strategies for deleting duplicate Morse variables, monomials, and polynomials: a sequential method, and a pairwise method. Further details of the two methods are included in Section S1 of the Supporting Information.

The second goal of our program is to provide an efficient approach for calculating derivatives of the energy with respect to all Cartesian coordinates of all atoms at the geometry corresponding to the energy. The straightforward method for doing this is relatively easy but inefficient; that is, one mechanically takes the derivatives of all monomials (assigned to variables dm_i) and polynomials (assigned to variables dp_i) and uses these to evaluate the derivative of the potential energy with respect to each of the Cartesian coordinates of the atoms. This is the method used successfully by Qu and Bowman[32] and by Nandi, Qu and Bowman.[38] One main reason why the inefficiency comes about is that, while there are up to $N(N-1)/2$ interatomic distances (where N is the number of atoms in the parent), the coordinates of a given atom occur only in $N-1$ of these. Thus, most of the derivatives

with respect to a Cartesian coordinate of a particular atom are zero; there is no reason to calculate them, especially because each calculation involves evaluating many members of the basis set. Specifically, if $a_{j,k}$ with $j = 1, N$ and $k = 1, 3$ represents the k^{th} Cartesian component of the j^{th} atom, then the derivative of Eq. (1) can be written as

$$\frac{dV}{da_{j,k}} = \sum_{i=1}^{n_c} c_i \sum_l \frac{dp_i}{dx_l} \frac{dx_l}{da_{j,k}}, \tag{3}$$

where l enumerates the Morse variables. If the Morse variable x_l do not depend on $a_{j,k}$ then $\frac{dx_l}{da_{j,k}}$ is zero and the l-th contribution to the second sum does not need to be evaluated. Our program avoids these "zero" calculations by branching to a separate calculation of Eq. (3) for each atom j. These separate calculations are initiated by setting all variables dp_i to zero and then calculating only the terms in the second sum for which $\frac{dx_l}{da_{j,k}} \neq 0$ for at least one combination of l, j and k.

Prior to running the Mathematica program one needs to assign a numbering scheme to the parent molecule; each atom should receive an integer number between 1 and N, the total number of atoms in the parent. The ordering of the atoms can be whatever is convenient, but, once established, that numbering needs to be consistent throughout the program and the simulations. The atoms of the fragments need to have the same numbers as they do in the parent, and the order in which the geometries are entered as input to the final Fortran program for calculating the energies and derivatives must be the same as the numerical order of the atoms. The inputs and outputs of the Mathematica program are described in Section S2 of the Supporting Information. A Mathematica Notebook showing examples and a Wolfram Language Function set are also available in the Supporting Information

An important point concerns the selection of the permutational symmetry for the fragments in input (3); it can rarely be the highest symmetry for the fragment considered individually. Consider a parent molecule with a fragment that has among others one atom A and two atoms B. Considered by itself, this fragment would produce a monomial in which the two AB bonds were treated as symmetrical and interchangeable. Now consider another

fragment of the same parent that has among others the same atom A and one of the atoms B. Considered by itself, this fragment would produce a monomial that treats the AB bond without any symmetry; it would not be interchangeable with any other bond. When these two fragments are combined, the composite basis set could not have permutational symmetry concerning the two B atoms. This leads to the following rule: In order to maintain permutational invariance for the final basis set, atoms that are assigned to permute with one another must appear together whenever they appear in any of the fragments.

We end this section by noting that once one has a fit, generated either from the original MSA software or from the new MSA_FRAG software here introduced, it is relatively easy to either add or subtract polynomials in order to increase the accuracy or increase the speed, respectively. The principle for selecting the polynomials is as follows. First, determine the maximum value of all the Morse variables among all the geometries in the *ab initio* data set. In order to add polynomials for increased accuracy, consider all those polynomials already used in the existing fit, make all combinations of these up to the desired order, discard those combinations already in use, evaluate all the new combinations using the maximum values of the Morse variables, and add those combinations to the basis set that have the largest values. In order to delete polynomials for increased speed, evaluate all the existing polynomials using the maximum values of the Morse variables and discard as many as desired starting from the lowest and moving to the highest. In the former case, the new set of polynomials will need to be examined for (new) duplicates, and in both cases the polynomials and monomials will need to be renumbered. The new set will retain any permutational invariance that was present in the original set. It is important to point out that for large molecules it is typically not possible to generate the MSA output for the parent compound; one needs to fragment the molecule to get the number of original polynomials (coefficients) down to a manageable number.

3 Results and Discussion

The fragmentation approach[32] is relatively new, so that at this point determination of the best fragmentation approach is still an art rather than a science. One objective of this paper is to make the approach easier for others to use so that collective intuition can be developed. Nonetheless, a few principles are clear, although these involve more mathematical intuition than chemical. The goal is to include those Morse variables with large values and to exclude those that will always have small values. Recall that the Morse variables have values between 1 and 0 and fall off exponentially with the distance between the two atoms. First, we note that there is no particular need for the fragments to be connected in a way that reflects the parent molecule; what one seeks is to have those atoms that are close to one another (and whose pairs have relatively large Morse values) included in at least one fragment. Morse variables involving distant atoms will be small, and these can be omitted by never including the relevant atoms in the same fragment. We note, however, that even if all Morse variables are included, the number of monomials and polynomials in the fragmented, permutationally invariant basis set will still be smaller in fragmentation than in the parent compound because the fragmentation reduces the cross terms between these variables. If one desires particular cross terms, then the relevant atoms for the Morse variables should be included in at least one fragment. Finally, fragments will often have atoms in common, atoms that have large Morse variables with others in each fragment, even though other atoms between the fragments may not have large Morse values. The overlap is often needed, and should not be shunned. The Delete Duplicates program ensures that Morse variables, monomials, and polynomials based on the same pair of atoms are not included more than once.

3.1 $CH_3{-}NH{-}CO{-}CH_3$ (N-methyl Acetamide, NMA)

In order to test the outcomes of our program we chose to calculate results for $CH_3{-}NH{-}CO{-}CH_3$ (NMA) because this molecule had already been studied by Qu and Bowman and results

on accuracy were available from their work.[32] They reported results for the full parent molecule, for two fragments, $CH_3-NHC-CO$ and $C-NH-CO-CH_3$, and for three fragments, CH_3-NH-C, $N-CO-CH_3$, and $C-NH-CO-C$. We also here report results for a 5-fragment calculation where the fragments are CH_3-NH-C, $N-CO-CH_3$, $C-NH-CO-C$, H_3-O-C, $C-H-H_3$. We followed the same numbering scheme, shown in Fig. 1, and used exactly the same *ab initio* data set for fitting as Qu and Bowman.[32] The data set consisted of 3000 energies and $3000 \times 36 = 108000$ Cartesian gradient components. Both energies and gradients in the data set have been fitted.

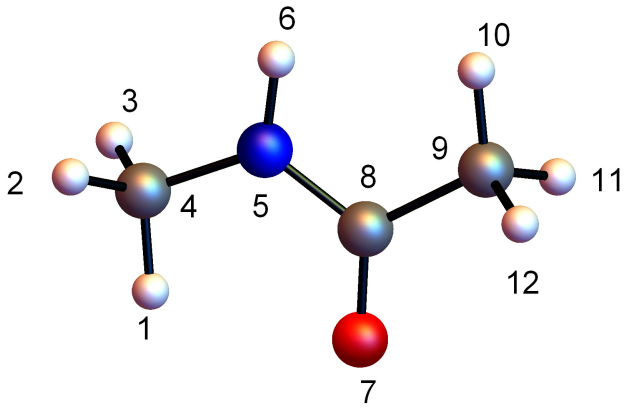

Figure 1: Numbering scheme used in text for NMA. H, C, O, and N atoms are white, gray, red, and blue, respectively.

By the example of NMA, we also hope to clarify issues concerning inputs (3) and (4) by considering in detail the permutational symmetry and atom entries for the two-fragment and three-fragment cases. In the two-fragment case, we have CH_3NHCOC and $CNHCOCH_3$. In order to maintain permutational invariance for the composite fragmented system, the H atoms on either methyl can be assigned to permute with one another because they always appear together, but those on the two methyls may not be assigned to permute between one another because one methyl is missing from each fragment. By our rule (see above), it would be possible to allow the three carbons in the set of common atoms to permute with one another because they appear in both fragments, but, we may not allow the H on the N to permute with any other H atoms because in each fragment some of the other H atoms are

missing. Following Qu and Bowman, we choose to ignore the carbon permutation, so that only the three H atoms on the methyl group of either fragment permute with one another; there is no permutation of H atoms between the two ends. The permutational symmetry for each fragment is thus {3,1,1,1,1,1,1}, and a possible atom listing is {1,2,3,4,5,6,7,8,9} for the first fragment and {10,11,12,4,5,6,7,8,9} for the second fragment. The notation here indicates that in the first fragment, for example, the atoms 1,2, and 3, the hydrogens on the left-hand methyl, permute among one another and that none of the other atoms permutes. Note that there are many possible atom listings because the ordering within one symmetry element is irrelevant and the ordering between groups of the same symmetry is also irrelevant. Thus, for the first fragment the atom listing {2,1,3,5,9,7,4,8,6} would also work. The two-fragment case essentially assumes that there is no interaction between the hydrogens on one end and those on the other; thus 9 Morse variables are neglected.

For the three-fragment case, we have CH_3−NH−C, N−CO−CH_3, and C−NH−CO−C. In order to maintain permutational invariance for the composite fragmented system, we may allow the H atoms on either methyl to permute with one another, since they always appear together when they are present, but no permutation may be assigned between the H atoms on opposite methyls, as in the 2-fragment case. The carbons may not be allowed to permute, because the first fragment has the left and middle one, which do not appear in fragment 2, while the second fragment has the middle and right one, which do not appear in fragment 1. The H on the NH is similarly not assigned to permute with any other H atoms, because they are not present, for example, in fragment 3. The permutational symmetry for the first and second fragments is {3,1,1,1,1}, with atom assignments of, for example, {1,2,3,4,5,6,8} and {10,11,12,5,7,8,9}, while the third fragment has symmetry {1,1,1,1,1,1} with an atom assignment of, for example, {4,5,6,7,8,9}.

We note that Nandi et al.[46] have recently extended the previous study of the NMA PES to the cis- isomer and transition states. They used the above two-fragment case and different three-fragment case which consists of the previous and present two-fragment case plus a non-

overlapping fragments consisting of the the three H atoms on each methyl group. This six-atom fragment basis was denoted as {3,3}. This work demonstrates that the fragmentation approach can describe isomerization.

Table 1: NMA Results. All RMS energies are in cm^{-1}.

	ref. 32	This Wk.	ref. 32	This Wk.	ref. 32	This Wk.
Fit Order	3	3	3	3	3	3
Frags	full	2	2	3	3	5
Morse vars	66	57	57	45	45	57
Polys	8040	5240	6056	1806	1974	1936
RMSE	26.8	34.3	34.3	148.9	148.9	93.1
RMSG	54.7	67.4	67.4	171.8	171.9	141.8
Time/s	6.611	2.450	3.281	0.830	1.056	0.914

Table 1 shows the results of our work in comparison to those of the previous study.[32] For the full molecule, there are 66 Morse variables and 8040 polynomials. The root-mean-square errors for the energies and gradients, as compared to the *ab initio* data set are provided. The times listed are those for evaluating 3000 energies and 3000 $\times$ 36 gradients on an intel i7 (2.7 GHz) processor. The 2-fragment results are given in columns 3 and 4. The results from the previous work[32] take about half the evaluation time as the full set. The results of our basis set and derivative evaluation method provide exactly the same RMSE and RMSG but take only about 75% as much time as the previous 2-fragment calculation. There are 57 Morse variables in both cases, but there are 13.5% fewer polynomials in our basis set due to the deletion of duplicates. Similarly, for the 3-fragment case the RMSE and RMSG results are identical, but the size of our basis set is 8.5 % smaller and the time required is 24% lower. The 5-fragment result provides a basis set that is between that of the 2-fragment and 3-fragment ones, with correspondingly intermediate RMS error values and timings.

It appears from these tests that our program has achieved the desired results of deleting duplicates and accelerating the calculation. In all cases, our program tested the results to be

sure that there were no remaining duplicated monomials or polynomials, and it also tested to confirm that the basis set maintained permutational invariance.

3.2 NH_2-CH_2-COOH (Glycine)

Glycine is the smallest amino acid and one of the building blocks of proteins. Its biological importance as a precursor of life has triggered the search for it in the interstellar medium through exprimental spectroscopic investigations which require theoretical validation.[47] For this reason, different theoretical approaches have been undertaken in the attempt to clarify glycine spectral features and describe accurately its elusive isomers and complicated potential energy surface. The surface and frequencies of vibration have been described in several ways ranging from reduced-dimensional models[48] to harmonic approximations,[49] from semiempirical electronic structure methods[50] to static and dynamical *ab initio* approaches.[51–53] In spite of all the interest around this amino acid, to best of our knowledge an accurate (and possibly permutationally invariant) potential energy surface is missing, probably due to its topological complexity and the non-trivial number of atoms to deal with. In the following we will contribute to partially fill this gap by focusing on the global minimum well.

The glycine minimum energy structure and the (arbitrary) numbering used here are shown in Fig. 2. The equilibrium geometry is of C_s symmetry and often conformationally referred to as an *all-trans* (ttt) structure,[54,55] with reference respectively to the relative positions of hydrogens (atoms 2-5) in the H-N-C-H bond chain (2-1-4-5), NH_2 (1-2-3) and OH (9-10) groups with respect to the C-C bond (4-7), and the hydroxy hydrogen (10) relative to the C atom (4) with respect to the C-O bond involving the hydroxy group (7-9).

We considered six potential energy surfaces for glycine, three with basis set polynomials to 3^{rd} order and three with basis set polynomials to 4^{th} order. In both 4^{th}-order and 3^{rd}-order potential energy surfaces we considered basis sets corresponding to the complete molecule, to 2 fragments, and to 3 fragments. For each surface, the minimum energy geometry was determined by a gradient method and the harmonic vibrational frequencies were calculated

through diagonalization of the mass-scaled Hessian matrix of the potential at the equilibrium geometry.

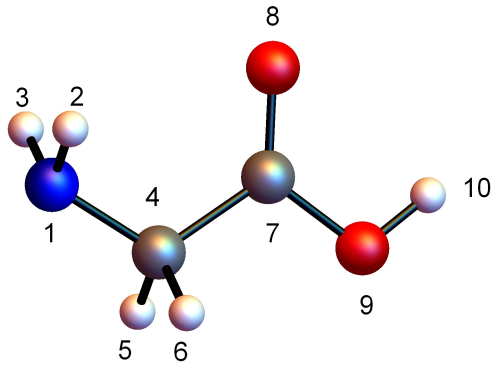

Figure 2: Numbering scheme used in the text for Glycine. H, C, O, and N atoms are white, gray, red, and blue, respectively.

The dataset used for fitting the surfaces consists of 3100 energies and $3100 \times 30 = 93000$ gradient components. Three thousand geometries were chosen from *ab initio* molecular dynamics simulations at energies ≤ 10000 cm^{-1}, while 100 geometries were chosen randomly on a grid centered at the temporary minimum located on the surface fitted to the preliminary 3000 geometries. A histogram of the 3100 energies along with the approximate energies of the minimum energy structure, the low-energy isomers, and the transition states between them is shown in Fig. 3. The molecular dynamics was performed under conditions similar to those used previously for NMA,[32] and the Molpro calculations were performed using DFT with hybrid B3LYP functional and Dunning's aug-cc-pVDZ basis set; both energies and gradients were obtained. The same Molpro software was used to get the *ab initio* minimum energy geometry, its energy, the set of gradients, and the benchmark harmonic vibrational frequencies.

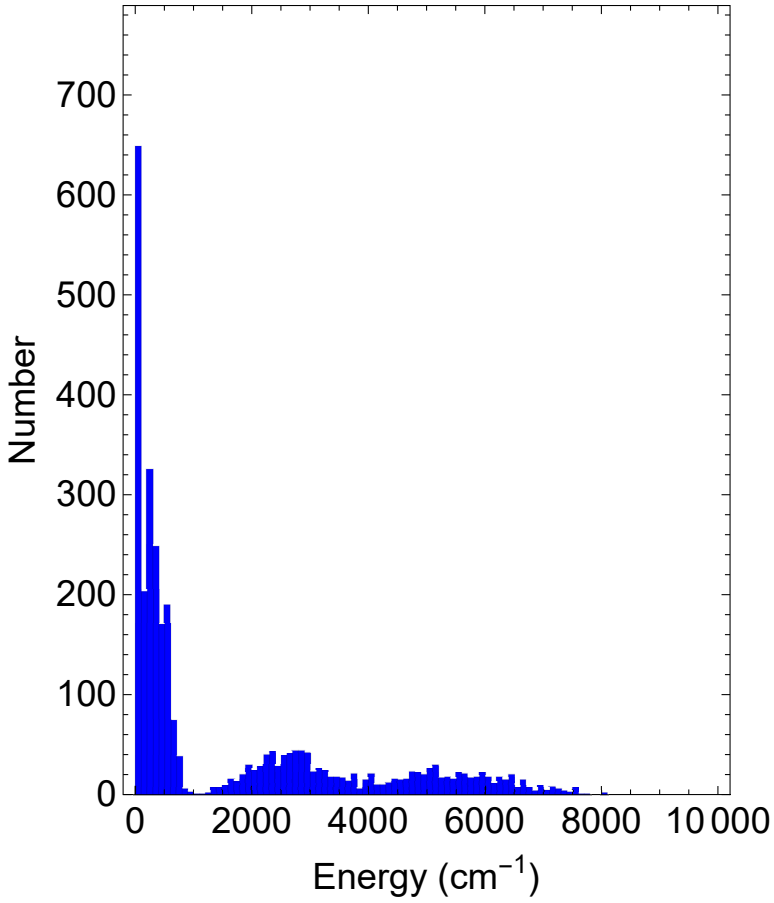

Figure 3: Histogram of energies for geometries used as the dataset for glycine. The bin size for the abscissa is 100 cm^{-1}. Most energies are for configurations close to that of the global minimum

For the surfaces corresponding to the non-fragmented molecule, we took the symmetry basis set to be {2,2,2,1,1,1,1}, allowing the H atoms on NH_2 to permute with one another, the H atoms on CH_2 to permute with one another, and the two O atoms to permute with one another. We omitted the carbon atom permutations and permutation of the COOH hydrogen in order to be consistent with the fragmentation symmetries to follow. In the 2-fragment cases, the fragments are NH_2-CH_2 and CH_2-COOH. The H atoms on CH_2 may permute with one another, because when they appear, they always appear together. Similarly, the H atoms on NH_2 may permute with one another for the same reason, but the H atoms on NH_2 may not be assigned to permute with those on CH_2 because the NH_2

atoms do not appear in fragment 2, whereas the CH_2 ones do. The COOH H atom may not be allowed to permute with the H atoms on either the CH_2 or the NH_2 because the second fragment lacks the NH_2 and the first fragment lacks the COOH. The two oxygens, when they appear, always appear together, so they may be allowed to permute. The two carbon atoms may not be assigned to permute because only one of them appears in each fragment. The permutational symmetry of the first fragment is {2,2,1,1} with an atom assignment, for example, of {2,3,5,6,1,4}. The permutational symmetry of the second fragment is {2,2,1,1,1} with an atom assignment, for example, of {5,6,8,9,4,7,10}.

In the 3-fragment cases, we add the fragment NH_2-COOH to the two from the 2-fragment case. Again, the H atoms on NH_2 may be allowed to permute with one another, the H atoms on the CH_2 may be allowed to permute with one another, and the O atoms of COOH may be allowed to permute with one another. The permutational symmetry of the third fragment is {2,2,1,1,1} with an atom assignment, for example, of {2,3,8,9,1,7,10}; the permutation symmetries and atom assignments for the first two of the three fragments remains unchanged from the 2-fragment case.

Table 2 shows the results of our glycine calculations. The RMS values for energies and gradients, as compared to those of the fitting data set, are listed. Vibrational frequencies were calculated for the minimum energy structure of each surface. The comparison of these frequencies to those determined by Molpro for the minimum-energy structure is given by the Mean Absolute Error (MAE) row, based on the 24 frequencies calculated. The RMS values were used for the energies and gradients because these are determined by least-squares methods. We used MAE values for the vibrational frequencies because these do not over emphasize a few worse matches. The vibrational frequencies for the *ab initio* surface, for the 3^{rd}-order full molecule surface, and for the 3^{rd}-order, 3-fragment surface are provided in Table 3.

As expected, the 4^{th}-order, full molecule calculation (Table 2, column 5) provides the most accurate results but also takes the most time for the evaluation of 3100 energies and

3100 × 30 gradients (compare across the last row). The 3^{rd}-order, full molecule fit is more than ten times faster with only modest loss of accuracy.

The advantages of fragmentation in glycine (10 atoms) are not as prominent as those in NMA (12 atoms) partly because the full molecule basis set can still be calculated reasonably rapidly, at least at 3^{rd} order. However, if one were to consider a somewhat larger amino acid or even a small peptide, it would be impractical to calculate the basis for the molecule without fragmentation. One aim of research in the future should be to see how fragmentation might help in these larger systems.

Table 2: Glycine Results. All RMS and MAE energies are in cm^{-1}.

Fit Order	3	3	3	4	4	4
Frags	full	2	3	full	2	3
Morse vars	45	33	45	45	33	45
Polys	4683	1022	1704	46654	5348	9337
RMSE	20.1	131.6	82.9	7.5	109.3	30.8
RMSG	59.6	354.7	189.3	1.3	257.6	77.2
MAEvib	11.1	31.2	17.5	7.2	27.3	13.6
Time/s	2.193	0.408	0.708	23.869	2.192	3.982

As mentioned in the Section on Computational Details, it is possible once a fit has been found to add or delete polynomials to or from the basis set in order to achieve increased accuracy or shorter execution times, respectively. Figure 4 shows an example, based on the 4^{th}-order, 3-fragment fit in the last column of Table 2. From the left, the number of coefficients for each set of points at a particular time is 6443, 7986, 9337, 10001, and 11001. As can be seen from the figure, the time and accuracy generally both increase with the number of coefficients. One can thus suggest a "figure of merit" for the fit as the product of the average of the RMSE values for the potential and the gradients in cm^{-1} times the time for execution of the computational test in s; the smallest value is desired. This figure of

Table 3: Glycine Results: vibrational frequencies for the *ab initio* frequencies for the glycine global minimum provided by Molpro compared to those for the 3^{rd}-order, full molecule and the 3^{rd}-order, 3-fragment surfaces. All energies are in cm^{-1}.

Mode	*ab initio*	3^{rd}-order, full mol.	3^{rd}-order, 3-frag
1	61.7	100.1	34.7
2	211.6	231.2	209.9
3	256.0	260.6	238.6
4	461.3	466.1	459.3
5	512.3	526.0	531.6
6	630.2	631.6	574.2
7	648.4	647.7	641.6
8	816.9	820.6	810.7
9	908.3	906.5	898.5
10	912.2	918.4	942.2
11	1122.6	1124.4	1129.3
12	1160.9	1161.0	1148.4
13	1176.2	1178.2	1158.1
14	1297.8	1291.0	1288.2
15	1371.8	1370.1	1357.7
16	1385.0	1379.0	1374.2
17	1437.3	1444.4	1443.7
18	1656.6	1664.2	1674.6
19	1803.6	1793.3	1787.7
20	3045.9	3037.3	3017.7
21	3083.7	3041.3	3058.7
22	3493.8	3466.4	3457.3
23	3567.1	3547.5	3540.6
24	3735.2	3704.1	3719.5

merit for the five cases is, from left to right, 188, 201, 215, 75, and 78, suggesting that the fit with a computational time near 4.4 s and with 10001 coefficients gives the best compromise between accuracy and time. Of course, depending on the application, one might still prefer a faster fit, or perhaps a more accurate one. The adding and pruning of polynomials offers a method to determine a potential fit that has the desired trade-off between accuracy and time.

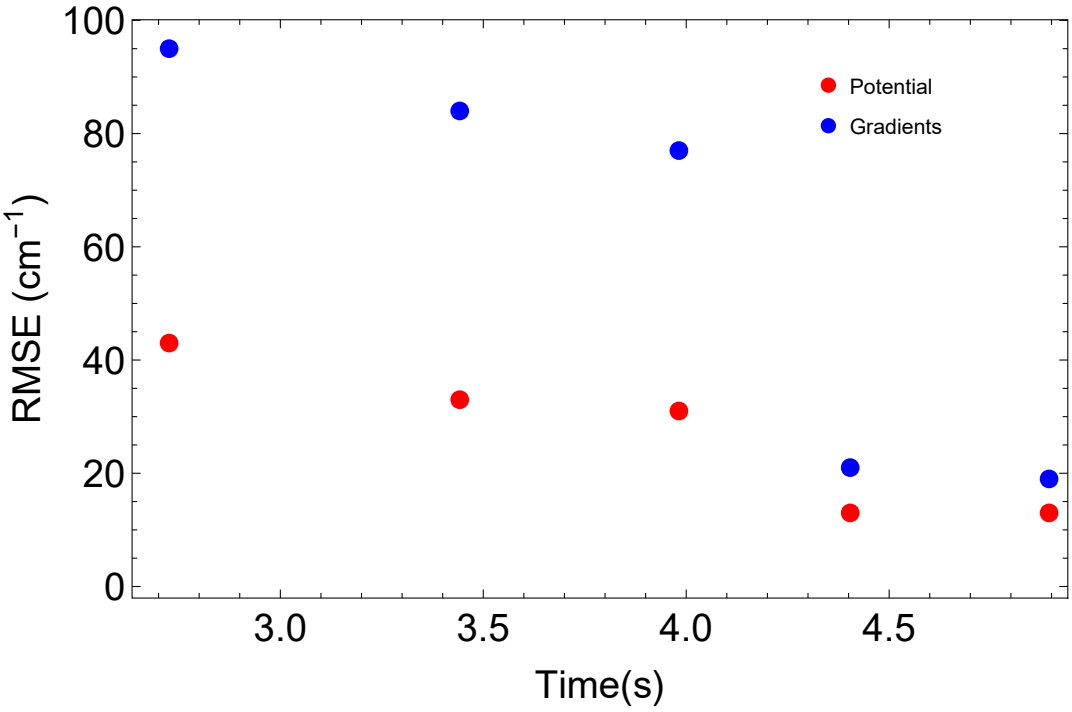

Figure 4: RMSE values for the 4^{th}-order, 3-fragment fit for Glycine vs the time for calculation of energy and gradients (see text for details). The original fit from the last column of Table 2 is shown by the points near 4 s on the abscissa. Points to the left of these are for pruning the number of coefficients back from 9337 to, from the left, 6433, or 7986, or by adding polynomials to give the points to the right, with 10001 and 11001 coefficients.

4 Conclusions

The program we have created provides an efficient and practical way to incorporate fragmentation in the choice of basis functions for calculating large-molecule potential energy surfaces. We have tested it against previous (and less efficient) methods for NMA with en-

couraging results; energies and gradients are exactly reproduced in less time. We have used it to calculate and compare several potential energy surfaces for glycine global minimum conformer. It is practical both because it incorporates all fragments in a single code with relatively simple input parameters and because the outputs include a Fortran program that can be used for fitting the coefficients or using them to calculate energies and gradients.

Glycine was chosen because it is the simplest amino acid, and it is often studied in supramolecular systems of biological interest involving other glycines, water, or hydrogen molecules. For our glycine database we have employed a set of energies up to 10000 cm^{-1} sampled from trajectories mainly confined within the global minimum well. This choice was driven by the goal of demonstrating the effectiveness of the new software rather than by the necessity to develop a global surface for glycine. The latter would require additional characterization of a number of shallow wells and energetically low-lying transition states, which is not reported here and left for future work. We believe that the present work opens a route to the study of other amino acids, peptides, nucleobases, and more complicated biological structures.

We note that in the examples we have used the increase in speed is modest, but this is because the molecules and fragments are relatively small. For example, there are only 816 duplicated polynomials in the 2-fragment NMA case, so the decrease in cpu time is only a bit over 25%. For large molecules with more fragments, the savings in cpu time is expected to be much greater.

More work is needed to develop an understanding of which fragmentation schemes are most effective for different systems. Isomerization has already been shown to be described by the fragmentation approach. Chemical reactions may be more difficult to tackle with fragmentation, as these can involve very large amplitude motion. Of course, the applicability of the approach would be system dependent and we see no reason in principle why it cannot succeed. It is fairly obvious, as shown by both molecular examples here, that when the molecule can bend back on itself it is important to include interactions between the two

ends of the molecule. Thus, treating the molecule as a circle and dividing into overlapping fragments around the circle is a good strategy. On the other hand, when the molecule is fairly rigid, it is more likely that end-to-end interactions can be neglected, so that dividing the molecule into overlapping fragments along the line and neglecting long-range interactions will be most efficient. It will take some practice and experience to uncover more subtle aspects of the fragmentation method.

Acknowledgement

The authors thank Apurba Nandi for useful discussions during the writing of this paper.

Supporting Information Available

The Supporting Information contains three sections and a set of Mathematica programs (NMA-FRAG.zip). Section S1 describes briefly the sequential and pairwise methods for deleting duplicates. Section S2 summarizes the inputs and outputs of the Mathematica program. Section S3 provides a detailed analysis of fragmentation of a ficitious eight atom molecule A_2BCDEF_2 using a two-fragment scheme with a single Morse variable in common. Results for three more complicated fragmentation schemes are also given in summary fashion. The NMA-FRAG Mathematica programs include a notebook with examples, called DeleteDuplicatesMainProgramAndExamples.nb and a set of functions, called DeleteDuplicates.wl

This information is available free of charge via the Internet at http://pubs.acs.org

References

(1) Braams, B. J.; Bowman, J. M. Permutationally invariant potential energy surfaces in high dimensionality. *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.

(2) Handley, C. M.; Popelier, P. L. A. Potential energy surfaces fitted by artificial neural networks. *J. Phys. Chem. A* **2010**, *114*, 3371–3383.

(3) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

(4) Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **2011**, *134*, 074106.

(5) Sergei, M.; Richard, D.; Carrington, T. Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces. *Int. J. Quantum Chem.* **2014**, *115*, 1012–1020.

(6) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

(7) Bartók, A. P.; Csányi, G. Gaussian approximation potentials: A brief tutorial introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.

(8) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.

(9) Uteva, E.; Graham, R. S.; Wilkinson, R. D.; Wheatley, R. J. Interpolation of intermolecular potentials using Gaussian processes. *J. Chem. Phys.* **2017**, *147*, 161706.

(10) Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. *J. Chem. Phys.* **2013**, *139*, 054112.

(11) Shao, K.; Chen, J.; Zhao, Z.; Zhang, D. H. Communication: Fitting potential energy surfaces with fundamental invariant neural network. *J. Chem. Phys.* **2016**, *145*, 071101.

(12) Jiang, B.; Li, J.; Guo, H. Potential energy surfaces from high fidelity fitting of ab initio points: The permutation invariant polynomial - neural network approach. *Int. Rev. Phys. Chem.* **2016**, *35*, 479–506.

(13) Fu, B.; Zhang, D. H. Ab initio potential energy surfaces and quantum dynamics for polyatomic bimolecular reactions. *J. Chem. Theory Comput.* **2018**, *14*, 2289–2303.

(14) Qu, C.; Yu, Q.; Bowman, J. M. Permutationally invariant potential energy surfaces. *Annu. Rev. Phys. Chem.* **2018**, *69*, 6.1–6.25.

(15) Kamath, A.; Vargas-Hernández, R. A.; Krems, R.; Carrington, T.; Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy. *J. Chem. Phys.* **2018**, *148*, 241702.

(16) Doughan, D. I.; Raff, L. M.; Rockley, M. G.; Hagan, M.; Agrawal, P. M.; Komanduri, R. Theoretical investigation of the dissociation dynamics of vibrationally excited vinyl bromide on an ab initio potential-energy surface obtained using modified novelty sampling and feed-forward neural networks. *J. Chem. Phys.* **2006**, *124*, 054321.

(17) Manzhos, S.; Carrington, T. A random-sampling high dimensional model representation neural network for building potential energy surfaces. *J. Chem. Phys.* **2006**, *125*, 084109.

(18) Pukrittayakamee, A.; Malshe, M.; Hagan, M.; Raff, L. M.; Narulkar, R.; Bukkapatnum, S.; Komanduri, R. Simultaneous fitting of a potential-energy surface and its corresponding force fields using feed-forward neural networks. *J. Chem. Phys.* **2009**, *130*, 134101.

(19) Nongnuch, A.; Behler, J. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B* **2012**, *85*, 045439.

(20) Li, J.; Jiang, B.; Guo, H. Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems. *J. Chem. Phys.* **2013**, *139*, 204103.

(21) Guan, Y.; Yang, S.; Zhang, D. H. Application of Clustering Algorithms to Partitioning Configuration Space in Fitting Reactive Potential Energy Surfaces. *J. Phys. Chem. A* **2018**, *122*, 3140–3147.

(22) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(23) Qu, C.; Yu, Q.; Van Hoozen, B. L.; Bowman, J. M.; Vargas-Hernández, R. A. Assessing Gaussian Process Regression and Permutationally Invariant Polynomial Approaches To Represent High-Dimensional Potential Energy Surfaces. *J. Chem. Theory Comput.* **2018**, *14*, 3381–3396.

(24) Bowman, J. M.; Braams, B. J.; Carter, S.; Chen, C.; Czakó, G.; Fu, B.; Huang, X.; Kamarchik, E.; Sharma, A. R.; Shepler, B. C.; Wang, Y.; Xie, Z. Ab-initio-based potential energy surfaces for complex molecules and molecular complexes. *J. Phys. Chem. Lett.* **2010**, *1*, 1866–1874.

(25) Xie, Z.; Bowman, J. M. Permutationally invariant polynomial basis for molecular energy surface fitting via monomial symmetrization. *J. Chem. Theory Comput.* **2010**, *6*, 26–34.

(26) Bowman, J. M.; Czakó, G.; Fu, B. High-dimensional ab initio potential energy surfaces for reaction dynamics calculations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 8094–8111.

(27) Brown, A.; Braams, B. J.; Christoffel, K.; Jin, Z.; Bowman, J. M. Classical and quasiclassical spectral analysis of CH_5^+ using an ab initio potential energy surface. *J. Chem. Phys.* **2003**, *119*, 8790–8793.

(28) Shepler, B. C.; Braams, B. J.; Bowman, J. M. "Roaming" dynamics in CH_3CHO photodissociation revealed on a global potential energy surface. *J. Phys. Chem. A* **2008**, *112*, 9344–9351.

(29) Kidwell, N. M.; Li, H.; Wang, X.; Bowman, J. M.; Lester, M. I. Unimolecular dissociation dynamics of vibrationally activated CH_3CHOO Criegee intermediates to OH radical products. *Nat. Chem.* **2016**, *8*, 509–514.

(30) Wang, Y.; Braams, B. J.; Bowman, J. M.; Carter, S.; Tew, D. P. Full-dimensional quantum calculations of ground-state tunneling splitting of malonaldehyde using an accurate ab initio potential energy surface. *J. Chem. Phys.* **2008**, *128*, 224314.

(31) Qu, C.; Bowman, J. M. An ab initio potential energy surface for the formic acid dimer: zero-point energy, selected anharmonic fundamental energies, and ground-state tunneling splitting calculated in relaxed 1–4-mode subspaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24835–24840.

(32) Qu, C.; Bowman, J. M. A Fragmented, Permutationally Invariant Polynomial Approach for Potential Energy Surfaces of Large Molecules: Application to N-methyl acetamide. *J. Chem. Phys.* **2019**, *150*, 141101.

(33) Li, J.; Guo, H. Communication: An accurate full 15 dimensional permutationally invariant potential energy surface for the $OH + CH_4 \longrightarrow H_2O + CH_3$ reaction. *J. Chem. Phys.* **2015**, *143*, 221103.

(34) Conte, R.; Qu, C.; Bowman, J. M. Permutationally Invariant Fitting of Many-body, Non-covalent Interactions with Application to Three-body Methane-water-water. *J. Chem. Theory Comput.* **2015**, *11*, 1631–1638.

(35) Jasper, A. W.; Harding, L. B.; Knight, C.; Georgievskii, Y. Anharmonic Rovibrational Partition Functions at High Temperatures: Tests of Reduced-Dimensional Models for Systems with up to Three Fluxional Modes. *J. Phys. Chem. A* **2019**, *123*, 6210–6228.

(36) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(37) Li, J.; Song, K.; Behler, J. A critical comparison of neural network potentials for molecular reaction dynamics with exact permutation symmetry. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9672–9682.

(38) Nandi, A.; Qu, C.; Bowman, J. M. Using Gradients in Permutationally Invariant Polynomial Potential Fitting: A Demonstration for CH_4 Using as Few as 100 Configurations. *J. Chem. Theory Comput.* **2019**, *15*.

(39) MSA Software with Gradients. `https://github.com/szquchen/MSA-2.0`, Accessed: 2019-01-20.

(40) Conte, R.; Houston, P. L.; Bowman, J. M. Communication: A bench-mark-quality, full-dimensional ab initio potential energy surface for Ar-HOCO. *J. Chem. Phys.* **2014**, *140*, 151101.

(41) Qu, C.; Conte, R.; Houston, P. L.; Bowman, J. M. "Plug and play" full-dimensional ab initio potential energy and dipole moment surfaces and anharmonic vibrational analysis for CH4-H2O. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8172–8181.

(42) Xie, Z.; Bowman, J. M. Permutationally Invariant Polynomial Basis for Molecular Energy Surface Fitting vis Monomial Symmetrization. *J. Chem. Theor. Comput.* **2010**, *6*, 26–34.

(43) Original MSA Software. `https://www.mcs.anl.gov/research/projects/msa/`, Accessed: 2019-12-20.

(44) Nandi, A.; Qu, C.; Bowman, J. M. Using Gradients in Permutationally Invariant Polynomial Potential Fitting: A Demonstration for CH_4 Using as Few as 100 Configurations. *J. Chem. Theory Comput.* **2019**, *15*, 2826–2835.

(45) Wolfram Research, . Inc., Mathematica, Version 12.0. Champaign, IL, 2019.

(46) Nandi, A.; Qu, C.; Bowman, J. M. Full and Fragmented Permutationally Invariant Polynomial Potential Energy Surfaces for trans and cis N-methyl Acetamide and Isomerization Saddle Points. *J. Chem. Phys.* **2019**, *151*, 084306.

(47) Ehrenfreund, P.; Glavin, D. P.; Botta, O.; Cooper, G.; Bada, J. L. Extraterrestrial amino acids in Orgueil and Ivuna: Tracing the parent body of CI type carbonaceous chondrites. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2138.

(48) Senent, M. L.; Fernandez-Herrera, S.; Smeyers, Y. G. Ab initio determination of the roto-torsional energy levels of hydrogen peroxide. *Spectrochim. Acta Mol. Biomol. Spectrosc.* **2000**, *56*, 1457–1468.

(49) Stepanian, S. G.; Reva, I. D.; Radchenko, E. D.; Rosado, M. T. S.; Duarte, M. L. T. S.; Fausto, R.; Adamowicz, L. Matrix-Isolation Infrared and Theoretical Studies of the Glycine Conformers. *J. Phys. Chem. A* **1998**, *102*, 1041–1054.

(50) Brauer, B.; Chaban, G. M.; Gerber, R. B. Spectroscopically-tested, improved, semi-empirical potentials for biological molecules: Calculations for glycine, alanine and proline. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2543–2556.

(51) Bludsky, O.; Chocholousova, J.; Vacek, J.; Huisken, F.; Hobza, P. Anharmonic treatment of the lowest-energy conformers of glycine: A theoretical study. *J. Chem. Phys.* **2000**, *113*, 4629–4635.

(52) Barone, V.; Biczysko, M.; Bloino, J.; Puzzarini, C. Characterization of the Elusive Conformers of Glycine from State-of-the-Art Structural, Thermodynamic, and Spectroscopic Computations: Theory Complements Experiment. *J. Chem. Theory Comput.* **2013**, *9*, 1533–1547.

(53) Gabas, F.; Conte, R.; Ceotto, M. On-the-Fly ab Initio Semiclassical Calculation of Glycine Vibrational Spectrum. *J. Chem. Theory Comput.* **2017**, *13*, 2378–2388.

(54) Csaszar, A. G. Conformers of gaseous glycine. *J. Am. Chem. Soc.* **1992**, *114*, 9568–9575.

(55) Balabin, R. M. Conformational Equilibrium in Glycine: Experimental Jet-Cooled Raman Spectrum. *J. Phys. Chem. Lett.* **2010**, *1*, 20–23.

Graphical TOC Entry