The Formic Acid-Ammonia Heterodimer: A New △-Machine Learning CCSD(T)-Level Potential Energy Surface Allows Investigation of the Double Proton Transfer

Paul L. Houston,[*,†] Chen Qu,[‡] Qi Yu,[¶] Priyanka Pandey,[¶] Riccardo Conte,[§] Apurba Nandi,[¶,‖] Joel M. Bowman,[*,¶] and Stephen G. Kukolich[*,⊥]

†*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*

‡*Independent Researcher, Toronto, CA*

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

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

‖*Department of Physics and Materials Science, University of Luxembourg, L-1511, Luxembourg City, Luxembourg*

⊥*Department of Chemistry and Biochemistry, University of Arizona, 1306 E. University Avenue, Tucson AZ 85721, United States*

E-mail: plh2@cornell.edu; jmbowma@emory.edu; kukolich@arizona.edu

Abstract

The formic acid-ammonia dimer is an important example of a hydrogen-bonded complex in which a double proton transfer can occur. Its microwave spectrum has recently been reported and rotational constants and quadrupole coupling constants were determined. Calculated estimates of the double-well barrier and the internal barriers to rotation were also reported. Here we report a full-dimensional potential energy surface (PES) for this complex, using two closely related Δ-machine learning methods to bring it to the CCSD(T) level of accuracy. The PES dissociates smoothly and accurately. Using a 2d quantum model the ground vibrational-state tunneling splitting is estimted to be less than 10^{-4} cm^{-1}. The dipole moment along the intrinsic reaction coordinate is calculated along with a Mullikan charge analysis and supports mildly ionic character of the minimum and strongly ionic character at the double-well barrier.

Introduction

Hydrogen bonds are ubiquitous in nature. They critically stabilize the structure of proteins and DNA; they are essential to the structure, high heat capacity, high solvency, high boiling and sublimation points of condensed phases of water; they are responsible for the structural properties of such polymers as cellulose, nylon, and cotton; and they are important in the design of orally administered drugs. To the quantum chemist, the movement of a hydrogen-bonded atom or proton by tunneling through a barrier from one acceptor to another remains a source of active investigation, especially in multidimensional systems.

One can distinguish several types of hydrogen transfer in which tunneling might be important. The simplest is where one hydrogen or proton is transferred between a donor and acceptor. Good examples are hydrogen transfers in malonaldehyde[1,2] and in tropolone.[3] In both cases the tunneling frequency has been accurately calculated.

A more complicated type is where a complex is doubly hydrogen bonded and in which

the synchronous transfer of the two hydrogen bonds maintains a double hydrogen bond but switches the locations of the hydrogen bonds. This is often called concerted proton tunneling. The formic acid dimer is a good example of a symmetric system which has been studied both experimentally[4] and theoretically.[5–7] Many other experiments and calculations on doubly hydrogen-bonded complexes have been performed.[8] Other complexes composed of two carboxylic acids with C_{2v}(m) symmetry[9] have a symmetric double-well potential for this concerted proton motion. Early microwave measurements showed effects of the tunneling[10,11] and provided accurate measurements of the tunneling frequency.[12] This double hydrogen bonding is similar to the hydrogen bonding in the adenine - thymine nucleic acid-base pair. Many other carboxylic acid complexes, for examples, formic acid – nitric acid,[13] benzoic acid – formic acid[14] and acetic acid – formic acid,[15] have the C_{2v}(m) symmetry and have been shown to exhibit resolvable concerted proton tunneling splittings in microwave spectra. Two other doubly hydrogen bonded dimers with C_{2v}(m) symmetry, fomamidine - formate[16] and tropolone - formic acid,[17] did not show measurable splitting. It is likely that the barrier to the proton tunneling is too high for these complexes. Doubly-bonded complexes with asymmetric double-well potentials and lacking the C_{2v}(m) symmetry are much less likely to show tunneling splittings in the spectra. Additional doubly hydrogen-bonded complexes are are: formic acid triflouro acetic acid,[18] formic acid – malemide,[19] formic acid - glyoxylic acid,[20] formic acid – diflouro acetic acid,[21] formic acid - acrylic acid,[22] and formic acid - perfluorobutyric acid.[23] Double proton hopping is debated also regarding DNA nucleobase pairs, like the guanine-cytosine pair, due to its importance in genetic transcription and mutagenesis. Optical spectroscopy experiments and calculations could not point out the presence of a tunneling-mediated double hydrogen hopping mechanism for the guanine-cytosine pair,[24,25] while kinetics calculations[26,27] or NMR experiments[28] point to the fact that tunneling may not be negligible for the system.

In the current study, we focus on a third type of hydrogen transfer system where there is only one hydrogen bond, but in which the transfer between the two donors involves a change

in the H atom identity and the binding site. An example which we will study in detail here is the formic acid - ammonia (FA-NH_3) heterodimer, and the transfer is illustrated in Figure 1 (the dipole moments in the figure will be discussed later). Other prominent examples are formic acid complexed with CH_4, H_2O, HF and H_2,[29] as well as with $(CH_3OH)_n, n = 1-3$.[30] The doubly-bonded structure for the formic acid water complex was measured by Bauder,[31] but tunneling was not observed.

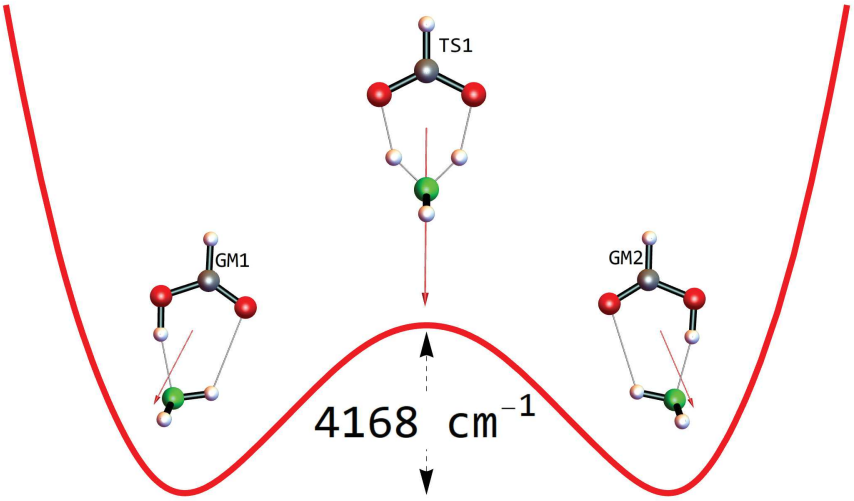

Figure 1: Global minimum (GM) and transition state (TS1) structures of the formic acid – NH_3 heterodimer. Motion along the reaction coordinate involves rocking and translation of the NH_3 from GM1 to TS1 to GM2, the last of which is equivalent to GM1 by 180° rotation around the vertical axis. The two out-of-plane N-H hydrogens at the bottom of each structure overlap in this view. The red arrows indicate the direction and relative magnitude of the dipole moment.

Microwave measurements and calculations for the formic acid-ammonia dimer were recently published.[32] Initial measurements were made in the 7-17 GHz range, followed by double resonance measurements between 20 and 23 GHz. Accurate rotational constants and quadrupole coupling parameters were reported by fitting the observed transitions for two different states. These two states were assigned to possible 0^+ and 0^- tunneling states. Tunneling transitions were not observed. This could be due to the high potential barrier or further complication from the low barrier to internal rotation.

Normally, double proton transfer is interpreted as the concerted proton tunneling in a doubly hydrogen-bonded dimer, but Minyaev et al.[29] consider the FA-NH_3 case as a proton migration through the ionic complex $HCOO^-$ - NH_4^+. We examine this migration here by a calculation of dipole moment along the Intrinsic Reaction Coordinate as well as standard charge analysis and find a strong dependence of both the dipole moment and the charge distribution on the motion.

Figure 1 shows the structures involved as well as the direction and relative magnitude of the dipole moment, discussed in detail later. The calculations described below indicate a more ionic character of the complex at the transition state. The IRC calculation shows a fairly sharp increase in the dipole moment of the complex as the transition state is approached. Evidence for this ionic form is also suggested from the observation of $HCOO^-$ in the IR studies of FA-NH_3 ice complex paper by Hellebust, O'Riordan, and Sodeau.[33] In that study, the complex was modeled as a singly hydrogen-bonded complex leading to proton transfer to produce ammonium formate.

The intriguing nature of this type of hydrogen bond led us to develop a full-dimensional potential energy surface (PES) for this system. The surface was constructed using the Δ-machine learning method[34] that allowed us to bring the surface to CCSD(T)-level accuracy. We then used the surface to examine the tunneling and dissociation energy in the complex. We show below that there is substantial ionic character to FA-NH_3, not only at the transition state to double proton transfer but also at the global minimum, where there is one hydrogen bond. We also investigate the dissociation energy of this hetero-dimer and show that its high value is also consistent with substantial ionic character at the global minimum. Finally, we investigate tunneling through the transition state barrier and show that it is negligibly small.

$\triangle$-machine learning potential

The goal of Δ-machine learning applied to potential surfaces is the correct a potential fit to low-level electronic energies and gradients so as to bring it to the CCSD(T) level of accuracy. The original implementation of Δ-machine learning method, as used in the present context, has been described previously.[34,35] The key equation for it is given by

$$V_{LL\rightarrow CC} = V_{LL} + \Delta V_{CC-LL}, \tag{1}$$

where $V_{LL\rightarrow CC}$ is the corrected PES, V_{LL} is a PES fit to low-level DFT electronic data, and ΔV_{CC-LL} is the correction PES based on high-level coupled cluster energies. It is noted that the difference between CCSD(T) and DFT energies, ΔV_{CC-LL}, is not as strongly varying as V_{LL} with respect to the nuclear configurations and therefore just a small number of high-level electronic energies are adequate to fit the correction PES. Subsequently Liu and co-workers suggested a similar strategy which results in a single expression for $V_{LL\rightarrow CC}$ which is fit to the DFT data plus difference energies.[36] To summarize, in the original approach two potentials are generated. In the second approach a data from the $V_{LL\rightarrow CC}$ PES is use to generate a new final fit $V_{LL\rightarrow CC}$ Both approaches should lead to very similar final potentials $V_{LL\rightarrow CC}$.

The next section describes the data sets used for the fits as well as the specific permutational symmetry used for FA-NH_3.

Computational Methods

The potential energy surfaces used in this work were constructed using permutational invariant polynomials (PIPs).[37–40] These ensure that permutations of identical atoms in a molecule that should permute at the energies considered, do permute. The basis sets used for the current surfaces have the permutational symmetry 42111, meaning that the potential is explicitly invariant with respect to the permutations of the four hydrogens associated with

the nitrogen atom, and similarly for the two oxygens. Permutations involving the hydrogen on the carbon atom are infeasible at the energies of this study. We used polynomials of order 2 and 3 for the ΔV_{CC-LL} correction CCSD(T) surfaces and of order 4 for the V_{LL} DFT surface as well as for the $V_{LL\rightarrow CC}$ PES surfaces. The fits will be described in more detail below, but first we describe the data sets.

A DFT-B3LYP/aug-cc-pVDZ data set was made from several direct molecular dynamics trajectories as well as from grids of points around important paths or stationary points. All electronic structure calculations were performed in Molpro.[41,42] For the trajectories we typically used every fifth geometry (5 a.u. step time) and set the kinetic energy at a particular value. Trajectories were terminated in any interatomic distance exceeded 20 bohr. The trajectories were started either at the global minimum, GM, or at the cyclic transition state, TS1, shown in Fig. 1. The trajectory starting position, kinetic energy (in cm^{-1}), and yield of points are as follows: (GM, 1000, 1200), (GM, 2000, 1200), (TS1, 3500, 800), (GM, 4000, 1200), (GM, 4000, 1200), (GM, 7000, 1200), (GM, 8000, 1200), (GM, 9000, 1200), (GM, 10000, 1200), (TS1, 10000, 1020), (TS1, 13000, 1200), (GM, 18000, 373), (GM 19000, 1200), and (GM, 22000, 451), (GM, 25000, 415).

Grids were used to choose geometries near the GM, near TS1, or near points on the imaginary coordinate (qim) or the intrinsic reaction coordinate (irc), discussed below. Typically, each grid consisted of 100 grid points near a starting geometry. For each grid point, geometries were chosen by setting each of the 27 Cartesian coordinates to be a value chosen at random between $c \pm del$, where c is the Cartesian coordinate in Å and del, also in Å, is selected for a particular grid. The following starting points and values of del were used: (TS1: 0.001, 0.005, 0.01), (point qim1: 0.001), (point qim2: 0.001), (GM: 0.001, 0.005, 0.01), (point qim1: 0.005, 0.01), (point qim2: 0.005, 0.01), (point irc2: 0.005, 0.01, 0..05, 0.1), (point irc7: 0.005, 0.01), (point irc12: 0.005, 0.01, 0.05,, 0.1), (point irc17: 0.005, 0.01, 0.05, 0.1), (point irc22: 0.005, 0.01, 0.05, 0.1), (point irc27: 0.005, 0.01, 0.05, 0.1), (point irc32: 0.005, 0.01, 0.05, 0.1), where the geometries of the starting points are given in the

Supporting Information.

In addition to the calculations described above, a few dozen or so more geometries were calculated in order to ensure that the potential function did not produce an energy lower than that of the global minimum. This situation can occur when there is over-fitting or when there are two few calculated energies in a critical region. The geometries where new points were needed were identified by using diffusion Monte Carlo techniques, as, for example, in previous work.[43] These techniques are normally used to calculate the energy and wavefunction of the zero point vibrational level, but they are also useful for identifying places where the fit produces unreasonable results.

The Δ-ML method also requires a smaller dataset of high-level calculations, in this case using CCSD(T)/aug-cc-pVDZ. Geometries were chosen at random from those used in an early version of the larger, DFT data set. We calculated 3532 geometries from that set. An even smaller dataset, with 472 energies and geometries was taken from a random subset of the set of 3532. Figure 2 shows histograms of the energy distribution for the energies of the DFT and the 3532 CCSD(T) data sets.

Table 1 describes properties of the potential energy surfaces fit to the datasets. Surface $\Delta V_{CC-LL}(1)$ is a fit to a small data set of CCSD(T)-DFT/B3LYP difference energies calculated at geometries of what we will call the "DFT" data set. Surface $\Delta V_{CC-LL}(2)$ is a somewhat larger set of similar difference energies. All fits use 42111 permutational symmetry. Weighting was used for the V_{LL} DFT surface: the weight as a function of energy was given by the formula $wt = dwt/(dwt + dE)$, where $dwt = 0.02$ hartree, and the weight of the fit to the gradients was 1/3 that of the fit to the potential energies. Both weighted and unweighted rms values for the energy and gradients are reported in the table. Given the much improved rmse for the $\Delta V_{CC-LL}(2)$ fit as compared to the $\Delta V_{CC-LL}(1)$ fit, we did not further pursue the $\Delta V_{CC-LL}(1)$ fit.

Figure 3 shows correlation plots comparing the potential energies calculated using the PES to the ab initio energies in the data set. Fig. 3a is for the $\Delta V_{CC-LL}(2)$, while Figs

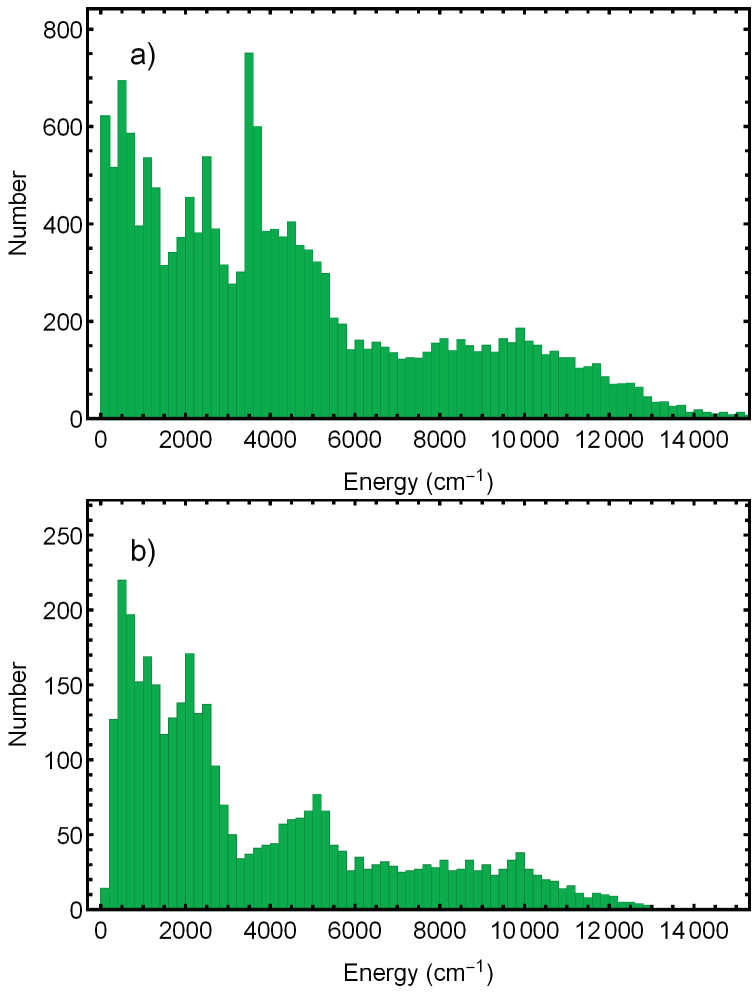

Figure 2: Energy Histograms. a) data set of 17664 DFT/B3LYP geometries; b) data set of 3532 CCSD(T) geometries

Table 1: Properties of the potential energy surfaces

PES	$\Delta V_{CC-LL}(1)$	$\Delta V_{CC-LL}(2)$	V_{LL}
Polynomial order	2	3	4
Num. geometries	472	3532	17664
Level	CCSD(T)	CCSD(T)	DFT/B3LYP
Num. coefficients	118	883	5882
Datasize	472	3532	395662
Num. E with G	0	0	14000
Total Number E	472	3532	17664
rmse	21.1	5	37.7
rmsg	-	-	43.4
Weighting	no	no	yes
wrmse	-	-	22.1
wrmsg	-	-	37.5

3b and c show, for the V_{LL} DFT surface, the correlation plots for energies and gradients, respectively. The CCSD(T) calculation for the difference energy does not provide gradients, but these can be calculated analytically from the fit and used to construct the Δ-ML surfaces discussed below.

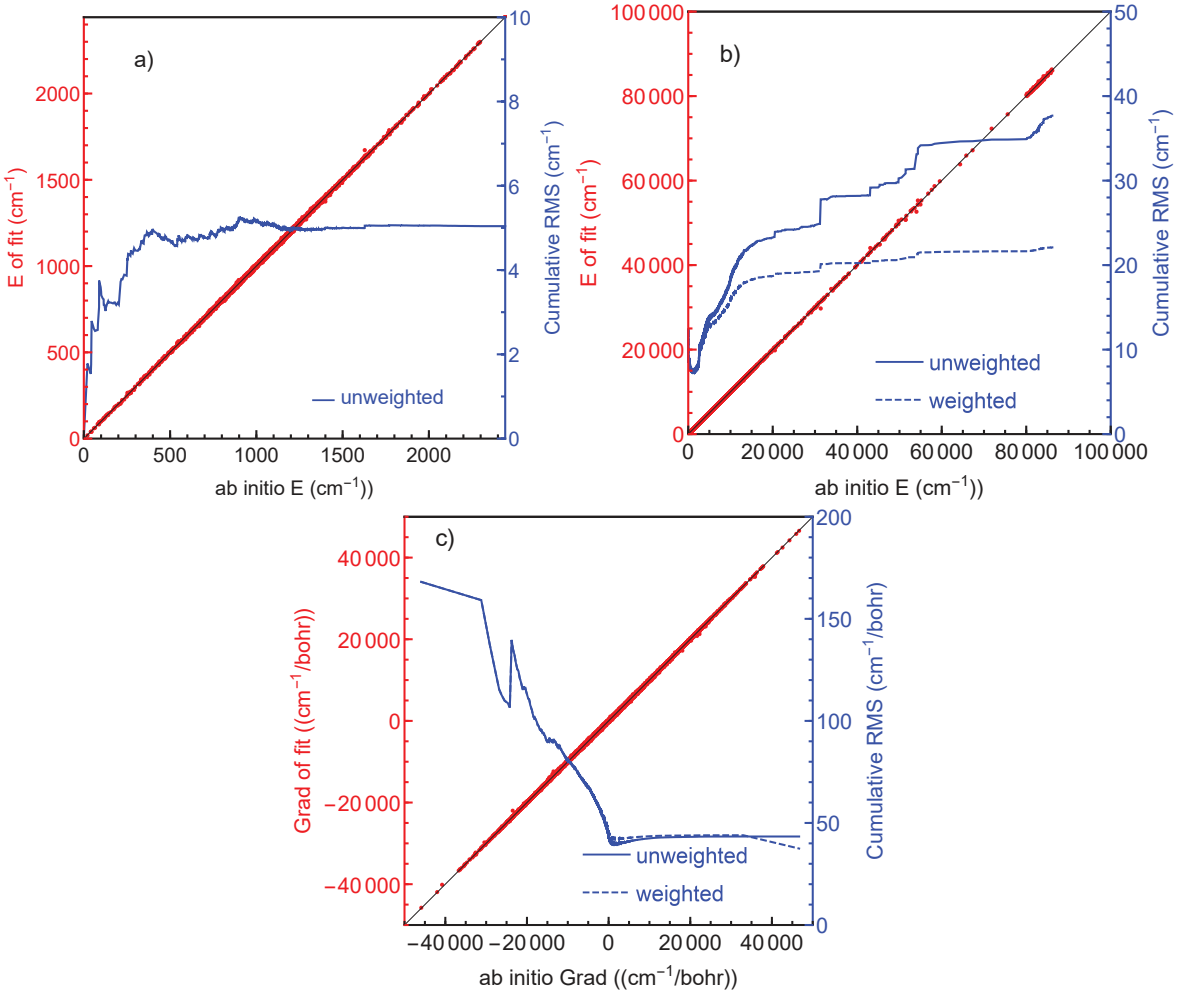

Figure 3: Correlation plots. a) for 42111_3 fit to 3532 CCSD(T)-B3LYP difference energies; b) and c) for 43111_4 fit to 17664 B3LYP energies and gradients, respectively. R^2 coefficients for the correlations are a) 0.999732 b) 0.999991, c) 0.999889.

Results and Discussion

Properties of the V_{LL} DFT and $V_{LL\rightarrow CC}$ PESs

The V_{LL} DFT and the $\Delta V_{CC-LL}(2)$ surfaces described above were used to form two CCSD(T)-corrected surfaces. Metrics with respect to DFT/B3LYP or CCSD(T) benchmark calculations are shown in Table 2. We considered three stationary states on the PES, the global minimum (GM), the transition-state TS1, mentioned above, and a transition-state to hindered rotation of the GM NH_3 around the N-O bond, TS2. For each stationary state, we found the optimized geometry and used it in a normal mode analysis to calculate the harmonic vibrational frequencies at that geometry. These form an excellent test of the first and second derivatives of the energy with respect to the Cartesian coordinates. We then compared the mean absolute errors (MAEs) of the frequencies from a benchmark calculation to those found for the geometry on the PES. As can be seen from the first two rows of the table, the DFT PES has quite good fidelity with respect to the B3LYP benchmark, both for the GM energy, the energies of transition states TS1 and TS2, and the vibrational frequencies.

We calculated the Δ-ML surfaces by two related methods. The basic equation common to both is given by equation 1

In the first, we used the method described by Nandi et al.[3,34,44] At each desired geometry, the results for the difference energy and gradients predicted by the $\Delta V_{CC-LL}(2)$ fit were added to the energy and gradients predicted by the DFT fit to produce the result. This surface is called $\Delta-$ML PES-1 and can be described briefly as the "sum-of-fits" method represented by equation 1.

In the second method, at each geometry in the DFT data set, the difference energy and gradients were calculated from the $\Delta V_{CC-LL}(2)$ fit. These results were added to those in the DFT data base to form a new, CCSD(T)-corrected, database. This database was then fit using the same fourth-order 4221 basis. The fit is called $\Delta-$ML PES-2 and can be described briefly as a "fit-of-the-sum" method, a method first used by Liu and Li.[36] As can

be seen in the table, these two methods give nearly identical results, and both are in good agreement with the CCSD(T) benchmark values. The supporting information (SI) provides the potential energy surfaces for $\Delta V_{CC-LL}(2)$, DFT, Δ-ML PES-1 and Δ-ML PES-2, as well as datasets and CCSD(T) geometries, energies, and vibrational frequencies for the three stationary points.

Table 2: Properties of the potential energy surfaces

Property: Benchmark or PES:	GM Energy (hartree)	TS1 Barrier (cm^{-1})	TS2 Barrier (cm^{-1})	MAE GM GM freqs. (cm^{-1})	MAE TS1 freqs. (cm^{-1})	MAE TS2 freqs. (cm^{-1})
B3LYP	-246.257815	3573.9	138.9	-	-	-
V_{LL} DFT PES	-246.257885	3604.8	148.4	3.1	17.3	4.9
CCSD(T)	-245.800588	4168.3	156.2	-	-	-
Δ-ML PES-1*	-245.800601	4228.4	136.2	5.8	40.4	10.0
Δ-ML PES-2**	-245.800602	4229.5	137.7	5.9	40.5	9.9

* Sum of fits to DFT (energies and gradients) and $\Delta V_{CC-LL}(2)$ (energies and gradients)
** Fit to sum of DFT and $\Delta V_{CC-LL}(2)$ results (both energies and gradients)

Dissociation to Formic Acid and Ammonia

We calculated the CCSD(T)/aug-cc-pVDZ optimized potential energies and structures of the GM for FA-NH_3, an isolated NH_3 and an isolated FA. The dissociation energy, D_e, is given, as usual, by the difference between the energies of the isolated fragments and the GM, and this equals 4143.7 cm^{-1}. It is of interest to investigate the behavior of the Δ-ML PES 2 as the fragments separate, This is shown in Fig. 4. We started from the energy and structure of the GM and then separated the NH_3 from the formic acid along the direction of the bond between the N and the hydrogen-bonded H; the zero of energy is taken to be that of the GM, at which geometry the distance between the N and the hydrogen-bonded H atom is about 1.774 Å.The geometries along this path were used to calculate both CCSD(T)/aug-cc-pVDZ potential energies (red points) and potential energies from the Δ-ML PES-2 surface

(blue line). It is clear that the asymptote is slightly higher than D_e. . This difference is expected and is due to the fact that the structures of the FA and NH_3 were fixed at the geometries in the complex and are not the minimum structures of the free NH_3 and FA. The asymptote thus provides an upper limit on the dissociation energy. Because no points along this potential energy cut were intentionally used in the PES fit, the agreement between the CCSD(T) calculations and the PES results serves as a test of the accuracy of the PES.

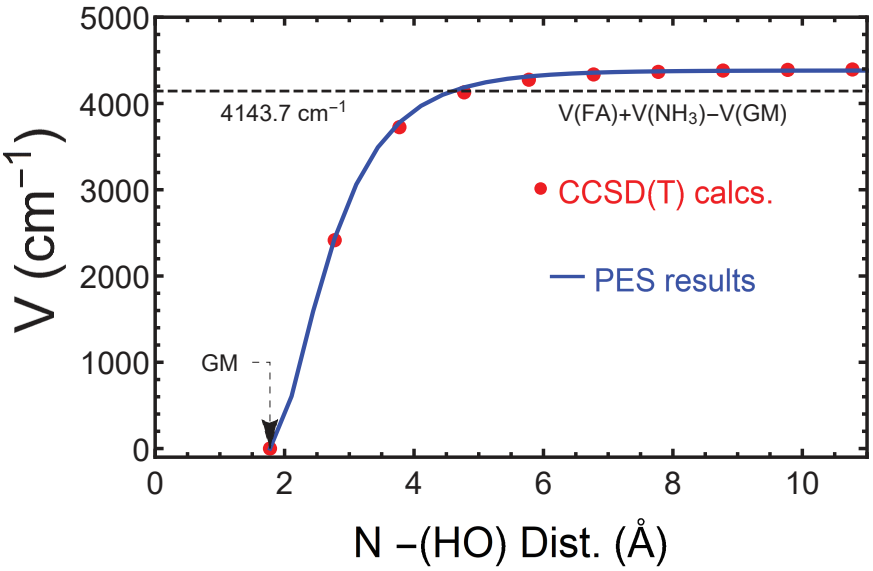

Figure 4: Potential energy cut showing the dependence on the distance between the N and the hydrogen-bonded H as the NH_3 is displaced along the N-(HO) bond direction. The red points are the results of direct CCSD(T)/aug-cc-pVDZ calculations, whereas the blue line is the prediction of the Δ-ML PES2 surface. The black dotted line shows the dissociation energy, D_e obtained from CCSD(T) calculations of the potentials of FA-NH_3, free NH_3, and free FA.

A first look at multidimensional tunneling

We now address the motion involving TS1, shown in Fig. 1. The reaction through and across this saddle point could take place by tunneling or by synchronous double proton transfer, both are possible and appear to be present. We first address tunneling. As can be surmised by the sudden change in the potential along te IRC there is strong multi-dimensional tunneling. To address tunneling in a simple way, we use m normal modes of the saddle point, and,

ignoring the vibrational angular momentum, the reduced-dimensional Hamiltonian can be written as

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2}{\partial Q_i^2} + V(Q_1, \cdots, Q_m), \tag{2}$$

where $V(Q_1, \cdots, Q_m)$ is the m-mode potential. In this work, we used $m = 2$, and the two normal modes used are Q_{21} (i.e., the imaginary-frequency mode) and Q_4. These two modes are chosen because they have the largest values at the global minimum (see Table S1 in the SI). This 2-dimensional Schroedinger equation is solved by using a direct-product of particle-in-a-box basis functions. The eigen energies, and thus the tunneling splittings, are obtained by diagonalizing the corresponding H-matrix.

We used two different 2-mode potentials for this calculation: unrelaxed and relaxed potential. By unrelaxed, we mean all the remaining 19 modes are kept at 0. Because of this, the minima on this unrelaxed 2-mode potential are not the global minimum geometry, and they are only about 1700 cm^{-1} below TS1. Therefore, the splitting would be overestimated. To make a better estimate of the tunneling splitting, we relaxed the remaining normal modes approximately, resulting in a relaxed 2-mode potential. The details of the relaxation are given in the SI. Fig. 5 shows contour plots of unrelaxed and relaxed 2-mode potentials as a function of Q_{21} and Q_4. The locations of the GM and TS1 are indicated by black dots. It can be seen from the upper panel (the unrelaxed potential) that the GM does not overlap with the minimum of the unrelaxed potential, and the barrier is significantly lower than the CCSD(T) barrier. Very strong curvature is seen in both plots, indicating strong correlation between these two modes.

The splitting obtained using the unrelaxed potential is 0.0015 cm^{-1}, which is already very small. This number significantly overestimates the splitting because the barrier is about 2500 cm^{-1} lower than the actual one. By using the relaxed potential, the splitting decreases to 3×10^{-5} cm^{-1}, which is almost is two orders of magnitude smaller. We do not claim that this very small splitting is a quantitative prediction. Rather, we simply state that the splitting is likely less than 10^{-4} cm^{-1}.

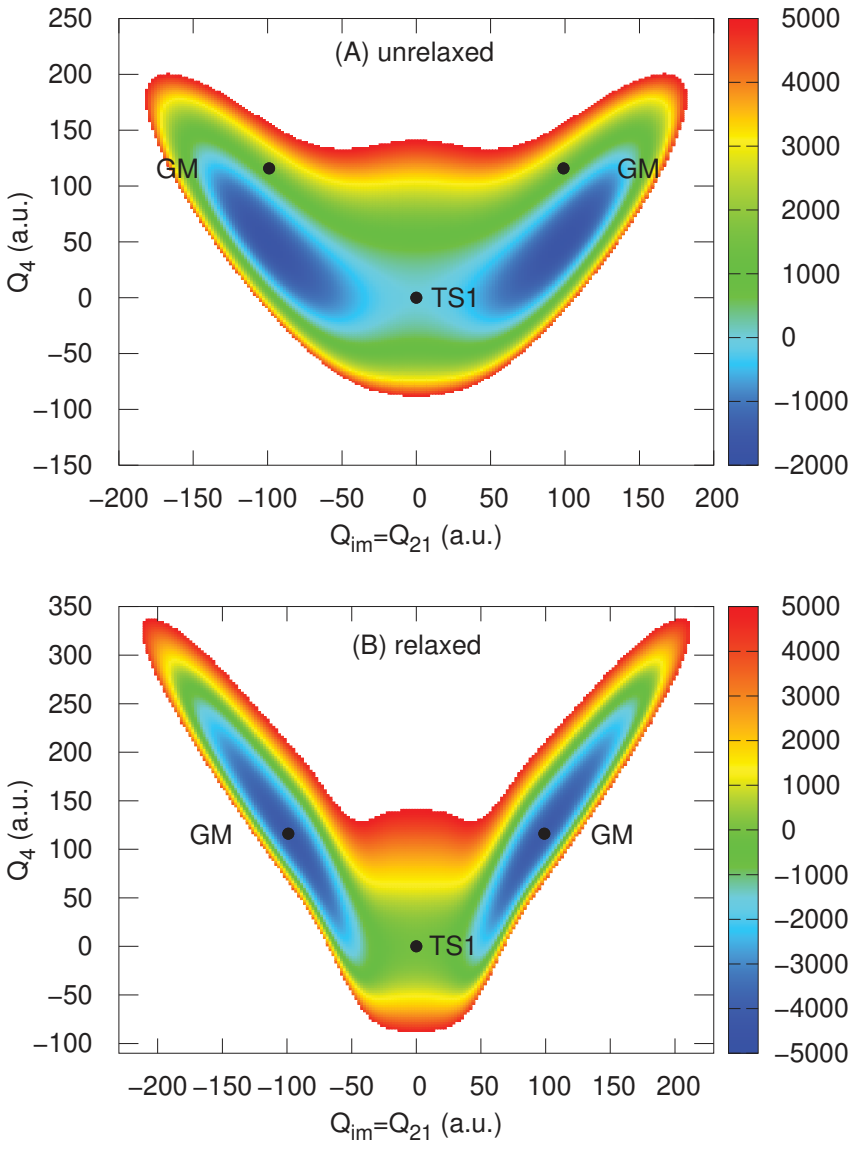

Figure 5: Contour plots of the potential energy (in cm^{-1}) as a function of Q_{21} and Q_4: (A) unrelaxed potential; (B) relaxed. The locations of the GM and TS1 are indicated by black dots.

The dipole moment and charge transfer on the IRC

We also calculated the IRC motion from TS1 using Gaussian 09 and B3LYP/aug-cc-pVDZ. The geometries were then used to calculate the energies using the same Molpro B3LYP/aug-cc-pVDZ that we used for the DFT benchmarks and the DFT PES. The barrier height was found to be 3573 cm^{-1} by this method, very similar to the benchmark and the value found on the DFT PES. The IRC is shown by the black line in Fig. 6, which also shows in blue points the dipole moment for these geometries calculated with Molpro using B3LYP/aug-cc-pVDZ. The ordinate shows the dipole moment, while the abscissa shows the IRC coordinate. The IRC path in black is scaled to have the same height on the diagram as the dipole. Thus, the minima to the right and left of the barrier have the geometry of the GM, while TS in the center has the geometry corresponding to the saddle point; these geometries are show in the figure.

It is obvious that there is a rapid change in the dipole moment on the IRC. The geometry at the barrier has four hydrogens near the nitrogen and far from the remaining atoms of the complex, so that the structure looks more like ammonium formate than the complex of formic acid and ammonia. Of course, the dipole moment of $NH_4{}^+HCOO^-$ is expected to be larger than that of $NH_3-HCOOH$.

A quantitative estimate of the degree of charge transfer is provided by a standard Mullikan charge analysis. This was done at the global minimum and the saddle point barrier and the results are shown in Figure 7 As seen there is large (0.72) charge transfer at the saddle point but less than a third that amount at the global minimum.

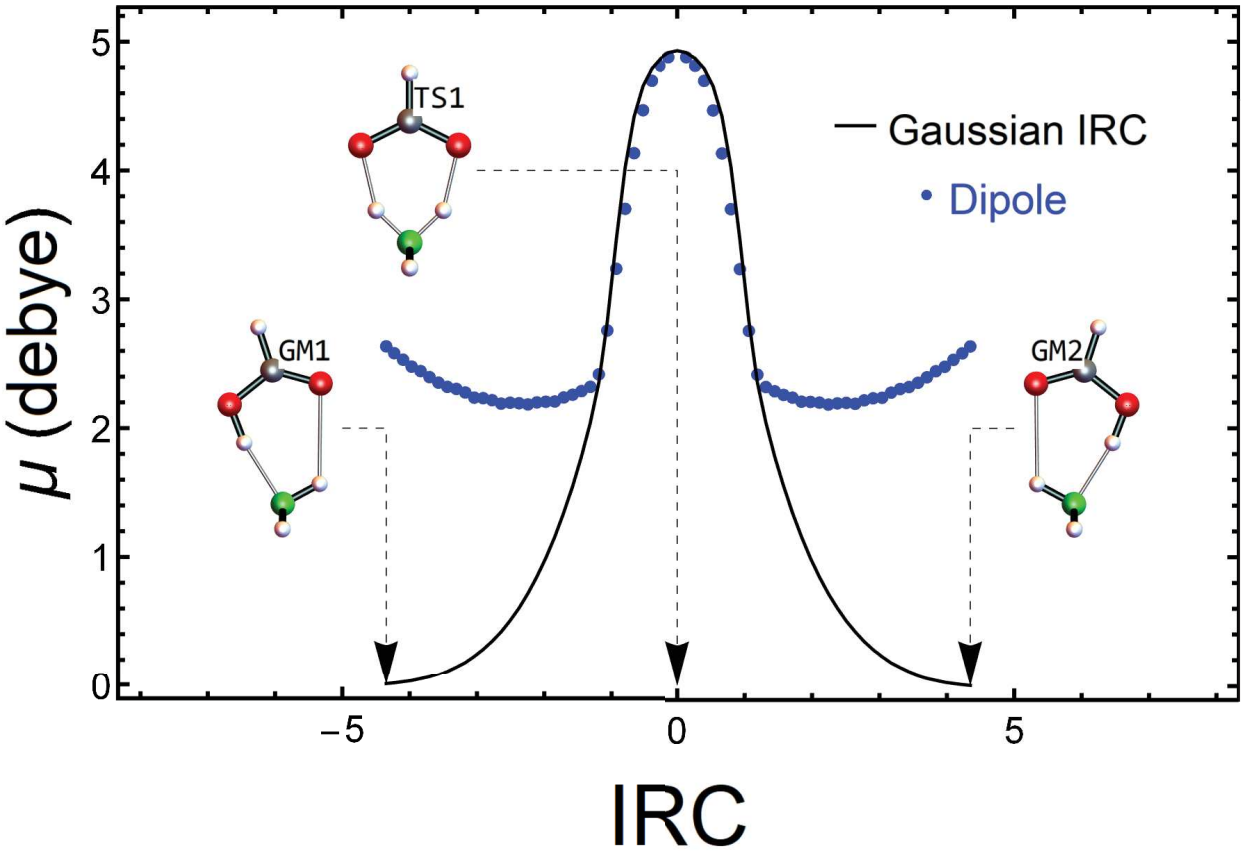

Figure 6: IRC path through the TS1, showing the change in dipole moment. The black line gives the reaction coordinate calculated by Gaussian; the potential energy maximum, 3577 cm^{-1}, has been scaled to the maximum dipole moment, 4.9 debye. The blue dots give the dipole moment as a function of this reaction coordinate. The larger dipole moment at the transition state is evidence that motion over the barrier involves proton transfer.

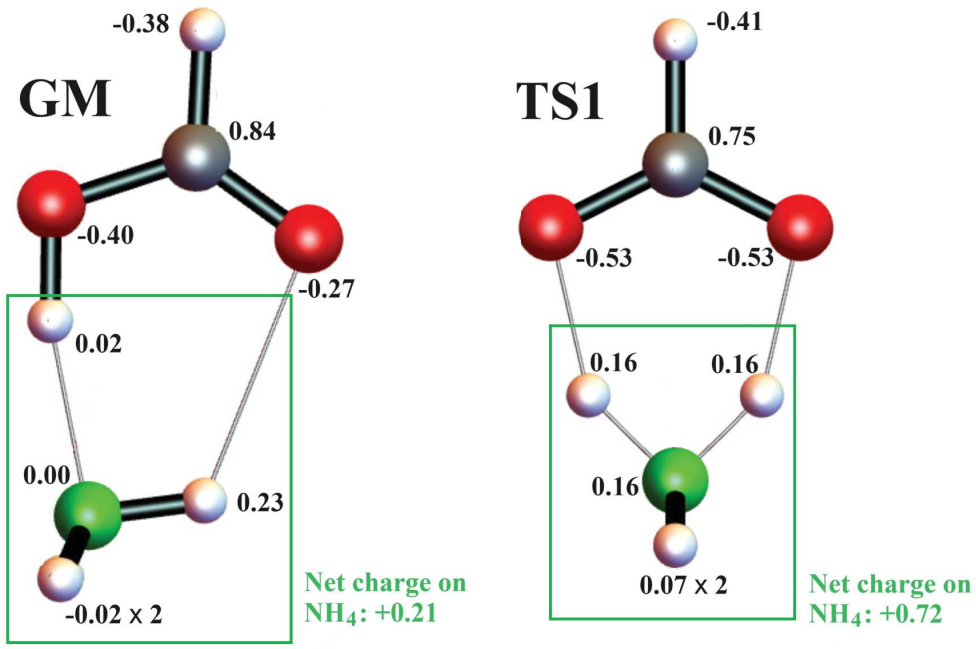

Figure 7: Mulliken population analysis of charges of formic acid-ammonia at the global minimum and saddle point barrier separating two equivalent minima.

Summary and Conclusions

The two Δ-ML methods to bring a low-level DFT to CCSD(T) accuracy produce very similar final results. This is not unexpected as both use the same approach to fit a difference potential ΔV_{CC-LL}. In the original approach this potential is just added to the fit to the DFT data.[34] In the subsequent approach[36] ΔV_{CC-LL} is used to create a dataset that is fit to get a single corrected PES. Provided ΔV_{CC-LL} is a precise fit to difference energies, predictions from it should be very close to direct calculation of the difference energies. That is evidently the case here.

We found that the dissociation energy of the formic acid-ammonia complex is 4144 cm^{-1} or about 11.8 kcal/mol. Because most N· · ·H-O hydrogen bonds have an energy of about 7.65 kcal/mol, our determination raises the question of why the dissociation energy should be higher. A possible answer to this involves the ionic contribution to the binding energy. The qualitative effect is seen from the dipole moments of the GM and TS1 structures in Figs.

1 and 6. It is clear that the dipole moment increases rapidly as one approaches TS1 from the GM on either side of the barrier. The stronger dipole implies more charge separation between the positive NH_4 end of the heterodimer and the negative HCOO end. This is verified using a Mulliken charge analysis. However, a dipole is also present at the structure of the GM and can contribute to the bonding. That it does is also supported by a symmetry adapted perturbation theory (SAPT) analysis of the GM, performed using Molpro, which shows that the largest contributions and the largest net contribution to the bonding are due to E(1) polarization and E(1) exchange.

The large barrier of around 4200 cm^{-1} the SP separating the two minima immediately suggests a small ground state tunneling splitting. (The corresponding splitting in the formic acid dimer is 0.016 cm^{-1} with a barrier height of roughly 2850 cm^{-1}.) Indeed, a 2d quantum calculation of the tunneling splitting provides an approximation to the splitting of 10^{-5} cm^{-1}. It is safe to assume that the actual splitting is smaller than 10^{-4} cm^{-1}.

Acknowledgement

JMB thanks the Army Research Office, DURIP grant (W911NF-14-1-0471), for funding a computer cluster where most of the calculations were performed. JMB and AN acknowledge current support from NASA grant (80NSSC22K1167). SGK thanks the NSF for support. RC thanks Università degli Studi di Milano for financial support under grant PSR2022_DIP_005_PI_RCONT. This material is partially based upon work supported by the National Science Foundation under Grant No. CHE-1952289 at the University of Arizona.

Supporting Information Available

- Structural Information for the GM, TS1, and TS2
- Vibrational Frequencies and mode displacement vectors for the GM and TS1

- Computer files including those for the $\Delta V_{CC-LL}(2)$, DFT, Δ-ML PES-1, and Δ-ML PES-2 potential energy surfaces, the IRC structures and energies, and the data sets used for the fits.

References

(1) 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.

(2) Wang, Y.; Bowman, J. M. Mode-specific Tunneling Using the Qim Path: Theory and an Application to Full-dimensional Malonaldehyde. J. Chem. Phys. **2013**, 139, 154303:1–5.

(3) Nandi, A.; Laude, G.; Khire, S. S.; Gurav, N. D.; Qu, C.; Conte, R.; Yu, Q.; Li, S.; Houston, P. L.; Gadre, S. R.; Richardson, J. O.; Evangelista, F. A.; Bowman, J. M. Ring-Polymer Instanton Tunneling Splittings of Tropolone and Isotopomers using a Δ-Machine Learned CCSD(T) Potential: Theory and Experiment Shake Hands. J. Amer. Chem. Soc. **2023**, 145, 9655–9664.

(4) Birer, Ö.; Havenith, M. High-Resolution Infrared Spectroscopy of the Formic Acid Dimer. Annu. Rev. Phys. Chem. **2009**, 60, 263–275.

(5) 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.

(6) Qu, C.; Bowman, J. M. High-dimensional fitting of sparse datasets of CCSD(T) electronic energies and MP2 dipole moments, illustrated for the formic acid dimer and its complex IR spectrum. J. Chem. Phys. **2018**, 148, 241713.

(7) Richardson, J. O. Full- and reduced-dimensionality instanton calculations of the tunnelling splitting in the formic acid dimer. Phys. Chem. Chem. Phys. **2017**, 19, 966–970.

(8) Novick, S. E. Bibliography of Rotational Spectra of Weakly Bound Complexes. https://snovick.faculty.wesleyan.edu/files/2020/09/vdw.pdf

(9) Bunker, R. P.; Jensen, P. Molecular Symmetry and Spectroscopy, 2nd ed.; NRC Research Press: Ottawa, 2006.

(10) Daly, A. M.; Bunker, P. R.; Kukolich, S. G. Communications: Evidence for proton tunneling from the microwave spectrum of the formic acid-propriolic acid dimer. J. Chem. Phys. **2010**, 132, 201101.

(11) Daly, A. M.; Bunker, P. R.; Kukolich, S. G. Erratum: Communications: Evidence for proton tunneling from the microwave spectrum of the formic acid: Propriolic acid dimer [J. Chem. Phys. 132, 201101 (2010)]. J. Chem. Phys. **2010**, 133, 079903.

(12) Daly, A. M.; Douglass, K. O.; Sarkozy, L. C.; Neill, J. L.; Muckle, M. T.; Zaleski, D. P.; Pate, B. H.; Kukolich, S. G. Microwave measurements of proton tunneling and structural parameters for the propiolic acid–formic acid dimer. J. Chem. Phys. **2011**, 135, 154304.

(13) Mackenzie, R. B.; Dewberry, C. T.; Leopold, K. R. The Formic Acid–Nitric Acid Complex: Microwave Spectrum, Structure, and Proton Transfer. J. Phys. Chem. A **2014**, 118, 7975–7985.

(14) Evangelisti, L.; Écija, P.; Cocinero, E. J.; Castaño, F.; Lesarri, A.; Caminati, W.; Meyer, R. Proton Tunneling in Heterodimers of Carboxylic Acids: A Rotational Study of the Benzoic Acid–Formic Acid Bimolecule. J. Phys. Chem. Lett. **2012**, 3, 3770–3775.

(15) Tayler, M. C. D.; Ouyang, B.; Howard, B. J. Unraveling the spectroscopy of coupled

intramolecular tunneling modes: A study of double proton transfer in the formic-acetic acid complex. J. Chem. Phys. **2011**, 134, 054316.

(16) Jiménez Hoyos, C. A.; Zhou, Z.; Kukolich, S. G. Calculated molecular properties and microwave spectrum analysis for formamidinium formate. J. Molec. Spectro. **2020**, 372, 111331.

(17) Pejlovas, A. M.; Serrato, r., A.; Lin, W.; Kukolich, S. G. Microwave Measurements of the Tropolone - Formic Acid Doubly Hydrogen Bonded Dimer. J. Chem. Phys. **2016**, 144, 044306.

(18) Martinache, L.; Kresa, W.; Wegener, M.; Vonmont, U.; Bauder, A. Microwave spectra and partial substitution structure of carboxylic acid bimolecules. Chemical Physics **1990**, 148, 129–140.

(19) Pejlovas, A. M.; Kukolich, S. G. Rotational spectra and gas phase structure of the maleimide – Formic acid doubly hydrogen bonded dimer. Journal of Molecular Spectroscopy **2016**, 321, 1–4.

(20) Nichols, J. L.; Roehling, K. K.; Daly, A. M.; Kukolich, S. G. Microwave measurements and calculations for the glyoxylic acid – Formic acid hydrogen-bonded complex. Journal of Molecular Spectroscopy **2023**, 395, 111806.

(21) Gou, Q.; Feng, G.; Evangelisti, L.; Caminati, W. Conformers of dimers of carboxylic acids in the gas phase: A rotational study of difluoroacetic acid–formic acid. Chemical Physics Letters **2014**, 591, 301–305.

(22) Feng, G.; Gou, Q.; Evangelisti, L.; Xia, Z.; Caminati, W. Conformational equilibria in carboxylic acid bimolecules: a rotational study of acrylic acid–formic acid. Phys. Chem. Chem. Phys. **2013**, 15, 2917–2922.

(23) Thomas, J.; Carrillo, M. J.; Serrato, A.; Lin, W.; Jäger, W.; Xu, Y. Rotational spectroscopic and theoretical study of the perfluorobutyric acid $\cdots$ formic acid complex. Journal of Molecular Spectroscopy **2017**, 335, 88–92.

(24) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; De Vries, M. Pairing of the nucleobases guanine and cytosine in the gas phase studied by IR–UV double-resonance spectroscopy and ab initio calculations. Phys. Chem. Chem. Phys. **2002**, 4, 732–739.

(25) Botti, G.; Ceotto, M.; Conte, R. Investigating the Spectroscopy of the Gas Phase Guanine–Cytosine Pair: Keto versus Enol Configurations. J. Phys. Chem. Lett. **2023**, 14, 8940–8947.

(26) Slocombe, L.; Sacchi, M.; Al-Khalili, J. An open quantum systems approach to proton tunnelling in DNA. Commun. Phys. **2022**, 5, 109.

(27) Angiolari, F.; Huppert, S.; Pietrucci, F.; Spezia, R. Environmental and Nuclear Quantum Effects on Double Proton Transfer in the Guanine–Cytosine Base Pair. J. Phys. Chem. Lett. **2023**, 14, 5102–5108.

(28) Slocombe, L.; Winokan, M.; Al-Khalili, J.; Sacchi, M. Quantum Tunnelling Effects in the Guanine-Thymine Wobble Misincorporation via Tautomerism. J. Phys. Chem. Lett. **2022**, 14, 9–15.

(29) Minyaev, R. M. Double proton transfers in complexes of formic acid with CH_4, NH_3, H_2O, HF and H_2. Chemical Physics Letters **1996**, 262, 194–200.

(30) Schweer, S. M.; Gawrilow, M.; Nejad, A.; Suhm, M. A. Formic acid–methanol complexation vs. esterification: elusive pre-reactive species identified by vibrational spectroscopy. Phys. Chem. Chem. Phys. **2023**, 25, 29982–29992.

(31) Priem, D.; Ha, T.-K.; Bauder, A. Rotational spectra and structures of three hydrogen-

bonded complexes between formic acid and water. The Journal of Chemical Physics **2000**, 113, 169–175.

(32) Roehling, K. K.; Nichols, J. L.; Daly, A. M.; Kukolich, S. G. Microwave measurements, calculations, and analysis for the gas phase ammonia-formic acid dimer. Journal of Molecular Spectroscopy **2023**, 393, 111772.

(33) Hellebust, S.; O'Riordan, B.; Sodeau, J. Cirrus cloud mimics in the laboratory: An infrared spectroscopy study of thin films of mixed ice of water with organic acids and ammonia. The Journal of Chemical Physics **2007**, 126, 084702.

(34) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ-machine Learning for Potential Energy Surfaces: A PIP Approach to Bring a DFT-based PES to CCSD(T) Level of Theory. J. Chem. Phys. **2021**, 154, 051102:1–8.

(35) Bowman, J. M.; Qu, C.; Conte, R.; Nandi, A.; Houston, P. L.; Yu, Q. Δ-Machine Learned Potential Energy Surfaces and Force Fields. J. Chem. Theory. Comput. **2023**, 19, 1–17.

(36) Liu, Y.; Li, J. Permutation-Invariant-Polynomial Neural-Network-Based Δ-Machine Learning Approach: A Case for the HO_2 Self-Reaction and Its Dynamics Study. J. Chem. Phys. Lett. **2022**, 13, 4729–4738.

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

(38) 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.

(39) Houston, P. L.; Qu, C.; Nandi, A.; Conte, R.; Yu, Q.; Bowman, J. M. Permutationally invariant polynomial regression for energies and gradients, using reverse differentiation,

achieves orders of magnitude speed-up with high precision compared to other machine learning methods. J. Chem. Phys. **2022**, 156, 044120.

(40) Houston, P. L.; Qu, C.; Yu, Q.; Conte, R.; Nandi, A.; Li, J. K.; Bowman, J. M. PESPIP: Software to fit complex molecular and many-body potential energy surfaces with permutationally invariant polynomials. J. Chem. Phys. **2023**, 158, 044109.

(41) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; others MOLPRO, version 2010.1, a package of ab initio programs. 2010; see http://www.molpro.net.

(42) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Scchutz, M. WIREs Comput. Mol. Sci. **2012**, 2, 242–253.

(43) Houston, P. L.; Conte, R.; Qu, C.; Bowman, J. M. Permutationally Invariant Polynomial Potential Energy Surfaces for Tropolone and H and D atom Tunneling Dynamics. J. Chem. Phys. **2020**, 153, 024107.

(44) Nandi, A.; Conte, R.; Qu, C.; Houston, P. L.; Yu, Q.; Bowman, J. M. Quantum Calculations on a New CCSD(T) Machine-Learned Potential Energy Surface Reveal the Leaky Nature of Gas-Phase Trans and Gauche Ethanol Conformers. J. Chem. Theor. Comput. **2022**, 18, 5527–5538.