






# Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature

Huziel E. Saucedo <sup>1,2,3✉</sup>, Valentin Vassilev-Galindo <sup>1</sup>, Stefan Chmiela <sup>2</sup>, Klaus-Robert Müller <sup>2,4,5,6✉</sup> & Alexandre Tkatchenko <sup>1✉</sup>

Nuclear quantum effects (NQE) tend to generate delocalized molecular dynamics due to the inclusion of the zero point energy and its coupling with the anharmonicities in interatomic interactions. Here, we present evidence that NQE often enhance electronic interactions and, in turn, can result in dynamical molecular stabilization at finite temperature. The underlying physical mechanism promoted by NQE depends on the particular interaction under consideration. First, the effective reduction of interatomic distances between functional groups within a molecule can enhance the  $n \rightarrow \pi^*$  interaction by increasing the overlap between molecular orbitals or by strengthening electrostatic interactions between neighboring charge densities. Second, NQE can localize methyl rotors by temporarily changing molecular bond orders and leading to the emergence of localized transient rotor states. Third, for noncovalent van der Waals interactions the strengthening comes from the increase of the polarizability given the expanded average interatomic distances induced by NQE. The implications of these boosted interactions include counterintuitive hydroxyl-hydroxyl bonding, hindered methyl rotor dynamics, and molecular stiffening which generates smoother free-energy surfaces. Our findings yield new insights into the versatile role of nuclear quantum fluctuations in molecules and materials.

<sup>1</sup>Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg. <sup>2</sup>Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany. <sup>3</sup>BASLEARN, BASF-TU joint Lab, Technische Universität Berlin, 10587 Berlin, Germany. <sup>4</sup>Department of Artificial Intelligence, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, Korea. <sup>5</sup>Max Planck Institute for Informatics, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany. <sup>6</sup>Google Research, Brain team, Berlin, Germany. ✉email: [saucedo@tu-berlin.de](mailto:saucedo@tu-berlin.de); [klaus-robot.mueller@tu-berlin.de](mailto:klaus-robot.mueller@tu-berlin.de); [alexandre.tkatchenko@uni.lu](mailto:alexandre.tkatchenko@uni.lu)

Nuclear delocalization is a fundamental feature of quantum mechanics resulting from Heisenberg's uncertainty principle. In molecules, light elements such as protons and first row atoms are especially prone to delocalization. Even molecules or materials with heavier atoms and strong bonds can exhibit significant nuclear quantum effects (NQE)<sup>1–6</sup>. Within the Born-Oppenheimer (BO) approximation, NQE tend to lower energy barriers and stimulate tunneling. In addition, the inclusion of NQE promotes a delocalized sampling of the molecular configuration space, consequently exploring regions of the potential-energy surface (PES) inaccessible by classical dynamics. As a result, this can enhance or inhibit certain molecular interactions<sup>7</sup>. A clear example is the hydrogen bond, where the NQE affect interactions in biological systems and molecular crystals by delocalizing protons. In the case of bulk water, NQE can even qualitatively change its fundamental physical and chemical properties<sup>8,9</sup>. In general, the study of NQE in molecular and biological systems is a thriving research field covering from rigid and fluxional molecules<sup>10–13</sup> to liquids and DNA base pairs<sup>14–19</sup>, allowing the analysis of, for example, interactions between neighboring molecules via hydrogen bonding<sup>7,8,20</sup>, spectroscopic properties<sup>21,22</sup> and proton transport<sup>23</sup>. Nevertheless, previous works have been mainly focused on the general implications of proton delocalization and much less is known about how NQE influence other types of covalent and non-covalent interactions. Particularly, biological systems often use combinations of covalent and non-covalent interactions for carrying out a wide variety of different processes. Therefore, it is crucial to understand whether NQE can also play an important role beyond hydrogen bonding.

In this work we report counterintuitive effects induced by NQE: nuclear delocalization can lead to a dynamical strengthening of covalent and non-covalent molecular interactions at finite temperature. Our conclusions are shown to be valid for a wide range of molecules. We have found that the explicit mechanism responsible of enhancing each interaction depends on its underlying physical or chemical origin. For example, NQE-induced increase in interatomic distances can modify the local environment of methyl rotors affecting their dynamics. At the same time, NQE increase the molecular polarizability, enhancing the dispersion interactions between molecules or molecular fragments. In other cases, NQE can shorten distances between molecular fragments, thereby boosting interactions that depend on such distances. In order to demonstrate these results, we have selected representative mechanisms ubiquitously occurring in biological systems:  $n \rightarrow \pi^*$  interactions, methyl rotors, as well as electrostatic and van der Waals interactions. The faithful description of such weak molecular interactions require high levels of theory (e.g., coupled cluster) which is not always computationally affordable when performing long ab initio path integral molecular dynamics (PIMD) simulations. In this study, we have performed PIMD simulations using machine learned molecular force fields constructed using the sGDML framework<sup>11,21,24–28</sup> and trained on coupled cluster reference data [CCSD(T) or CCSD depending on the size of the molecule] (See Supplementary Note 1 for more details).

## Results

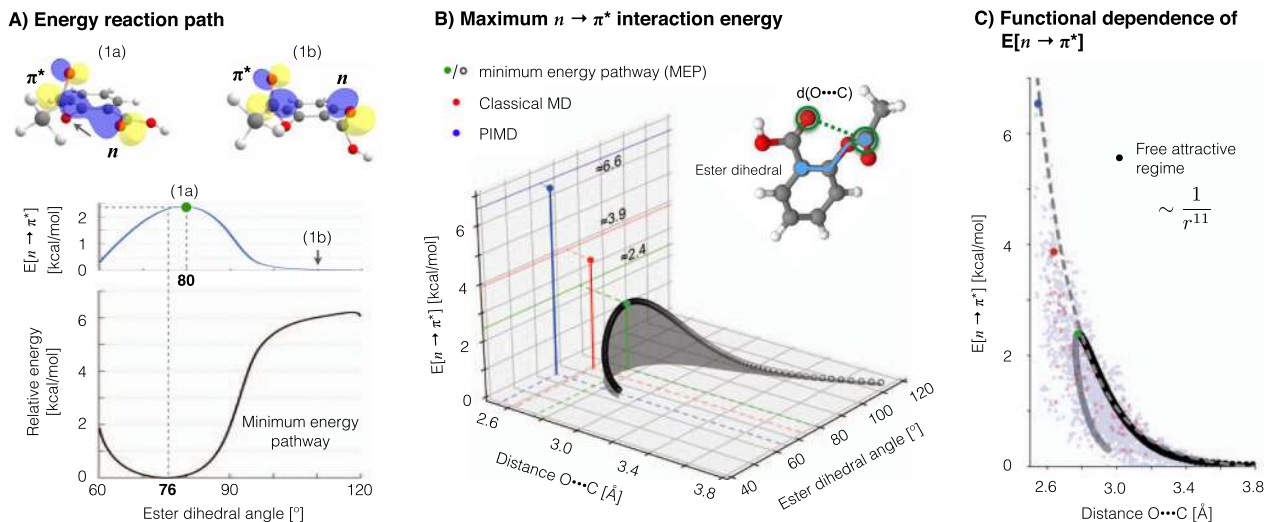
We have chosen to study the role of NQE in a series of small molecules that serve as fundamental examples of mechanisms that are present in larger chemical and biological systems. We start by analyzing  $n \rightarrow \pi^*$  interactions in aspirin as an example of local electronic orbital effects influenced by nuclear quantum delocalization. We then proceed to study the unexpected NQE-induced localization of methyl rotor in toluene as a model

for methyl groups in biomolecules. Finally, we analyze the strengthening of electrostatic interactions in paracetamol and van der Waals interactions in different conformations of benzene dimer.

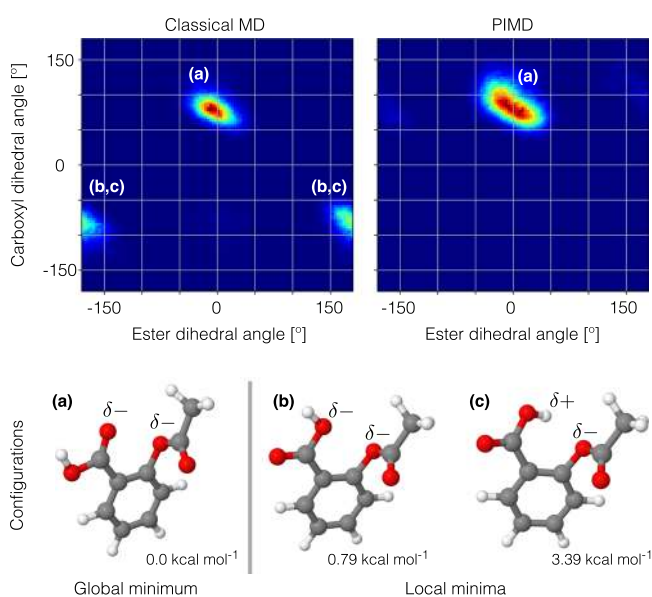
To do so, we employ the symmetric gradient domain machine learning (sGDML) framework to reconstruct the molecular force fields<sup>11,21</sup>. This approach makes full use of physical symmetries (i.e. permutational invariance and energy conservation) as priors, which enables particularly data-efficient models. In previous work, we have demonstrated that sGDML yields accurate descriptions of quantum effects in small flexible molecules<sup>25</sup>. Performing predictive calculations require a series of stringent demands on the setting of the simulations, such as highly accurate description of the quantum interactions, extensive sampling of the potential energy surface, as well as the incorporation of the nuclear quantum fluctuations. Thereby, in this article we have used coupled-cluster reference data as well as accurate density-functional approximations to train the sGDML molecular force fields and the NQE were included via the path integral molecular dynamics formalism. The molecular dynamics simulations were performed at room temperature using the i-PI package and its interface to the sGDML force field<sup>29</sup>, and all the simulations presented here were done for at least 500 ps with time steps of 0.2 fs. Additional post-processing ab initio calculations were done using methods such as natural bond orbital (NBO)<sup>30</sup> to compute the  $n \rightarrow \pi^*$  interactions and symmetry-adapted perturbation theory (SAPT)<sup>31</sup> to study the non-covalent intermolecular interactions. The SAPT and coupled cluster calculations were done using the Psi4 package<sup>32</sup>, while the NBO calculations we done using the ORCA package<sup>33</sup>. See Methodology section and Supplementary Notes 1 and 3 for an extended description of the methods used in this work.

**Enhanced  $n \rightarrow \pi^*$  interaction.** A particularly important type of interaction often occurring between pairs of neighboring carbonyl groups is the so-called  $n \rightarrow \pi^*$  interaction. It arises from the delocalization of lone-pair electrons on electronegative atoms (e.g., oxygen atom) into an antibonding  $\pi^*$  orbital of an aromatic ring or a carbonyl group (see Fig. 1A)<sup>34</sup>. First discussed in early 1970, the  $n \rightarrow \pi^*$  interaction has attracted significant attention in recent years and it is hypothesized to impart substantial structural stability to proteins<sup>35–38</sup> and molecules<sup>39–42</sup>, as well as define reactivity<sup>40</sup>, regulate isomerisation<sup>34</sup> and energy barriers<sup>43</sup>, and promote charge transfer<sup>44</sup>. Nevertheless, the actual dynamical implications at finite temperature of such interaction have not been explicitly studied.

To elucidate this matter, here we study the aspirin molecule as a proof of concept. For this molecule the  $n \rightarrow \pi^*$  interaction is the main contribution to the relative energy of the global minimum (Fig. 2a) and two other local minima (Fig. 2b, c), thereby defining their energetic ordering<sup>39,40</sup>. Figure 2 shows the configuration space sampling obtained from classical MD and PIMD simulations at room temperature, where the dynamical implications of the NQE on aspirin's behavior are evident: NQE constrains the dynamics of the molecule to the global minimum in contrast to the results from classical MD. Hence, the NQE must be promoting a particular intramolecular interaction and, given the evidence provided by Choudhary et al.<sup>40</sup>, the  $n \rightarrow \pi^*$  interaction between the ester and carboxyl groups is the main candidate. To investigate further the contribution of the  $n \rightarrow \pi^*$  interaction to the total energy, we have computed the  $n \rightarrow \pi^*$  interaction energy  $E_{n \rightarrow \pi^*}$  along the ester's minimum energy pathway (MEP) trajectory (Fig. 1A) using NBO analysis<sup>30,45</sup> (see section Methods for computational details). In what follows, we use the NBO definition for the  $n \rightarrow \pi^*$  energy<sup>30,45</sup>, whereby a



**Fig. 1** Enhancement of the  $n \rightarrow \pi^*$  interaction by nuclear delocalization in Aspirin. **A**—Bottom Aspirin's potential energy surface and **A**—Middle  $n \rightarrow \pi^*$  interaction energy  $E_{n \rightarrow \pi^*}$  along the minimum energy pathway (MEP) of the ester's dihedral angle. The energy  $E_{n \rightarrow \pi^*}$  was computed using natural bond orbital (NBO) method<sup>30,45</sup> at CCSD/cc-pVDZ level of theory, whereby a positive energy value means stabilization and zero energy means absence of overlap between  $n$  and  $\pi^*$  orbitals. Hence, a positive value implies a stabilization of the molecule. Both plots are in the same energy scale. **A**—Top The configuration (1a) (marked by the green circle) defines the maximum interaction (stabilization) energy  $E_{n \rightarrow \pi^*}$  along the MEP, while (1b) represents a configuration where overlap between lone-pair electrons  $\phi_{(n)}$  and the antibonding  $\phi_{(\pi^*)}$  orbitals, and therefore the energy  $E_{n \rightarrow \pi^*}$ , has gone to zero. **B** Estimations of the maximum  $E_{n \rightarrow \pi^*}$  interaction energy values reached while running PIMD (blue circle,  $\approx 6.6$  kcal mol<sup>-1</sup>) and classical MD (red circle,  $\approx 3.9$  kcal mol<sup>-1</sup>) at 300K using the sGDML@CCSD model. As a reference, the energy  $E_{n \rightarrow \pi^*}$  curve is plotted along the MEP trajectory as a function of its two main degrees of freedom, the interatomic distance  $d_{O \dots C}$  and the ester's dihedral angle. The maximum energy  $E_{n \rightarrow \pi^*}$  value along the MEP is  $\approx 2.4$  kcal mol<sup>-1</sup> (green circle). **C** Approximate functional dependence of the interaction energy  $E_{n \rightarrow \pi^*}$  on the oxygen (in hydroxyl) and carbon (in ester) interatomic distance  $d_{O \dots C}$ . The  $1/r^{11}$  function was fitted to the free-attractive-regime part of  $E_{n \rightarrow \pi^*}$  (black circles) along the MEP starting from 3.8 Å, giving a value of  $n - 11$ .



**Fig. 2** Classical (MD) and path integral molecular dynamics (PIMD) simulations at room temperature of aspirin described by the sGDML@CCSD molecular force field. The plots are projections of the dynamics to the two main degrees of freedom of aspirin: carboxyl and ester dihedral angles. Structures of three relevant molecular configurations are shown: **a** global minimum and **b**, **c** two of the lowest local minima.

positive energy value means stabilization and zero energy means absence of overlap between  $n$  and  $\pi^*$  orbitals. Hence, a positive value of the  $n \rightarrow \pi^*$  energy in Fig. 1 implies stabilization of the molecule. The visual representation of the  $n$  electron delocalization into  $\pi^*$  is presented in Fig. 1-A-Top. These results show that

the  $E_{n \rightarrow \pi^*}$  is maximum near the global minimum of aspirin (green circle in Fig. 1A-Middle, 1a), such value quickly vanishes as the molecule moves away towards the transition state (at 180°). From these results we see that the  $n \rightarrow \pi^*$  interaction contributes  $\sim 40\%$  of the global energy minimum relative to the transition state. Additionally, we also found that the two main degrees of freedom describing  $E_{n \rightarrow \pi^*}$  in aspirin are the  $d_{O \dots C}$  distance, also known as Bürgi-Dunitz parameter<sup>34,35,37,40,41</sup>, and the ester's dihedral angle, as shown in Fig. 1B.

Now, in order to understand the results from finite temperature simulations (Fig. 2) in the context of the  $n \rightarrow \pi^*$  interaction (Fig. 1A-Middle), we have computed the  $E_{n \rightarrow \pi^*}$  energy for a set of configurations sampled from the classical MD and PIMD trajectories on the global minimum (Fig. 2a), and then we have plotted their maximum  $E_{n \rightarrow \pi^*}$  respective values in Fig. 1-B. All the computed samples are plotted against the Bürgi-Dunitz parameter  $d_{O \dots C}$  in Fig. 1C.

The results displayed in Fig. 1B already provide a clear picture of the behavior of the  $n \rightarrow \pi^*$  interaction at finite temperatures: The maximum  $E_{n \rightarrow \pi^*}$  energy along the MEP (i.e. 0K) is of 2.4 kcal mol<sup>-1</sup> (green circle in Fig. 1), but this value can be enhanced by 160% due to pure thermal fluctuations (red circle in Fig. 1) and up to 270% by NQE at room temperature (blue circle in Fig. 1). This means that the NQE alone could strengthen the attractive interaction energy between the carbonyl and the ester functional groups by up to  $\sim 2.7$  kcal mol<sup>-1</sup> at room temperature. Consequently, given the evidence of such a considerable increment of the  $n \rightarrow \pi^*$  interaction energy and the configurational localization resulting from the molecular dynamics simulations, both originated by the NQE, we have found that nuclear quantum delocalization can stabilize intramolecular interactions and selected molecular conformations.

It is worth to analyze the underlying dynamics created by NQE that lead to such a prominent increase of the  $E_{n \rightarrow \pi^*}$  energy, which

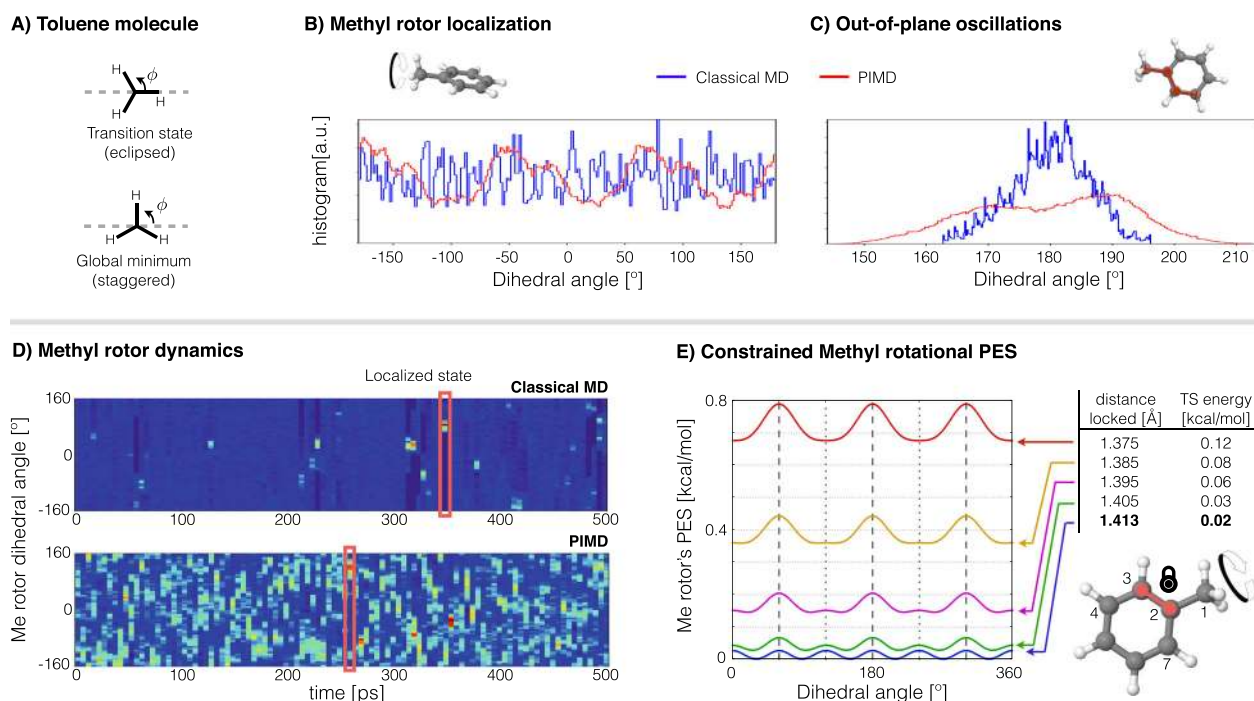
could suggest ways to generalize the results found here to other systems. From the MEP trajectory in Fig. 1B we can see the approximate dependence of the  $E_{n \rightarrow \pi^*}$  energy as a function of the  $d_{O \dots C}$  distance and the ester's dihedral angle. Given the nature of the  $n \rightarrow \pi^*$  interaction, i.e., its increase with the orbital overlap<sup>38</sup>, variations of the  $d_{O \dots C}$  distance should generate the steepest changes of  $E_{n \rightarrow \pi^*}$ . This can be seen in Fig. 1B, where a small decrease of the  $d_{O \dots C}$  and dihedral values increase the interaction energy. More interestingly, if we only focus on the free-attractive-regime of the interaction energy as displayed in Fig. 1C, e.g., qualitatively described by  $E_{n \rightarrow \pi^*}(r) \sim \int \phi_{(n)}^*(r-x)\phi_{(\pi^*)}(x)dx$  with  $\phi_{(x)}$  being the respective molecular orbitals shown in Fig. 1A-Top, the fitting of a  $r^{-n}$  function to the attractive part of the MEP trajectory suggests that the  $n \rightarrow \pi^*$  interaction energy can be approximated by  $E_{n \rightarrow \pi^*}(r) \sim r^{-11}$  (dashed line in Fig. 1C). This approximation serves as an upper limit envelope to the out-of-equilibrium configurations sampled from classical MD and PIMD simulations (red and blue circles in Fig. 1, respectively), and even extrapolates to the more extreme cases such as the maximum energy value reached by the quantum dynamics. Such a steep dependence on the distance between functional groups reveals that even a minor nuclear quantum delocalization leads to a substantial increase in stability.

From these results, and based on the fact that  $n \rightarrow \pi^*$  interactions have been consistently reported to occur in different molecular and biological systems<sup>34–40,40–44</sup>, we can hypothesize that the strengthening of such interaction by the NQE at finite temperature could prompt similar localization effects in biological systems. Hence, we conclude that nuclear quantum fluctuations are not only the source of the enhanced sampling in atomic systems, but also they can promote molecular and intermolecular rigidity in systems with prominent  $n \rightarrow \pi^*$  interactions

such as polyproline helices in protein fragments which displays a double carbonyl–carbonyl interaction<sup>41</sup>.

**Methyl rotor hindering.** The methyl (Me) functional group is a pervasive fragment in chemical and biological systems, playing a fundamental role in, for example, genetics<sup>46</sup> and protein synthesis<sup>47</sup>. The immediate chemical neighborhood of the Me group can drastically modify its energy landscape, going from a free rotor to a localized one with large energetic rotational barriers.

In general, NQE are known to play an important role in lowering energetic barriers when these are of the order of  $k_B T$ . In the particular case of the Me group, rotational barriers can be much lower than  $k_B T$ . For this reason the Me group is often considered to be a nearly free rotor at room temperature ( $k_B T \approx 0.6$  kcal mol<sup>-1</sup> at  $T = 300$  K). The toluene molecule is one of the simplest representative examples of a molecule with a Me group. The Me rotor in toluene has a sixfold rotational PES whose best experimental estimates for the energetic barrier range from  $\approx 0.014$  to  $0.028$  kcal mol<sup>-1</sup> ( $\approx 4.9$ – $9.8$  cm<sup>-1</sup>)<sup>48–50</sup>, while theoretical results at the CCSD(T) level of theory give  $\approx 0.024$  kcal mol<sup>-1</sup> (blue curve in Fig. 3E). After performing MD simulations at room temperature (classical MD and PIMD) at the sGDM@CCSD(T) level of theory and analyzing the Me rotor's dynamics, we found contrasting results to what can be trivially assumed given the nature of the system. One would expect that NQE lower the rotational energy barriers even more, but Fig. 3B shows that NQE actually hinder the Me group rotations (red) contrary to the free rotation obtained from classical MD (blue). In fact, an incremental inclusion of the NQE via increasing the number of beads in PIMD simulations demonstrates that nuclear delocalization systematically localizes the Me rotor dynamics. Additionally, the PIMD results show that the Me group no longer



**Fig. 3** Hinderance of methyl rotor dynamics by nuclear delocalization in toluene molecule. **A** Global minima and transition state of Me rotor. Histogram of the Me rotor's dihedral angle (**B**) and out-of-plane dihedral angle (**C**) computed from classical MD (blue) and PIMD (red). **D** Time evolution of the Me rotor for classical MD (top) and PIMD (bottom) simulations. The red rectangles indicate some of the localized states in the dynamics. The size of the bins are  $5$  ps  $\times$   $6^\circ$ . The simulations were done at room temperature at the sGDM@CCSD(T) level of theory. **E** Methyl rotor's PES and its transition state (TS) energy for different fixed values of the  $C_2 - C_3$  distance (marked in red). In a localized state, there is a quasi-linear correlation between  $C_2 - C_3$  and  $C_2 - C_7$  distances,  $d_{C_2-C_7} \sim -d_{C_2-C_3}$  (See Supplementary Note 7 for further details).

stays in the plane defined by the benzene ring as in the classical case, instead higher amplitude out-of-plane oscillations are observed due to the NQE (Fig. 3C).

To understand the origin of this localization, we first focus on the time evolution of the Me rotor shown in Fig. 3D. Here we can see that the classical description of the rotor is indeed a free rotor most of the time, nevertheless an interesting phenomenon emerges: The Me rotor can suddenly stop rotating for up to 4 ps. Still this is not apparent from the cumulative histograms in Fig. 3B, C due to the rare nature of this event. Contrasting with the classical model, PIMD results show a qualitatively different picture. In this case the rotor localization is much more frequent but in general the lifetime of the localized state is shorter. From this, we can hypothesize that nuclear quantum delocalization promotes the localization of the Me rotor, but at the same time the NQE tunnel the system out of the localized state. In contrast, Me rotor localization is a rare event in classical dynamics, but when it occurs, it can take a longer time for purely thermal fluctuations to bring the system out of such state. An important remark is that the Me rotor hindering mechanism in toluene is a general phenomenon independent of the employed level of electronic structure theory. We found similar results for Hartree-Fock and density-functional calculations at the PBE0 and PBE0+MBD<sup>51,52</sup> levels of theory even though their rotational energy barriers are only ~0.001, 0.008 and 0.009 kcal mol<sup>-1</sup>, respectively (See Supplementary Section 6 and Supplementary Fig. 2).

The hindering of the rotations of the Me group has a dynamical origin, and it is determined by the delocalization of the benzene carbon-carbon bonds generated by the NQE. Bond length delocalization is a well known implication of NQE, which, in this particular case, transforms the Me rotor's PES from a six-fold energy surface to a three-fold energy surface as shown in Fig. 3E. Furthermore, the magnitude of the transition state (TS) energy is determined by the two benzene ring bonds C<sub>2</sub>–C<sub>3</sub> and C<sub>2</sub>–C<sub>7</sub> near to the Me group (see Fig. 3A, E). Consequently, PIMD will tend to generate much higher rotation energy barriers given the extra dilation of such bond lengths induced by the NQE beyond the thermal dilation generated by classical MD (See also Supplementary Fig. 3). According to our results, the rotor described by PIMD experiences energetic barriers of up to 0.55 kcal mol<sup>-1</sup>, energy comparable to  $k_B T$ , therefore hindering Me rotations. From here we can conclude that the intricate quantum dynamics exhibited by the Me rotor in toluene is due to two competing NQE: On one side the nuclear delocalization of the carbon atoms promotes higher rotational energetic barriers hindering the rotor, but the quantum fluctuation of the hydrogen atoms in the Me takes the rotor out of the localized state.

Even though the results shown here are for toluene, the electronic origin of the rotational energetic barrier of Me rotor is very similar in different molecular systems (See Supplementary Fig. 3)<sup>50,53–55</sup>. Hence, similar dynamical effects are to be expected in large biological systems given the ubiquity of methyl groups in macro-molecules and protein fragments.

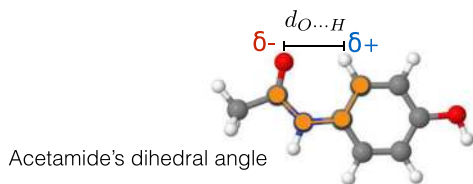
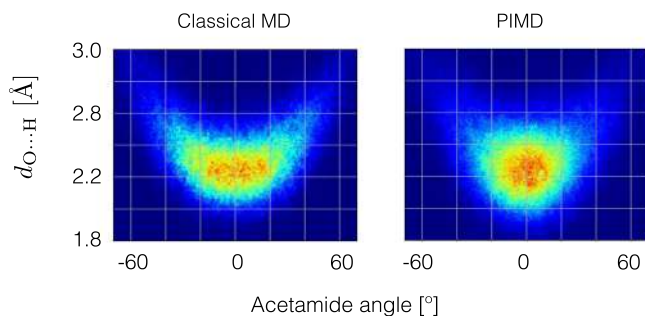
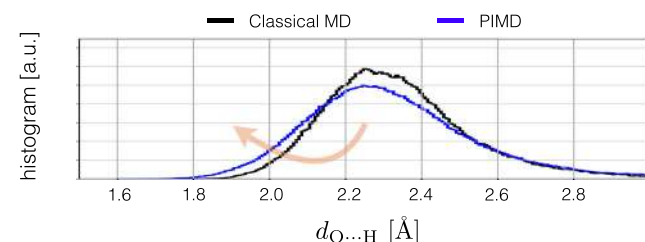
**Electrostatic interaction strengthening.** Up to now, we have shown examples of how NQE enhance covalent interactions, either by local perturbations of the electronic structure or by promoting electronic transitions. Now we turn to study how NQE can affect electrostatic interactions between charged fragments or molecules. For example, it could happen that nuclear delocalization would lead to a decrease in the distance between functional groups with opposite effective charges, thereby strengthening their direct electrostatic attractive interaction. In general, electrostatic interactions are known to play a fundamental role in biological systems given their long range effects. At the

intramolecular level, the interaction between effective atomic charges is weak compared to covalent forces. Nevertheless, the non-bonded fragments within the same molecule can substantially affect each other due to the closeness between them.

Here, we illustrate this effect by analyzing the case of the paracetamol molecule. Figure 4A shows the charge-charge interaction between the oxygen atom in the acetamide functional group and one of the hydrogen atoms in the benzene ring. By performing classical MD and PIMD simulations at room temperature, we found that the interatomic distance between the two atoms  $d_{O\dots H}$  and the acetamide's main dihedral angle  $\phi_{Ace}$  display correlated dynamics (Fig. 4B). Effectively, the surfaces displayed in Fig. 4B represent the free energies of the two simulations, clearly showing that the inclusion of NQE considerably reshapes the free energy surface. The NQE localize both the  $d_{O\dots H}$  and  $\phi_{Ace}$  degrees of freedom, and therefore, the average configuration of the molecule changes. Carefully examining the rest of the molecule's degrees of freedom during the dynamics, we see that  $d_{O\dots H}$  is the one that changes the most, mainly due to the proton delocalization to regions closer to the oxygen atom (see Fig. 4C). Therefore, such delocalization of  $d_{O\dots H}$  and the strong localization of  $\phi_{Ace}$  observed in PIMD results compared to their weak coupling in classical MD, leads us to conclude that the dynamical localization is due to the strengthened electrostatic interaction between the partially charged atoms,  $O^{\delta-} \dots H^{\delta+}$ . Similarly to the  $n \rightarrow \pi^*$  interaction where the reduction of the interatomic distance increases the  $E_{n \rightarrow \pi^*}$  energy, here the electrostatic energy is also increased, thereby effectively stabilizing the molecule.

**Impact of NQE on van der Waals interactions.** The previous examples dealt with intramolecular interactions. Now we proceed to analyze the potential role of NQE on van der Waals dispersion interactions. There is a plethora of literature on the importance of van der Waals forces in nature, from small molecules to water solvation of proteins<sup>56</sup> and from materials science to cohesion in asteroids<sup>57</sup>. As a result, it is natural to wonder if such omnipresent interaction can be affected by NQE. By considering a simple model for vdW energies, we can see that there is an immediate relationship between them. Starting from the well known  $R^{-6}$  approximation to the dispersion interaction energy between two non-bonded fragments A and B:  $E_{vdW} \sim C_6^{A,B} R_{AB}^{-6}$ . The  $C_6^{A,B}$  coefficient is computed via the Casimir-Polder integral,  $C_6^{A,B} = 3/\pi \int_0^\infty d\omega \alpha_A(i\omega) \alpha_B(i\omega)$ , where  $\alpha(i\omega)$  is the frequency-dependent polarizability evaluated at imaginary frequencies<sup>58</sup>. In the case of atoms, it was recently demonstrated that the dipole polarizability is given by  $\alpha \sim R_{vdW}^7$  where  $R_{vdW}$  is the vdW radius<sup>59</sup>. This implies that subtle increments of the molecular volume driven by temperature or NQE will considerably increase the polarizability, and consequently strengthen vdW interactions. In this regard, Aguado et al.<sup>60</sup> and Sharipov et al.<sup>61</sup> reported studies where they estimated measurable corrections to polarizability by zero-point vibrations for sodium clusters and polyatomic molecules, respectively. In consequence, NQE should have an inevitable impact on dispersion forces according to the Casimir-Polder formula.

As a proof of concept, we have considered here the benzene dimer in its three most studied conformations (see Fig. 5). But before analysing their intermolecular interaction, we test our hypothesis that NQE increase the molecular polarizability. To do so, we have computed the relative increment on the benzene isotropic polarizability compared to the optimized molecular structure. We created two ensembles of 50 molecular configurations, each sampled from classical MD and PIMD simulations (see Supporting Information for more details) and then computed

**A) Electrostatic attraction****B) Acetamide group localization****C) Hydrogen atom delocalization**

**Fig. 4 Strengthening of electrostatic interactions in paracetamol molecule by proton delocalization.** **A** Graphical representation of the paracetamol molecule. **B** Effective free energy surface generated from classical MD and PIMD at room temperature using the sGDML@DFT(PBE0+MBD) force field. **C** Interatomic distance distribution  $d_{O...H}$  generated from classical MD (black) and PIMD (blue).

the polarizability tensor using CCSD/aug-cc-pVDZ calculations. Such level of theory recovers 87% of the experimental molecular polarizability for the benzene molecule. Given that here we are interested in analyzing the relative gain in the polarizability by the thermal fluctuations and NQE relative to the equilibrium molecular configuration, such semi-quantitative calculations are sufficient. From this, we obtained that the increment on the mean isotropic polarizability  $\langle \frac{1}{3} \text{Tr}(\alpha) \rangle_{ensemble}$  is of  $\sim 1\%$  and  $\sim 3.4\%$  for the classical and quantum dynamics, respectively, which indeed proves our initial hypothesis.

Now, we proceed to investigate the implications of such NQE-induced increase of the molecular polarizability on noncovalent intermolecular interactions. In order to approximate the intermolecular interaction between the benzene pair at finite temperature (300 K), we have performed a noncovalent energy surface scanning calculation using molecular configurations sampled from classical MD and PIMD simulations. The interaction energy was computed using symmetry-adapted perturbation theory (SAPT)<sup>31</sup> as implemented in Psi4<sup>32</sup>. Each one of the points in the curves displayed in Fig. 5 is the result of an ensemble average over 50 molecular pair configurations. See Supplementary Fig. 4 for more details on how the dataset was generated. As a reference, we have also performed a normal scanning of the benzene dimer inter-molecular interaction energy (zero Kelvin) to see what is the gain in the interaction energy due to thermal and nuclear quantum fluctuations (labeled as *Relaxed*

in Fig. 5). The cases of the methane dimer and methane–benzene pair are also analyzed in the Supplementary section 8.

From all the cases considered in this study, our main result is the consistent trend displayed by the inclusion of nuclear quantum fluctuations in enhancing the total interaction energy relative to the classical simulations. The results in Fig. 5 show a very revealing behavior which can help understand further the experimental findings for the benzene dimer fluxional dynamics. First, we compare the well known relaxed-configurations energy curves for this dimer. The shape of the curves and energetic ordering of the structures agrees with previously reported results, and in particular the almost degenerate energy values for the parallel-displaced and T-shaped configurations<sup>62</sup>. If we consider the thermal dilation of the molecules, we see that the interaction energy increases as expected, but it is only the inclusion of the nuclear quantum fluctuations which reveals the trends in the benzene–benzene interaction at finite temperature. From Fig. 5 we see that NQE at finite temperature break the energetic quasi-degeneracy between the parallel-displaced and T-shaped configurations by increasing the energy of the parallel-displaced structure by  $\sim 10\%$  relative to the T-shaped one. This suggests that the parallel-displaced benzene dimer configuration would be slightly favored at room temperature. Still it would be expected that the most sampled region of the configuration space is the intermediate state between the two minima as reported by the experiment<sup>63</sup>. In fact, we notice that the T-shaped arrangement of the dimer barely profits (energetically) from the thermal or nuclear quantum fluctuations, contrary to the case of the parallel configuration which is the one that benefits the most ( $\sim 20\%$ ).

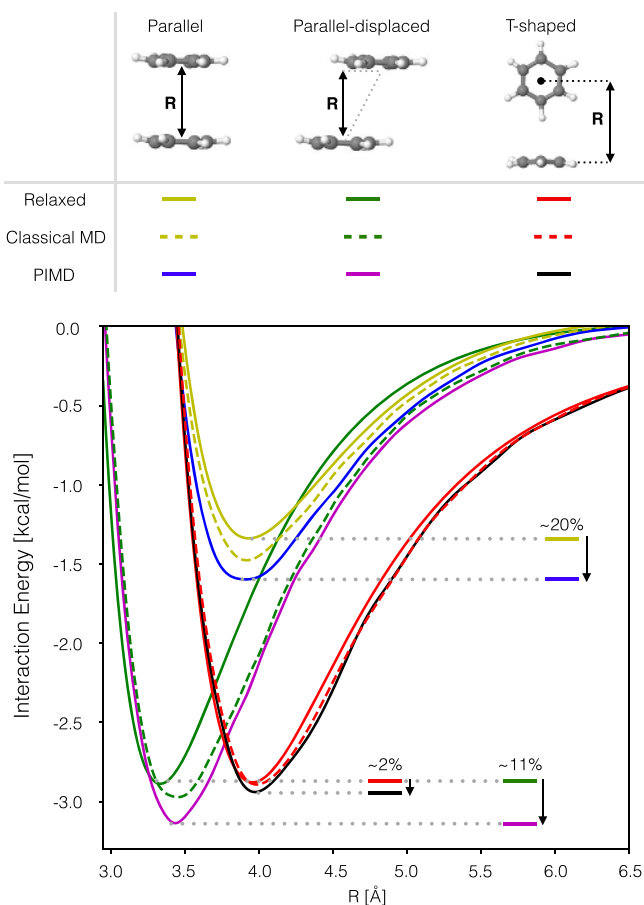
From the results presented here, we can conclude that indeed the NQE increase the molecular polarizability and therefore enhance noncovalent interactions. Furthermore, these results suggest that NQE could have a strong impact on large dispersion-dominated biological systems such as proteins and solvated molecules due to their cumulative nature.

Here we studied a model where the benzene molecules are maintained at a fixed distance from each other. While the presented model might be considered idealized, such situations are abundant in nature when the mutual distance between polarizable moieties is effectively fixed by intramolecular constraints or external conditions such as pressure or by confinement of molecules in nanostructures.

**Discussion**

The results presented in this study show convincing evidence of intra- and inter-molecular interactions enhanced by NQE, which then promote localized dynamics in molecular configuration space. This finding is counterintuitive, given that it is commonly assumed that NQE tend to delocalize molecular dynamics by expanding covalent bond lengths, increasing non-covalent distances or lowering energetic barriers. Nevertheless, the effects of NQE delocalization are far from being trivial in complex molecular conformations that are stabilized by an interplay between covalent and non-covalent interactions. These interactions can conspire to reshape the molecular bond length distribution and the free energy landscape. Even relatively small changes in bond lengths can, in a cumulative manner, bring together non-bonded fragments of the molecules altering their local environments and facilitating the occurrence of unexpected processes. The underlying mechanism responsible for the NQE-induced localization varies depending on the particular interaction. Localization can happen by breaking molecular symmetry as in the case of the Me rotor via increasing the total potential energy and creating a transient well-like potential which confines its dynamics (see Fig. 3E). Alternatively, in the case of non-bonded

## Benzene dimer noncovalent interaction



**Fig. 5 Strengthening of noncovalent interactions in benzene dimer by NQE.** Interaction energy calculations scan along the inter-molecular distance  $R$  for the parallel, parallel-displaced and T-shaped configurations of the benzene dimer. Each of these curves was calculated for the relaxed configurations and for configurations sampled from classical MD and PIMD trajectories of single benzene molecule. The total noncovalent interaction energies were computed using SAPT0/jul-cc-pVDZ<sup>31</sup> as implemented in Psi4 package<sup>32</sup>. The molecular configurations were sampled from classical MD and PIMD trajectories computed using i-PI software coupled with the sGDML@CCSD(T) force field. See Supplementary Note 8 for more details.

functional groups, NQE can decrease the distance between neighboring fragments thereby promoting orbital overlap as in the case of  $n \rightarrow \pi^*$  interaction or increase the electrostatic energy between partial charges, which indeed drastically lowers the energy of the system. On the other hand, the NQE-induced dilation of the molecular effective volume and polarizability increase is the underlying mechanism behind the strengthening of noncovalent van der Waals interactions.

It is clear that in large and complex molecules (for example biological systems) there are many more interactions than the ones discussed here, but the diverse selection of interactions and processes in this study represents several important classes of phenomena that are commonly present in biological systems. A valid argument in this regard is the fact that many of these interactions are weak compared to covalent forces or even hydrogen bonds. Interestingly, this assumption is mostly based on theoretical calculations where the finite temperature effects are ignored. More importantly, here we have shown that nuclear quantum fluctuations can contribute considerably to the enhancement of these seemingly weak interactions. In the case of

dispersion interactions, it is worth to highlight that here we have found that they can be considerably strengthened by NQE. This result suggests that we could have been underestimating the true contribution of vdW forces, given their ubiquity and paramount role in the description of chemical and biological mechanisms.

It is worth to remark that even for efficient machine-learning models, it is not yet computationally feasible to describe real biological systems with tens of thousands of atoms while preserving the accuracy needed to describe many of the subtle quantum effects. Thereby a proper understanding of NQE and their implications on covalent and noncovalent molecular interactions in biologically relevant fragments (e.g., amino acids, DNA base pairs, peptides and small proteins) could be the key to design atomistic molecular force fields capable of accurately capturing complex electronic and nuclear quantum-mechanical effects at finite temperature.

## Methods

**Machine learned force fields: sGDML.** The molecular interactions used for the simulations performed in this work were generated with the symmetric Gradient-Domain Machine Learning (sGDML) framework<sup>11,21</sup>. The sGDML method has been used to obtain accurate reconstructions of flexible molecular force fields from reference datasets of high-level ab initio calculations<sup>25</sup>. More specifically, the different levels of generalization accuracies and applicability of the sGDML framework have been assessed by (1) reproducing the exact trajectories from electronic structure-based MD simulations starting from the same initial conditions<sup>26</sup>, (2) recovering the interatomic distance distribution for several molecules relative to ab initio simulations<sup>11,26</sup>, and (3) by corroborating experimental molecular vibrational spectra and the statistical occupation values of molecular conformers<sup>21</sup>.

Unlike traditional classical force fields, the sGDML approach imposes no hypothesized interaction pattern for the atoms and can thus model any complex molecular interactions<sup>25,64</sup>. Instead, sGDML imposes energy conservation as inductive bias, a fundamental property of closed classical and quantum mechanical systems that does not limit generalization. This allows highly data efficient reconstruction of molecular force fields, without sacrificing generality.

There have been many attempts to accurately learn molecular force fields by using mainly neural networks<sup>65–69</sup> and kernel-based models<sup>70–77</sup>, cf. also refs. <sup>78</sup> and <sup>79</sup>. Nevertheless, the sGDML framework offers many advantages desirable when performing highly accurate PIMD simulations, such as high data efficiency during training for reaching state-of-the-art prediction accuracies and computational efficiency<sup>24</sup>. Therefore, to fully reach the limits of the sGDML models' accuracy, an appropriate number of training data points has to be used. In particular, sGDML models display a fast saturation of the generalization accuracy with the size of the training set<sup>21</sup>. In this work, we have used models trained on 1000 molecular configurations, clearly reaching the saturation limit<sup>21</sup>. The generalization error across the potential energy surface for all the molecules included in this study is below 0.2 kcal/mol for energies and below 0.7 kcal mol<sup>-1</sup> Å<sup>-1</sup> for forces <sup>24,25</sup>.

**Reference molecular databases.** For this study we have used a variety of quantum chemistry levels of theory. The selection of the reference level of theory for each molecule in this study was done based on its accuracy relative to CCSD(T)/cc-pVDZ method. The results for aspirin were generated with models trained on CCSD calculations using the Dunning's correlation-consistent basis set cc-pVDZ, while in the cases of toluene and benzene, we have used CCSD(T)/cc-pVDZ data directly. It is important to highlight that for the small molecules considered in this work, the convergence of their thermodynamical and statistical properties do not require large basis sets, and in particular for the above mentioned molecules this means that their intramolecular interactions are already well captured with cc-pVDZ basis set. As an example, we have performed classical MD and PIMD simulations using a sGDML model for aspirin trained on CCSD/cc-pVTZ reference calculations, results that display the same behavior as the ones reported in Fig. 2 generated using a smaller cc-pVDZ basis set (See Supplementary Figure 1 for a comparison). In the case of the paracetamol molecule, we have used DFT at the PBE0 level of theory with many body dispersion (MBD) method to account for van der Waals interactions<sup>51</sup> using the *really\_tight* basis set settings as implemented in the FHI-aims code<sup>80</sup>. Such reference level of theory was used given that PBE0 + MBD yields better relative energies of equilibrium configurations with respect to CCSD(T)/cc-pVDZ compared to CCSD/cc-pVDZ for this specific molecule. See Supplementary Table 1 for a comparison of the energies for the different levels of theory. For further details regarding the accuracy of the *really\_tight* basis set refer to the Supplementary Note 2.

To confirm the robustness of our conclusions, we also carried out a range of classical MD and PIMD simulations using sGDML models trained on different levels of theory, such as Hartree-Fock/aug-cc-pVQZ and PBE0+MBD/*really\_tight* reference calculations for toluene and aspirin. These additional reference

calculations are free from any basis set artifacts and are converged to meV-level accuracy. Overall, PIMD simulations with these additional sGDML models fully support the results and conclusions presented in the main manuscript. The generated results and their description can be found in the Supplementary Figs. 1 and 2.

All the datasets used in this work were generated by the representative sampling method as reported in ref. <sup>26</sup>, page 284.

**Calculation of  $E_{n \rightarrow \pi^*}$ .** Natural Bond Orbital (NBO) second order perturbative energies  $E_{n \rightarrow \pi^*}$  obtained from NBO 7.0<sup>45</sup> calculations coupled with ORCA 4.1.2<sup>33</sup> at CCSD/cc-pVDZ level of theory were taken as the stabilization energies due to  $n \rightarrow \pi^*$  interactions. Such interactions, which play an important role in molecular reactivity and conformation (for instance, the Bürgi-Dunitz trajectory<sup>81</sup> preferred during nucleophilic attacks at a carbonyl carbon), comprise delocalization of lone-pair electrons ( $n$ ) of an electronegative atom into an empty  $\pi^*$ -antibonding orbital of an aromatic ring or a carbonyl group<sup>40,42,82</sup>.

## Data availability

All datasets used in this work are available at <http://www.sgdml.org> or <http://quantum-machine.org/datasets/>. Additional data related to this paper may be requested from the authors.

## Code availability

The full documentation of the sGDML software can be found at <http://quantum-machine.org/gdml/doc/> and the code can be downloaded from <https://github.com/stefanch/sGDML>.

Received: 12 June 2020; Accepted: 12 November 2020;

Published online: 19 January 2021

## References

- Merchant, H. D., Srivastava, K. K. & Pandey, H. D. Equations of state and thermal expansion of alkali halides. *Crit. Rev. Solid State* **3**, 451–504 (1973).
- Kirchner, V., Heinke, H., Hommel, D., Domagala, J. Z. & Leszczynski, M. Thermal expansion of bulk and homoepitaxial GaN. *Appl. Phys. Lett.* **77**, 1434–1436 (2000).
- Hermet, P., Koza, M. M., Ritter, C., Reibel, C. & Viennois, R. Origin of the highly anisotropic thermal expansion of the semiconducting ZnS and relations with its thermoelectric applications. *RSC Adv.* **5**, 87118–87131 (2015).
- Kim, D. S. et al. Nuclear quantum effect with pure anharmonicity and the anomalous thermal expansion of silicon. *Proc. Natl Acad. Sci. USA* **115**, 1992–1997 (2018).
- Poltavsky, I., Zheng, L., Mortazavi, M. & Tkatchenko, A. Quantum tunneling of thermal protons through pristine graphene. *J. Chem. Phys.* **148**, 204707 (2018).
- Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2**, 0109 (2018).
- Rossi, M., Fang, W. & Michaelides, A. Stability of complex biomolecular structures: van der Waals, hydrogen bond cooperativity, and nuclear quantum effects. *J. Phys. Chem. Lett.* **6**, 4233–4238 (2015).
- Li, X.-Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl Acad. Sci. USA* **108**, 6369–6373 (2011).
- Shrestha, B. R. et al. Nuclear quantum effects in hydrophobic nanoconfinement. *J. Phys. Chem. Lett.* **10**, 5530–5535 (2019).
- Häberhohn, S., Markland, T. E. & Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **131**, 024501 (2009).
- Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
- Schran, C., Briec, F. & Marx, D. Converged colored noise path integral molecular dynamics study of the Zundel cation down to ultralow temperatures at coupled cluster accuracy. *J. Chem. Theory Comput.* **14**, 5068–5078 (2018).
- Novikov, I. S., Shapeev, A. V. & Suleimanov, Y. V. Ring polymer molecular dynamics and active learning of moment tensor potential for gas-phase barrierless reactions: Application to  $s + h_2$ . *J. Chem. Phys.* **151**, 224105 (2019).
- Pérez, A., Tuckerman, M. E., Hjalmarsen, H. P. & von Lilienfeld, O. A. Enol tautomers of Watson-Crick base pair models are metastable because of nuclear quantum effects. *J. Am. Chem. Soc.* **132**, 11510–11515 (2010).
- Hay, S. & Scrutton, N. S. Good vibrations in enzyme-catalysed reactions. *Nat. Chem.* **4**, 161–168 (2012).
- Ceriotti, M. et al. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* **116**, 7529–7550 (2016).
- Fang, W. et al. Inverse temperature dependence of nuclear quantum effects in DNA base pairs. *J. Phys. Chem. Lett.* **7**, 2125–2131 (2016).
- Berger, A., Ciardi, G., Sidler, D., Hamm, P. & Shalit, A. Impact of nuclear quantum effects on the structural inhomogeneity of liquid water. *Proc. Natl Acad. Sci. USA* **116**, 2458–2463 (2019).
- Clark, T., Heske, J. & Kühne, T. D. Opposing electronic and nuclear quantum effects on hydrogen bonds in H<sub>2</sub>O and D<sub>2</sub>O. *ChemPhysChem* **20**, 2461–2465 (2019).
- Wang, L., Fried, S. D., Boxer, S. G. & Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl Acad. Sci. USA* **111**, 18454–18459 (2014).
- Chmiela, S., Sauceda, H. E., Müller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **9**, 3887 (2018).
- Litman, Y., Richardson, J. O., Kumagai, T. & Rossi, M. Elucidating the nuclear quantum dynamics of intramolecular double hydrogen transfer in porphycene. *J. Am. Chem. Soc.* **141**, 2526–2534 (2019).
- Wang, L., Fried, S. D. & Markland, T. E. Proton network flexibility enables robustness and large electric fields in the ketosteroid isomerase active site. *J. Phys. Chem. B* **121**, 9807–9815 (2017).
- Chmiela, S., Sauceda, H. E., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. sgdml: Constructing accurate and data efficient molecular force fields using machine learning. *Comput. Phys. Commun.* **240**, 38–45 (2019).
- Sauceda, H. E., Chmiela, S., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *J. Chem. Phys.* **150**, 114102 (2019).
- Sauceda, H. E., Chmiela, S., Poltavsky, I., Müller, K.-R. & Tkatchenko, A. Construction of Machine Learned Force Fields with Quantum Chemical Accuracy: Applications and Chemical Insights, 277–307 (Springer International Publishing, 2020).
- Chmiela, S., Sauceda, H. E., Tkatchenko, A. & Müller, K.-R. Accurate Molecular Dynamics Enabled by Efficient Physically Constrained Machine Learning Approaches, 129–154 (Springer International Publishing, 2020).
- Wang, J., Chmiela, S., Müller, K.-R., Noé, F. & Clementi, C. Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.* **152**, 194106 (2020).
- Kapil, V. et al. i-pi 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **236**, 214–223 (2019).
- Reed, A. E., Curtiss, L. A. & Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **88**, 899–926 (1988).
- Parker, T. M., Burns, L. A., Parrish, R. M., Ryno, A. G. & Sherrill, C. D. Levels of symmetry adapted perturbation theory (sapt). i. efficiency and performance for interaction energies. *J. Chem. Phys.* **140**, 094106 (2014).
- Smith, D. G. A. et al. Psi4numpy: an interactive quantum chemistry programming environment for reference implementations and rapid development. *J. Chem. Theory Comput.* **14**, 3504–3511 (2018).
- Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1327 (2018).
- Gorske, B. C., Bastian, B. L., Geske, G. D. & Blackwell, H. E. Local and tunable  $n \rightarrow \pi^*$  interactions regulate amide isomerism in the peptoid backbone. *J. Am. Chem. Soc.* **129**, 8928–8929 (2007).
- Bartlett, G. J., Choudhary, A., Raines, R. T. & Woolfson, D. N.  $n \rightarrow \pi^*$  interactions in proteins. *Nat. Chem. Biol.* **6**, 615–620 (2010).
- Newberry, R. W., VanVeller, B., Guzei, I. A. & Raines, R. T.  $n \rightarrow \pi^*$  interactions of amides and thioamides: implications for protein stability. *J. Am. Chem. Soc.* **135**, 7843–7846 (2013).
- Jakobsche, C. E., Choudhary, A., Miller, S. J. & Raines, R. T.  $n \rightarrow \pi^*$  interaction and  $n(\pi)$  Pauli repulsion are antagonistic for protein stability. *J. Am. Chem. Soc.* **132**, 6651–6653 (2010).
- Wenzell, N. A. et al. Electronic and steric control of  $n \rightarrow \pi^*$  interactions: stabilization of the  $\alpha$ -helix conformation without a hydrogen bond. *Chem. Bio. Chem.* **20**, 963–967 (2019).
- Cabezas, C., Alonso, J. L., López, J. C. & Mata, S. Unveiling the shape of aspirin in the gas phase. *Angew. Chem. Int. Ed.* **51**, 1375–1378 (2012).
- Choudhary, A., Kamer, K. J. & Raines, R. T. An  $n \rightarrow \pi^*$  interaction in aspirin: Implications for structure and reactivity. *J. Org. Chem.* **76**, 7933–7937 (2011).
- Rahim, A., Saha, P., Jha, K. K., Sukumar, N. & Sarma, B. K. Reciprocal carbonyl-carbonyl interactions in small molecules and proteins. *Nat. Commun.* **8**, 78 (2017).



42. Singh, S. K., Mishra, K. K., Sharma, N. & Das, A. Direct spectroscopic evidence for an  $n \rightarrow \pi^*$  interaction. *Angew. Chem. - Int. Ed.* **55**, 7801–7805 (2016).
43. Blanco, S., Macario, A. & López, J. C. Pyridine-acetaldehyde, a molecular balance to explore the  $n \rightarrow \pi^*$  interaction. *Phys. Chem. Chem. Phys.* **21**, 20566–20570 (2019).
44. Zhao, Y.-L., Ullah, N., Chen, S. & Zhang, R.-Q.  $n \rightarrow \pi^*$  interaction promoted charge carrier transfer between helical swnts and a 4-(1-pyrenyl)phenyl group. *J. Phys. Chem. C* **123**, 13976–13982 (2019).
45. Glendening, E. D. et al. NBO 7.0 (2018).
46. Feherm, J. 2.2 - dna and protein synthesis. In Feher, J. (ed.) *Quantitative Human Physiology (Second Edition)*, 120–129 (Academic Press, Boston, 2017), second edition edn.
47. Harris, C. J. et al. A dna methylation reader complex that enhances gene transcription. *Science* **362**, 1182–1186 (2018).
48. Gough, K. M., Henry, B. R. & Wildman, T. A. Conformational preferences and internal rotation in toluene, o-xylene and hexamethylbenzene. *J. Mol. Structure: THEOCHEM* **124**, 71–85 (1985).
49. Gough, K. M. & Henry, B. R. Gas-phase overtone spectral investigation of inequivalent aryl and alkyl carbon-hydrogen (c-h) bonds in toluene and the xylenes. *J. Phys. Chem.* **88**, 1298–1302 (1984).
50. Breen, P. J., Warren, J. A., Bernstein, E. R. & Seeman, J. I. A study of nonrigid aromatic molecules by supersonic molecular jet spectroscopy. i. toluene and the xylenes. *J. Chem. Phys.* **87**, 1917–1926 (1987).
51. Ambrosetti, A., Reilly, A. M., DiStasio, R. A. & Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **140**, 18A508 (2014).
52. Tkatchenko, A., DiStasio, R. A. Jr., Car, R. & Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **108**, 236402 (2012).
53. Liljefors, T. & Allinger, N. L. A note on torsional force constants in molecular mechanics for a methyl group attached to a conjugated system. *J. Comput. Chem.* **6**, 478–480 (1985).
54. George, P., Bock, C. W., Stezowski, J. J., Hildenbrand, T. & Glusker, J. P. Molecular orbital studies of methylation effects in aromatic hydrocarbons. 1. ab initio calculations of the structure, electronic properties and energy of toluene, 1-methylnaphthalene, and 2-methylnaphthalene. *J. Phys. Chem.* **92**, 5656–5666 (1988).
55. Kundu, T., Pradhan, B. & Singh, B. P. Origin of methyl torsional potential barrier — an overview. *J. Chem. Sci.* **114**, 623–638 (2002).
56. Stöhr, M. & Tkatchenko, A. Quantum mechanics of proteins in explicit water: the role of plasmon-like solute-solvent interactions. *Sci. Adv.* **5**, eaax0024 (2019).
57. Scheeres, D., Hartzell, C., Sánchez, P. & Swift, M. Scaling forces to asteroid surfaces: the role of cohesion. *Icarus* **210**, 968–984 (2010).
58. Tkatchenko, A. & Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **102**, 073005 (2009).
59. Fedorov, D. V., Sadhukhan, M., Stöhr, M. & Tkatchenko, A. Quantum-mechanical relation between atomic dipole polarizability and the van der Waals radius. *Phys. Rev. Lett.* **121**, 183401 (2018).
60. Aguado, A., Vega, A. & Ballbás, L. C. Structural and zero-point vibrational effects on the electric dipole moments and static dipole polarizabilities of sodium clusters. *Phys. Rev. B* **84**, 165450 (2011).
61. Sharipov, A. S., Loukhovitski, B. I. & Starik, A. M. The influence of vibrations of polyatomic molecules on dipole moment and static dipole polarizability: theoretical study. *J. Phys. B* **50**, 165101 (2017).
62. Janowski, T. & Pulay, P. High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chem. Phys. Lett.* **447**, 27–32 (2007).
63. Schnell, M. et al. Structure of the benzene dimer-governed by dynamics. *Angew. Chem. Int. Ed.* **52**, 5180–5183 (2013).
64. Saucedo, H. E., Gastegger, M., Chmiela, S., Müller, K.-R. & Tkatchenko, A. Molecular force fields with gradient-domain machine learning (gdml): Comparison and synergies with classical force fields. *J. Chem. Phys.* **153**, 124109 (2020).
65. Behler, J. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
66. Gastegger, M., Behler, J. & Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **8**, 6924–6935 (2017).
67. Schütt, K. T. et al. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In Guyon, I. et al. (eds.) *Advances in Neural Information Processing Systems 30*, 991–1001 (Curran Associates, Inc., 2017).
68. Schütt, K. T., Saucedo, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet – a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
69. Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 16.1–16.30 (2020).
70. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
71. Bartók, A. P. & Csányi, G. Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051–1057 (2015).
72. Botu, V. & Ramprasad, R. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B* **92**, 094306 (2015).
73. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).
74. Glielmo, A., Sollich, P. & De Vita, A. Accurate interatomic force fields via machine learning with covariant kernels. *Phys. Rev. B* **95**, 214302 (2017).
75. Brockherde, F. et al. Bypassing the kohn-sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).
76. von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358 (2020).
77. Bogojeski, M., Vogt-Maranto, L., Tuckerman, M. E., Müller, K.-R. & Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nat. Commun.* **11**, 5223 (2020).
78. Schütt, K. T. et al. Machine Learning Meets Quantum Physics, vol. 968 (Springer Lecture Notes in Physics, 2020).
79. Unke, O. T. et al. Machine learning force fields. Preprint at <https://arxiv.org/abs/2010.07067> (2020).
80. Blum, V. et al. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **180**, 2175–2196 (2009).
81. Bürgi, H. B., Dunitz, J. D. & Shefter, E. Chemical reaction paths. IV. Aspects of  $O \cdots C=O$  interactions in crystals. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **30**, 1517–1527 (1974).
82. Newberry, R. W. & Raines, R. T. The  $n \rightarrow \pi^*$  interaction. *Acc. Chem. Res.* **50**, 1838–1846 (2017).

## Acknowledgements

V.V.G. and A.T. were supported by the Luxembourg National Research Fund (DTU PRIDE MASSENA) and by the European Research Council (ERC-CoG BeStMo). K.R.M. was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (No. 2019-0-00079, Artificial Intelligence Graduate School Program, Korea University), and was partly supported by the German Ministry for Education and Research (BMBF) under Grants 01IS14013A-E, 01GQ1115, 01GQ0850, 01IS18025A and 01IS18037A; the German Research Foundation (DFG) under Grant Math+, EXC 2046/1, Project ID 390685689.

## Author contributions

H.E.S. and A.T. conceived the research and designed the analyses. H.E.S. and V.V.G. performed quantum chemical calculations. S.C., H.E.S., A.T., and K.R.M. developed the machine learning methodology. H.E.S. performed the molecular dynamics simulations. H.E.S. created the figures with help from other authors. H.E.S., K.R.M. and A.T. wrote the paper. H.E.S., V.V.G., S.C., K.R.M., and A.T. discussed results and commented on the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-20212-1>.

Correspondence and requests for materials should be addressed to H.E.S., K.-R.M. or A.T.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021