

# Nuclear quantum effects in molecular liquids across chemical space

Received: 9 January 2025

Baris E. Ugur  & Michael A. Webb  

Accepted: 4 June 2025

Published online: 01 July 2025

 Check for updates

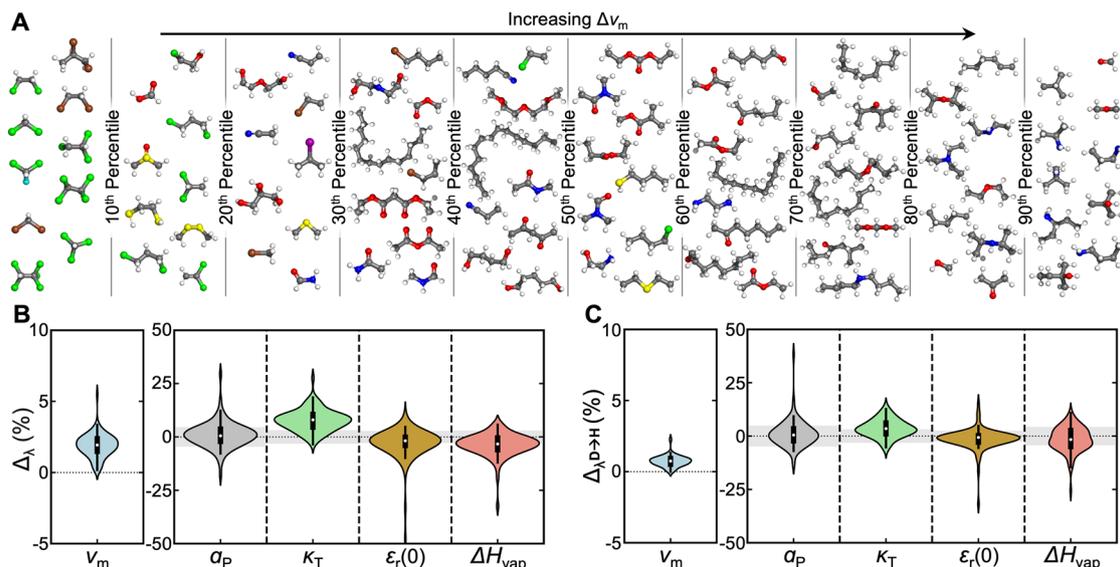
Nuclear quantum effects (NQE) influence many physical and chemical phenomena, particularly those involving light atoms or occurring at low temperatures. However, their impact has been carefully quantified in few systems-like water-and is rarely considered more broadly. Here we use path-integral molecular dynamics to systematically investigate NQEs on thermophysical properties of 92 organic liquids at ambient conditions. Depending on chemical constitution, we find substantial impact across thermal expansivity, compressibility, dielectric constant, enthalpy of vaporization, and notably molar volume, which shows consistent, positive quantum-classical differences up to 5%; similar, less pronounced trends manifest as isotope effects from deuteration. Using data-driven analysis, we identify three features-molar mass, classical hydrogen density, and classical thermal expansivity-that accurately predict NQEs and facilitate understanding of how characteristics like branching and heteroatom content influence behavior. This work highlights the broad relevance of NQEs in molecular liquids, while also providing a conceptual and practical framework to anticipate their impact.

The quantum nature of nuclei influences every material and chemical process. Resulting nuclear quantum effects (NQEs) can arise as zero-point energy,<sup>1</sup> quantum tunneling<sup>2,3</sup>, and quantization of proton energy levels. NQEs are generally expected to be substantial under certain conditions, such as processes occurring at low temperature or involving light nuclei. They have been found to be consequential in genetic stability of DNA<sup>4,5</sup>, zeolite catalysts<sup>6–9</sup>, superconductor materials<sup>10,11</sup>, and enzymatic reactions<sup>12–15</sup>. NQEs also directly impact experimental studies that investigate or depend on isotope substitution<sup>16–22</sup>. Therefore, NQEs are important for many biological, chemical, and physical processes, even those containing heavier atoms or strong interactions<sup>6,10,23–26</sup>, yet their extent and significance remain not well characterized for many systems.

Nuclear quantum effects are often deduced through equilibrium or kinetic isotope effects. In classical mechanics, kinetic isotope effects are expected but depend solely on the mass differences between isotopes; deviations from such results highlight quantum mechanical phenomena-such as nuclear tunneling. This is common in enzymatic reactions<sup>13,27–32</sup>; for instance, isotope substitution can alter the oxidation rate of yeast alcohol dehydrogenase by up to 33%.<sup>15</sup> In a very

different scenario, the conductivity of deuterated versus normal phosphoric acid differs by orders of magnitude, while classical expectations suggest a ratio of ca. 1.4.<sup>33</sup> Meanwhile, classical statistical thermodynamics does not predict isotope-dependent changes in equilibrium properties, yet isotope effects on such properties are widely observed<sup>34–39</sup>. For example, the solid-vapor isotope fractionation ratio between deuterated and normal water reaches values as high as 1.208<sup>35</sup>. Replacing H<sub>2</sub>O with D<sub>2</sub>O similarly impacts the structure and properties of biopolymer solutions<sup>40,41</sup>, despite D<sub>2</sub>O being chemically equivalent to H<sub>2</sub>O under classical assumptions. The temperature-dependence of intramolecular isotopic equilibria has even been used to estimate dinosaur body temperatures via isotope-ratio mass spectrometry measurements on fossilized sauropod teeth<sup>42</sup>. The presence of such isotope effects signals the relevance of NQEs but does not fully characterize them.

Path-integral molecular dynamics (PIMD) enables explicit treatment of NQEs and calculation of isotope effects. This is in contrast to conventional classical MD, which treats nuclei as point particles evolving on a potential energy surface. Meanwhile, electronic structure methods, like density functional theory, provide a quantum



**Fig. 1 | Impact on thermophysical properties.** **A** Molecular structures of studied systems. Molecules are categorized into deciles based on  $\Delta v_m$ . Within each decile, higher vertical position indicates higher  $\Delta v_m$ . Atoms are colored as follows: carbon (gray), hydrogen (white), oxygen (red), nitrogen (blue), sulfur (yellow), chlorine (green), bromine (brown), iodine (purple). Structures were rendered with PyMOL (v3.1.0)<sup>82</sup>. The same data are shown with skeletal formula in Supplementary Fig. 1. **B** The magnitude of NQEs on molar volume, thermal expansion coefficient, isothermal compressibility, static dielectric constant, and enthalpy of vaporization of each system. **C** The calculated isotope effects on each property. The gray regions

around  $y = 0$  under each plot indicate the average standard error of the mean for the magnitude of NQEs on each property obtained from four independent simulations. The violin plots in **(B, C)** represent the distribution of  $\Delta_\lambda$  and  $\Delta_{\lambda^{D \rightarrow H}}$  values, with the white dot representing the median, the black box representing the interquartile range, and the inner lines representing the  $1.5\times$  interquartile range. The spread along the x-axis represents the density obtained from Gaussian kernel density estimation. All displayed properties are calculated at 298.15 K and 1 atm. Source data are provided as a Source Data file.

mechanical treatment of electronic degrees of freedom but generally rely on the Born-Oppenheimer approximation, under which the nuclei are held fixed or propagated classically. PIMD incorporates NQEs via a mathematical isomorphism, mapping each quantum nucleus to a classical ring polymer, thereby capturing quantum delocalization and thermal fluctuations by sampling in an extended phase space<sup>43</sup>. For water, PIMD has demonstrated how intramolecular zero-point motion and intermolecular tunneling affect properties like translational diffusion and orientational relaxation rates<sup>44,45</sup>. PIMD has also revealed the significance of NQEs in DNA base pairs, acetylene:ammonia co-crystals, and electrolyte transport in confined aqueous systems<sup>46–48</sup>. Nevertheless, in addition to the computational overhead of the extended phase space, PIMD can be practically limited by the availability of force fields that accurately represent the Born-Oppenheimer potential energy surface. Most conventional classical MD simulations typically use efficient force fields based on analytical equations. However, these force fields have parameters often calibrated to experimental data, implicitly including NQEs in uncontrolled ways, such that their use in PIMD risks “double counting” NQEs<sup>44,49</sup>. As a result, broad application of PIMD remains limited, and the role of NQEs across many systems and conditions is still largely unexplored.

Here, we characterize NQEs in 92 molecular liquids, spanning a wide organic chemical space, at ambient conditions. In particular, by comparing classical and path-integral MD simulations, we quantify NQEs on several thermodynamic properties including molar volume ( $v_m$ ), thermal expansion coefficient ( $\alpha_p$ ), isothermal compressibility ( $\kappa_T$ ), static dielectric constant ( $\epsilon_r(0)$ ), and enthalpy of vaporization ( $\Delta h_{\text{vap}}$ ). Equilibrium isotope effects on the same systems are also examined by simulating deuterated systems. This investigation spans a diverse set of chemistries—including amines, ethers, ketones, alcohols, alkanes, and more—and is enabled by the Topology Automated Force-Field Interactions (TAFFI) framework<sup>50</sup>, an efficient analytical force field parameterized solely from quantum chemical calculations at the  $\omega$ B97X-D3/def2-TZVP level. Because TAFFI does not utilize

experimental calibration, it is well-suited for conducting PIMD across the numerous systems studied. We find a meaningful influence of NQEs for every substance and highlight key chemical features that correlate with NQEs using unsupervised and supervised machine learning. This reveals competing effects between system stability and hydrogen-atom densities on NQEs, which particularly explains trends related to hydrogen-bonding groups and molecular branching. By combining physics-based simulation and data-driven analysis, this work provides a deeper understanding of NQEs in common liquid organic systems and the conditions under which explicit consideration of NQEs may be needed.

## Results

### Varied effects across molecules and properties

To understand how NQEs manifest across chemical space, we simulate 92 organic molecules at ambient conditions. Of the 92 molecules studied, 87 were previously simulated using classical MD with various force fields<sup>51</sup>, including TAFFI<sup>50</sup>. Benchmarking against experimental thermophysical properties showed that TAFFI compared favorably to other widely used empirical force fields in terms of reproducing experimental thermophysical properties across this set, which features amines, ethers, nitriles, ketones, alcohols, aldehydes, esters, amides, as well as some sulfur- and halogen-containing compounds. We further augmented this set with five n-alkanes ( $C_9$ ,  $C_{10}$ ,  $C_{11}$ ,  $C_{14}$ , and  $C_{15}$ ) to have representation of nonpolar hydrocarbons. The extent of NQEs is then quantified via

$$\Delta_\lambda(\%) = 100 \times \frac{\lambda^{\text{PI}} - \lambda^{\text{cl}}}{\lambda^{\text{PI}}} \quad (1)$$

where  $\lambda^{\text{cl}}$  and  $\lambda^{\text{PI}}$  are the properties calculated via classical and path-integral MD simulations.

Figure 1 summarizes the results, with Fig. 1A qualitatively organizing all systems based on influence on molar volume,  $v_m$ . While it is clear and expected that heavy atoms like chlorine, bromine, and

sulfur lead to smaller NQEs, other relativistic organizing principles based on molecular constitution are less evident. More quantitatively, NQEs affect all systems and properties (Fig. 1B). Accounting for NQEs markedly increases molar volume by up to 5.5% and also increases  $\kappa_T$  for most systems, with an average effect of nearly 8%. These systematic increases may signify weaker cohesive intermolecular interactions due to nuclear delocalization, in line with prior studies on water and other liquid organic systems<sup>52–59</sup>. In prior investigation featuring linear alkanes, the NQEs were found to decrease density by as much as 11%;<sup>56</sup> for water, NQEs induce a much smaller effect on the density, but its direction is consistent with our observations<sup>56,60,61</sup>. For reference, trends in density are opposite but similar in magnitude to molar volume, as analytically  $\Delta\rho = -\frac{\Delta v_m}{1-\Delta v_m} \approx -\Delta v_m$  (Supplementary Fig. 2). Quantum treatment of nuclei yields less consistent directional changes from classical results on the thermal expansion coefficient, enthalpy of vaporization and the static dielectric constant. For these properties, the average effect is small but also comparable to the statistical error in their calculation. These average effects are slightly positive for  $\alpha_p$  and negative for  $\varepsilon_r(0)$  and  $\Delta h_{\text{vap}}$ . Overall, these results indicate that NQEs can significantly and detectably influence certain thermodynamic properties across diverse molecular chemistries.

For the original TAFFI framework, all bonds are fit to purely harmonic functions, but anharmonicity in the potential energy surface—especially in bond-stretching modes—has been identified as a key factor influencing the magnitude of NQEs<sup>44,62–64</sup>. For example, in water, anharmonicity in the O-H bond contributes to competing effects: intermolecular zero-point energy and tunneling weaken hydrogen bonding, while intramolecular zero-point motion enhances the dipole moment and strengthens interactions, resulting in relatively small net NQEs<sup>44</sup>.

To evaluate then whether the harmonic description in TAFFI may be responsible for the large NQEs relative to water, we reparameterized bonds involving hydroxyls, amines, and thiols (29 molecules) with an anharmonic Morse potential<sup>44</sup> based on expanded mode scans at  $\omega$ B97X-D3/def2-TZVP level of theory. We then evaluated  $\Delta v_m$  with this anharmonic bond description and compared it to the harmonic results (Supplementary Fig. 3). Overall, we find that anharmonicity in bond-stretching slightly suppresses NQEs, though the overall magnitudes remain comparable to the harmonic description. We speculate that this mitigating effect is weaker in the organic molecules studied than in water, likely due to smaller dipole moment changes and correspondingly less enhancement of intermolecular interactions following introduction of anharmonicity. Intuitively, as most of the molecules in this study are larger than water, they are expected to exhibit smaller overall changes in dipole moment due to anharmonicity, as local bond distortions contribute less to the total molecular polarization. Although anharmonicity appears to have little impact for the molecules and conditions studied, its role in other phases, molecules, or force fields warrants careful consideration.

### Isotope effects versus nuclear quantum effects

Building on the preceding results that directly compare quantum and classical systems, we probe isotope effects of deuteration using

$$\Delta_{\lambda^{\text{D-H}}}(\%) = 100 \times \frac{\lambda^{\text{PI}} - \lambda^{\text{PI,D}}}{\lambda^{\text{PI,D}}} \quad (2)$$

where  $\lambda^{\text{PI,D}}$  is the property for a fully-deuterated system simulated via PIMD; deuterated systems are often expected to approximate classical behavior. The quantity  $\Delta_{\lambda^{\text{D-H}}}$  is equivalent to  $1 - \frac{\lambda^{\text{H}}}{\lambda^{\text{D}}}$  where  $\frac{\lambda^{\text{H}}}{\lambda^{\text{D}}}$  is often noted as the (experimentally accessible) equilibrium isotope effect. Trends in  $\Delta_{\lambda^{\text{D-H}}}$  (Fig. 1C) align directionally with  $\Delta_{\lambda}$  (Fig. 1B) but with reduced magnitudes. However, beyond consistent shifts in molar

volume, the effect of deuteration is often within the range of statistical error. This change is most notable for  $\kappa_T$ , where the average effect shifts from 7.8% when comparing to purely classical systems to 3.5% when comparing to quantum mechanical but deuterated ones. This highlights that deuteration does not fully replicate a purely classical treatment. A system with a negligible isotope effect may still display significant NQEs, which partially cancel between the isotopically normal and deuterated systems.

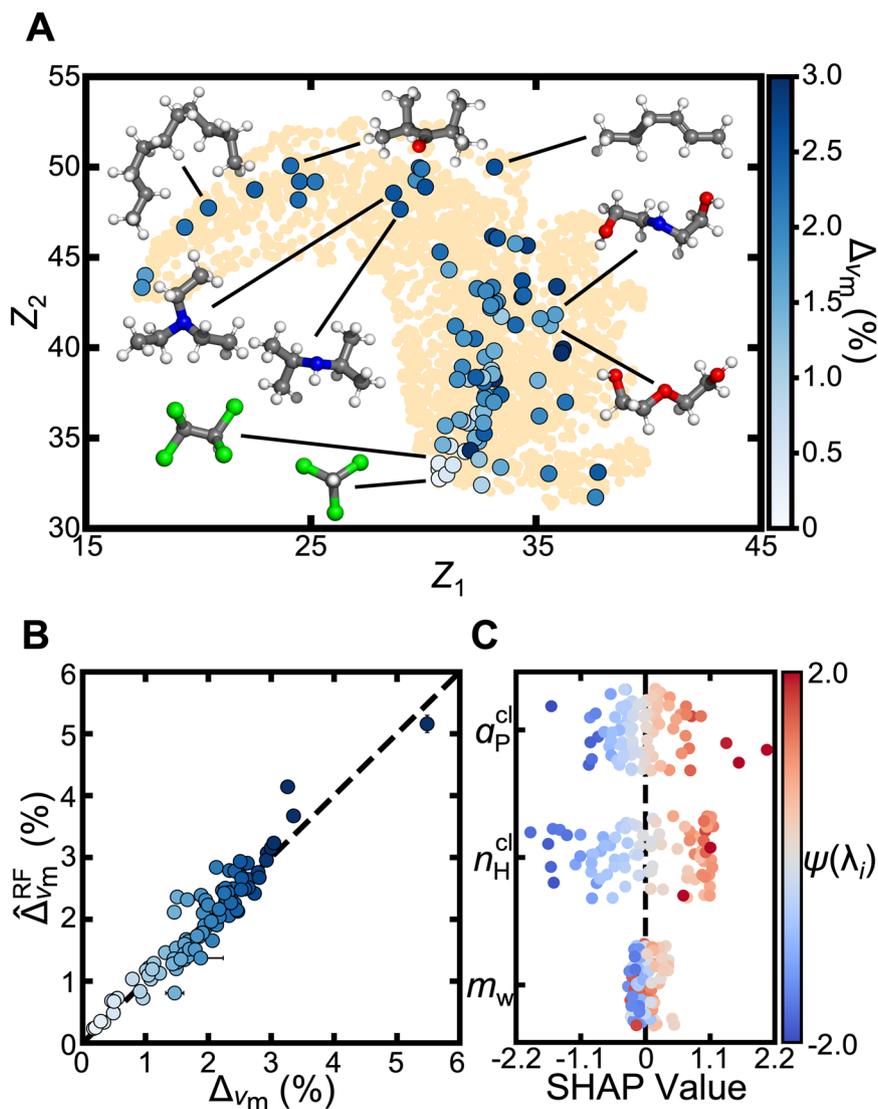
### Key molecular determinants

We next identify key molecular features that influence the strength of NQEs in liquid systems. Given its consistent and statistically resolvable effects, in addition to  $v_m$  being a fundamental thermophysical property relevant to equations of state and other material properties, our discussion focuses on  $\Delta v_m$ , while analyses of other properties are included in Supplementary Fig. 4.

To visualize how NQEs vary across a broad chemical space, we apply dimensionality reduction from a high-dimensional molecular feature space to two dimensions using unsupervised machine learning. Specifically, we use the Uniform Manifold Approximation and Projection (UMAP) algorithm<sup>65,66</sup> on a dataset that includes our 92 investigated molecules and an extended set of 2874 molecules from the ChEMBL database<sup>67,68</sup>. Each molecule is initially represented by a 34-dimensional atom-type feature vector generated using the Merck Molecular Force Field (MMFF94)<sup>69</sup>, and then UMAP constructs a neighborhood graph from the collection of such vectors, models pairwise relationships probabilistically, and optimizes a low-dimensional embedding that preserves both local and global structure. The resulting collective variables,  $Z_1$  and  $Z_2$ , are non-linear transformations of the high-dimensional feature values, which obfuscates facile interpretation. Nevertheless,  $Z_1$  and  $Z_2$  define a coordinate space that can be easily visualized and for which chemical similarity trends with distance.

Within the broader chemical context of the ChEMBL database, we find that the simulated substances indeed span a wide chemical space (Fig. 2A). In terms of readily identifiable trends, chemical structures displaying lower  $\Delta v_m$  are seemingly clustered in the  $Z_1$ - $Z_2$  space in the vicinity of molecules with generally higher molecular weights, such as chloroform. In contrast, molecules with larger  $\Delta v_m$  are distributed across the manifold, indicating that diverse molecular features may enhance NQEs. This is further supported by the fact that different functional groups can be associated with population-level effects on  $\Delta v_m$  (Supplementary Fig. 6). For instance, molecules with amines tend to exhibit higher  $\Delta v_m$ , and those containing heavier atoms exhibit lower  $\Delta v_m$ . At the same time, there is significant variation of  $\Delta v_m$  within each population of functional groups, such that attributing NQEs purely to functional groups is difficult. This highlights the necessity to consider additional factors to ultimately understand the manifestation of NQEs.

To gain insight into what factors correlate with NQEs, we create a data-driven model for  $\Delta v_m$  from simple descriptors, with the rationale that an effective model would highlight the importance of those descriptors. We find that a model using just three inputs—average atomic mass ( $m_w$ ), hydrogen density ( $n_{\text{H}}^{\text{cl}}$ ), and thermal expansion coefficient ( $\alpha_p^{\text{cl}}$ )—accurately predicts  $\Delta v_m$  (Fig. 2B). The coefficient of determination over all predictions is  $R^2 = 0.881 \pm 0.002$ . Here, predictions are made as  $\hat{\Delta v}_m = \hat{T}_{v_m} \alpha_p^{\text{cl}}$ , where  $\hat{T}_{v_m} = \text{RF}(m_w, n_{\text{H}}^{\text{cl}})$  is the output of a random forest (RF) regressor trained to predict a defined quantity  $T_{v_m} \equiv (\alpha_p^{\text{cl}})^{-1} \Delta v_m$ , and the RF model is trained on all data except that of the molecule being predicted. We empirically find that this approach outperforms models formulated as  $\hat{\Delta v}_m = \text{RF}(m_w, n_{\text{H}}^{\text{cl}}, \alpha_p^{\text{cl}})$ , likely due to the large variability in  $\alpha_p^{\text{cl}}$  across molecules and its nature as a response function, unlike the more intrinsic descriptors  $m_w$  and  $n_{\text{H}}^{\text{cl}}$ . As a point of possible interest, we note that  $T_{v_m}$  has units of temperature and, to leading order, can be interpreted as the temperature shift required for a classical system to match the molar volume of its quantum



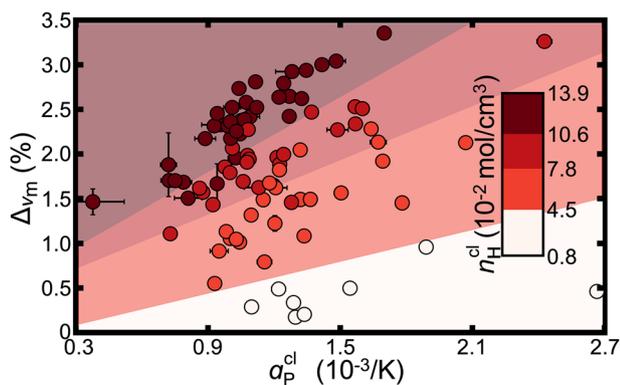
**Fig. 2 | Correlation of chemical features with the magnitude of NQEs.** **A** A two-dimensional manifold organization of the 92 organic molecules (blue, larger markers) and 2874 small molecules (tan, smaller markers) obtained from the ChEMBL database. The coordinates  $Z_1$  and  $Z_2$  are derived by applying the Uniform Manifold Approximation and Projection (UMAP) algorithm to a 34-dimensional input vector based on MMFF94 atom typing<sup>69</sup>. Inset lines highlight representative molecules from different regions of the manifold. **B** Comparison of predicted versus simulated effect of NQEs on molar volume ( $\Delta v_m$ ). Predicted values are obtained from a data-driven model with three chemically interpretable input features: average atomic mass ( $m_w$ ), hydrogen density ( $n_H^{cl}$ ), and thermal expansion coefficient ( $\alpha_p^{cl}$ ); the latter two are determined from classical MD simulations. Error bars

reflecting the standard error of the mean, determined using four independent simulations and from five different training and testing cycles for the data-driven model, are generally smaller than the symbol size. The markers (blue) use the same color scale as those in (A). **C** Impact of feature contributions based on Shapley Additive Explanations (SHAP) analysis. The position on the x-axis indicates the impact of each feature on the model output, and the marker color indicates feature value. Feature values ( $\lambda_i$ ) are displayed following a Yeo-Johnson power transformation, denoted by  $\psi$ , to present data on similar scales<sup>83</sup>. For visual clarity, the SHAP data are shown over a restricted range; Supplementary Fig. 5 depicts the full observed range of data. Source data are provided as a Source Data file.

counterpart (i.e.,  $v_m^{cl}(T + T_{v_m}) \approx v_m^{pl}(T)$ ). This interpretation does not imply a true physical equivalence between the quantum system and a classical system at elevated temperature—a comparison that has been cautioned against as potentially misleading in prior work<sup>70</sup>. Across the molecules studied,  $T_{v_m}$  ranges from a few to several tens of degrees (Supplementary Fig. 7).

We use Shapley Additive Explanations (SHAP) analysis<sup>71</sup> to attribute feature contributions to predictions made by our machine learning model. This technique assesses the impact of each feature by computing the difference in model output with and without its inclusion for all feature combinations. This analysis reveals some intuitive trends: abundance of hydrogens and

greater  $n_H^{cl}$  enhance  $\Delta v_m$  (Fig. 2C). These effects reflect a general dependence on molecular composition, as systems with low mass-density or hydrogen-rich systems would generally imply greater sensitivity to NQEs due to the presence of lighter nuclei. However,  $\alpha_p^{cl}$  correlates positively with  $\Delta v_m$ . SHAP analysis also indicates that  $\alpha_p^{cl}$  and  $n_H^{cl}$  features have greater impact on model output than  $m_w$ , which has a non-monotonic relationship with  $\Delta v_m$ . Although variable importance was established within the context of 92 organic molecules, it is useful to consider how these principles translate to a well-studied system, like water. Using classical simulation data from the q-TIP4P/F water model<sup>44</sup>



**Fig. 3 | Analysis of hydrogen density and thermal expansion coefficient.** Four groups of data are generated with *k*-means clustering algorithm on the respective hydrogen densities  $n_H$ ; the marker colors indicate being within the range specified by the color bar. The colored regions are distinguished by support vector machine (SVM) margin lines. Error bars (black) represent the standard error of the mean obtained from four independent simulations. For visual clarity of trends, the data are shown over a restricted range; Supplementary Fig. 8 depicts the full observed range of data. Source data are provided as a Source Data file.

as input, the data-driven model predicts  $\hat{\Delta}_{v_m}^{RF} = 0.47\%$ , in good agreement with the simulation result  $\Delta_{v_m} = 0.25\%$ . Despite water being chemically distinct from the other molecules studied, the model captures the small impact on  $v_m$  due to NQEs—an outcome largely driven by its low  $\alpha_P$ .

Overall, these results indicate a complex interplay between NQEs, fluid composition, and its stability that is yet predictable through simple properties readily computed with classical simulation.

### Competing effects of system stability and hydrogen density

At first glance, the trends for  $n_H$  and  $\alpha_P$  appear to reflect competing effects: higher  $n_H$  suggests stronger intermolecular forces, which would typically reduce  $\alpha_P$ . To explore this trade-off, we leverage the dataset to create controlled comparisons by fixing either  $n_H$  or  $\alpha_P$  and examining the effect of the other on  $\Delta_{v_m}$ . This is accomplished through *k*-means clustering to group molecules with similar  $n_H$  values and examining correlations of the groups with  $\Delta_{v_m}$  (Fig. 3). Within groups with similar  $n_H$ , a decrease in fluid stability (indicated by higher  $\alpha_P$ ) tends to enhance NQEs. Conversely, across groups with comparable stability (similar  $\alpha_P$ ), an increase in  $n_H$  (and thus more light nuclei) leads to stronger NQEs. Nevertheless, the overlapping  $\alpha_P$  ranges across different  $n_H$  groups (with some weak negative correlation) suggest that both properties facilitate predicting the magnitude of NQEs.

### Trends with hydrogen bonding and branching

To elucidate how  $\Delta_{v_m}$  correlates with molecular constitution, we systematically compare groups of molecules varying in hydrogen bonding (Fig. 4A) and branching (Fig. 4B) and examine how key descriptors,  $n_H$  and  $\alpha_P$ , are affected.

In Fig. 4A, comparing 1-bromobutane (i), 1-chlorobutane (ii), butane-1-thiol (iii), and butan-1-ol (iv) highlights the positive impact of hydrogen bonding on  $n_H$  and  $\Delta_{v_m}$ . Among these molecules—which share the same number of heavy atoms, the same chemical topology, and comparable  $\alpha_P$ —the hydrogen-bonding thiol and alcohol exhibit higher  $n_H$ . However,  $n_H$  and  $\Delta_{v_m}$  do not scale monotonically with additional hydrogen-bonding groups. This is revealed by comparison of butan-1-ol (iv), butane-1,4-diol (v), pentane-1,5-diol (vi), and propane-1,2,3-triol (vii). Among these alcohols, increasing the number of hydrogen-bonding groups minimally affects hydrogen density but increases the stability of the fluid (decreases  $\alpha_P$ ). Evaluation of

different hydrogen-bonding groups reveals similar trends (Supplementary Fig. 9) where the balance of  $n_H$  and  $\alpha_P$  elucidates the resulting  $\Delta_{v_m}$ . For example, relative to alcohols with similar  $n_H$ , amines display larger NQEs and associated larger  $\alpha_P$ .

In Fig. 4B, comparing multiple linear molecules and their branched analogues, we observe consistent impact on NQEs. In particular, the branched molecule exhibits larger  $\Delta_{v_m}$  than the linear one. Across all four pairs, branching reduces  $n_H$ , but there is a coupled increase in  $\alpha_P$ . For these molecules, the net result is stronger NQEs in the branched molecules. We therefore suggest that NQEs, at least as manifested through  $\Delta_{v_m}$ , are likely to be more substantial in branched molecules owing to relatively diminished fluid stability. This observation is reminiscent of an expected trend: boiling points for branched molecules are typically lower than those of linear molecules of comparable molecular weight and composition.

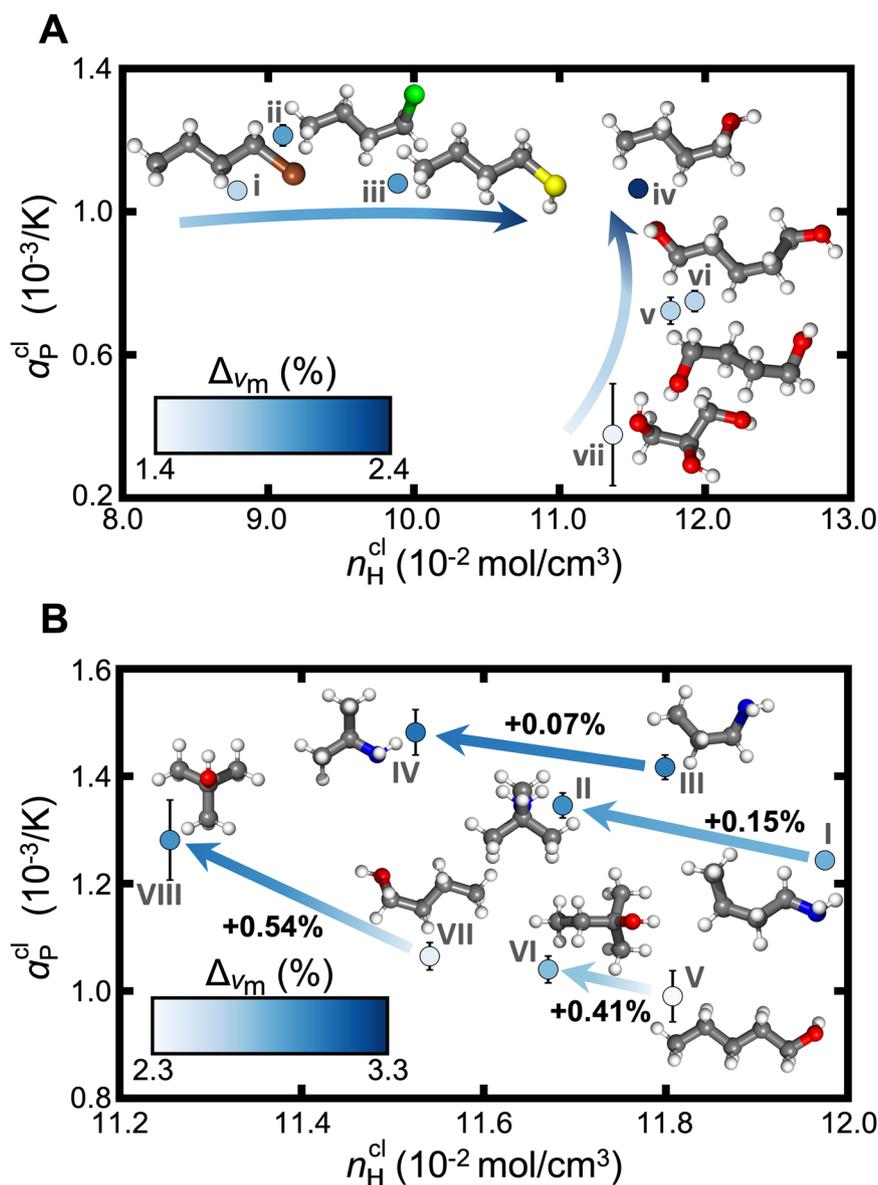
### Detailed physics of linear versus branched systems

The conventional rationale for the trend involving boiling temperatures is that branched molecules have reduced surface area and less efficient packing, which weakens intermolecular forces and impacts the magnitude of NQEs. Intriguingly, the linear and branched molecules studied here exhibit comparable interaction patterns with respect to the number of hydrogen bonds, their strength, and average nearest-neighbor distances (Supplementary Fig. 10). This renders the importance of hydrogen-bonding groups on trends in  $\alpha_P$ , and thus their impact on  $\Delta_{v_m}$ , unclear.

Resolving interaction patterns as a function of distance between neighboring molecules, however, illustrates key differences between butan-1-ol with its branched analogue, 2-methylpropan-2-ol (Fig. 5). The distribution of hydrogen bonds (Fig. 5A) shows that butan-1-ol forms bonds over a broad range of intermolecular distances, while methylpropan-2-ol, with less conformational flexibility, forms bonds only at separations of approximately 4.6 Å. A further distinction emerges in energetics of molecular interactions within the fluid (Fig. 5B): 2-methylpropan-2-ol exhibits a pronounced minimum in interaction energy near the distance associated with hydrogen bonding, while butan-1-ol shows steadily diminishing attractive interactions with distance. These nuanced changes in intermolecular interactions due to branching result in a pronounced increase in  $\alpha_P$ , which tends to enhance  $\Delta_{v_m}$ . Notably, the minimum interaction energy for butan-1-ol occurs at a distance where few or no hydrogen bonds form. Additionally, butan-1-ol lacks a strong preference for relative molecular orientation, unlike 2-methylpropan-2-ol (Fig. 5C). These findings suggest that the cohesive energy of 2-methylpropan-2-ol relies more heavily on hydrogen bonding whereas butan-1-ol exhibits overall stronger dispersion forces. This comparison also serves to demonstrate the sensitivity of NQEs to subtle changes in nanoscale interactions.

### Discussion

Nuclear quantum effects (NQEs) are present in all physical systems, but their significance is rarely known beyond certain well-studied systems (e.g., water) or expected scenarios (e.g., low temperatures). In this study, we quantified the impact of NQEs at ambient conditions on various properties across 92 organic molecules spanning diverse chemical space. By comparing path-integral and classical simulations, we observed that NQEs tend to significantly and measurably enhance molar volumes and isothermal compressibilities. Effects on thermal expansion coefficients, dielectric constants, and enthalpies of vaporization lack clear general directional trends relative to the uncertainty of the calculations themselves, though individual systems can still be significantly impacted. Simulations of deuterated systems reveal similar but reduced effects, with isotope-induced changes generally within the statistical uncertainty, except for molar volume. This suggests deuteration often underestimates the full



**Fig. 4 | Impact of hydrogen bonding and branching on the extent of NQEs.**

**A** Hydrogen densities and thermal expansion coefficients for systems of various hydrogen-bonding groups and counts. Visualized molecules are labeled as (i) 1-bromobutane, (ii) 1-chlorobutane, (iii) butane-1-thiol, (iv) butan-1-ol, (v) butane-1,4-diol, (vi) pentane-1,5-diol, (vii) propane-1,2,3-triol for reference. **B** Comparison between four pairs including linear molecules and their branched analogues of chemically similar structures. The impact of branching on the thermal expansion coefficient and hydrogen density is shown on the y- and x-axis, respectively. The marker color

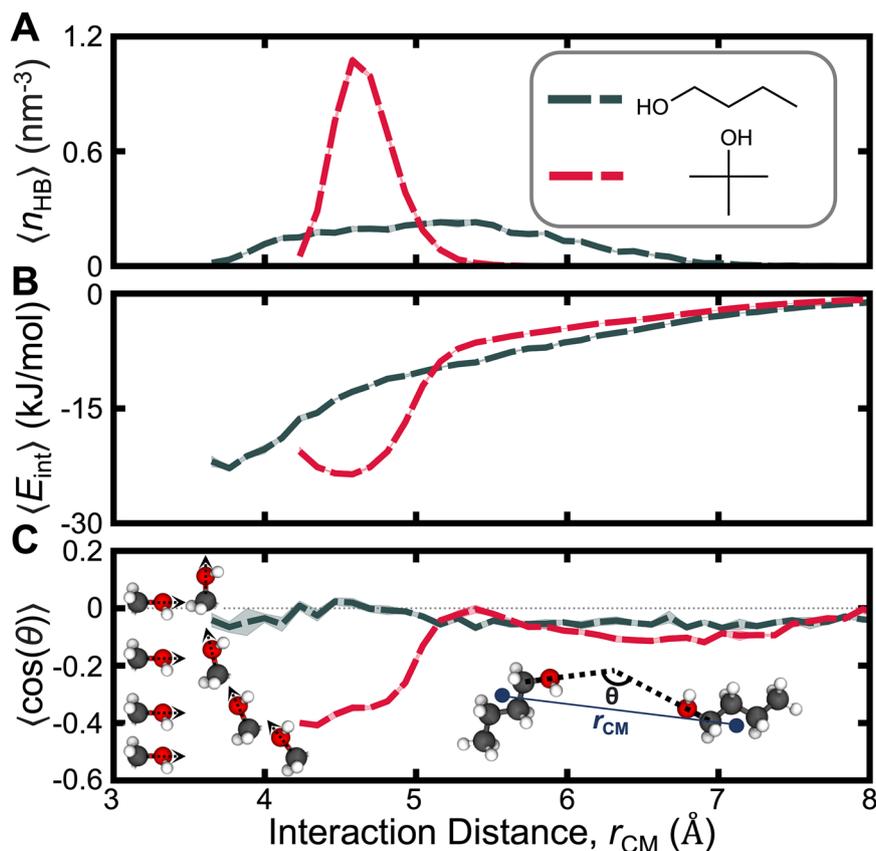
intensity denotes the magnitude of NQEs, indicated by the color bars and arrow directions in each panel. Visualized molecules are labeled as (I) butan-1-amine, (II) 2-methylpropan-2-amine, (III) propan-1-amine, (IV) propan-2-amine, (V) pentan-1-ol, (VI) 2-methylbutan-2-ol, (VII) butan-1-ol, (VIII) 2-methylpropan-2-ol. It should be noted that while the two color bars have different bounds, the range of the bounds is equal. Error bars (black) represent the standard error of the mean obtained from four independent simulations. Source data are provided as a Source Data file.

extent of NQEs, as a quantum treatment of deuterium still impacts properties.

We further identified physical properties that effectively correlate with the magnitude of NQEs. This was particularly demonstrated by a data-driven model with interpretable inputs of the average mass of composite atoms, hydrogen densities, and thermal expansion coefficients. We posit this as a pragmatic approach to anticipate the relevance of NQEs and the necessity of computationally intensive path-integral simulations. Beyond prediction, these property trends highlight key physical factors, some intuitive and others less so, affecting NQEs. An intriguing finding from the analysis is the competing effect between hydrogen density and fluid stability (linked to the thermal

expansion coefficient), where both factors enhance NQEs but are somewhat inversely correlated.

By exploiting the chemical diversity of our dataset, molecular-level insights were obtained by examining sets of molecules with variations in hydrogen content, hydrogen-bonding capabilities, and comparisons between linear and branched topologies. Grouping data based on functional groups revealed that NQEs are generally more significant in amines compared to ethers and alcohols, for example. Broadly, systems with high hydrogen density (and therefore light nuclei) but weaker interactions, leading to lower stability, exhibit the largest NQEs. This explains a potentially unintuitive result that systems with multiple hydrogen-bonding groups may not necessarily exhibit strong NQEs because the hydrogen bonding tends to enhance fluid



**Fig. 5 | Effect of branching on intermolecular interactions through comparison of linear butan-1-ol with branched 2-methylpropan-2-ol.** Ensemble averages of (A) the hydrogen-bond density, (B) pairwise intermolecular interaction energies, and (C) interaction orientations as a function of distance between molecules. The inset images of panel (C) illustrate calculations of the interaction distance and

orientation; the renderings adjacent to the y-axis illustrate configurations consistent with the nearest tick marks. Error regions indicate the standard error obtained from four simulations for each bin. All molecular renderings were generated using the Visual Molecular Dynamics (VMD) software<sup>54</sup> (version 1.9.4). Source data are provided as a Source Data file.

stability, reflected in a smaller thermal expansion coefficient. By contrast, molecular branching has a general tendency to both reduce fluid stability and hydrogen density. This study advances our understanding of NQEs in commonly used substances, guiding experimental and computational approaches and providing a framework for when explicit treatment of NQEs is necessary to capture essential physical phenomena.

## Methods

### General simulation protocols

Systems were modeled with the quantum chemistry-based Topology Automated Force-Field Interactions (TAFFI) framework<sup>30</sup> with Waldman-Hagler mixing rules. The TAFFI framework was chosen for this study due to its efficiency as an analytical force field, its strict parameterization using quantum chemical calculations at the  $\omega$ B97X-D3/def2-TZVP level, and the generally good agreement for a range of thermophysical properties, comparing values calculated from classical simulations and those reported experimentally<sup>50</sup>. Real-space non-bonded interactions were truncated at 14 Å, unless otherwise noted. Long-range electrostatics calculations used the Particle Mesh Ewald (PME) algorithm; tail corrections were applied for Lennard-Jones interactions beyond the real-space cutoff, unless otherwise noted. All PIMD simulations used a ring-polymer bead count of  $P = 32$ . All simulations used a timestep of 0.5 fs. Constant-temperature conditions of 298.15 K were achieved using a Langevin thermostat for classical simulations and the path-integral Langevin equation (PILE) thermostat<sup>72</sup> for path-integral simulations; both utilize a friction coefficient of 1 ps<sup>-1</sup>. Constant-pressure conditions at 1 atm were

achieved using a Monte Carlo barostat with an attempt frequency of 25 fs.

The force field in tandem with aforementioned settings enables accurate reproduction of liquid-phase experimental densities<sup>51</sup> for 87 systems with available empirical data (Supplementary Fig. 11). Benchmarking of classical and PIMD simulation results also shows that path-integral treatment of most systems brings predicted molar volumes closer to experiment. To gain quantitative insight, we apply a one-tailed paired t-test on the differences with experiment and report  $p = 4.8 \times 10^{-17} < 0.05$ , indicating that PIMD results are systematically closer to experiment. These results, along with good alignment between force-field intramolecular normal-mode frequencies and DFT results, suggest that TAFFI can effectively represent and discriminate the chemically diverse set of systems in this study.

Property calculations and analyses were performed for four independent classical and path-integral MD simulations of each molecular system. Each simulation contained the minimum number of molecules to reach 5000 atoms, initialized with random positions and orientations within a simulation cell of  $V = 50 \text{ nm}^3$ . The initial configuration was subjected to energy minimization with an energy tolerance of 10 kJ/mol followed by 0.1 ns simulation in the microcanonical ensemble. The system was then equilibrated in the canonical ensemble at 400 K for 0.2 ns and cooled to 298.15 K over 0.8 ns with a linear temperature schedule. Subsequently, systems were equilibrated in the isothermal-isobaric ensemble at 1 atm for 3 ns followed by a production run of 10 ns, during which thermodynamic data were collected, and system configurations were saved every 10 ps. Simulations of q-TIP4P/F water systems followed a similar procedure with 1000 water

molecules and with a real-space non-bonded cutoff distance of 10 Å. The above procedure was found to be insufficient to equilibrate C<sub>14</sub> and C<sub>15</sub> systems. For these, initial structures were generated using a 100 ns classical simulation in the isothermal-isobaric ensemble at 298.15 K and 1 atm, followed by 3 ns of additional equilibration and 10 ns production runs using either classical MD or PIMD. These protocols were sufficient to achieve adequate energy conservation and effectively converge the system density (see Supplementary Figs. 12 and 13).

All simulations were performed using the GPU implementation of the OpenMM 7.7.0 software package<sup>73</sup>. PIMD simulations were performed on NVIDIA A100 and P100 GPUs, with mean simulation times of 46.5 and 89.3 h, respectively, in contrast with classical MD simulations of 1.7 h on A100 GPUs. Total simulation time throughout the study was 55,383 GPU hours.

### Gas-phase simulations

Single-molecule simulations of each system were used to calculate the gas-phase energy to obtain  $\Delta h_{\text{vap}}$ . A molecule was placed in a large box with a volume of 125 nm<sup>3</sup> and simulated in the microcanonical ensemble for 1 ns, followed by the canonical ensemble at 298.15 K for 0.4 ns using an Andersen thermostat with a collision probability of 10 ps<sup>-1</sup>. For these simulations, all pairwise interactions were evaluated in real-space and restricted to within 15 Å to ensure that there are no interactions with periodic images. Linear and angular momenta of each system were removed, the particle velocities were rescaled, and the potential energies were saved every 1 ps.

### Analysis of system properties

Molar volumes were calculated using

$$v_m = \frac{M}{\langle \rho \rangle} \quad (3)$$

where  $M$  is the molar mass of the molecule and  $\langle \rho \rangle$  is the average (mass) density.

Thermal expansion coefficients,  $\alpha_P$ , were calculated using

$$\alpha_P = \frac{1}{\langle V \rangle} \left( \frac{\partial V}{\partial T} \right)_P \quad (4)$$

where  $\left( \frac{\partial V}{\partial T} \right)_P$  was numerically approximated using average volumes from simulations at 293.15, 298.15, and 303.15 K. The additional simulations at 293.15 and 303.15 K were performed using the equilibrated configurations from 298.15 K. The systems were equilibrated in the isothermal-isobaric ensemble at for 0.5 ns, followed by 1.5 ns production runs where the system volume was saved every 1 ps.

Isothermal compressibilities,  $\kappa_T$ , were calculated using the fluctuations in the system volume using

$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle} \quad (5)$$

where  $k_B$  is the Boltzmann constant. Similarly, static dielectric constants,  $\epsilon_r(0)$  were calculated using

$$\epsilon_r(0) = \frac{1}{P} \sum_{i=1}^P 1 + \frac{4\pi \langle \mathbf{M}_i^2 \rangle - \langle \mathbf{M} \rangle^2}{3 V k_B T} \quad (6)$$

where  $\mathbf{M}$  is the total dipole moment of the simulation box ( $P=1$  for classical MD simulations).

Molar heats of vaporization,  $\Delta h_{\text{vap}}$ , were calculated using

$$\Delta h_{\text{vap}} = \frac{1}{P} \sum_{i=1}^P E_i^{(g)} + k_B T - \frac{1}{P} \sum_{i=1}^P E_i^{(l)} \quad (7)$$

where  $E^{(v)}$  is the total internal energy per mole (including both potential and kinetic energy) in phase  $v$ , either liquid or gas. This expression

assumes the ideal gas approximation and that the vapor-phase molar volume is much greater than that of the liquid. Often, the kinetic energy contributions in the gas and liquid phases are expected to cancel, allowing  $\Delta h_{\text{vap}}$  to be approximated from potential energies alone<sup>51</sup>. However, in PIMD simulations, quantum kinetic energy arising from harmonic coupling of systems in the extended phase space can differ between phases. Therefore, we retain all energy contributions in the calculation for simplicity.

### Unsupervised Learning

Unsupervised learning via Uniform Manifold Approximation and Projection (UMAP) dimension reduction was performed using the Python umap-learn package (version 0.5.3), with MMFF94 atom types to describe the 34-dimensional feature space. The MMFF94 atom types were obtained using the rdkit Python package (version 2022.9.3)<sup>74</sup>. The UMAP analysis includes additional small organic molecules obtained from the ChEMBL database. Molecules containing rings, net charge, fluorine atoms, or more than two double or triple bonds were excluded, resulting in 2874 molecules. For the UMAP hyperparameters, the analysis used 100 for the size of local neighborhood, 1.0 for the minimum distance between embedded points and Euclidean distances. The  $k$ -means clustering algorithm to categorize the systems into four groups based on their  $n_H$  was generated using the scikit-learn Python package (version 1.0.2) with default hyperparameters<sup>75</sup>.

### Supervised Learning

Supervised learning via random forest regression used the scikit-learn Python package. Model inputs included  $m_w$  and  $n_H$  from classical MD simulations. Model training and assessment used a leave-one-out split. Each molecule was tested by training the model on the other 91 molecules. This process was repeated for all 92 molecules. The reported  $R^2$  value and its associated error were calculated from five iterations with different random seeds. Each random forest model used 20 decision trees and required two samples per split. Shapley Additive Explanations (SHAP) analysis was performed using the shap Python package (version 0.44.1) on each of the three input parameters for the data-driven model.

The LIBSVM implementation of C-Support Vector Classification algorithm<sup>76,77</sup> in scikit-learn Python package (version 1.0.2) was used to generate support vector machine boundary lines for four groups of molecules based on their  $n_H$ . A linear kernel type and a regularization parameter of 1.0 were chosen as the algorithm hyperparameters.

### Analysis of molecular interactions

All interaction analyses were performed and averaged over configurations sampled every 1 ns from four independent simulations of each molecule. The interaction distance,  $r_{\text{CM}}$ , was defined as the distance between the centers of mass of two molecules. The array of interaction distances was discretized into 50 bins between 3.3 and 9 Å. Hydrogen-bond densities ( $n_{\text{HB}}$ ) were defined as the number of hydrogen bonds in each bin per volume. Hydrogen bonds were defined using geometric criteria with a cutoff distance of 3.6 Å between the donor and the acceptor and a cutoff angle of 150 degrees between the three participating atoms. These calculations were facilitated using the MDAnalysis Python library (version 2.1.0)<sup>78,79</sup>. Interaction energies,  $E_{\text{int}}$ , were computed as the sum of all pairwise non-bonded energy contributions between two molecules. Interaction orientations,  $\cos(\theta)$ , were calculated based on the angle ( $\theta$ ) between the bond vectors adjoining the carbon to the oxygen on the alcohol.

### Impact of force field

As previously discussed, TAFFI is used because it relies purely on quantum chemical calculations for its parameterization and computational efficiency; however, we acknowledge that the reported magnitude of NQEs could differ based on the underlying force field. To gain

some insight into the sensitivity of NQEs, classical and PIMD simulations using two different force fields—the all-atom optimized potentials for liquid simulations (OPLS-AA) force field and the Open Force Field (OpenFF, version 2.0.0 with unconstrained bonds)—were also performed and analyzed for a subset of 11 molecules, following the procedures described in Methods.

Relative to the TAFFI results, OPLS-AA and OpenFF consistently find larger  $\Delta\nu_m$  across all systems (Supplementary Fig. 15). It is important to note that OPLS-AA and OpenFF include empirical adjustments, potentially inherited from prior developments, making them strictly unsuitable for PIMD simulations. As a result, this analysis is merely suggestive rather than definitive regarding differences from alternative representations of the Born-Oppenheimer potential energy surface. However, the results do not indicate significant exaggeration of effect sizes from TAFFI. Future work could explore variations arising from entirely different interaction models, such as machine learning potentials.

### Data availability

The processed data generated in this study<sup>80</sup> have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.15236880>.

### Code availability

Analysis scripts and source code used in this research are publicly accessible at <https://github.com/webbtheosim/chem-space-nqes>, including instructions for reproducing the results. The code<sup>81</sup> has been deposited in the Zenodo database under accession code <https://zenodo.org/records/15465591>.

### References

1. Benoit, M., Marx, D. & Parrinello, M. Tunnelling and zero-point motion in high-pressure ice. *Nature* **392**, 258–261 (1998).
2. Cendagorta, J. R. et al. Competing quantum effects in the free energy profiles and diffusion rates of hydrogen and deuterium molecules through clathrate hydrates. *Phys. Chem. Chem. Phys.* **18**, 32169–32177 (2016).
3. Meier, T., Petitgirard, S., Khandarkhaeva, S. & Dubrovinsky, L. Observation of nuclear quantum effects and hydrogen bond symmetrisation in high pressure ice. *Nat. Commun.* **9**, 2766(2018).
4. Slocombe, L., Sacchi, M. & Al-Khalili, J. An open quantum systems approach to proton tunnelling in DNA. *Commun. Phys.* **5**, 109 (2022).
5. 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.* **14**, 5102–5108 (2023).
6. Bocus, M. et al. Nuclear quantum effects on zeolite proton hopping kinetics explored with machine learning potentials and path integral molecular dynamics. *Nat. Commun.* **14**, 1008 (2023).
7. Suzuki, K. et al. Nuclear quantum effect on hydrogen adsorption site of zeolite-templated carbon model using path integral molecular dynamics. *J. Alloy. Compd.* **509**, S868–S871 (2011).
8. McCardle, K. Accounting for quantum effects in catalysis. *Nat. Comput. Sci.* **3**, 189–189 (2023).
9. Gao, L. G., Zhang, R. M., Xu, X. & Truhlar, D. G. Quantum effects on H<sub>2</sub> diffusion in zeolite RHO: Inverse kinetic isotope effect for sieving. *J. Am. Chem. Soc.* **141**, 13635–13642 (2019).
10. Qin, X. et al. Zero-point quantum diffusion of protons in the hydrogen-rich superconductor LaH<sub>10</sub> from first principles. *Phys. Rev. B* **108**, 064102 (2023).
11. van de Bund, S. & Ackland, G. J. Competition between superconductivity and molecularization in the quantum nuclear behavior of lanthanum hydride. *Phys. Rev. B* **108**, 184102 (2023).
12. Cha, Y., Murray, C. J. & Klinman, J. P. Hydrogen tunneling in enzyme reactions. *Science* **243**, 1325–1330 (1989).
13. Prah, A., Ogrin, P., Mavri, J. & Stare, J. Nuclear quantum effects in enzymatic reactions: simulation of the kinetic isotope effect of phenylethylamine oxidation catalyzed by monoamine oxidase A. *Phys. Chem. Chem. Phys.* **22**, 6838–6847 (2020).
14. Wang, M., Lu, Z. & Yang, W. Nuclear quantum effects on an enzyme-catalyzed reaction with reaction path potential: proton transfer in triosephosphate isomerase. *J. Chem. Phys.* **124**, 124516 (2006).
15. Klinman, J. P. & Kohen, A. Hydrogen tunneling links protein dynamics to enzyme catalysis. *Annu. Rev. Biochem.* **82**, 471–496 (2013).
16. Bates, F. S., Koehler, W. C., Wignall, G. D. & Fetters, L. J. Impact of isotope effects on the SANS analysis of amorphous polymers. *MRS Online Proc. Libr.* **79**, 159–166 (1986).
17. Yang, H., Hadziioannou, G. & Stein, R. S. The effect of deuteration on the phase equilibrium of the polystyrene/poly(vinyl methyl ether) blend system. *J. Polym. Sci. Polym. Phys. Ed.* **21**, 159–162 (1983).
18. Strazielle, C. & Benoit, H. Some thermodynamic properties of polymer-solvent systems. Comparison between deuterated and undeuterated systems. *Macromolecules* **8**, 203–205 (1975).
19. Bates, F. S., Keith, H. D. & McWhan, D. B. Isotope effect on the melting temperature of nonpolar polymers. *Macromolecules* **20**, 3065–3070 (1987).
20. Bates, F. S., Wignall, G. D. & Koehler, W. C. Critical behavior of binary liquid mixtures of deuterated and protonated polymers. *Phys. Rev. Lett.* **55**, 2425–2428 (1985).
21. Bates, F. S. & Wignall, G. D. Isotope-induced quantum-phase transitions in the liquid state. *Phys. Rev. Lett.* **57**, 1429–1432 (1986).
22. Eiler, J. M. "Clumped-isotope" geochemistry—the study of naturally-occurring, multiply-substituted isotopologues. *Earth Planet. Sci. Lett.* **262**, 309–327 (2007).
23. Saucedo, H. E., Vassilev-Galindo, V., Chmiela, S., Müller, K.-R. & Tkatchenko, A. Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature. *Nat. Commun.* **12**, 442 (2021).
24. Patrick, C. E. & Giustino, F. Quantum nuclear dynamics in the photophysics of diamondoids. *Nat. Commun.* **4**, 2006 (2013).
25. Poltavsky, I., Zheng, L., Mortazavi, M. & Tkatchenko, A. Quantum tunneling of thermal protons through pristine graphene. *J. Chem. Phys.* **148**, 204707 (2018).
26. Kim, D. S. et al. Nuclear quantum effect with pure anharmonicity and the anomalous thermal expansion of silicon. *Proc. Natl Acad. Sci.* **115**, 1992–1997 (2018).
27. Wang, L., Isborn, C. & Markland, T. Computational approaches for studying enzyme mechanism part A. In *Methods in Enzymology* (ed. Voith, G. A.), Vol. 577 of, 389–418 <https://www.sciencedirect.com/science/article/pii/S0076687916300830> (Academic Press, 2016).
28. Sutcliffe, M. J. & Scrutton, N. S. A new conceptual framework for enzyme catalysis: hydrogen tunneling coupled to enzyme dynamics in flavoprotein and quinoprotein enzymes. *Eur. J. Biochem.* **269**, 3096–3102 (2002).
29. Harris, T., Zhao, Q. & Mildvan, A. NMR studies of strong hydrogen bonds in enzymes and in a model compound. *J. Mol. Struct.* **552**, 97–109 (2000).
30. Vardi-Kilshtain, A., Nitoker, N. & Major, D. T. Nuclear quantum effects and kinetic isotope effects in enzyme reactions. *Arch. Biochem. Biophys.* **582**, 18–27 (2015).
31. Brookes, J. C. Quantum effects in biology: golden rule in enzymes, olfaction, photosynthesis and magnetodetection. *Proc. R. Soc. A Math Phys. Eng. Sci.* **473**, 20160822 (2017).
32. 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.* **111**, 18454–18459 (2014).

33. Heres, M., Wang, Y., Griffin, P., Gainaru, C. & Sokolov, A. Proton conductivity in phosphoric acid: The role of quantum effects. *Phys. Rev. Lett.* **117**, 156001 (2016).
34. Horita, J. & Wesolowski, D. J. Liquid-vapor fractionation of oxygen and hydrogen isotopes of water from the freezing to the critical temperature. *Geochim. et. Cosmochim. Acta* **58**, 3425–3437 (1994).
35. Merlivat, L. & Nief, G. Fractionnement isotopique lors des changements d'état solide-vapeur et liquide-vapeur de l'eau à des températures inférieures à 0°C. *Tellus A Dyn. Meteorol. Oceanogr.* **19**, 122 (1967).
36. Bottinga, Y. Calculated fractionation factors for carbon and hydrogen isotope exchange in the system calcite-carbon dioxide-graphite-methane-hydrogen-water vapor. *Geochim. et. Cosmochim. Acta* **33**, 49–64 (1969).
37. Majoube, M. Fractionnement en oxygène 18 et en deutérium entre l'eau et sa vapeur. *J. de. Chim. Phys.* **68**, 1423–1436 (1971).
38. Kakiuchi, M. & Matsuo, S. Direct measurements of D/H and  $^{18}\text{O}/^{16}\text{O}$  fractionation factors between vapor and liquid water in the temperature range from 10 to 40°C. *Geochem. J.* **13**, 307–311 (1979).
39. Webb, M. A. & Miller, T. F. Position-specific and clumped stable isotope studies: comparison of the urey and path-integral approaches for carbon dioxide, nitrous oxide, methane, and propane. *J. Phys. Chem. A* **118**, 467–474 (2014).
40. Giubertoni, G., Bonn, M. & Woutersen, S. D<sub>2</sub>O as an imperfect replacement for H<sub>2</sub>O: Problem or opportunity for protein research? *J. Phys. Chem. B* **127**, 8086–8094 (2023).
41. Gittings, M. R. et al. Structure of guar in solutions of H<sub>2</sub>O and D<sub>2</sub>O: an ultra-small-angle light-scattering study. *J. Phys. Chem. B* **104**, 4381–4386 (2000).
42. Eagle, R. A. et al. Dinosaur body temperatures determined from isotopic ( $^{13}\text{C}$ - $^{18}\text{O}$ ) ordering in fossil biominerals. *Science* **333**, 443–445 (2011).
43. Chandler, D. & Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.* **74**, 4078–4095 (1981).
44. Habershon, S., Markland, T. E. & Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **131**, 024501 (2009).
45. Chen, B., Ivanov, I., Klein, M. L. & Parrinello, M. Hydrogen bonding in water. *Phys. Rev. Lett.* **91**, 215503 (2003).
46. Fang, W. et al. Inverse temperature dependence of nuclear quantum effects in DNA base pairs. *J. Phys. Chem. Lett.* **7**, 2125–2131 (2016).
47. Thakur, A. C. & Remsing, R. C. Nuclear quantum effects in the acetylene: ammonia plastic co-crystal. *J. Chem. Phys.* **160**, 024502 (2024).
48. Ganeshan, K. et al. Importance of nuclear quantum effects on aqueous electrolyte transport under confinement in Ti<sub>3</sub>C<sub>2</sub> MXenes. *J. Chem. Theory Comput.* **18**, 6920–6931 (2022).
49. Spura, T., John, C., Habershon, S. & Kühne, T. D. Nuclear quantum effects in liquid water from path-integral simulations using anab initioforce-matching approach. *Mol. Phys.* **113**, 808–822 (2014).
50. Seo, B., Lin, Z.-Y., Zhao, Q., Webb, M. A. & Savoie, B. M. Topology automated force-field interactions (TAFFI): a framework for developing transferable force fields. *J. Chem. Inf. Model.* **61**, 5013–5027 (2021).
51. Caleman, C. et al. Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **8**, 61–74 (2011).
52. Ceriotti, M. et al. Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* **116**, 7529–7550 (2016).
53. Kuharski, R. A. & Rossky, P. J. A quantum mechanical study of structure in liquid H<sub>2</sub>O and D<sub>2</sub>O. *J. Chem. Phys.* **82**, 5164–5177 (1985).
54. Wallqvist, A. & Berne, B. Path-integral simulation of pure water. *Chem. Phys. Lett.* **117**, 214–219 (1985).
55. Morrone, J. A. & Car, R. Nuclear quantum effects in water. *Phys. Rev. Lett.* **101**, 017801 (2008).
56. Pereyaslavets, L. et al. On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proc. Natl. Acad. Sci. USA* **115**, 8878–8882 (2018).
57. 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).
58. Ceriotti, M., Cuny, J., Parrinello, M. & Manolopoulos, D. E. Nuclear quantum effects and hydrogen bond fluctuations in water. *Proc. Natl. Acad. Sci. USA* **110**, 15591–15596 (2013).
59. Markland, T. E. & Berne, B. J. Unraveling quantum mechanical effects in water using isotopic fractionation. *Proc. Natl. Acad. Sci. USA* **109**, 7988–7991 (2012).
60. Stolte, N., Daru, J., Forbert, H., Behler, J. & Marx, D. Nuclear quantum effects in liquid water are marginal for its average structure but significant for dynamics. *J. Phys. Chem. Lett.* **15**, 12144–12150 (2024).
61. Eltareb, A., Lopez, G. E. & Giovambattista, N. Nuclear quantum effects on the thermodynamic, structural, and dynamical properties of water. *Phys. Chem. Chem. Phys.* **23**, 6914–6928 (2021).
62. Li, X.-Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. USA* **108**, 6369–6373 (2011).
63. Kapil, V., Engel, E., Rossi, M. & Ceriotti, M. Assessment of approximate methods for anharmonic free energies. *J. Chem. Theory Comput.* **15**, 5845–5857 (2019).
64. Rossi, M., Gasparotto, P. & Ceriotti, M. Anharmonic and quantum fluctuations in molecular crystals: a first-principles study of the stability of paracetamol. *Phys. Rev. Lett.* **117**, 115702 (2016).
65. Sainburg, T., McInnes, L. & Gentner, T. Q. Parametric UMAP embeddings for representation and semisupervised learning. *Neural Comput.* **33**, 2881–2907 (2021).
66. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
67. Zdrzil, B. et al. The ChEMBL database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2023).
68. Davies, M. et al. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* **43**, W612–W620 (2015).
69. Halgren, T. A. Merck molecular force field. i. basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519 (1996).
70. Li, C., Paesani, F. & Voth, G. A. Static and dynamic correlations in water: comparison of classical ab initio molecular dynamics at elevated temperature with path integral simulations at ambient temperature. *J. Chem. Theory Comput.* **18**, 2124–2131 (2022).
71. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Advances in Neural Information Processing Systems* (ed. Guyon, I.), Vol. 30, 4765–4774. <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf> (Curran Associates, Inc., 2017).
72. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostating of path integral molecular dynamics. *J. Chem. Phys.* **133**, 124104 (2010).
73. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
74. Landrum, G. et al. rdkit/rdkit: 2022\_09\_3 (Q3 2022) release, Zenodo, <https://doi.org/10.5281/zenodo.7415128> (2022).
75. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

76. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).
77. Platt, J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, 61–74 (MIT Press, 1999).
78. Gowers, R. J. et al. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. In *Proc. 15th Python in Science Conference* (eds. Benthall, S. & Rostrup, S.) 98–105 (2016).
79. Michaud Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
80. Ugur, B. E. & Webb, M. A. NQELiq-298: dataset of nuclear quantum effects and equilibrium isotope effects for 92 chemically diverse molecular liquids. *Zenodo*, <https://doi.org/10.5281/zenodo.15236880> (2025).
81. Ugur, B. E. & Webb, M. A. Nuclear quantum effects in molecular liquids across chemical space (this paper). v1.1 web-theosim/chem-space-nqes: nuclear quantum effects across chemical space, *Zenodo*, <https://doi.org/10.5281/zenodo.15465591> (2025).
82. Schrödinger, LLC. The PyMOL molecular graphics system, version 3.1.0 (2015).
83. Yeo, I.-K. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959 (2000).
84. Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

## Acknowledgements

B.E.U. and M.A.W. acknowledge support from the National Science Foundation under Grant No. 2237470 (M.A.W.). The authors also acknowledge support from the “Chemistry in Solution and at Interfaces” (CSI) Center funded by the U.S. Department of Energy through Award No. DE-SC0019394 (M.A.W.). Simulations and analyses were performed using resources from Princeton Research Computing at Princeton University, which is a consortium led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology’s Research Computing. These resources include a GPU-based computing cluster purchased with support from the National Science Foundation (Grant No. NSF-MRI: OAC-2320649) (M.A.W.). We thank Hang Zhang for his valuable assistance with quantum chemistry mode scans for force field parameterization.

## Author contributions

M.A.W. and B.E.U. designed research; B.E.U. performed research; B.E.U. and M.A.W. analyzed data; B.E.U. and M.A.W. discussed results; B.E.U. and M.A.W. wrote the paper; B.E.U. and M.A.W. edited the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60850-x>.

**Correspondence** and requests for materials should be addressed to Michael A. Webb.

**Peer review information** *Nature Communications* thanks Davide Donadio, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025