

Nuclear Quantum Effects Across Chemical Space

Baris E. Ugur¹ and Michael A. Webb^{1*}

Princeton, NJ 08544, USA

¹Department of Chemical and Biological Engineering, Princeton University
Princeton, NJ 08544, USA

*To whom correspondence should be addressed: mawebb@princeton.edu

Abstract

Nuclear quantum effects (NQEs) are important in many physical and chemical processes—particularly those involving lighter nuclei or occurring at low temperatures. Nevertheless, NQEs have been carefully quantified in few systems, such as water, and their importance is rarely considered. Here, using path-integral molecular dynamics, we critically examine NQEs for a range of molecular properties (e.g., density, thermal expansion coefficients, isothermal compressibility, static dielectric constant, and the heat of vaporization) across 87 molecular liquids. We discover substantial NQEs in this broad chemical space, with molar volumes exhibiting differences of up to 5% between classical and quantum treatments; similar magnitudes are revealed in equilibrium isotope effects from deuteration. Using machine learning, we identify that four molecular properties that are readily known or easily computed with classical simulations—density, molar mass, hydrogen density, and thermal expansion coefficient—are strong predictors of NQEs. Further data-driven analysis reveals how molecular factors, such as branching and heteroatom composition, influence intermolecular interactions and fluid stability and thereby affect observed NQEs. This work offers new insights into the relationship between NQEs and molecular chemistry and refines expectations for when rigorous treatment of NQEs is necessary.

Introduction

The quantum nature of nuclei influences every material and chemical process. Resulting nuclear quantum effects (NQEs) can arise as zero-point energy, (1) quantum tunneling, (2) 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, (3, 4) zeolite catalysts, (5, 6, 7, 8) superconductor materials, (9, 10) and enzymatic reactions, (11, 12, 13, 14) amongst many other biological, chemical, and physical processes. NQEs also directly impact experimental studies that investigate or depend on isotope substitution (15, 16, 17, 18, 19, 20, 21). Therefore, NQEs are important for a wide range of processes and conditions, yet their extent and significance is 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; (12, 22, 23, 24, 25, 26, 27) for instance, isotope substitution can alter the oxidation rate of yeast alcohol dehydrogenase by up to 33%. (14) 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. (28) Meanwhile, classical statistical thermodynamics does not predict isotope-dependent changes in equilibrium properties, yet isotope effects are widely observed. (29, 30, 31, 32, 33, 34) For example, the solid-vapor isotope fractionation ratio between deuterated and normal water reaches values as high as 1.208. (30) Replacing H₂O with D₂O similarly impacts the structure and properties of biopolymer solutions, (35, 36) despite D₂O being chemically equivalent to H₂O 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 (37). The presence of such isotope effects signal the relevance of NQEs but do not fully characterize them.

Path-integral molecular dynamics (PIMD) enables explicit treatment of NQEs and calculation of isotope effects. For water, PIMD has demonstrated how intramolecular zero-point motion and intermolecular tunneling affect properties like translational diffusion and orientational relaxation rates. (38, 39) PIMD has also revealed the significance of NQEs in DNA base pairs, acetylene:ammonia co-crystals, and electrolyte transport in confined aqueous systems. (40, 41, 42) Unlike more conventional methods, such as density functional theory or classical molecular dynamics, which do not directly account for NQEs, PIMD uses an isomorphism between the statistical thermodynamics of a quantum system and a classical system in an extended phase space (43). However, the extended phase space increases computational costs, relative to classical MD, and it requires a force field that describes the Born-Oppenheimer potential energy surface, which is often expensive or unavailable. Most force fields deployed in classical simulations often implicitly account for NQEs in uncontrolled ways by fitting to experimental data. Thus, despite the prospective utility of PIMD, the role of NQEs for many systems and conditions remains largely unexplored.

Here, we characterize NQEs in 87 organic liquids spanning a wide chemical space. By comparing classical and path-integral MD simulations using a quantum chemical-

based force field, we quantify NQEs on several thermodynamic properties including molar volume (v_m), thermal expansion coefficient (α_P), isothermal compressibility (κ_T), static dielectric constant ($\epsilon_r(0)$), and heat of vaporization (ΔH_{vap}). Equilibrium isotope effects on the same systems are also examined by simulating deuterated systems. We find a meaningful influence of NQEs for every substance and highlight key chemical features that correlate with NQEs using unsupervised and unsupervised 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.

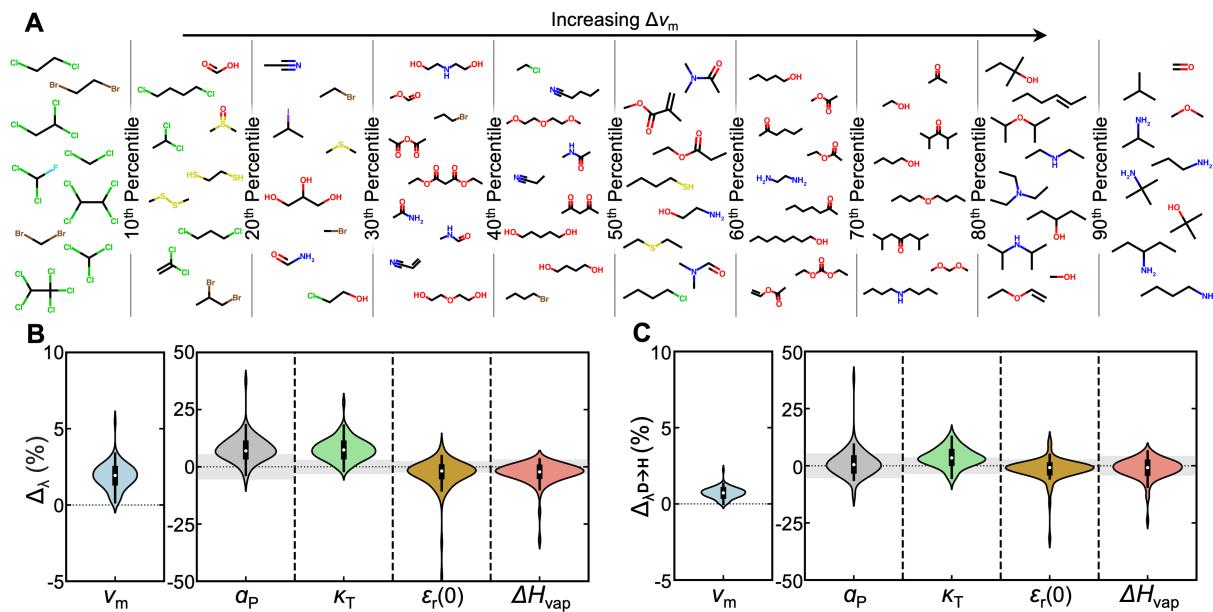


Figure 1: Impact of NQEs on various macroscopic properties. (A) Chemical structure drawings of every simulated system. Molecules are categorized into 10 percentile bins based on their Δv_m . Within each bin, higher vertical order also denotes higher Δv_m . (B) The magnitude of NQEs on the molar volume, thermal expansion coefficient, isothermal compressibility, static dielectric constant, and heat of vaporization of each liquid system. (C) The calculated isotope effects on each property of the investigated liquid organic systems. The gray regions around $y = 0$ under each plot indicates the average standard error of the mean for the magnitude of NQEs on each property obtained from four independent simulations. All displayed properties are calculated at 298.15 K and 1 atm.

Results

Varied effects across molecules and properties

To understand how NQEs manifest across chemical space, we simulate 87 organic molecules at ambient conditions and quantify the extent of NQEs via

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

where λ^{cl} and λ^{PI} are the properties calculated via classical and path-integral MD simulations. Figure 1 summarizes the results, with Figure 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 (Figure 1B). Accounting for NQEs markedly increases molar volume by up to 5.5% and also increases α_P and κ_T for a majority of the systems, with an average effect of nearly 8%. These systematic increases signify weaker cohesive intermolecular interactions due to nuclear delocalization, consistent with prior studies on water and other liquid organic systems. (44, 45, 46, 47, 48, 49, 50, 51) Quantum treatment of nuclei yields less consistent directional changes from classical results on the enthalpy of vaporization and the static dielectric constant. For both properties, the average effect is small and slightly negative but also comparable to the statistical error in their calculation. Overall, these results indicate that NQEs can significantly and detectably influence certain thermodynamic properties across diverse molecular chemistries.

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

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

where $\lambda^{\text{PI,D}}$ is the property for a fully-deuterated system simulated via PIMD, and deuterated systems are expected to approximate classical behavior. The quantity $\Delta_{\lambda^{\text{D} \rightarrow \text{H}}}$ is equivalent to $1 - \lambda^{\text{H}}/\lambda^{\text{D}}$ where $\lambda^{\text{H}}/\lambda^{\text{D}}$ is often noted as the (experimentally accessible) equilibrium isotope effect. Trends in $\Delta_{\lambda^{\text{D} \rightarrow \text{H}}}$ (Figure 1C) align directionally with Δ_λ (Figure 1B) but with reduced magnitudes. Beyond consistent shifts in molar volume, the effect of deuteration is often within the range of statistical error. This change is most notable for α_P and κ_T , where the average effect shifts from 7.7% and 7.6% when comparing to purely classical systems to 1.0% and 3.6% when comparing to quantum mechanical but deuterated ones. This highlights that deuteration does not fully replicate a purely classical treatment, and a null deuteration effect does not necessarily indicate negligible NQEs

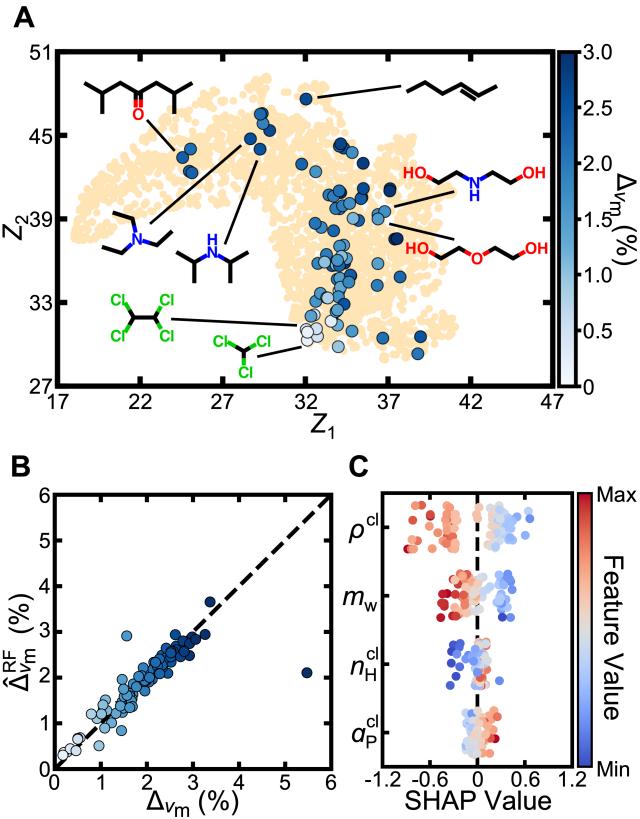


Figure 2: Correlation of chemical features with the magnitude of NQEs. (A) A two-dimensional manifold organization of the 87 organic molecules (blue, larger markers) and 2,879 small molecules (tan, smaller markers) obtained from the ChEMBL database. The manifold is generated using the Uniform Manifold Approximation and Projection (UMAP) unsupervised learning algorithm using the 34-dimensional Merck molecular force field (52) (MMFF94) atom typing as the input feature vector. Inset lines highlight representative molecules from different regions of the manifold. (B) Comparison of predicted versus simulated effect of NQEs on molar volume (Δ_{v_m}). Predicted values are obtained from random forest regression with four chemically interpretable input features: average atomic mass (m_w), substance density (ρ^{cl}), hydrogen density (n_H^{cl}), and thermal expansion coefficient (α_P^{cl}) where the latter three 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 RF data, are generally smaller than the symbol size. (C) Impact of feature contributions to random forest predictions 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 are normalized using the Yeo-Johnson power transformation and then min-max scaling for consistency. (53)

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, our discussion focuses on Δ_{v_m} , while analyses of other properties are included in Supplemental Information, Section S1.

Unsupervised machine learning allows us to reduce a high-dimensional molecular feature space into a two-dimensional representation that contextualizes the studied molecules with a broader set from the ChEMBL database (54, 55). Using the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) algorithm (56, 57) for the dimensionality reduction illustrates how the simulated substances span a wide chemical space (Figure 2A). Chemical structures displaying lower Δ_{v_m} are seemingly clustered on the two-dimensional manifold in the vicinity of high molecular weight systems such as chloroform. In contrast, molecules with larger Δ_{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 Δ_{v_m} (Supplemental Information, Figure S5). For instance, molecules with amines tend to exhibit higher Δ_{v_m} .

To gain insight into what factors might correlate with NQEs, supervised machine learning is used to model Δ_{v_m} from simple molecular descriptors. We find that a random forest (RF) regressor using only four input features—average atomic mass m_w and classically determined substance density ρ^{cl} , hydrogen density n_H^{cl} , and thermal expansion coefficient α_P^{cl} —effectively predicts Δ_{v_m} (Figure 2B). Shapley Additive Explanations (SHAP) analysis (58), which enables the attribution of feature contributions to predictions made by a machine learning model, reveals intuitive trends: lighter nuclei enhance Δ_{v_m} , higher ρ^{cl} reduces it, and greater n_H^{cl} increases it (Figure 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 the presence of lighter nuclei. However, both ρ^{cl} and substances n_H^{cl} also have some non-trivial dependence on underlying intermolecular interactions. These interactions also directly influence the thermal expansion coefficient α_P^{cl} , which correlates positively with Δ_{v_m} . Overall, the 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 α_P appear to reflect competing effects: higher n_H suggests stronger intermolecular forces, which would typically reduce α_P . To explore this tradeoff, we leverage the dataset to create controlled comparisons by fixing either n_H or α_P and examining the effect of the other on Δ_{v_m} . This is accomplished through k -means clustering to group molecules with similar n_H values and examining correlations

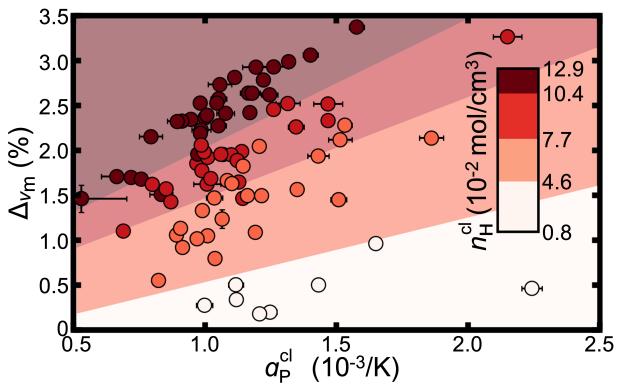


Figure 3: Analysis of hydrogen density and thermal expansion coefficient. Four groups of data are generated with k -means clustering algorithm on the respective hydrogen densities n_H ; the marker colors indicate being within the range specified by the color bar. The colored regions are distinguished by support vector machine (SVM) margin lines. Methanol is excluded as an outlier, as determined by the random forest regressor of Fig. 2. Error bars (black) represent the standard error of the mean obtained from four independent simulations.

of the groups with Δ_{v_m} (Figure 3). Within groups with similar n_H , a decrease in fluid stability (indicated by higher α_P) tends to enhance NQEs. Conversely, across groups with comparable stability (similar α_P), an increase in n_H (and thus more light nuclei) leads to stronger NQEs. Nevertheless, the overlapping α_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 the correlation between Δ_{v_m} and n_H and α_P manifests on the basis of molecular constitution, we systematically compare groups of molecules varying in hydrogen bonding (Figure 4A) and branching (Figure 4B).

In Figure 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 Δ_{v_m} . Among these molecules— which share the same number of heavy atoms, the same chemical topology, and comparable α_P —the hydrogen-bonding thiol and alcohol exhibit higher n_H . However, simply increasing the number of hydrogen-bonding groups does not necessarily enhance n_H and Δ_{v_m} . This is revealed by comparing the group of butan-1-ol (iv), butane-1,4-diol (v), pentane-1,5-diol (vi), and propane-1,2,3-triol (vii). Within this group, increasing the number of hydrogen-bonding groups minimally affects hydrogen density but increases the stability of the fluid (decreases α_P). Evaluation of different hydrogen-bonding groups reveals similar trends (Supplemental Information, Figure S6)

where the balance of n_H and α_P elucidates the resulting Δ_{v_m} . For example, relative to alcohols with similar n_H , amines display larger NQEs and associated larger α_P .

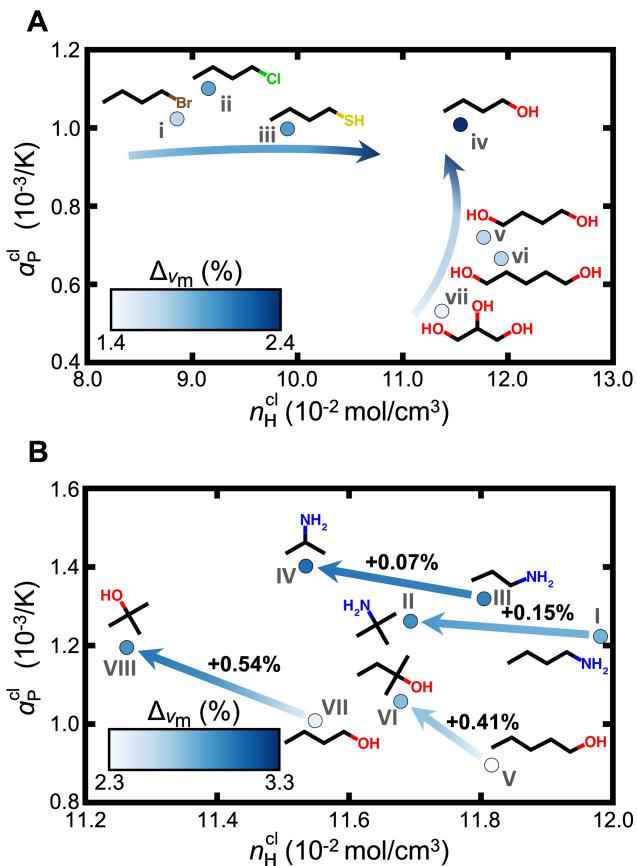


Figure 4: Impact of hydrogen bonding and branching on the extent of NQEs. (A) Hydrogen densities and thermal expansion coefficients for systems of various hydrogen-bonding groups and counts. The darker blue tone of marker colors denote higher magnitudes. Visualized molecules are labeled as (i) 1-bromobutane, (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 are 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 are equal.

In Figure 4B, comparing multiple linear molecules and their branched analogues, we observe consistent impact on NQEs. In particular, the branched molecule exhibits larger Δ_{v_m} than the linear one. Across all four pairs, branching reduces n_H , but there is a coupled increase in α_P . For these molecules, the net result is stronger NQEs in the branched molecules. We therefore suggest that NQEs, at least as manifest through Δ_{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 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. 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 (Supplemental Information, Figure S7). This renders the importance of hydrogen-bonding groups on trends in α_P 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 (Figure 5). The distribution of hydrogen bonds (Figure 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 (Figure 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. 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 (Figure 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.

Discussion & Conclusions

Nuclear quantum effects (NQEs) are present in all physical systems, but their significance is rarely known. In this study, we quantified the impact of NQEs on various properties across 87 organic molecules spanning diverse chemical space. By comparing path-integral and classical simulations, we observed that NQEs can significantly and measurably influence molar volumes, thermal expansion coefficients, and isothermal compressibilities.

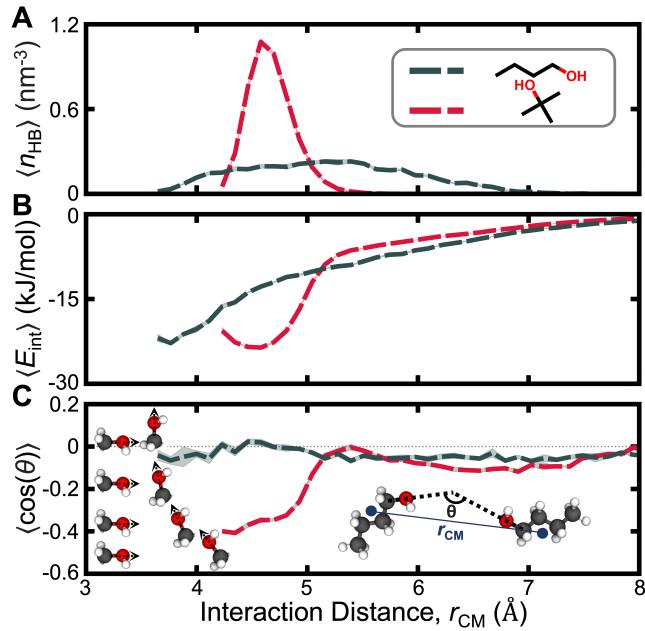


Figure 5: Effect of branching on intermolecular interactions through comparison of linear butan-1-ol with branched 2-methylpropan-2-ol. Ensemble averages of (A) the hydrogen-bond density, (B) pairwise intermolecular interaction energies, and (C) interaction orientations as a function of distance between molecules. The inset images of panel (C) illustrate calculations of the interaction distance, which is measured between the centers of mass of two molecules, and the orientation, which is based on the angle between the bond vectors adjoining the carbon to the oxygen on the alcohol; the renderings adjacent to the y-axis illustrate configurations consistent with the nearest tick marks. The interaction energy is the sum of all pairwise non-bonded energy contributions. 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 (59) (version 1.9.4).

Effects on dielectric constants and enthalpies of vaporization lack clear trends relative to the uncertainty of the calculations themselves. 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 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 random forest model with interpretable inputs of the average mass of composite atoms and classically computed densities, 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 exhibit strong NQEs because of the enhanced fluid stability. 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 (60) with Waldman-Hagler mixing rules. Real-space non-bonded interactions were truncated at 14 Å. Long-range electrostatics calculations used the Particle Mesh Ewald (PME) algorithm; tail corrections were applied for Lennard-Jones interactions beyond the real-space cutoff. All PIMD simulations used a ring-polymer bead count of $P=32$. All simulations used a timestep of 0.5 fs. Constant-temperature conditions were achieved using a Langevin thermostat for classical simulations and the path-integral Langevin equation (PILE) thermostat (61) for path-integral simulations; both utilize a friction coefficient of 1 ps⁻¹. Constant-pressure conditions were achieved using a Monte Carlo barostat with an attempt frequency of 25 fs. All PIMD simulations were employed with a ring-polymer bead count of $P=32$. The force field in tandem with these settings enables accurate reproduction of liquid-phase experimental densities (62), effective energy conservation, and sufficient convergence of physical properties (see Supplemental Information, Figure S8, S9, and S10).

Property calculations and analyses follow utilized 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 $L = 50$ nm³. The initial configuration was sub-

jected 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 was collected every 1 ps system configurations were saved every 10 ps.

All simulations were performed using the GPU implementation of the OpenMM 7.7.0 software package. (63) PIMD simulations were performed on NVIDIA A100 and P100 GPUs, with mean simulation times of 46.5 and 89.3 hours, respectively, in contrast with classical MD simulations of 1.7 hours 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 Δh_{vap} . A molecule was placed in a large box with a volume of 125 nm³, 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⁻¹. 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, α_P , were calculated using

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

where $\frac{\partial V}{\partial T}$ 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 1 atm for 0.5 ns, followed by 1.5 ns production runs where the system volume was saved every 1 ps.

Isothermal compressibilities, κ_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, $\varepsilon_r(0)$ were calculated using

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

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

Molar heats of vaporization, Δh_{vap} , were calculated using

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

where $U^{(\nu)}$ is the potential energy per mole in phase ν , either liquid or gas.

Hydrogen-bond densities, n_H , were computed by analyzing configurations sampled every 1 ns. Hydrogen bonds were identified using geometric criteria: a cutoff distance of 3.6 Å between participating atoms and a cutoff angle of 150° for the alignment of the three involved atoms. These calculations were facilitated using the MDAnalysis Python library (64, 65) (version 2.1.0).

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 a 34-dimensional feature space. The MMFF94 atom types were obtained using the rdkit Python package (version 2022.9.3). (66) 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 2,879 molecules. For the UMAP hyperparameters, 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. (67)

Supervised Learning

Supervised learning via random forest regression used the scikit-learn Python package. Model inputs included ρ , m_w , n_H , and α_P 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 86 molecules. This process was repeated for all 87 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 four samples per split. Shapley Additive Explanations (SHAP) analysis was performed using the shap

Python package (version 0.44.1) on each of the four input parameters for the random forest regression model.

The LIBSVM implementation of C-Support Vector Classification algorithm (68, 69) 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.

Impact of force field

We use TAFFI because it relies purely on quantum chemical calculations for its parameterization and for its computational efficiency; however, we acknowledge that the reported magnitude of NQEs could differ based on the underlying force field. To gain some insight into this sensitivity, classical and PIMD simulations using two different force fields—the all-atom optimized potentials for liquid simulations (OPLS-AA) force field and Open Force Field (OpenFF, version 2.0.0 with unconstrained bonds)—were also performed and analyzed for a subset of 11 molecules. Simulation procedures were the same as described in *General Simulation Protocols*.

Relative to the TAFFI results, OPLS-AA and OpenFF consistently find larger Δv_m across all systems (Supplemental Information, Figure S12). 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.

Acknowledgments

B.E.U. and M.A.W. acknowledge support from the National Science Foundation under Grant No. 2237470. 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. 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)

Supporting Information

Analysis of NQEs across various properties. Impact of functional groups on Δv_m . Impact of various hydrogen-bonding groups on Δv_m . Effects of molecular branching on system characteristics. Benchmarking of TAFFI force field with experiment. NVE energy conservation in PIMD simulations. Convergence of system properties via ring-polymer bead count. Impact of force field on Δv_m .

References

1. M. Benoit, D. Marx, M. Parrinello, *Nature* **392**, 258 (1998).
2. J. R. Cendagorta, *et al.*, *Physical Chemistry Chemical Physics* **18**, 32169 (2016).
3. L. Slocombe, M. Sacchi, J. Al-Khalili, *Communications Physics* **5** (2022).
4. F. Angiolari, S. Huppert, F. Pietrucci, R. Spezia, *The Journal of Physical Chemistry Letters* **14**, 5102 (2023).
5. M. Bocus, *et al.*, *Nature Communications* **14** (2023).
6. K. Suzuki, *et al.*, *Journal of Alloys and Compounds* **509**, S868 (2011).
7. K. McCardle, *Nature Computational Science* **3**, 189 (2023).
8. L. G. Gao, R. M. Zhang, X. Xu, D. G. Truhlar, *Journal of the American Chemical Society* **141**, 13635 (2019).
9. X. Qin, *et al.*, *Physical Review B* **108**, 064102 (2023).
10. S. van de Bund, G. J. Ackland, *Physical Review B* **108**, 184102 (2023).
11. Y. Cha, C. J. Murray, J. P. Klinman, *Science* **243**, 1325 (1989).
12. A. Prah, P. Ogrin, J. Mavri, J. Stare, *Physical Chemistry Chemical Physics* **22**, 6838 (2020).
13. M. Wang, Z. Lu, W. Yang, *The Journal of Chemical Physics* **124** (2006).
14. J. P. Klinman, A. Kohen, *Annual Review of Biochemistry* **82**, 471 (2013).
15. F. S. Bates, W. C. Koehler, G. D. Wignall, L. J. Fetters, *MRS Proceedings* **79** (1986).
16. H. Yang, G. Hadzioannou, R. S. Stein, *Journal of Polymer Science: Polymer Physics Edition* **21**, 159 (1983).

17. C. Strazielle, H. Benoit, *Macromolecules* **8**, 203 (1975).
18. F. S. Bates, H. D. Keith, D. B. McWhan, *Macromolecules* **20**, 3065 (1987).
19. F. S. Bates, G. D. Wignall, W. C. Koehler, *Physical Review Letters* **55**, 2425 (1985).
20. F. S. Bates, G. D. Wignall, *Physical Review Letters* **57**, 1429 (1986).
21. J. M. Eiler, *Earth and Planetary Science Letters* **262**, 309 (2007).
22. L. Wang, C. Isborn, T. Markland, *Simulating Nuclear and Electronic Quantum Effects in Enzymes* (Elsevier, 2016), pp. 389–418.
23. M. J. Sutcliffe, N. S. Scrutton, *European Journal of Biochemistry* **269**, 3096 (2002).
24. T. Harris, Q. Zhao, A. Mildvan, *Journal of Molecular Structure* **552**, 97 (2000).
25. A. Vardi-Kilshtain, N. Nitoker, D. T. Major, *Archives of Biochemistry and Biophysics* **582**, 18 (2015).
26. J. C. Brookes, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20160822 (2017).
27. L. Wang, S. D. Fried, S. G. Boxer, T. E. Markland, *Proceedings of the National Academy of Sciences* **111**, 18454 (2014).
28. M. Heres, Y. Wang, P. Griffin, C. Gainaru, A. Sokolov, *Physical Review Letters* **117**, 156001 (2016).
29. J. Horita, D. J. Wesolowski, *Geochimica et Cosmochimica Acta* **58**, 3425 (1994).
30. L. Merlivat, G. Nief, *Tellus A: Dynamic Meteorology and Oceanography* **19**, 122 (1967).
31. Y. Bottinga, *Geochimica et Cosmochimica Acta* **33**, 49 (1969).
32. M. Majoube, *Journal de Chimie Physique* **68**, 1423 (1971).
33. M. Kakiuchi, S. Matsuo, *Geochemical Journal* **13**, 307 (1979).
34. M. A. Webb, T. F. Miller, *The Journal of Physical Chemistry A* **118**, 467 (2014).
35. G. Giubertoni, M. Bonn, S. Woutersen, *The Journal of Physical Chemistry B* **127**, 8086 (2023).
36. M. R. Gittings, *et al.*, *The Journal of Physical Chemistry B* **104**, 4381 (2000).
37. R. A. Eagle, *et al.*, *Science* **333**, 443 (2011).

38. S. Habershon, T. E. Markland, D. E. Manolopoulos, *The Journal of Chemical Physics* **131** (2009).
39. B. Chen, I. Ivanov, M. L. Klein, M. Parrinello, *Physical Review Letters* **91**, 215503 (2003).
40. W. Fang, *et al.*, *The Journal of Physical Chemistry Letters* **7**, 2125 (2016).
41. A. C. Thakur, R. C. Remsing, *The Journal of Chemical Physics* **160** (2024).
42. K. Ganeshan, *et al.*, *Journal of Chemical Theory and Computation* **18**, 6920 (2022).
43. D. Chandler, P. G. Wolynes, *The Journal of Chemical Physics* **74**, 4078 (1981).
44. M. Ceriotti, *et al.*, *Chemical Reviews* **116**, 7529 (2016).
45. R. A. Kuharski, P. J. Rossky, *The Journal of Chemical Physics* **82**, 5164 (1985).
46. A. Wallqvist, B. Berne, *Chemical Physics Letters* **117**, 214 (1985).
47. J. A. Morrone, R. Car, *Physical Review Letters* **101**, 017801 (2008).
48. L. Pereyaslavets, *et al.*, *Proceedings of the National Academy of Sciences* **115**, 8878 (2018).
49. A. Berger, G. Ciardi, D. Sidler, P. Hamm, A. Shalit, *Proceedings of the National Academy of Sciences* **116**, 2458 (2019).
50. M. Ceriotti, J. Cuny, M. Parrinello, D. E. Manolopoulos, *Proceedings of the National Academy of Sciences* **110**, 15591 (2013).
51. T. E. Markland, B. J. Berne, *Proceedings of the National Academy of Sciences* **109**, 7988 (2012).
52. T. A. Halgren, *Journal of Computational Chemistry* **17**, 490 (1996).
53. I.-K. Yeo, *Biometrika* **87**, 954 (2000).
54. B. Zdrazil, *et al.*, *Nucleic Acids Research* **52**, D1180 (2023).
55. M. Davies, *et al.*, *Nucleic Acids Research* **43**, W612 (2015).
56. T. Sainburg, L. McInnes, T. Q. Gentner, *Neural Computation* **33**, 2881 (2021).
57. L. McInnes, J. Healy, N. Saul, L. Großberger, *Journal of Open Source Software* **3**, 861 (2018).

58. S. M. Lundberg, S.-I. Lee, *Advances in Neural Information Processing Systems 30*, I. Guyon, *et al.*, eds. (Curran Associates, Inc., 2017), pp. 4765–4774.
59. W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **14**, 33 (1996).
60. B. Seo, Z.-Y. Lin, Q. Zhao, M. A. Webb, B. M. Savoie, *Journal of Chemical Information and Modeling* **61**, 5013 (2021).
61. M. Ceriotti, M. Parrinello, T. E. Markland, D. E. Manolopoulos, *The Journal of Chemical Physics* **133** (2010).
62. C. Caleman, *et al.*, *Journal of Chemical Theory and Computation* **8**, 61 (2011).
63. P. Eastman, *et al.*, *PLOS Computational Biology* **13**, e1005659 (2017).
64. R. Gowers, *et al.*, *Proceedings of the 15th Python in Science Conference*, SciPy (SciPy, 2016).
65. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, *Journal of Computational Chemistry* **32**, 2319 (2011).
66. G. Landrum, *et al.*, rdkit/rdkit: 2022_09_3(q32022)release(2022).
67. F. Pedregosa, *et al.*, *Journal of Machine Learning Research* **12**, 2825 (2011).
68. C.-C. Chang, C.-J. Lin, *ACM Transactions on Intelligent Systems and Technology* **2**, 1 (2011).
69. J. Platt, *Adv. Large Margin Classif.* **10** (2000).