Assessing non-bonded aggregates populations: application to the concentration-dependent IR O–H band of phenol

J. Pablo Galvez,[†] José Zúñiga,[†] and Javier Cerezo[*,‡]

†Departamento de Química Física, Universidad de Murcia, 30100 Murcia, Spain

‡Departamento de Química and Institute for Advanced Research in Chemical Sciences (IAdChem), Universidad Autónoma de Madrid, 28049 Madrid, Spain

E-mail: javier.cerezo@uam.es

Abstract

In this work, we present two alternative computational strategies to determine the populations of non-bonded aggregates. One approach extracts these populations from molecular dynamics (MD) simulations, while the other employs quantum mechanical partition functions for the most relevant minima of the multimolecular potential energy surfaces (PESs), identified by automated conformational sampling. In both cases, we adopt a common graph-theory-based framework, introduced in this work, for identifying aggregate conformations, which enables a consistent comparative assessment of both methodologies and provides insight into the underlying approximations.

We apply both strategies to investigate phenol aggregates, up to the tetramer, at different concentrations in phenol/carbon tetrachloride mixtures. Subsequently, we simulate the concentration-dependent OH stretching IR region by averaging the harmonic IR spectra of aggregates using the populations predicted by each strategy. Our results indicate that the populations extracted from MD trajectories yield OH stretching

1

signals that closely follow the experimental trends, outperforming the spectra from populations obtained by systematic conformational searches. Such a better performance of MD is attributed to a better description of the entropic contributions. Moreover, the proposed protocol not only successfully addresses a very challenging problem but also offers a benchmark to assess the accuracy of intermolecular force fields.

1 Introduction

Non-covalent interactions, such as dispersion, hydrogen or halogen bonding, and π-π stacking, play a pivotal role in the condensed phase. These interactions are fundamental to the supramolecular organization of biologically and technologically relevant systems, including protein structures, DNA base-pairing, and liquid crystals[1]. Detailed understanding of these interactions is therefore crucial for advancements in material design, drug discovery, or protein folding.

Molecular aggregates, which are held together by such non-bonded interactions, may present a remarkable challenge, as their relative populations are governed by a delicate balance of enthalpic and entropic contributions. Indeed, predicting these populations requires (i) an accurate description of non-bonded interactions, which determine the enthalpic component, and (ii) a thorough accounting of the extensive range of conformers arising from complex potential energy surfaces (PESs) with numerous energetically accessible minima, which govern the entropic contribution.

Accurate modeling of non-bonded interactions necessitates electronic structure methods that incorporate non-local correlation effects, often requiring computationally demanding post-Hartree Fock techniques. However, density functional theory (DFT), enhanced by suitable correction schemes such as Grimme's empirical dispersion corrections[2,3], provides a computationally feasible alternative for systems dominated by non-covalent interactions. Combined with implicit solvent models[4], DFT has significantly expanded its applicability to biologically and technologically important systems. Still, such approaches become imprac-

2

tical for large systems requiring numerous potential evaluations, as in molecular dynamics (MD) simulations. In such cases, interatomic interactions are typically described using empirical molecular mechanics (MM) potentials, where non-bonded interactions are represented by Coulombic terms for electrostatics and Lennard-Jones (LJ) potentials for van der Waals forces. While polarizable force fields offer improved descriptions [5,6], simpler point charges and LJ potentials still remain the standard in MD simulations, whose quality can be improved by reparameterizing them against quantum mechanical (QM) energy databases [7].

Moving to the exploration of conformational flexibility, automated conformational searches are commonly employed to identify the most relevant solute configurations [8,9], particularly when interactions are described at QM level. The PES around these minima is typically approximated using harmonic models, with specific adaptations available to account for non-harmonic regions, such as torsional degrees of freedom. Solvent effects are generally included through implicit solvation models, and aggregate populations are ultimately derived by constructing partition functions that incorporate a quantum mechanical treatment of vibrations.

Alternatively, classical MD simulations, which explicitly represent the solvent, provide a highly effective approach for exploring the configurational space of systems, encompassing both solute and solvent. However, few studies have applied MD simulations to directly calculate the populations of clusters of varying sizes within the context of organic aggregates [10–13]. Unbiased MD simulations have primarily been limited to obtaining statistical insights into aggregate sizes and structural characteristics. These analyses typically involve tracking the average number of noncovalent interactions throughout the trajectory [11,14–19], computing radial or pair distribution functions [20–23], or monitoring the evolution of structural parameters [24,25].

A critical step in these strategies is the identification of unique conformers from the vast number of molecular configurations typically generated. This challenge can be effectively addressed using Graph Theory tools. Specifically, adjacency matrix analysis has been suc-

3

cessfully applied to classify molecular configurations into different topological conformers in MD snapshots of individual solute molecules in solution[26,27] or to distinguish topologically distinct, QM-optimized isomers[28]. However, a comprehensive top-down approach that identifies individual aggregates coexisting in solution—potentially within the same snapshot—in terms of their isomorphic conformations, is still lacking. Such a framework is essential for accurately calculating populations in systems involving thousands of particles.

A notable strength of MD approaches for thermodynamic predictions is their independence from prior knowledge of aggregate structures, unlike strategies that compute partition functions that necesitation the identification of the relevant minima. Nevertheless, several challenges persist for MD-based approaches. Accurate determination of population distributions requires a dynamic evaluation of chemical identities at each simulation step, which typically relies on geometric criteria that may involve some arbitrariness. Furthermore, for species present at low concentrations, achieving sufficient conformational sampling demands extended simulation times and/or large system sizes. In contrast, first-principles quantum chemistry calculations are well-suited to these scenarios, employing efficient reaction equilibrium algorithms to compute species concentrations without the need for extensive sampling[29–31]. Still, a comprehensive and reliable framework to compare the performance of MD-based and partition function-based approaches remains an open challenge.

The effects of noncovalent interactions, which influence the populations of aggregates, are ultimately manifested in the spectroscopic signals of the system. Infrared (IR) spectra, in particular, contain exquisite information about microscopic characteristics and provide detailed insights into key noncovalent interactions, notably hydrogen bonds[32,33]. Recent advances have reported simulations of IR and vibrational circular dichroism (VCD) spectra based on aggregate populations for pure solvents[34–42] and solvent mixtures with various solutes[43–52]. These approaches involve averaging the spectra of optimized clusters and calculating their populations using either the quantum cluster equilibrium theory[53–57] or the clusters-in-a-liquid framework[43]. While most studies focus on the fingerprint region, simu-

lations of the O−H stretching region for hydrogen-bonded aggregates[36,37,39,40,42,52] have revealed limitations in addressing the challenging broadband that is observed. The formation of hydrogen bonds significantly affects the vibrational spectrum, shifting the O−H stretching bands of the acceptor to lower wavenumbers. Indeed, this band is highly sensitive to the number and arrangement of hydrogen bonds, whether intra- or intermolecular. Consequently, these characteristics make the simulations of such O−H stretching region (3100–3800 cm^{-1}) a stringent test to compare populations derived from MD and conformational search-based methods.

Building on these ideas, we first develop two methodologies to evaluate the populations of non-bonded aggregates–one based on the analysis of MD trajectories and the other on conformational searches followed by a statistical thermodynamic treatment using model potential energy surfaces–and subsequently assess their accuracy by simulating the IR spectrum, focusing on the O−H stretching region. Our methodologies are applied to phenol/carbon tetrachloride mixtures, which presents marked changes in the O−H IR band with concentration. In these mixtures, phenol monomers coexist with various aggregates, whose distinct peaks in the 3000–3800 cm^{-1} range reflect the structure and conformational landscape of the system[58]. Self-association of phenol in CCl$_4$ indeed remains a challenging problem and, despite extensive experimental studies over the past decades[59–71], no general consensus exists regarding the degree of aggregation or the structures of the resulting clusters.

In this study, we analyze the populations of phenol clusters in solution through atomistic simulations, extending our analysis to tetramers. The two developed strategies are applied to compute the populations of phenol aggregates, which are eventually used to simulate the average IR spectra at varying phenol/CCl$_4$ ratios, focusing on the challenging O−H broadband. Our results offer valuable insights into the strengths and limitations of both approaches in capturing enthalpic and entropic contributions to aggregate formation.

This paper is organized as follows: Section 2 provides a detailed description of the newly developed approaches for determining aggregate populations and outlines the method for

https://doi.org/10.26434/chemrxiv-2025-70p06 ORCID: https://orcid.org/0000-0003-4820-4371 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

calculating averaged IR spectra. Section 3 presents the computational details. Section 4 discusses the main findings, and the conclusions are presented in Section 5.

2 Theoretical methods

This section begins by introducing the graph-theory-based approach used to identify equivalent conformers, which applies to both methodologies for determining aggregate populations. The two strategies, referred to as $\mathbf{Q/s}$ and $\mathbf{C/d}$, are then described in detail, eventually providing a concise comparison between them. Finally, we outline the procedure to compute IR spectra based on the populations derived from these two approaches.

2.1 Unique classification of conformers based on graph theory

In this study, we have developed and implemented a systematic procedure for identifying and classifying macromolecular aggregates based on the powerful algorithms of graph theory. Our method is primarily intended to analyze MD trajectories, which usually contain thousands of particles. Indeed, their analysis is far from straightforward, which explains why traditional MD analysis tools do not directly provide information on these aggregates that may appear for short periods of time during the simulation. Moreover, as we will show, the top-down nature of our approach makes it generally applicable to any set of configurations, for instance, those arising from a systematic conformational search.

A molecular graph is a 2D representation of a molecule where atoms and bonds are typically simplified as vertices and edges of the graph, respectively. Graphs can be represented in matrix form, leading to computationally efficient methods to extract information from the system. In this work, we resort to the adjacency matrix, which contains topological information, i.e., the covalent and non-covalent bonds within the atoms. The knowledge of this matrix at each snapshot ultimately provides us with the required information about the isolated molecular clusters present during the simulation.

In the initial step, the analysis of the adjacency matrix of the entire system determines the number of aggregates of each size, n. This is achieved by transforming the adjacency matrix of the system into a block-diagonal form, where each diagonal block represents the individual adjacency matrix of a distinct aggregate in the system. The dimension of this matrix determines the size of the aggregate. In this way, from the number of blocks of identical dimensions and the total number of blocks, we can calculate the population of aggregates of varying sizes.

For the sake of simplicity, we consider only the connectivity between the functional groups of the molecules, represented as the nodes in an undirected graph. In the case of phenol, we focus on the hydroxyl group (OH) and the phenyl group (π) of the monomer. Aggregate configurations, up to tetramers, are classified according to their interaction patterns, i.e., the topology according to the functional groups. Although interactions are determined based on the position (or center of mass) of atoms (or groups of atoms), we employ a simplified representation based on functional groups to classify the conformations explored during the simulation according to their interaction patterns in topologically distinct isomers. While using this simplified graphical representation does not retain information about the specific hydrogen donor and acceptor molecules forming a hydrogen bond, it offers a streamlined approach for evaluating the effectiveness and limitations of this clustering-like method in analyzing molecular configurations.

The interaction patterns considered for the phenol dimer with the associated molecular graphs are illustrated in Figure 1. Hereafter, we refer to the set of geometric configurations that can be associated with the same interaction pattern as conformers of a given isomer. One interaction (edge) between hydroxyl groups corresponds to a hydrogen bond (H-bond) between them. The presence of an edge between two phenyl groups can indicate either a pure π⋯π interaction between the aromatic rings or $C-H$⋯π interactions between a $C-H$ bond in one molecule and the π-electron cloud of an aromatic ring. Finally, OH-π bonded isomers correspond either to an H-bond between an aromatic hydrogen atom and the hydroxyl group

7

or a hydroxyl proton with the π-electron cloud.

Naturally, the number of possible molecular graphs increases substantially in higher-order clusters. To simplify our analysis, trimer and tetramer configurations are classified attending to the topology of the H-bonds among molecules. If no H-bond is present, they are categorized as either $\pi-\pi$ or OH-π bonded clusters, depending on whether any monomers interact solely through OH$\cdots\pi$ interactions or not. This results in five distinct interaction patterns for trimers and ten for tetramers, which are schematized in Figures S15 and S16 in the Supporting Information (SI) and listed in the first column of Table 1.
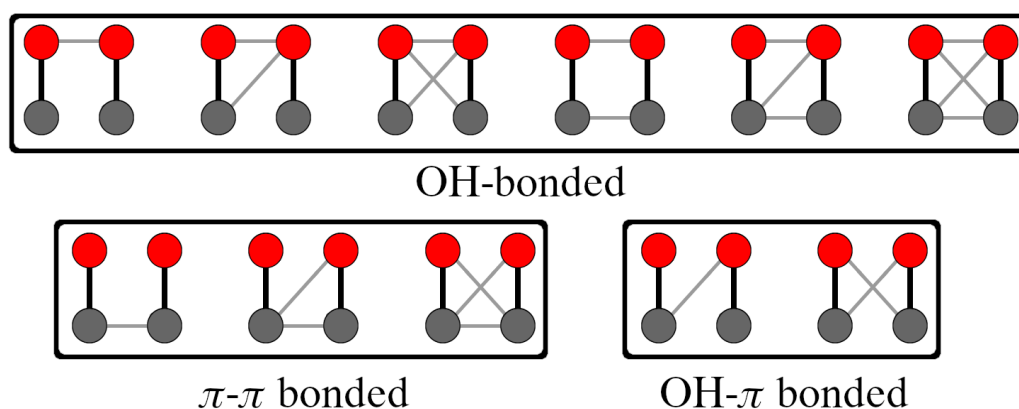

Figure 1: Molecular graphs used to classify interaction patterns in phenol dimers. Red and gray circles represent hydroxyl groups and phenyl groups, respectively. Black edges indicate the existence of covalent bonds between atoms within each node, while gray edges represent non-covalent interactions.

In our approach, intermolecular interactions are defined based on specific geometrical criteria (detailed information on the selection of the criteria used to determine the non-covalent interactions in the system is provided in Section A of the SI). Two phenol molecules are considered to form a hydrogen bond when the O$\cdots$H distance is less than 2.6 Å and the O$\cdots$H$-$O angle is between 135 and 180°. The π-electronic clouds of two phenyl groups are considered to interact when the distance between the centers of mass of the carbon atoms in each molecule is less than 4.0 Å. C$-$H$\cdots$O interactions are present if the H$\cdots$O distance is lower than 2.8 Å and the angle with the adjacent C atom is between 120 and 180°. Finally,

8

X−H⋯π interactions are defined when the H⋯π distance is lower than 3.0 Å with a C−H⋯π or O−H⋯π angle in the range of 120° to 180° or 135° to 180°, respectively. The effect on the populations of using different thresholds has been proven to be rather limited, as shown in the SI (Section D), which supports the robustness of our approach.

This protocol can be applied to the snapshots generated through MD simulations, which may include all or just part of the interaction patterns. The total population would be given by the average over all snapshots, which would generally include all interaction patterns for sufficiently long trajectories. The same protocol can be applied to an ensemble of configurations generated through a systematic conformational search. In this case, the number of interaction patterns identified will depend on the extent of the automated search. All algorithms described here have been implemented in `Aggregate`[72], a code developed by one of us to identify and classify interaction patterns.

2.2   The Quantum/static (Q/s) approach

The **Q/s** strategy starts from a systematic conformational search to locate all possible conformers for monomers to tetramers, which are subsequently analyzed by means of statical thermodynamics, in line with quantum cluster equilibrium theory[53–57] or the clusters-in-a-liquid framework[43]. Here, we used *AutoMeKin*[73] software, with some adaptations as explained in Section 3.2, to generate the conformational space. This results in a set of unique isomers for each aggregation size, containing 1 monomer, 38 dimers, 174 trimers and 144 tetramers. Each of them is optimized, and their second derivatives (frequencies) are subsequently computed at DFT/PCM level (see Section 3.1 for details).

All the aforementioned isomers may be grouped into macroscopically differentiable conformers. Specifically, we here adopt the interaction-pattern conformers identified with the graph-theory tool described in Section 2.1, `Aggregate`, in terms of hydrogen bonds, π-π interactions and π-OH contacts. Therefore, each interaction-pattern conformer includes, in general, several isomers and we account for such a multimolecular description of the con-

9

former treating the molecular system as an equilibrated mixture of all the minimum energy structures with a total free energy given by Equation 1 [31,74–76],

$$G\left(\{\mathbf{X}\}\right) = G\left(X_{ref}\right) - RT \ln \left[\sum_{X_i \in \{\mathbf{X}\}} m_i e^{-\frac{\Delta G\left(X_i\right)}{RT}}\right] \qquad (1)$$

where the sum runs over the set of all energy-minimized molecular conformations (isomers) that are characterized by a given interaction pattern, $\{\mathbf{X}\}$. In the previous expression, $\Delta G\left(X_i\right) = G\left(X_i\right) - G\left(X_{ref}\right)$, with X_{ref} being some arbitrarily chosen reference configuration and $G\left(X_i\right)$ the absolute free energy corresponding to the ith conformation. m_i is the number of equivalent energy wells in the PES of the system compatible with species i (conformers and rotamers), which can be obtained using either a direct counting method or automatic conformational search algorithms. When dealing with non-covalent complexes, this PES degeneracy must also account for the indistinguishability of monomer molecules [76].

At this point, we still have to determine the Gibbs free energy of each individual isomer $(G(X_i))$. This corresponds to the standard free energy at temperature T and pressure p, $G(X_i) \equiv G_i^*\left(T, p\right)$, of solute i, which is determined as the sum of the following terms [77]

$$G_i^*\left(T, p\right) = E_{i,gas}^{elec} + \Delta G_{i,solv}^*\left(T, p^\circ\right) + \delta G_{i,VRT}^* + RT \ln\left(\frac{M_{solv} C^\circ}{1000 \rho_{solv}\left(T, p\right)}\right) \qquad (2)$$

In this equation, $E_{i,gas}^{elec}$ is the gas phase total electronic energy of the solute, and $\Delta G_i^{solv}\left(T, p^\circ\right)$ its solvation Gibbs free energy (defined to transfer a solute from an ideal gas at a concentration of 1 mol L^{-1} to an ideal solution at a concentration of 1 mol L^{-1}). The third term of this equation is usually referred to as the vibrational-rotational-translational contribution to the gas-phase Gibbs free energy. It can be written as $\delta G_{i,VRT}^* = \delta G_{i,VRT}^\circ + RT \ln\left(\frac{RT}{p^\circ}\right)$, and it includes the free energy change associated to bring the solute from its ideal gas reference state at $p^\circ = 1$ bar to a density of unit molarity [78], *i.e.* a reference state concentration of 1 mol L^{-1}. Note that omitting this factor would lead to a systematic error of 1.89 kcal mol^{-1}

at 298.15 K to the absolute free energies[79]. The last term involves the density and, despite being systematically omitted in the literature, has a notable impact on the free energies at high temperatures, especially for nonaqueous solvents, as shown by Noroozi and Smith[78].

Equation (2) provides the absolute free energies in solution at the standard state of 1 mol L^{-1} used in this work. Thermal contributions to solution-phase free energies ($\delta G^* VRT$) are computed by referencing the partition function to the ground-state electronic energy, which inherently includes the zero-point vibrational energy (ZPE). These contributions are calculated using ideal gas partition functions under the quasi-rigid rotor harmonic oscillator (quasi-RRHO) approximation[80,81], employing scaled[82] solution-phase harmonic vibrational frequencies, with the temperature and pressure fixed at 298.15 K and 1 bar, respectively. The vibrational frequencies were scaled for thermal correction calculations using scaling factors derived in this work from experimentally measured fundamental frequencies of phenol monomer in the gas phase and in CCl_4 solution (see Section C of the SI).

Once the free energies of each interaction pattern, α, are evaluated, we can now obtain the populations, which we here define as the molar fraction over the total amount of the aggregate with the same size, n, as

$$P_{n\alpha}^{Q/s} = \frac{e^{-\dfrac{G_{n\alpha}}{RT}}}{\sum\limits_{\alpha} e^{-\dfrac{G_{n\alpha}}{RT}}} \tag{3}$$

where $G_{n\alpha}$ is computed according to Eq. (1) including all the isomers classified within the interaction pattern α.

Equation 3 provides the populations of each conformer relative to the aggregate of size n. The complete elucidation of the system composition thus requires the macroscopic concentration of each aggregate, which is computed from the self-association equilibria,

$$nPhOH \rightleftharpoons (PhOH)_n \tag{4}$$

The thermodynamic constant of the above equilibrium is calculated from the global free energy of the aggregate of size n,

$$K_n^{\mathbf{Q/s}} = e^{-\frac{\Delta G_n}{RT}} \tag{5}$$

where the free energies include the contribution of all interaction-pattern conformers of this size through Eq. (1).

As dictated by the self-aggregation equilibrium, Eq. 4, the concentrations of the monomer (C_1) and that of the n-th aggregate (C_n) at equilibrium must obey,

$$K_n = \frac{C_n \, (C^\circ)^{n-1}}{(C_1)^n} \tag{6}$$

where we have included the standard concentration, C°, which is equal to 1 mol L^{-1} and can be effectively omitted. In turn, the stoichiometric concentration of phenol, C_0, assuming that only aggregates up to tetramers exist, is given by $C_0 = \sum_{n=1}^{4} nC_n$. This mass-conservation condition, together with all three self-association equilibria, Eq. (6), complete a system of four equations whose solution provides the concentration of the monomer and the three aggregates.

2.3 The Classical/dynamic (C/d) approach

The $\mathbf{C/d}$ approach leverages the sampling provided by an MD simulation (details of the simulation are given in Section 3.3). Our novel methodology determines aggregate populations directly comparable to the $\mathbf{Q/s}$ approach by indentifying the same interaction pattern conformers also using the `Aggregate` tool.

Namely, the proposed protocol provides the number of conformational pattern conformers at each step, $N_{n\alpha j}$, from the block-diagonalization adjacency matrix and subsequent classification of the aggregate. The equilibrium concentrations are then computed according to

the following equation

$$\langle C_{n\alpha} \rangle = \frac{1}{N_{steps}} \sum_{j}^{N_{steps}} \frac{N_{n\alpha j}}{V_j} \frac{1}{N_A} \tag{7}$$

The population of the α-th interaction pattern is then given by

$$P_{n\alpha}^{\mathrm{C/d}} = \frac{\langle C_{n\alpha} \rangle}{\sum_{\alpha} \langle C_{n\alpha} \rangle} \tag{8}$$

The global concentration of each aggregate is, in this case, given by the sum over all the conformers of the same size, $\langle C_n \rangle = \sum_{n\alpha} \langle C_{n\alpha} \rangle$. Naturally, during the simulation, aggregation states larger than tetramers are also encountered, which will make the sum from the monomer to the tetramer lower than the stoichiometric phenol concentration, C_0. However, as their fraction in the mixture is below 5%, they can be safely disregarded, therfore keeping the same monomer-dimer-trimer-tetramer equilibrium model as in the $\mathbf{Q/s}$ case. In this sense, to maintain consistency with the $\mathbf{Q/s}$ protocol, the equilibrium concentrations considered here are not those measured directly from the MD simulation but rather those obtained from the mass-conservation law (assuming species up to the tetramer) and the self-association equilibria, Eq. 6. The required equilibrium constants are obtained directly from the ratio of the average equilibrium concentrations of the aggregates within the simulation box,

$$K_n^{\mathbf{C/d}} = \frac{\langle C_n \rangle \, (C_0)^{n-1}}{\langle C_1 \rangle^n} \tag{9}$$

In Table S7 of the SI, we report the concentrations computed directly from the MD simulations and those eventually used in this work. Although the overall trends are consistent, the MD populations reveal a noticeable presence of higher-order aggregates. In the most concentrated solution, these account for up to 15% of the total phenol population. Nevertheless, the molar concentration of these larger species remains a small fraction of all aggregates in solution. Indeed, our results suggest that beyond the tetramer, the influence of higher-order aggregates is quite limited. Since the focus of this study is on comparing the $\mathbf{Q/s}$ and $\mathbf{C/d}$

13

approaches, we have decided to maintain our simplified treatment, including only aggregates up to the tetramer. We acknowledge, however, that the presence of higher-order aggregates raises certain limitations, which could be addressed in future work by adopting generalized methodologies that randomly select individual phenol molecules and explicitly consider their surrounding environments.

2.4   Summary of the proposed methodologies

The two strategies described above are summarized in the diagram presented in Figure 2. On the one hand, the $\mathbf{C/d}$ approach (top) applies the `Aggregate` program to the snapshots generated along the MD trajectory. These snapshots typically contain aggregates of varying sizes, each spanning multiple interaction patterns. The `Aggregate` program identifies the size and interaction pattern of each aggregate, ultimately providing their populations. On the other hand, the $\mathbf{Q/s}$ approach (bottom), implies an initial systematic conformational search that generates isomers for aggregates of different sizes. The `Aggregate` program is subsequently run over these isomers to map them to their corresponding interaction-pattern conformers, and their populations are determined by accounting for the free energies of all isomers contributing to each interaction pattern.
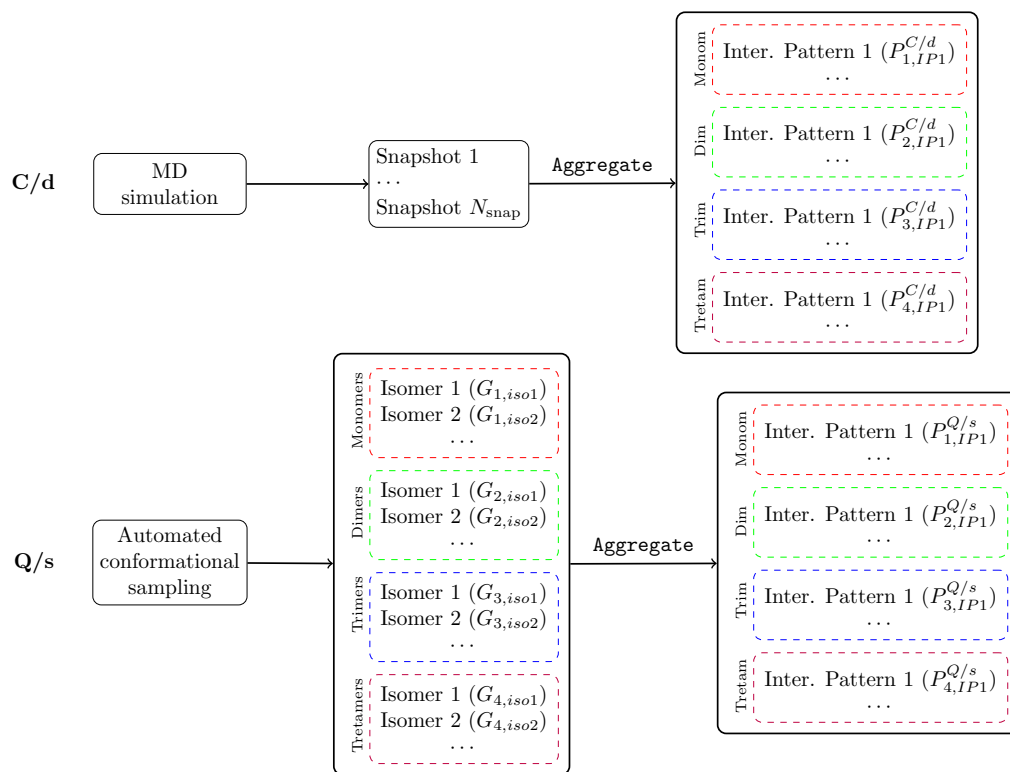
Figure 2: Diagram of the two protocols adopted in this work to obtain the populations of the different interaction-pattern conformers using the **C/d** (above) and **Q/s** (below) strategies.

It is important to emphasize that both approaches yield populations for the same set of interaction-pattern conformers, as identified by the `Aggregate` program. This ensures a consistent and reliable comparison between the populations derived from MD simulations (**C/d**) and those obtained via systematic conformational searches (**Q/s**).

Additionally, to maintain consistency, in all cases, we provide the populations for each interaction pattern corresponding to a specific aggregate size (up to tetramers). Then, the relative concentrations of monomers, dimers, trimers, and tetramers are computed from the equilibrium constants for aggregate formation. These constants are determined either from free energies (**Q/s** method, Eq. 5) or average populations (**C/d** method, Eq. 9).

15

2.5 Calculation of averaged IR spectra

Once the populations of the interaction-pattern conformers are obtained using both methodologies, their performance is assessed by evaluating the averaged broad IR band arising from the contributions of these conformers. A key step in this process is computing the IR spectra associated with each interaction-pattern conformer.

In this work, the IR spectra for each conformer are derived from the $\mathbf{Q}/\mathbf{s}$ approach, asuming that the flexibility within the interaction pattern is properly captured by this methodology. Specifically, the spectra are obtained by averaging the harmonic IR spectra of all isomers mapping to a given interaction-pattern conformer. Vibrational frequencies were scaled using the scaling factor computed in this work (0.96) to correct for errors stemming from anharmonicity and the inherent approximations of the DFT functional. Further details on the calculation of the scaling factor are provided in Section C of the SI. The Boltzmann-averaged molar absorption coefficient for the conformer α, of size n, $\overline{\varepsilon}_{n\alpha}(\omega)$, is computed as:

$$\overline{\varepsilon}_{n\alpha}(\omega) = \sum_i w_{n\alpha i}\varepsilon_{n\alpha i}(\omega) \tag{10}$$

where i runs over all isomers contributing to the interaction pattern, and the relative weight of each isomer, $w_{n\alpha i}$, is given by:

$$w_{n\alpha i} = \frac{e^{-\dfrac{G_{n\alpha i}}{RT}}}{\sum\limits_i e^{-\dfrac{G_{n\alpha i}}{RT}}} \tag{11}$$

These molar absorption coefficients are then used to compute the total absorbance of the sample by exploiting the additivity of the Beer–Lambert law. The total absorbance is the sum of contributions from all conformers, weighted by their populations and macroscopic concentrations. The absorbance spectrum is expressed as:

16

$$A(\omega) = l \sum_{n}^{4} C_n \sum_{\alpha}^{N_{IP}^{(n)}} P_{n\alpha} \bar{\varepsilon}_{n\alpha}(\omega) \tag{12}$$

Here, l is the cuvette path length, set to the experimental value[68]; C_n is the concentration of the n-th aggregate, related to the stoichiometric concentration via the self-association equilibria (Eq. 6), which accounts from the monomer ($n = 1$) to the tetramer ($n = 4$), and $N_{IP}^{(n)}$ represents the number of interaction-pattern conformers for each aggregate size. This approach enables a quantitative comparison with experimental data at different concentrations. To facilitate this comparison, the computed absorbance spectra are converted into transmittance to match the experimental format.

The individual IR spectra of conformers are convoluted using a Lorentzian function with a half-width at half-maximum (HWHM) of 10 cm⁻¹ for frequencies below 3200 cm⁻¹ and above 3650 cm⁻¹, a commonly adopted value. However, for frequencies in the 3200–3650 cm⁻¹ range, corresponding to the O−H stretching of phenol aggregates, a Gaussian function with a larger HWHM of 60 cm⁻¹ is applied. This increased broadening accounts for the inhomogeneous effects caused by the complex conformational flexibility of the aggregates. Frequencies below 3200 cm⁻¹ correspond to C−H stretching vibrations, while those above 3650 cm⁻¹ arise from free O−H stretching in bare phenol and chain- or branched-like aggregates. The large Gaussian broadening in the O−H region reflects the structural diversity of the aggregates, which induces significant inhomogeneous broadening.

3 Computational details

3.1 Electronic struture calculations

All calculations were performed using the B3LYP functional including D3 Grimme's dispersion correction for improved non-covalent interactions. Geometry optimizations were carried out using the N07D basis set, including the environment effect within the IEF-PCM[83] im-

17

plicit solvation model. Thermal corrections are computed from the geometries and frequencies obtained at this level of theory. Final energies are refined by increasing the basis set size (N07Tdiff), computing the electronic gas phase and solvation-free energy contributions in Eq. (2) independently. Moreover, the basis set superposition error was corrected using the counterpoise correction in the calculation of the total gas-phase energies. For the standard state correction in Eq. (2), we used the value of the experimental density of CCl_4.

3.2 Systematic conformational search

The PES of non-covalent complexes is characterized by the presence of many local, low-lying energy minima, which need to be taken into account effectively, as an insufficient conformational sampling could easily contribute several kcal mol^{-1} to the free-energy differences[76]. In this work, we have conducted such conformational sampling with *AutoMeKin* package[73], with some modifications to ensure an efficient and complete as possible search. We note that a reliable screening is essential to reduce as much as possible to the computation of vibrational frequencies for the clusters, which consume most of the computational time and represent by far the computational bottleneck of the entire procedure. The specific strategy employed to minimize the computational cost of conformational exploration is described in detail in the SI (Section B).

Of course, this modelling strategy based on partial optimizations can be coupled to other automated conformational search methods from the very beginning[9,84,85], or even MD simulations to obtain the candidate structures. In fact, after analyzing the MD trajectories, some interaction pattern isomers listed in Table 1 could not be identified using the procedure described above. To address this, we conducted a second round of conformational sampling *post hoc*, starting from a set of **C/d** structures classified as the missing conformers. Still, some isomers could not be optimized, but they can be safely ignored because their low populations make their overall contribution to the spectrum negligible. The effect of expanding the aggregate set on the **Q/s** populations and the computed IR spectra is discussed in the

18

SI (see Section E).

3.3 Molecular dynamics settings

The **C/d** approach involves conducting unbiased MD simulations of phenol in carbon tetrachloride mixtures that were conducted with Gromacs 5.0.7. The simulation box contains a mixture of both compounds, with approximately 11000 molecules in total. Specifically, we conducted MD simulations of three solutions with increasing solute concentrations (0.115, 0.45 and 1.1 M) incorporating 10000 solvent molecules and 111, 434 and 1060 solute molecules into a starting cubic simulation box with edges of 12.0, 12.5 and 13.0 nm respectively.

Over the simulation, bonds containing hydrogen atoms were constrained using the LINCS algorithm, allowing a time step of 1 fs for the integration of the equations of motion (using the leapfrog algorithm). Particle-mesh Ewald was considered for the calculation of long-range electrostatic forces, using a cutoff of 1.1 nm. The same cutoff was used for the short-range van der Waals forces. The general AMBER force field (GAFF) was used to model solvent and solute molecules. GAFF parameters were taken directly from virtualchemistry.org.

After constructing and minimizing the simulation box, the system is equilibrated by subsequent 2 ns of NVT simulation (using the modified Berendsen–velocity rescale–thermostat with a time constant for coupling of 0.1 ps to keep the temperature constant at 298.15 K) followed by 15 ns of NpT simulation. Production trajectories were then extended for a total of 100 ns, adopting a NpT ensemble at 298 K and 1 atm. The Nose-Hoover thermostat and the Parrinello-Rahman barostat were used in the NpT simulations with relaxation times of 1.0 and 5.0 ps, respectively. Configurations used in the subsequent analysis were extracted from the MD trajectory every 1 ps.

19

4 Results and discussion

In this work, we have deployed two alternative strategies to account for the populations of different phenol aggregates, i.e., the $\mathbf{Q/s}$ and $\mathbf{C/d}$ methodologies described in Sections 2.2 and 2.3. In the following sections, we first introduce and compare the populations of aggregates provided by each of them and eventually use them to compute the average IR spectra, focusing on the $O-H$ stretching region.

4.1 Conformational sampling and populations

The graph theory formalism adopted to identify the configurations of the aggregates provides a unified framework for analyzing the populations computed with both classical ($\mathbf{C/d}$) and quantum ($\mathbf{Q/s}$) methodologies. Namely, we characterize the different molecular configurations of aggregates based on connectivity patterns, allowing us to systematically classify aggregates of varying sizes (monomers, dimers, trimers, and tetramers), each with distinct interaction pattern conformers. These conformers, in turn, lead to different minima (isomers) that arise from the molecular flexibility.

In the $\mathbf{Q/s}$ approach, automated sampling generates many of these isomers, with all of them (and potentially more) appearing during MD sampling. However, tracking every possible isomer from the trajectory is infeasible. Instead, by using interaction-pattern conformers, which will generally encompass various minima in the $\mathbf{Q/s}$ approach and multiple configurations in the $\mathbf{C/d}$ MD sampling, we establish a common descriptor to classify configurations and enable a meaningful comparison of their populations with both methodologies.

As described in detail in the previous section, in the $\mathbf{Q/s}$ approach, the population of each conformer is derived from quantum mechanical partition functions, assuming a harmonic model around each minimum and summing the populations of all minima associated with the corresponding conformer. Conversely, in the $\mathbf{C/d}$ approach, the population is determined by counting the occurrences of the different conformers as identified through MD sampling.

The population of aggregates with a given size can then be determined by summing the populations of all conformers corresponding to each aggregate. This enables us to determine the populations from the monomer to the tetramer, which can then be compared with experimental results, offering an initial assessment of the performance of both $\mathbf{Q/s}$ and $\mathbf{C/d}$ approaches. Our findings are summarized in Figure 3, which presents the populations of monomers and aggregates (up to tetramers) in terms of the stoichiometric concentration of phenol (the equilibrium constants for the interconversion between aggregates corresponding to such populations are shown in Table S6 of the SI). Notably, the $\mathbf{Q/s}$ approach provides a continuous curve representing populations across concentrations, while the $\mathbf{C/d}$ approach is limited to the discrete systems simulated.

Our results clearly highlight the different performance of the methodologies. In both cases, the monomer is the most populated species; however, the relative populations of the different aggregates differ between the two strategies. Specifically, in the $\mathbf{C/d}$ approach, aggregate populations follow the order dimer > trimer > tetramer over the entire range of concentrations. In the $\mathbf{Q/s}$ approach, this order holds only at low concentrations, but the tetramer population increases more sharply, surpassing the trimer population at concentrations above approximately $0.1\,\mathrm{M}$ and the dimer population at concentrations above approximately $0.5\,\mathrm{M}$. Although no specific evidence on the relative populations of the aggregates is experimentally available, the monomer population derived from experiments[66] aligns more closely with the results from the $\mathbf{C/d}$ method. Experimental estimations are available[59–71], however, for the self-association equilibrium constants, which can be compared with those computed with $\mathbf{Q/s}$ and $\mathbf{C/d}$. Notably, equilibrium constants calculated with the $\mathbf{C/d}$ approach align better with experimental values. In contrast, those derived from $\mathbf{Q/s}$ consistently shift the self-association equilibria toward the monomer, compared to experimental findings. This explains the overpopulation of these species reported in Figure 3 with $\mathbf{Q/s}$. The equilibrium constants used to generate the distribution curves shown in Figure 3, along with their experimental counterparts, are provided in Table S6 of the SI.
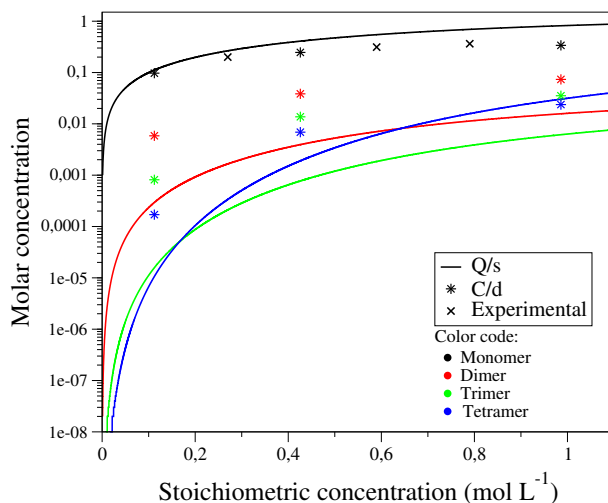
Figure 3: Overall population of aggregates (up to tetramers) with respect to phenol concentration, computed with $\mathbf{Q/s}$ (solid continuous lines) and $\mathbf{C/d}$ approach (stars). The experimental populations for the monomer from Schaefer *et al.*[66] are included for reference (crosses).

We now turn our attention to the relative distribution of interaction-pattern conformers for each aggregate. Table 1 lists all such interaction patterns in the first column, along with the populations of each species obtained using both $\mathbf{Q/s}$ and $\mathbf{C/d}$ approaches, allowing for a more detailed comparison of the performance of each strategy.

For dimers, the populations of different conformers are relatively similar across both methods, with the O−H bonded conformer being the most populated in both cases. However, notable differences emerge as aggregate size increases. For trimers, the most populated conformer changes from dimer-like with $\mathbf{Q/s}$ to linear with $\mathbf{C/d}$. This trend, favoring linear conformers in the $\mathbf{C/d}$ approach, continues with tetramers, where linear conformers remain the most populated with $\mathbf{C/d}$, whereas the cyclic conformer becomes dominant with $\mathbf{Q/s}$.

Notably, the cyclic conformer represents one of the most striking differences between the two methodologies: it is significantly populated in both trimer and tetramer aggregates with $\mathbf{Q/s}$, while $\mathbf{C/d}$ predicts a relatively low population. This high population of cyclic structures observed with $\mathbf{Q/s}$ may, in fact, be overestimated, as previously noted in similar studies[52]. The relative accuracy of each strategy will be evaluated further in the next section,

where these populations are used to simulate the broad IR O−H stretching band of phenol as a function of concentration.

Table 1: Populations of the different conformers computed with QM and classical MD methodologies.

	Dimer[a]	
	C/d	**Q/s**
OH-bonded	68.65	51.00
$\pi - \pi$	8.58	33.43
OH - π	22.76	15.56
	Trimer[b]	
	C/d	**Q/s**
Dimer-like	41.42	55.78
Linear	48.67	23.82
Cyclic	1.40	17.79
$\pi - \pi$ bonded	4.13	2.55
OH-π bonded	4.41	-
	Tetramer[b]	
	C/d	**Q/s**
Dimer-like	15.41	0.18
Double dimer	10.59	3.67
Trimer-like linear	26.80	5.24
Trimer-like cycle	0.47	0.11
Star	0.38	0.04
Linear	34.53	32.32
Cocktail	0.09	3.51
Cyclic	9.68	54.96
$\pi - \pi$ bonded	1.30	0.01
OH-π bonded	0.75	-

[a] Computed averaging the populations of dimer conformations of every MD simulation.

[b] Computed averaging the populations of the aggregate conformations of the two MD simulations at higher concentration.,

23

4.2 Simulation of the IR spectrum

We focus on simulating the OH region of the IR spectra for the mixtures. The spectral shapes in this region arise from the variety of aggregates formed, each exhibiting distinct IR features, whose populations depend on the relative concentrations of phenol and carbon tetrachloride, as shown in the previous section. This makes calculating this spectral region particularly challenging, as it requires careful consideration of the interplay of non-bonding interactions, especially hydrogen bonding, that influence the relative populations of the aggregates[52]. For this reason, such a simulation serves as a valuable benchmark to assess the accuracy of the populations computed using the two methodologies proposed in this work, $\mathbf{Q/s}$ and $\mathbf{C/d}$.

To isolate the effects of relative populations, we evaluate the spectra from a weighted average over the interaction-pattern conformers, using identical individual spectra for each of them. In turn, these individual spectra are computed as weighted averages of the IR spectra for each optimized isomer, calculated at the DFT level, with relative weights provided by the $\mathbf{Q/s}$ method. The spectra for each conformer are included in Section H of the SI.

Let us begin by analyzing the shape and position of the different aggregates displayed in Figure 4 for both $\mathbf{Q/s}$ and $\mathbf{C/d}$ strategies. In general, we find the expected behavior[58,86–88], with the position of the band going towards lower frequencies as the size of the aggregate increases. It should be noted, however, that the average bands for each aggregate generally consist of a complex superposition of different conformers. In this sense, since not all conformers are solely dictated by hydrogen bonds, e.g., those with π-π stacking interactions, there are some features that lie at higher frequencies. Moreover, it is worth noting that the shape of each aggregate also depends on the adopted strategy. Namely, $\mathbf{C/d}$ generally results in broader spectra, suggesting the inclusion of a wider variety of species in the average, which follows the distribution of relative populations displayed in Table 1.

We continue by identifying the contributions of monomers, dimers, trimers, and tetramers to the final spectra. It is noteworthy that the spectra for each aggregate differ not only in intensity between methods but also in shape. The relative weight of each aggregate aligns

24

with the populations shown in Figure 3. For example, at the lowest concentration, the spectrum is clearly dominated by the monomer in the $\mathbf{Q/s}$ method, while the contribution of the dimer, although lower than the monomer, is still notable in the $\mathbf{C/d}$ approach.

As the phenol concentration increases to 0.450 M, $\mathbf{Q/s}$ reveals a predominant contribution of the tetramer, even if the monomer remains as major component. In contrast, $\mathbf{C/d}$ shows a more balanced contribution across the different aggregates. Finally, in the most concentrated sample, the tetramer band is the most intense with $\mathbf{Q/s}$, while $\mathbf{C/d}$ provides again a more evenly distributed contribution among the aggregates, with an increased contribution of the tetramer compared to lower concentrations.
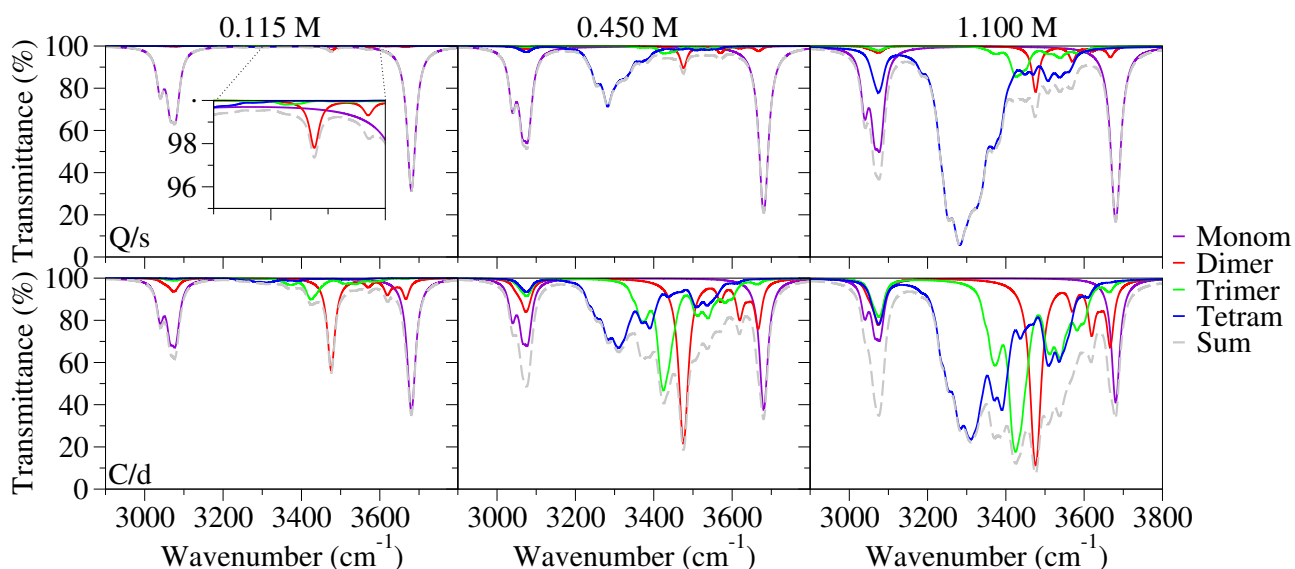

Figure 4: IR spectrum of phenol-tetrachloride mixtures at different concentrations, simulated through $\mathbf{C/d}$ (bottom panels) and $\mathbf{Q/s}$ (top panels) strategies. The contributions of monomers to tetramers are depicted individually. Inset: zoom of the plot to make evident the small bands appearing in the intermediate region in the case of a concentration of 0.115 M with the $\mathbf{Q/s}$ approach.

The spectra in Figure 4 display a complex band structure with numerous well-resolved peaks in the 3150–3600 cm^{-1} region. This effect, observed with both the $\mathbf{Q/s}$ and especially the $\mathbf{C/d}$ strategies, contrasts with the broad, structureless band seen in experimental data[68]. This discrepancy likely arises from conformational flexibility associated with a relatively

25

flat potential surface and the sensitivity of OH stretching frequencies to hydrogen bonding in this region. Accurately capturing this effect would require incorporating contributions from out-of-equilibrium structures, which, while accessible through the $\mathbf{C/d}$ protocol, lies beyond the scope of this study. Instead, we address this phenomenon phenomenologically by applying larger Gaussian broadening specifically in this region to account for inhomogeneous broadening. This adjustment results in smoother curves that more closely resemble the experimental band. Specifically, we apply Gaussian broadening with a half-width at half maximum (HWHM) of 60 cm^{-1} in the 3150–3600 cm^{-1} region while retaining a Lorentzian broadening with HWHM of 10 cm^{-1} in the rest of the spectral region.

The resulting spectra, simulated with both $\mathbf{Q/s}$ and $\mathbf{C/d}$, alongside the experimental one[68], are shown in Figure 5. Focusing first on the experimental data, it is evident that for all concentrations, there are a number of bands that remain nearly unchanged. These can be associated with the monomer, at approximately 3600 cm^{-1}, and CH stretching vibrations (the double-peaked band around 3000 cm^{-1}). Both of these features are identifiable in the simulated spectra with $\mathbf{Q/s}$ and $\mathbf{C/d}$ methods, with slight deviations in position that can be attributed to the usual inaccuracies in DFT-computed frequencies at the harmonic level.

The main concentration-dependent changes in the spectrum occur in the intermediate region between these relatively stable bands. Signals in this region correspond to the aggregates as discussed above, and in the following, we will focus our discussion on them, comparing the experiments with the theoretical simulations at each concentration.

Starting with the most diluted sample, the experimental spectrum shows a band in the intermediate region located just below the band of the monomer. While appreciable, this band is significantly less intense than the monomer one. In the simulations, the $\mathbf{Q/s}$ method does not produce any significant signal in this region, whereas $\mathbf{C/d}$ displays a small contribution from the dimer, which is consistent with the experimental data, albeit shifted to lower frequencies.

Moving to the intermediate concentration, the experimental band in the intermediate

region broadens and increases in intensity. In this case, the $\mathbf{Q/s}$ approach produces a signal mainly associated with the tetramer, which is again notably less intense than the experimental band. In contrast, the $\mathbf{C/d}$ method is able to capture the broadening and intensity increase in this intermediate region, resulting in a simulated band that aligns quite well with the experimental data, although it is shifted to lower frequencies, suggesting a possible underestimation of tetramer-related bands.

Finally, for the most concentrated sample, the experimental band in the intermediate region shows a moderate increase in intensity compared to the previous concentration, but its shape remains mostly the same, with a slight enhancement on the low-frequency side. The $\mathbf{Q/s}$ method displays a significant increase in intensity, dominated by the tetramer-related band. Although the intensity is similar to the experiment, the band is narrower and overly centered on the tetramer. In contrast, the $\mathbf{C/d}$ method provides a simulated band that follows the experimental trend, showing an increase in intensity compared to the previous concentration. Regarding the shape, there is a change induced by the increased contribution of the tetramer, resulting in a band that aligns even more closely with the experimental data.

Following the analysis of the OH region in the IR spectrum (Figure 5), it is evident that the $\mathbf{C/d}$ strategy outperforms the $\mathbf{Q/s}$ approach, indicating that the populations derived from the classical MD trajectory better reflect the experimental behavior. But which factors account for the superiority of $\mathbf{C/d}$ populations over those from $\mathbf{Q/s}$? To address this question, we need to consider the two thermodynamic contributions to population distributions: the enthalpic and entropic terms associated with the formation of aggregate.

The enthalpic contribution is primarily dictated by intermolecular interactions, including hydrogen bonds and π-π interactions. In this sense, it is expected that the dispersion-corrected DFT calculations in the $\mathbf{Q/s}$ approach will outperform the force field calculations applied in $\mathbf{C/d}$. Indeed, the inaccuracies in the FF could partly explain some of the deficiencies in $\mathbf{C/d}$, such as its apparent underestimation of tetramer populations. Therefore, enthalpy differences cannot explain the better performance of the $\mathbf{C/d}$ strategy. The key
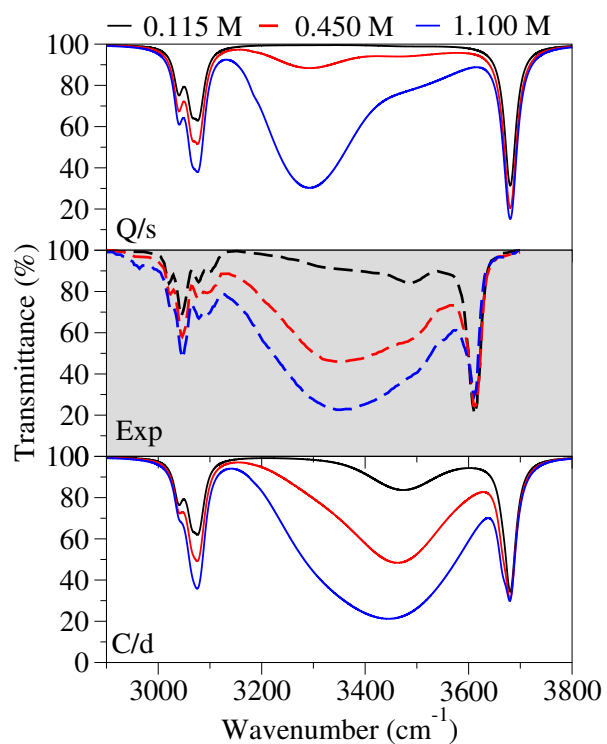
Figure 5: IR spectrum of phenol-tetrachloride mixtures at different concentrations (0.115, 0.450 and 1.100 M), simulated through $\mathbf{C/d}$ (top panel) and $\mathbf{Q/s}$ (bottom panel) strategies, with a phenomenological broadening in the O−H region using a Gaussian with a HWHM of $60\,\mathrm{cm^{-1}}$. The experimental spectra for these concentrations are included in the central panel.

28

to understand this difference would necessarily lie in the entropic term, which is linked to the vast number of configurations arising from the significant molecular flexibility in loosely bound aggregates. This implies that many configurations are necessary to achieve population distributions achieved by the experimental system, which becomes unmanageable for the automated conformational search of the $\mathbf{Q/s}$ approach. Conversely, this conformational flexibility is, by construction, effectively captured by the MD simulation adopted in the $\mathbf{C/d}$ strategy.

The entropy-based analysis further elucidates the prevalence of cyclic conformers in the $\mathbf{Q/s}$ approach, whereas linear conformers are more common with the $\mathbf{C/d}$ strategy, as shown in Table 1. The significantly larger number of configurations available to linear conformers makes it challenging to identify all possible configurations via the automated search used in the $\mathbf{Q/s}$ approach. This limitation leads to an underestimation of the entropic contributions that favor linear conformers over cyclic ones, which, while being enthalpically favorable, display much lower entropy.

It is important to note that once an optimized configuration of an isomer is identified in the $\mathbf{Q/s}$ method, conformational motions around that configuration are included through the partition function, which is based on harmonic model surfaces and a quantum mechanical treatment of vibrations. For intramolecular vibrations, the entropic terms are better described by this quantum description, compared to the classical treatment of MD trajectories in the $\mathbf{C/d}$ approach. However, aggregate formation is primarily governed by intermolecular non-bonded interactions, which lead to loose nuclear motions that are effectively described at the classical level. For the study of non-bonded species, the $\mathbf{C/d}$ strategy thus emerges as a very suitable approach.

Moreover, the $\mathbf{C/d}$ approach demonstrates a remarkably good agreement with experimental data. This is a notable achievement, as the shape of the OH band results from a complex interplay of non-bonding interactions governing aggregate formation, which in turn defines the band structure.

29

5   Conclusions

In this study, we have introduced and evaluated two distinct methodologies for sampling the nuclear degrees of freedom in a system governed primarily by non-bonded interactions: a liquid mixture of phenol and carbon tetrachloride. Our aim was to determine which protocol provides a more reliable description of the aggregate populations. Subsequently, we assess their performance by comparing the average IR spectra, computed with the populations resulting from each methodology, against the experimental one at varying concentrations, particularly focusing on the OH stretching region.

The first method employs a systematic conformational search, followed by a potential energy surface (PES) expansion around each identified structure. This approach, which we termed $\mathbf{Q}/\mathbf{s}$, computes the nuclear partition function by assuming a harmonic PES and treats vibrational modes at a quantum mechanical level (with rotational and translational contributions considered semiclassically and classically, respectively). The second method, termed $\mathbf{C}/\mathbf{d}$, calculates populations by conducting a classical molecular dynamics (MD) simulation, thus treating nuclear motion fully classically.

A rigorous framework was essential to identify and categorize aggregates consistently across both methodologies. We accomplished this through a graph-theory classification based on non-bonded interaction patterns that define each aggregate, including hydrogen bonds, π-π stacking and π-OH bonds. The $\mathbf{Q}/\mathbf{s}$ approach identifies several optimized structures for each conformer during the automated search, capturing some degree of conformational flexibility. The $\mathbf{C}/\mathbf{d}$ approach inherently includes this flexibility through the MD trajectories.

Upon comparing the aggregate populations obtained with the $\mathbf{Q}/\mathbf{s}$ and $\mathbf{C}/\mathbf{d}$ approaches, we found that they align closely for smaller aggregates, such as dimers, but diverge significantly as aggregate size increases. Specifically, $\mathbf{Q}/\mathbf{s}$ tends to favor cyclic conformers, while $\mathbf{C}/\mathbf{d}$ predicts a predominance of linear aggregates. This divergence arises due to the different entropic and enthalpic contributions captured by each method: while $\mathbf{Q}/\mathbf{s}$ benefits from quantum mechanical accuracy to evaluate the interactions, it underrepresents the vast

configurational entropy associated with loosely bound aggregates due to the limitations of systematic conformational search. In contrast, $\mathbf{C}/\mathbf{d}$ captures a more realistic entropic profile by exploring a broader ensemble of conformers through MD, even though it approximates intermolecular interactions using a molecular mechanics force field.

The ability of the $\mathbf{Q}/\mathbf{s}$ method to align with experimental observations is limited by the extent of its conformational search. While an exhaustive search could yield results closer to those of $\mathbf{C}/\mathbf{d}$, non-bonded aggregates exhibit high configurational entropy, imposing practical limits on systematic search-based approaches. Aggregates with such loosely bound, flexible structures inherently challenge even sophisticated automated search techniques, as capturing their full entropy becomes infeasible. Our findings thus indicate that the $\mathbf{C}/\mathbf{d}$ approach provides a more reliable representation of aggregate populations in this system, as demonstrated by its superior performance in simulating IR spectral regions sensitive to non-bonded interactions, such as the broad OH stretching band.

To achieve such complete conformational sampling within $\mathbf{Q}/\mathbf{s}$-based strategies, we recommend carrying out the screening of the structures not solely based on energy criteria (relative population of the aggregates) but an entropy-like criteria based on the convergence of the multi-conformational contributions to the free-energy, or equivalently, the conformational entropy[89]. Moreover, when possible, this convergence must be achieved for each isomer group. Naturally, this is associated with a significant increase in computational time, particularly for non-covalent complexes, where large basis sets or the counterpoise correction are required to mitigate basis set superposition error.

It is noteworthy that the $\mathbf{C}/\mathbf{d}$ approach not only outperforms $\mathbf{Q}/\mathbf{s}$ but also achieves impressive absolute accuracy. This method closely reproduces the experimental trends as phenol concentration changes. This is indeed a challenging task, given that the observed broadband reflects the complex network of non-bonding interactions that give rise to different populations of conformers as the concentrations increase. With minor deviations, $\mathbf{C}/\mathbf{d}$ captures these experimental variations nearly quantitatively. In essence, its capacity to

31

handle loosely bound, entropically driven conformers makes the $\mathbf{C/d}$ approach a compelling choice for systems governed by weak, non-covalent interactions. Of course, it is also possible to adapt $\mathbf{Q/s}$ approaches to incorporate the entropic effects in an effective manner, e.g. through Bayesian inference[52]. In this context, our results can also be helpful to improve such methodologies.

The observed deviations with the $\mathbf{C/d}$ approach are likely due, in part, to inaccuracies in the FF used to evaluate intermolecular interactions. In this sense, this system may serve as an effective benchmark to assess refined intermolecular FF quality. In this study, we adopted standard GAFF parameters for the intermolecular potential, though we are currently pursuing refined parameterizations based on quantum mechanical (QM) calculations, as implemented in Picky[7,90]. These adjustments may further improve agreement with experimental data. Moreover, another source of potential inaccuracies arises from the presence of aggregates larger than the tetramer, whose influence increases with concentration. One route for improving the simulations involves extending the $\mathbf{C/d}$ approach to avoid strict reliance on predefined interaction patterns, e.g., by randomly selecting individual phenol molecules and subsequently incorporating their environment in the analysis. This strategy is currently being investigated by our group.

Another limitation is that the method does not directly account for the inhomogeneous broadening that affects the region where aggregate signals emerge. In this work, we included this effect phenomenologically, adding specific Gaussian broadening to this region to mimic the experimental band closely. An alternative to incorporate such effects from first principles could involve mixed quantum-classical strategies similar to those used in electronic spectroscopy[91], where the QM spectra of stiff solute degrees of freedom are computed across configurations generated via MD simulations. For IR spectra, this would require spectral frequency calculations at non-stationary points, which is feasible today using curvilinear internal coordinates[92]. These strategies are currently being explored to achieve a fully ab initio simulation of the challenging broad OH IR band.

Acknowledgement

Supporting Information Available

Comprehensive explanation of the thresholds used to define non-covalent interactions and the interaction patterns employed to classify aggregate conformations. Details on the determination of the scaling factors for harmonic frequencies used in this work. Examination of the dependence of $\mathbf{C}/\mathbf{d}$ populations and computed spectra on the thresholds and broadening parameter (HWHM), as well as the convergence of $\mathbf{Q}/\mathbf{s}$ populations and both sets of spectra with the chosen QM set and HWHM parameter. Concentration measurements from the MD simulations, contributions to the free energy of QM-optimized conformers, and a comparison of the calculated equilibrium constants with values reported in the literature.

References

(1) Jena, S.; Dutta, J.; Tulsiyan, K. D.; Sahu, A. K.; Choudhury, S. S.; Biswal, H. S. Noncovalent interactions in proteins and nucleic acids: beyond hydrogen bonding and π-stacking. *Chem. Soc. Rev.* **2022**, *51*, 4261–4286.

(2) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(3) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for

intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

(4) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094, PMID: 16092826.

(5) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122–4139, PMID: 31136175.

(6) Inakollu, V. S.; Geerke, D. P.; Rowley, C. N.; Yu, H. Polarisable force fields: what do they add in biomolecular simulations? *Curr. Opin. Struct. Biol.* **2020**, *61*, 182–190, Theory and Simulation, Macromolecular Assemblies.

(7) Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-Corrected DFT Data: The Benzene Dimer as a Prototype for Aromatic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 5182–5196, PMID: 26574315.

(8) Ferro-Costas, D.; Fernández-Ramos, A. New computational tools for chemical kinetics: the Cathedral Package. *Theor. Chem. Acc.* **2023**, *142*, 76.

(9) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST-A program for the exploration of low-energy molecular chemical space. *J. Chem. Phys.* **2024**, *160*, 114110.

(10) Kosztolányi, T.; Bakó, I.; Pálinkás, G. Hydrogen bonding in liquid methanol, methylamine, and methanethiol studied by molecular-dynamics simulations. *J. Chem. Phys.* **2003**, *118*, 4546–4555.

(11) Hamad, S.; Hughes, C. E.; Catlow, C. R. A.; Harris, K. D. M. Clustering of Glycine

Molecules in Aqueous Solution Studied by Molecular Dynamics Simulation. *J. Phys. Chem. B* **2008**, *112*, 7280–7288, PMID: 18503273.

(12) Bernardes, C. E. S. AGGREGATES: Finding structures in simulation results of solutions. *J. Comput. Chem.* **2017**, *38*, 753–765.

(13) Choi, S.; Parameswaran, S.; Choi, J.-H. Understanding alcohol aggregates and the water hydrogen bond network towards miscibility in alcohol solutions: graph theoretical analysis. *Phys. Chem. Chem. Phys.* **2020**, *22*, 17181–17195.

(14) Saiz, L.; Padró, J. A.; Guàrdia, E. Structure and Dynamics of Liquid Ethanol. *J. Phys. Chem. B* **1997**, *101*, 78–86.

(15) Gavezzotti, A. Molecular Aggregation of Acetic Acid in a Carbon Tetrachloride Solution: A Molecular Dynamics Study with a View to Crystal Nucleation. *Chem. Eur. J.* **1999**, *5*, 567–576.

(16) Khurana, E.; Nielsen, S. O.; Ensing, B.; Klein, M. L. Self-Assembling Cyclic Peptides: Molecular Dynamics Studies of Dimers in Polar and Nonpolar Solvents. *J. Phys. Chem. B* **2006**, *110*, 18965–18972, PMID: 16986891.

(17) Chen, J.; Trout, B. L. Computational Study of Solvent Effects on the Molecular Self-Assembly of Tetrolic Acid in Solution and Implications for the Polymorph Formed from Crystallization. *J. Phys. Chem. B* **2008**, *112*, 7794–7802, PMID: 18529073.

(18) Gaines, E.; Maisuria, K.; Di Tommaso, D. The role of solvent in the self-assembly of m-aminobenzoic acid: a density functional theory and molecular dynamics study. *CrystEngComm* **2016**, *18*, 2937–2948.

(19) Gaines, E.; Di Tommaso, D. Solvation and Aggregation of Meta-Aminobenzoic Acid in Water: Density Functional Theory and Molecular Dynamics Study. *Pharmaceutics* **2018**, *10*.

(20) Tanaka, H.; Nakanishi, K.; Touhara, H. Computer experiments on aqueous solutions. VII. Potential energy function for urea dimer and molecular dynamics calculation of 8 mol % aqueous solution of urea. *J. Chem. Phys.* **1985**, *82*, 5184–5191.

(21) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. Kinetic Insights into the Role of the Solvent in the Polymorphism of 5-Fluorouracil from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 3323–3329, PMID: 16494346.

(22) Chaiwongwattana, S.; Sagarik, K. Structures and dynamics of phenol clusters in benzene solutions. *Chem. Phys.* **2009**, *355*, 103.

(23) Yani, Y.; Chow, P. S.; Tan, R. B. H. Glycine Open Dimers in Solution: New Insights into α-Glycine Nucleation and Growth. *Crystal Growth & Design* **2012**, *12*, 4771–4778.

(24) Gavezzotti, A.; Filippini, G.; Kroon, J.; van Eijck, B. P.; Klewinghaus, P. The Crystal Polymorphism of Tetrolic Acid (CH3C CCOOH): A Molecular Dynamics Study of Precursors in Solution, and a Crystal Structure Generation. *Chem. Eur. J.* **1997**, *3*, 893–899.

(25) Ramondo, F.; Bencivenni, L.; Caminiti, R.; Pieretti, A.; Gontrani, L. Dimerisation of urea in water solution: a quantum mechanical investigation. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2206–2215.

(26) Barth, D.; Bougueroua, S.; Gaigeot, M.-P.; Quessette, F.; Spezia, R.; Vial, S. A New Graph Algorithm for the Analysis of Conformational Dynamics of Molecules. Information Sciences and Systems 2015. Cham, 2016; pp 319–326.

(27) Bougueroua, S.; Spezia, R.; Pezzotti, S.; Vial, S.; Quessette, F.; Barth, D.; Gaigeot, M.-P. Graph theory for automatic structural recognition in molecular dynamics simulations. *J. Chem. Phys.* **2018**, *149*, 184102.

(28) Aida, M.; Akase, D. Hydrogen-bond pattern to characterize water network. *Pure Appl. Chem.* **2019**, *91*, 301–316.

(29) Leal, A. M. M.; Kulik, D. A.; Smith, W. R.; Saar, M. O. An overview of computational methods for chemical equilibrium and kinetic calculations for geochemical and reactive transport modeling. *Pure Appl. Chem.* **2017**, *89*, 597–643.

(30) Smith, W. R.; Qi, W. Molecular Simulation of Chemical Reaction Equilibrium by Computationally Efficient Free Energy Minimization. *ACS Cent. Sci.* **2018**, *4*, 1185–1193, PMID: 30276252.

(31) Noroozi, J.; Smith, W. R. An Efficient Molecular Simulation Methodology for Chemical Reaction Equilibria in Electrolyte Solutions: Application to CO2 Reactive Absorption. *J. Phys. Chem. A* **2019**, *123*, 4074–4086, PMID: 30950623.

(32) Watanabe, H.; Iwata, S. Theoretical studies of geometric structures of phenol-water clusters and their infrared absorption spectra in the O-H stretching region. *J. Chem. Phys.* **1996**, *105*, 420–431.

(33) Chaudret, R.; de Courcy, B.; Contreras-García, J.; Gloaguen, E.; Zehnacker-Rentien, A.; Mons, M.; Piquemal, J.-P. Unraveling non-covalent interactions within flexible biomolecules: from electron density topology to gas phase spectroscopy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9876–9891.

(34) Ludwig, R.; Reis, O.; Winter, R.; Weinhold, F.; Farrar, T. C. Quantum Cluster Equilibrium Theory of Liquids: Temperature Dependence of Hydrogen Bonding in Liquid N-Methylacetamide Studied by IR Spectra. *J. Phys. Chem. B* **1998**, *102*, 9312–9318.

(35) Blasius, J.; Kirchner, B. Cluster-Weighting in Bulk Phase Vibrational Circular Dichroism. *J. Phys. Chem. B* **2020**, *124*, 7272–7283.

(36) Malloum, A.; Fifen, J. J.; Conradie, J. Exploration of the potential energy surfaces of small ethanol clusters. *Phys. Chem. Chem. Phys.* **2020**, *22*, 13201–13213.

(37) Malloum, A.; Fifen, J. J.; Conradie, J. Theoretical infrared spectrum of the ethanol hexamer. *Int. J. Quantum Chem.* **2020**, *120*, e26234.

(38) Kirchner, B.; Blasius, J.; Esser, L.; Reckien, W. Predicting Vibrational Spectroscopy for Flexible Molecules and Molecules with Non-Idle Environments. *Adv. Theory Simul.* **2021**, *4*, 2000223.

(39) Teh, S.; Hsu, P.-J.; Kuo, J.-L. Size of the hydrogen bond network in liquid methanol: a quantum cluster equilibrium model with extensive structure search. *Phys. Chem. Chem. Phys.* **2021**, *23*, 9166–9175.

(40) Malloum, A.; Dhaouadi, Z.; Conradie, J. Quantum cluster equilibrium prediction of liquid ethanol. *J. Mol. Liq.* **2023**, *387*, 122562.

(41) Malloum, A.; Conradie, J. Structures of DMSO clusters and quantum cluster equilibrium (QCE). *J. Mol. Graph. Model.* **2024**, *126*, 108661.

(42) Maya, J.; Malloum, A.; Fifen, J. J.; Dhaouadi, Z.; Fouda, H. P. E.; Conradie, J. Quantum cluster equilibrium theory applied to liquid ammonia. *J. Comput. Chem.* **2024**, *45*, 1279–1288.

(43) Perera, A. S.; Thomas, J.; Poopari, M. R.; Xu, Y. The Clusters-in-a-Liquid Approach for Solvation: New Insights from the Conformer Specific Gas Phase Spectroscopy and Vibrational Optical Activity Spectroscopy. *Front. Chem.* **2016**, *4*.

(44) Perera, A. S.; Cheramy, J.; Merten, C.; Thomas, J.; Xu, Y. IR, Raman, and Vibrational Optical Activity Spectra of Methyl Glycidate in Chloroform and Water: The Clusters-in-a-liquid Solvation Model. *ChemPhysChem* **2018**, *19*, 2234–2242.

(45) Le Barbu-Debus, K.; Scherrer, A.; Bouchet, A.; Sebastiani, D.; Vuilleumier, R.; Zehnacker, A. Effect of puckering motion and hydrogen bond formation on the vibrational circular dichroism spectrum of a flexible molecule: the case of (S)-1-indanol. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14635–14646.

(46) Le Barbu-Debus, K.; Bowles, J.; Jähnigen, S.; Clavaguéra, C.; Calvo, F.; Vuilleumier, R.; Zehnacker, A. Assessing cluster models of solvation for the description of vibrational circular dichroism spectra: synergy between static and dynamic approaches. *Phys. Chem. Chem. Phys.* **2020**, *22*, 26047–26068.

(47) Le Barbu-Debus, K.; Zehnacker, A. Competition between inter and intramolecular hydrogen bond evidenced by vibrational circular dichroism spectroscopy: The case of (1S,2R)-(−)-cis-1-amino-2-indanol. *Chirality* **2021**, *33*, 858–874.

(48) Kutsyk, A.; Ilchenko, O.; Pilhun, Y.; Nikonova, V.; Obukhovsky, V. Complex formation in methanol-chloroform solutions: Vibrational spectroscopy and quantum cluster equilibrium study. *J. Mol. Liq.* **2022**, *367*, 120499.

(49) Yang, Y.; Cheramy, J.; Brehm, M.; Xu, Y. Raman Optical Activity of N-Acetyl-L-Cysteine in Water and in Methanol: The "Clusters-in-a-Liquid" Model and ab Initio Molecular Dynamics Simulations. *ChemPhysChem* **2022**, *23*, e202200161.

(50) Puente, A. R.; Polavarapu, P. L. Influence of microsolvation on vibrational circular dichroism spectra in dimethyl sulfoxide solvent: A Bottom-Up approach using Quantum cluster growth. *Spectrochim. Acta Part A* **2023**, *303*, 123231.

(51) Perera, A. S.; Carlson, C. D.; Cheramy, J.; Xu, Y. Infrared and vibrational circular dichroism spectra of methyl β-D-glucopyranose in water: The application of the quantum cluster growth and clusters-in-a-liquid solvation models. *Chirality* **2023**, *35*, 718–731.

(52) Blasius, J.; Drysch, K.; Pilz, F. H.; Frömbgen, T.; Kielb, P.; Kirchner, B. Efficient Prediction of Mole Fraction Related Vibrational Frequency Shifts. *J. Phys. Chem. Lett.* **2023**, *14*, 10531–10536.

(53) Weinhold, F. Quantum cluster equilibrium theory of liquids: General theory and computer implementation. *J. Chem. Phys.* **1998**, *109*, 367–372.

(54) Kirchner, B.; Spickermann, C.; Lehmann, S. B.; Perlt, E.; Langner, J.; von Domaros, M.; Reuther, P.; Uhlig, F.; Kohagen, M.; Brüssel, M. What can clusters tell us about the bulk?: Peacemaker: Extended quantum cluster equilibrium calculations. *Comput. Phys. Commun.* **2011**, *182*, 1428–1446.

(55) Brüssel, M.; Perlt, E.; von Domaros, M.; Brehm, M.; Kirchner, B. A one-parameter quantum cluster equilibrium approach. *J. Chem. Phys.* **2012**, *137*, 164107.

(56) von Domaros, M.; Perlt, E.; Ingenmey, J.; Marchelli, G.; Kirchner, B. Peacemaker2: Making clusters talk about binary mixtures and neat liquids. *SoftwareX* **2018**, *7*, 356–359.

(57) Frömbgen, T.; Drysch, K.; Zaby, P.; Dölz, J.; Ingenmey, J.; Kirchner, B. Quantum Cluster Equilibrium Theory for Multicomponent Liquids. *J. Chem. Theory Comput.* **2024**, *20*, 1838–1846.

(58) Laurence, C.; Berthelot, M.; Graton, J. *PATAI'S Chemistry of Functional Groups*; John Wiley & Sons, Ltd, 2009.

(59) Coggeshall, N. D.; Saier, E. L. Infrared Absorption Study of Hydrogen Bonding Equilibria. *J. Am. Chem. Soc.* **1951**, *73*, 5414–5418.

(60) Huggins, C. M.; Pimentel, G. C.; Shoolery, J. N. Proton Magnetic Resonance Studies of the Hydrogen Bonding of Phenol, Substituted Phenols and Acetic Acid. *J. Phys. Chem.* **1956**, *60,* 1311–1315.

(61) Saunders, M.; Hyne, J. B. Study of hydrogen bonding in systems of hydroxylic compounds in carbon tetrachloride through the use of NMR. *J. Chem. Phys.* **1958**, *29*, 1319.

(62) Maguire, M. M.; West, R. Hydrogen-bonding studies VII. Near infrared spectroscopic studies of the intermolecular hydrogen bonding of phenol, p-cresol and p-chlorophenol. *Spectrochim. Acta* **1961**, *17*, 369.

(63) Singh, S.; Rao, C. N. R. Spectroscopic studies of self-association due to hydrogen bonding. *J. Phys. Chem.* **1967**, *71*, 1074–1078.

(64) Whetsel, K. B.; Lady, J. H. *Spectrometry of fuels. Chapter 20: Self-association of phenol in nonpolar solvents*; Springer, 1970.

(65) Woolley, E.; Travers, J. G.; Erno, B. P.; Hepler, L. G. Molecular association of hydrogen-bonding solutes. Phenol in carbon tetrachloride. *J. Phys. Chem.* **1971**, *75*, 3591.

(66) Schaefer, T.; Rowbotham, J. B.; Chum, K. The proton magnetic resonance spectra of phenol in the absence of intermolecular proton exchange. *Can. J. Chem.* **1976**, *54*, 3666–3671.

(67) Lin, L.-N.; Christian, S. D.; Tucker, E. E. Solute activity study of the self-association of phenol in cyclohexane and carbon tetrachloride. *J. Phys. Chem.* **1978**, *82*, 1897.

(68) Petelenz, B. U.; Shurvell, H. Factor analysis as a complement to infrared band resolution. VII. The temperature dependence of the self association of phenol in carbon tetrachloride solution. *Can. J. Chem.* **1980**, *58*, 353.

(69) Josefiak, C.; Schneider, G. Determination of reaction volumes of hydrogen-bonding equilibria by high-pressure near-infrared spectroscopy. 2. Self-association of phenol in CCl4 up to 1 kbar. *J. Phys. Chem.* **1980**, *84*, 3004.

(70) Frohlich, H. Using infrared spectroscopy measurements to study intermolecular Hhy-drogen bonding: Calculating the degree of association, equilibrium constant, and bond energy for hydrogen bonding in benzyl alcohol and phenol. *J. Chem. Educ.* **1993**, *70*, A3.

(71) Lessinger, L. Hydrogen-bonding equilibrium in phenol analyzed by NMR spectroscopy. *J. Chem. Educ.* **1995**, *72*, 85.

(72) Galvez, J. P. `Aggregate`, a graph-theoretical tool to iden-tify interaction patterns in Molecular Dynamics simulations, see: `https://github.com/jpablogalvez/Aggregate`. 2024.

(73) Martínez-Núñez, E.; Barnes, G. L.; Glowacki, D. R.; Kopec, S.; Peláez, D.; Ro-dríguez, A.; Rodríguez-Fernández, R.; Shannon, R. J.; Stewart, J. J. P.; Tahoces, P. G.; Vazquez, S. A. AutoMeKin2021: An open-source program for automated reaction dis-covery. *J. Comput. Chem.* **2021**, *42*, 2036–2048.

(74) Smith, W. R.; Missen, R. W. The effect of isomerization on chemical equilibrium. *Can. J. Chem. Eng.* **1974**, *52*, 280–282.

(75) Ho, J.; Coote, M.; Cramer, C.; Truhlar, D. *Theoretical Calculation of Reduction Po-tentials*; 2015; pp 229–259.

(76) Jensen, J. H. Predicting accurate absolute binding energies in aqueous solution: ther-modynamic considerations for electronic structure methods. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12441.

(77) Noroozi, J.; Smith, W. R. Accurately Predicting CO2 Reactive Absorption Properties in Aqueous Alkanolamine Solutions by Molecular Simulation Requiring No Solvent Experimental Data. *Ind. Eng. Chem. Res.* **2020**, *59*, 18254–18268.

(78) Noroozi, J.; Smith, W. R. Prediction of Alkanolamine pKa Values by Combined Molecular Dynamics Free Energy Simulations and ab Initio Calculations. *Journal of Chemical & Engineering Data* **2020**, *65*, 1358–1368.

(79) Ho, J.; Coote, M. L. A universal approach for continuum solvent pKa calculations: are we there yet? *Theor. Chem. Acc.* **2009**, *125*, 3.

(80) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

(81) Li, Y.-P.; Gomes, J.; Mallikarjun Sharada, S.; Bell, A. T.; Head-Gordon, M. Improved Force-Field Parameters for QM/MM Simulations of the Energies of Adsorption for Molecules in Zeolites and a Free Rotor Correction to the Rigid Rotor Harmonic Oscillator Model for Adsorption Enthalpies. *J. Phys. Chem. C* **2015**, *119*, 1840–1850.

(82) Scott, A. P.; Radom, L. Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Möller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

(83) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132*, 114110.

(84) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.

(85) Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.

(86) Ebata, T.; Watanabe, T.; Mikami, N. Evidence for the cyclic form of phenol trimer:

Vibrational spectroscopy of the OH stretching vibrations of jet-cooled phenol dimer and trimer. *J. Phys. Chem.* **1995**, *99*, 5761.

(87) Ohta, K.; Tominaga, K. Vibrational population relaxation of hydrogen-bonded phenol complexes in solution: Investigation by ultrafast infrared pump probe spectroscopy. *Chem. Phys.* **2007**, *341*, 310.

(88) Fedor, A. M.; Toda, M. Investigating hydrogen bonding in phenol using infrared spectroscopy and computational chemistry. *J. Chem. Educ.* **2014**, *91*, 2191.

(89) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551–6568.

(90) Cacelli, I.; Cimoli, A.; Livotto, P. R.; Prampolini, G. An automated approach for the parameterization of accurate intermolecular force-fields: Pyridine as a case study. *J. Comput. Chem.* **2012**, *33*, 1055–1067.

(91) Cerezo, J.; Aranda, D.; Avila Ferrer, F. J.; Prampolini, G.; Santoro, F. Adiabatic-Molecular Dynamics Generalized Vertical Hessian Approach: A Mixed Quantum Classical Method To Compute Electronic Spectra of Flexible Molecules in the Condensed Phase. *J. Chem. Theory Comput.* **2020**, *16*, 1215–1231, PMID: 31855424.

(92) Cerezo, J.; Santoro, F.; Prampolini, G. Comparing classical approaches with empirical or quantum-mechanically derived force fields for the simulation electronic lineshapes: application to coumarin dyes. *Theor. Chem. Acc.* **2016**, *135*, 143.