THE EUROPEAN PHYSICAL JOURNAL D

Regular Article - Atomic and Molecular Collisions

Predicting dipole orientations in spontelectric methyl formate

Christian Kexel[1,2,3,a] and Andrey V. Solov'yov[3]

[1] Department of Physics, Goethe University, Max-von-Laue-Str. 1, 60438 Frankfurt, Germany
[2] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt, Germany
[3] MBN Research Center, Frankfurt Innovation Center Biotechnology, Altenhöferallee 3, 60438 Frankfurt, Germany

© The Author(s) 2021

Abstract. Capturing intermolecular interactions accurately is essential for describing, e.g., morphology of molecular matter on the nanoscale. When it reveals characteristics which are not directly accessible through experiments or *ab initio* theories, a model here becomes eminently beneficial. In laboratory astrochemistry, the intense study of ices has led i.a. to the exploration of the spontelectric state of nanofilms. Despite its success in biophysics or biochemistry and despite its predictive power, molecular modeling has however not yet been widely deployed for solid-state astrochemistry. In this article, therefore a pertinent hitherto unaddressed problem is tackled by means of the classical molecular-dynamics method, namely the unknown distribution of relative dipole orientations in spontelectric *cis*-methyl formate (MF). In doing so, from *ab initio* data, a molecular model is derived which confirms for the first time the anomalous temperature-dependent polarization of MF. These insights thus represent a further step toward understanding spontelectric behavior. Moreover, unprecedented first-principles predictions are reported regarding the ground-state geometry of the MF trimer and tetramer. In conjunction with the study of the binding to carbonaceous substrates, these additional findings can help to exemplarily elucidate molecular ice formation in astrochemical settings.

1 Introduction and background

The theoretical and computational characterization of objects on the nanoscale is a fundamental plus highly-demanding scientific endeavor. Albeit specific systems are dealt with in dedicated and seemingly unlike research disciplines, ranging, e.g., from pharmacology to astrochemistry, their appropriate description bears joint challenges. Common intertwined tasks are (i) the decision on a quantum-mechanical, semi-classical or classical ansatz, (ii) the complications introduced by many-body effects, (iii) the consistent linkage of multiple adjacent scales, defined either by time, characteristic lengths or the number of particles, covering the changeover from individual atoms to the solid state as well as the changeover between inanimate and living systems. There are shared underpinning phenomena, such as thermally driven structural transitions [1], chemical ordering [2], bonding or dissociation events [3,4] as well as cluster formation [5,6].

In spite of the low temperatures and the specific vacuum conditions in the interstellar medium, spectroscopic studies indicate intricate chemical activities, particularly in birth regions of planets and stars, involving diverse organics which reflect the evolution of complexity toward animate matter. Pure gas-phase chemistry is not able to account for the diversity, enhanced abundances and the existence of larger organics, such as MF (see Fig. 1). Hence, explanations additionally imply surface chemistry with molecules persisting in the solid state. Emissions of protostars and stars that move through molecular clouds generate broad infrared absorption bands which originate from condensed phases. Here, the infrared profiles of these ices carry information on temperature, composition or morphology of icy mantles. In the interstellar settings, the underlying processes are similar in nature to the already-mentioned phenomena above concerning nanosystems. They include, e.g., (i) deposition of molecules on dust particles within dense clouds, (ii) creation of ice films on the particles' surface, (iii) catalyzation of chemical reactions due to the presence of ice, (iv) shocks by protons and swift ions leading to sputtering and to morphological changes, (v) energetic processing by ultraviolet, X-ray and cosmic radiation, (vi) polymerization through thermal processing by adjacent protostars [7–9].

With regard to the theoretical description, in the realm of organic computational chemistry, density-functional theory (DFT) conventionally yields a bal-

---

On leave from A. F. Ioffe Physical-Technical Institute, St. Petersburg, Russia.

[a] e-mail: kexel@fias.uni-frankfurt.de (corresponding author)

anced trade-off between accuracy and numerical efficiency, e.g., for determination of molecular ground-state geometries. DFT is deployable for systems of, say, up to hundreds of atoms. Understandably, this limit is lowered if dynamical descriptions are needed. Moreover, weak intermolecular dispersion forces call for the inclusion of empirical long-range corrections [10] and possibly for more demanding *ab initio* approaches that treat electronic correlations adequately, such as many-body perturbation theory [11]. Furthermore, numerous problems on the nanoscale involving organics require the simulation of larger systems. Hence, a great variety of classical interaction potentials has been derived [12] which do not consider electronic structure explicitly. The analytic formulation of these force-fields is deduced from quantum-mechanical principles and provides a mapping from geometry of the systems to its potential energy. The accuracy of a molecular model relies largely on the deployed parameters and this model accuracy can be quantified with respect to distinct properties, such as transition temperatures [2,13] or binding energies. Regarding the development of force-fields, some requirements have to be fulfilled. In order to be considered accurate, the interaction potential should (i) exhibit reasonable complexity while being numerically practicable. It should also (ii) be transferable, i.e., accurate for geometries that were not utilized in its derivation [14].

It has become evident that astrochemistry and the study of ices can offer possible research opportunities related to nanoscopic systems. The large-scale production of metal clusters using an ice matrix is, e.g., an instructive example [15]. The scientific investigation of the coupled gas-surface chemistry, motivated by the ongoing discovery of more sophisticated species in the interstellar medium, represents a formidable challenge for ultrahigh-vacuum laboratory experiments and also molecular modeling. However, such models are only infrequently deployed in solid-state astrochemistry with, e.g., Ref. [16] being an exception. Astrochemical theory, e.g., centers around kinetic considerations on the one hand [17] or high-precision quantum chemistry methods for the calculation of finely structured rotational–vibrational spectra on the other hand [18,19]. Classical molecular modeling in conjunction with advanced computing can fill the gap between both aforementioned approaches.

The committed experimental examination of astrochemical ice analogs culminated i.a. in the description of a peculiar state, namely spontaneously electric matter [20]. When deposited at low temperature, certain species can spontaneously create a surficial electric potential which roots in the orientation of the molecular dipoles. The corresponding species possess a permanent dipole moment, but are diverse apart from that. The emerging electric potential depends on temperature, hinting at a competition between order and thermally induced disorder. Therefore, the alignment is usually reduced with increasing temperature. Nonetheless, this rule has one known exception, namely MF with a temperature-dependent polarization curve deviating
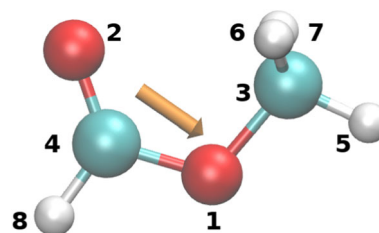

Fig. 1 Geometry of gas-phase MF monomer. Orientation of the electric dipole moment is according to the derived molecular model. For an improved presentation, the length of the dipole vector is shortened by a factor of two in the shown depiction. The provided numbering of atoms corresponds to the atomic identifiers given in Table 1

from this rule of monotonous decrease [21]. Spontelectric matter in general exhibits a Curie temperature, marking the breakdown of the molecular orientation which does not re-emerge upon cooling. The aligned dipole orientation appears to be an unfavorable state in terms of energy, because of the repulsive dipole–dipole interaction. Up to this date, only a semi-empirical analytical model has been reported which captures the spontaneously polarization. No atomistic model exists so far. Molecular modeling of the effect is challenging because of the subtle nature of the respective long-range and many-body forces [20].

In the following section, some recent studies are reviewed which characterize different aspects of MF. Nunes et al. [22] have carried out electronic state spectroscopy of MF using vacuum-ultraviolet photoabsorption and photoelectron spectroscopy alongside *ab initio* electronic structure calculations. Experiments are done in the gas phase and theory is considering here the single molecule. Feketeova et al. [23] have studied dissociation of MF molecules in an electron attachment spectrometer and have supported the dissociation study by *ab initio* calculations. Thakur et al. [24] have studied i.a. MF dimers with DFT (M06-2X). They found an antiparallel orientation to be the most stable. Modica and Palumbo [25] have studied the mid-infrared spectrum of icy films under different conditions. The deposited MF film has been heated and the amorphous--to-crystalline transition has been observed. Moreover, films of pure methanol and a mixture of methanol with carbon monoxide have been bombarded with protons and the formation of MF has been observed by looking at molecular vibrations. They suggest that gaseous MF observed in dense molecular clouds is formed in the solid state after cosmic ion irradiation of icy grain mantles. Burke et al. [26] have studied the adsorption and thermal desorption of MF on graphite-type substrate at low temperatures. MF behavior is explained by its intrinsic inability to form hydrogen bonds. MF wets the surface whereas its isomers preferentially cluster. Very recently, Roman et al. [27] combined IR spectroscopy and *ab initio* calculations to suggest that a certain non-polar MF dimer represents a structural motif of the crystalline

state. They ground their arguments on the absence of spontelectric dipole orientation.

In this work, we address the open question concerning the statistical distribution of orientation angles in spontaneously polarized MF using the classical molecular-dynamics method which could not be directly observed in experiments. By doing so, for the first time the anomalous polarization curve in MF is reproduced through a physical atomistic model. The molecular model has been derived from first-principles calculations. Moreover, in order to investigate the binding of MF, *ab initio* results are presented regarding the trimer, the tetramer and regarding MF monomers interacting with (poly)aromatic hydrocarbons molecules. Differently sized aromatic molecules are considered, because (i) they enable a systematic examination of the attachment to ever larger binding partners and because (ii) they can act as a proxy for other carbonaceous substrates [7–9].

In summary, on the one side the early stage of ice formation is elucidated from first principles and on the other side the behavior of bulky film is simulated in this work using a coarser molecular model which is consistently derived from the quantum results. The all-embracing simulation of the condensation onto a substrate until the multi-monolayer films is however beyond the scope of this present work.

The article is structured in the following manner: Sect. 2 provides the relevant background on the employed computational methods, in Sect. 3 results are presented and discussed. Lastly, in Sect. 4 some conclusions are formulated accompanied by an outlook for future research.

2 Computational methodology

Different computational tasks are considered in this article which can be divided into quantum chemistry calculations (Sect. 2.1), the molecular modeling approach (Sect. 2.2) and moreover the estimation of parameters by global optimization (Sect. 2.3).

2.1 Quantum treatment of non-covalent bonding

In MF, the underlying weak non-covalent bonding, that also can be encountered in other molecular complexes, poses difficulties for an efficient quantum-mechanical treatment with conventional DFT. The challenge of dispersion forces has been answered in recent years with inclusion of empirical corrections to functionals [10].

In order to be able to screen different candidate geometries in determining the ground-state oligomers of MF, i.e., trimer and tetramer, B97D has been used in this work. These candidate clusters are then subject to optimization with a higher level of theory. Additionally, for sake of comparison further dispersion-corrected DFT approaches have been selected, e.g., ωB97X-D.

Since it accounts naturally for some electronic correlation, second-order many-body Moller–Plesset per-turbation theory (MP2) [11] represents a sensible choice for treating intermolecular complexes while being computationally more demanding than DFT. Here, the optimization of tetramers of MF or binding studies of MF with the anthracene molecule, they mark the limit of what can be achieved currently with our computing resources at this level of theory. Depending on task, the double- and triple-ζ augmented correlation-consistent basis sets [28] are deployed: aug-cc-pvdz (ADZ) and aug-cc-pvtz (ATZ), respectively. Basis-set superposition errors are corrected using the counterpoise method. The GAUSSIAN 09 software package is deployed for all the quantum calculations [29]. The coupled-cluster methods, like CCSD(T), which are ideally desired and yield benchmark accuracy, are unfortunately beyond what is feasible in the considered scenarios at the moment.

2.2 Classical atomistic molecular modeling

For many years now, molecular-dynamics simulations based on classical interaction potential have been successfully employed in biophysics and biochemistry. The CHARMM (Chemistry at Harvard Molecular Mechanics) family of force-fields [30] is widely used. Among others, it provides all-atom interaction parameters for several organics through the CHARMM General Force Field (CGenFF) [31]. The potential energy of a molecular system is expressed as the sum of intra- and intermolecular terms, hence $U = U_{\mathrm{intra}} + U_{\mathrm{inter}}$. The focus of the present work lies on the interaction between molecules, therefore non-bonded intermolecular terms are of central relevance. They comprise Coulomb and van-der-Waals contributions which dependent on the interatomic distance r_{ij} between atom i and j:

$$U_{\mathrm{inter}} = U_{\mathrm{Cou}} + U_{\mathrm{vdW}} \tag{1}$$

$$= \sum_{i \neq j} \left[C \frac{q_i q_j}{r_{ij}} + \varepsilon \left(\left(\frac{R}{r_{ij}} \right)^{12} - 2 \left(\frac{R}{r_{ij}} \right)^{6} \right) \right] \tag{2}$$

Here, C is Coulomb's constant, the parameter q denotes the atomic partial charges, R an equilibrium distance and ε a potential well-depth. The parameter R and ε depend on the types of atom that are involved in the interaction. For dissimilar types, the Lorentz–Berthelot combination rule [32] is used and contributions of adjacent atoms are not accounted for according to the 1–4 exclusion scheme [30]. Furthermore, the intramolecular bonded contribution is given by the addition of bond, angle and dihedral terms:

$$U_{\mathrm{intra}} = U_{\mathrm{b}} + U_{\mathrm{a}} + U_{\mathrm{d}} \tag{3}$$

$$= \sum_{i \neq j} c_{\mathrm{b}} \left(r_{ij} - R_{\mathrm{b}} \right)^2 + \sum_{i \neq j \neq k} c_{\mathrm{a}} \left(\theta_{ijk} - \theta_{\mathrm{a}} \right)^2 \tag{4}$$

$$+ \sum_{i \neq j \neq k \neq l} c_{\mathrm{d}} \left(1 + \cos \left(n \cdot \varphi_{ijkl} - \delta \right) \right). \tag{5}$$
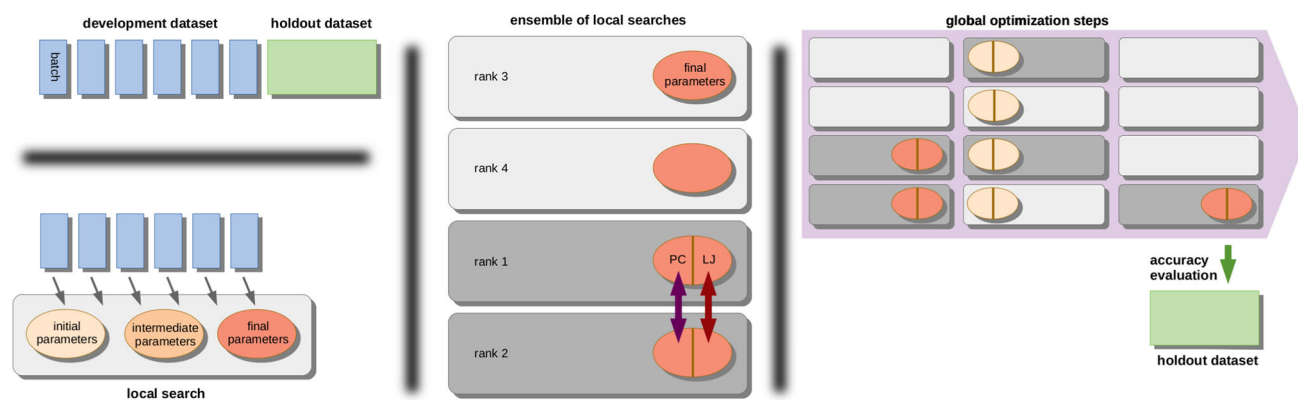
Fig. 2 Schematic outlining the parameter optimization procedure. Details are stated in the text. Top left: the generated *ab initio* dimer geometries are split into a development dataset and a holdout set. Bottom left: during a local search the model parameters are gradually improved based on data of successive batches (i.e., subsets of the development set). Middle: after local searches, among an ensemble of local searches the best-ranked solutions (highlighted through darker gray) are considered as input for the next step of the global optimization where partial charge (PC) parameters (one half of an ellipse) and Lennard-Jones (LJ) parameters (other half of an ellipse) are recombined to form new initial model parameters. Right: after several steps, from the best molecular parameters (ellipse at the right) the model's accuracy in terms of energy is assessed using the independent holdout dataset

Here, c_b, c_a, c_d denote spring constants, R_b is an equilibrium distance, θ_a an equilibrium angle, n the multiplicity and δ an equilibrium phase.

2.3 Global optimization strategy for intermolecular parameter estimation

The emphasis of this work lies on the accurate modeling of intermolecular interactions for the simulation of a subtle phenomenon, i.e., spontelectric behavior. While parameters provided by CGenFF are a reasonable guess regarding forces within a molecule, e.g., for common proteins, DNA (deoxyribonucleic acid) or conventional organic drug-like molecules, we aim at a dedicated set of non-bonded parameters for the mentioned purpose. The necessary covalent parameters will instead be taken from CGenFF. It is worth stressing that the resulting molecular model (i.e., intra-plus intermolecular interactions) represents a first attempt to model MF having therefore initially a reduced complexity as compared to an all-atom approach.

In order to fit non-bonded energy terms, first-principles calculations on the counterpoise-corrected MP2/ADZ level are considered. 10,245 random MF dimer geometries are created where the intermolecular distance between the centers-of-mass is sampled uniformly as well as the mutual orientation in terms of the rotation angles. Rigid molecules are considered for the single-point evaluation, thus with no intramolecular variation. The rigid molecule originates from monomer optimization on the same level of theory. Each geometry hence provides the interatomic distances r_{ij} with the corresponding dimer potential energy E_{dimer}. The *ab initio* interatomic energy can be obtained according to

$$E_{\mathrm{inter}} = E_{\mathrm{dimer}} - 2 \cdot E_{\mathrm{monomer}}. \tag{6}$$

The global optimization aims toward arriving at a set of parameters q, ε, R which minimizes the deviation Δ from the first-principles results. A common figure-of-merit in computational chemistry is the mean absolute deviation (MAD)

$$\Delta = |E_{\mathrm{inter}} - U_{\mathrm{inter}}|. \tag{7}$$

In chemical physics, basin-hopping optimization [33, 34] is a successful strategy, especially for finding minima in the intricate energy landscape of clusters. It operates through a combination of local minimization with the complementary option to escape local minima by stochastic global moves. Similar in spirit, an evolution strategy [35] is used in the present work for global minimization of the MAD Δ. Our heuristic incorporates Monte Carlo-type local search with the possibility for global moves. The scheme is sketched in Fig. 2. The procedure can be summarized through a list of stages:

1. Division of the *ab initio* data into a development set for parameter optimization and a holdout dataset for accuracy evaluation on similar, but previously unseen data
2. Running several global optimization steps for parameter refinement

   (a) At each global step, an ensemble of competing parallel local minimizations with differing parameter values is done in order to reasonably cover the parameter space
   (b) At the end of each global step from the population of local solutions, the best-performing

parameters are combined as input for the next global step so as to construct new good parameter combinations

(c) Each local search consists of many steps where each time varied parameter values are proposed and which are accepted if their MAD on batch is smaller than for the previous considered parameters

3. Model assessment with the holdout dataset after the last global step through calculating the MAD between the first-principles energies and the energy of the derived molecular model

More precisely, the details are as follows. First, the generated *ab initio* dimer geometries are randomly split into a so-called development set for performing the actual parameter optimization (8000 geometries) and an independent holdout set for final accuracy assessment of the developed classical model (2245 geometries; green box in Fig. 2). The development set is moreover divided into batches (blue boxes in Fig. 2) of which each comprises a bunch of geometries. Batch size is a meta-parameter of the optimization procedure. Batches are introduced so as to circumvent consideration of the full dataset for each parameter evaluation, therefore batches should be larger than one but significantly smaller than the development set.

The parameters q, ε as well as R are initialized by randomly selecting reasonable guesses. Then at each local search step, one random parameter is altered by a small increment. Here, the increment strength also represents a free meta-parameter that can be tuned for improving minimization. The new parameters are evaluated on a batch of geometries and the mean absolute deviation Δ_{batch} is calculated. If Δ_{batch} is smaller than for the previously best parameters, the new parameters are kept as current model parameters. The greedy local search corresponds to a Monte Carlo approach at low temperatures [36] which accepts mostly down-hill moves toward a smaller Δ_{batch}. The model is updated during the local search in order to find the final parameters of this local run.

So as to explore the wider parameter landscape several searches, i.e., an ensemble of S searches, are done in parallel with random initial parameters. S represents another meta-parameter. From the ensemble of S solutions a fraction of them, namely the $\sqrt{S}$ best-ranked ones (dark gray boxes in Fig. 2), is considered for the next round of local search. The local solutions are ranked based on the final error Δ_{batch}. Every solution consists of two parts: the block of parameters related to partial atomic charges and the block regarding van-der-Waals parameters (depicted as two halves of an ellipse in Fig. 2). Among the best solutions, the parts related to atomic charges and the part with van-der-Waals (also dubbed Lennard-Jones) parameters are recombined to generate a new ensemble of S initial parameters for the next local minimization. The recombination rule rests on the assumption that both blocks are physically independent and can be optimized sep-

arately, their recombination should therefore enable an efficient exploration of the promising portions of the parameter space. After several of those global optimization steps, the derived final parameters are assessed for their accuracy based on the sofar-unused holdout set: the MAD is finally calculated on the full holdout set which yields Δ_{holdout}. With respect to the performance measure Δ_{holdout}, molecular models of different complexity can be compared or different meta-parameters can be tested.

Like for other hard high-dimensional optimization problems, the outlined heuristic is ought to approximate a good global solution, but is not guaranteed to deliver a uniquely best parameter estimation.

3 Results and discussion

The following section consists of two major parts: The first is devoted to an analysis of novel first-principles results and the second part covers the molecular model for the simulation of bulk spontelectric MF.

3.1 Ab initio studies of intermolecular binding involving methyl formate

3.1.1 Interaction of monomers with the substrate

A series of experimental results [20] suggests that spontelectric behavior is independent of the type of substrate to which nanofilms are deposited. In order to better comprehend the potential early processes involved in the onset of MF ice film formation, binding to aromatic molecules (i.e., benzene, naphthalene and anthracene) is examined here deploying the MP2/ADZ level of theory with counterpoise-correction. Instead of an extended substrate, we restrict ourselves to the aforementioned finite systems in this work, in order to maintain the more accurate, but at the same time more demanding level of theory. The MP2 method is known to yield results comparable to expensive CCSD(T) calculations for hydrogen bonding [37] and for the interaction of alkanes [38]. Here, alkanes can be expected to represent a model system, e.g., for the methyl group of MF. Nonetheless, MP2 is overestimating the strength of $\pi - \pi$ interaction between aromatics [39]. The complexes studied here are mixtures of those three scenarios involving the methyl and carbonyl group of MF as well as the aromatic π-system. Also MP2 is thus a sensible choice for MF–MF interactions investigated in Sects. 3.1.2 and 3.1.3.

In laboratory settings, graphite is used in some experiments [7], moreover (poly)aromatics are observed in astrochemical scenarios and carbonaceous surfaces are generally discussed as interstellar spots of ice growth [8,9]. Hence, three different aromatic molecules are considered (see Fig. 3), because they also enable to systematically examine the attachment of MF to ever larger binding partners.
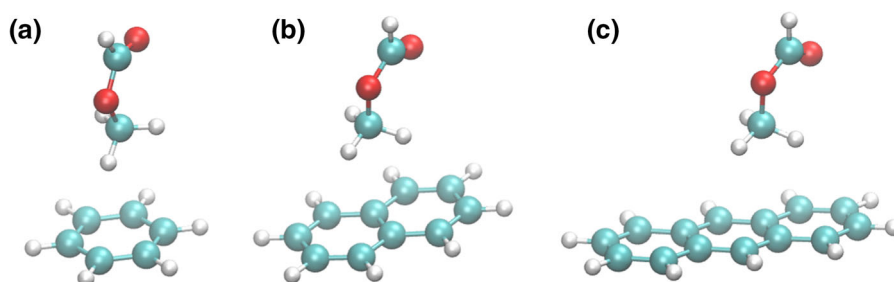
Fig. 3 Optimized geometries for MF in complex with different carbonaceous substrates: **a** benzene, **b** naphthalene, **c** anthracene. Hydrogen is drawn in white, carbon in cyan and oxygen in red

For each of the three substrate molecules, various random orientations of MF are probed as the MF monomer approaches the substrate. From this sequence of quantum single-point calculations, the most favorable orientation in terms of energy is finally subject to geometry optimization. The resulting binding energies of the complexes range between $E_{\mathrm{inter}}^{\mathrm{benzene}} = -2.8$ kcal/mol for benzene and $E_{\mathrm{inter}}^{\mathrm{anthracene}} = -4.3$ kcal/mol for anthracene. In each of the cases, the methyl group of MF is facing toward the substrate. Hence, during deposition of MF onto the substrate, if the monomer is free to settle toward its energy minimum, the carbonyl group will be directed upward enabling binding of arriving interaction partners. An alternative scenario is given by the situation when MF monomers just randomly stick to the surface without the possibility to re-orientate in a more favorable pose. In an experimental deposition and desorption study [7] MF is reported to wet the graphite surface prior to building up the layers of a nanofilm. This experimental finding indicates that MF has little tendency for clustering. In the following section, thus binding between MF molecules will be described.

3.1.2 Dimer motifs

In Ref. [27], the most stable MF dimers were identified with MP2 extrapolated to the complete basis set (CBS) limit. The general geometry of the three most stable ones is also depicted in Fig. 4. We here complement the analysis of Ref. [27] with additional arguments regarding the dimer formation when a MF monomer is already bound to the carbonaceous binding partner. Assuming the orientation of MF as demonstrated in Fig. 3, the access to the methyl hydrogens is restricted for the second incident MF monomer by the presence of the surface (i.e., the aromatic molecule). Hence, the aromatic molecule accordingly will complicate a hydrogen-bonding behavior in a fashion analogous to MF dimer D_1 in Fig. 4a. In contrast, dimer D_2 or D_3 can be realized with the carbonyl group of the already-bound MF molecule facing upward and away from the aromatic.

At the MP2/ADZ level of theory with counterpoise-correction, following Eq. 6, we separately compute binding energies for complex D_2 and subsequently for complex D_3. The difference in binding energies between both geometries is 0.13 kcal/mol ≈ 65 K with $D3$ being
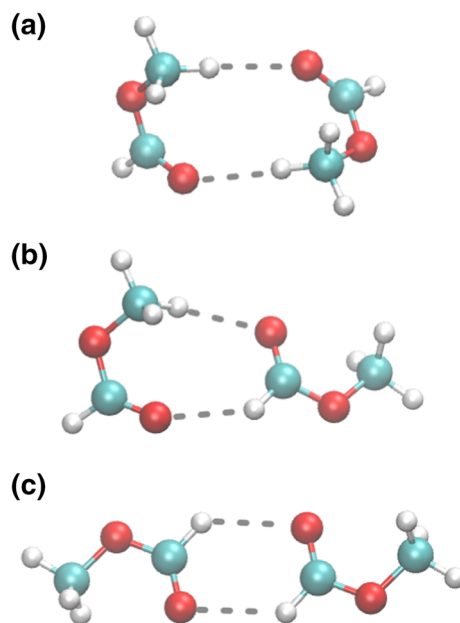

Fig. 4 Low-energy dimer geometries which are also discussed as structural motifs D_1 to D_3 (top to bottom) in Ref. [27]. Hydrogen bonding is highlighted with dashed lines

energetically more favorable. Hence, when bound to the aromatic surface, the monomer can readily engage in both types of conformations and moreover the barrier between both conformation can be overcome by a thermal energy and deposition temperature which is typical encountered in the study of analogs of astrochemical ices and studies of the spontelectric effect. At the MP2/ATZ level of theory with counterpoise-correction, the difference even vanishes almost and is below 0.01 kcal/mol. For MP2/ATZ both binding energies are circa -3.9 kcal/mol and accordingly below the binding energy of the complex with the larger polyaromatic hydrocarbons, such as anthracene. Substrate binding is hence preferred over dimerization.

The dimer formation is explained primarily with respect to differing hydrogen bonds, because in the three discussed dimers the resulting dipole orientation yields non-polar complexes, due to the antiparallel alignment of molecular electric dipole moments (compare Fig. 1 for the depiction of the dipole orientation in the monomer).

3.1.3 Trimer and tetramer geometries

Next, the study is extended to MF trimers. With counterpoise-corrected and dispersion-corrected DFT (B97D/6-311G**) first 60 initial geometries were setup which have then been optimized to obtain five lowest-energy candidate conformations. These remaining five geometries have been optimized at the MP2/ADZ level which is considered to be equivalent for MF–MF interaction to the highly accurate CCSD(T) [37,38]. Two of them converged to the same optimized geometry with lowest energy which is depicted in Fig. 5 from two different perspectives. The geometry with second lowest potential energy exhibits a barrier equivalent to 150 K relative to the ground state and is therefore dismissed for the further analysis. Apart from our presented predictions here, to the best of our knowledge, there is, up to now, no information available in the literature either from theory or experiment concerning the geometry of the MF oligomers beyond the dimer.

The trimer can be interpreted in terms of the already-presented dimers: Two of the molecules (green and brownish in Fig. 5; the two most right molecules in Fig. 5a) form together a motif that resembles the dimer D_1 in Fig. 4a, but with the ester oxygen being closer to the aldehyde hydrogen of the opposing molecule. The plane which is defined by the heavy atoms in each monomer is here almost parallel to the very same plane in the other molecule. The third monomer (red in Fig. 5) is stacked on top of the described complex and builds a deformed variant of the D_2 motif with one of the other monomers (brownish in Fig. 5; the most right one in Fig. 5a). The stacked third molecule (red) further stabilizes the position through its aldehyde hydrogen pointing toward the ester oxygen of the remaining partner (green in Fig. 5). In summary, the observed trimer can be explained as combination of dimers D_1 and D_2 with increased intermolecular proximity of aldehyde hydrogens and ester oxygen. Therefore, a possible route for clustering from a stable dimer to the stable trimer is evident, namely deformation of the dimer and stacking of an additional molecule. Due to the already-discussed antiparallel orientation of two of the three molecules which form dimer motif D_1, the total electric dipole moment of the cluster equals approximately the dipole moment of the remaining stacked monomer $|\mu_{\mathrm{trimer}}| \approx |\mu_{\mathrm{monomer}}|$.

The MF ground-state tetramer is found similar to the above methodology: with counterpoise-corrected and dispersion-amended DFT (B97D/6-311G**) 30 start geometries were set up which have subsequently been optimized to obtain two low-energy candidates. Both geometries have then been optimized at the MP2/ADZ level. The optimal geometry is shown in Fig. 6 from two distinct angles. The other geometry with the second lowest energy possesses a barrier of 595 K relative to the ground-state conformation. In the tetramer, two of the participating monomers (yellow and purple; the molecule on the bottom and the most right molecule in Fig. 6a) constitute a modified version of the D_3 motif with their carbonyl groups being adjacent. The remain-
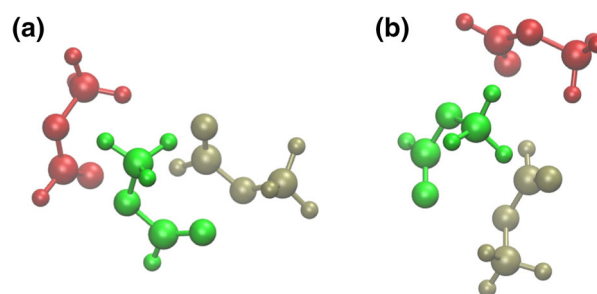

Fig. 5 Optimized trimer geometry. For enhanced clarity, the molecules are colored differently and moreover **b** represents a rotation of the geometry given in **a**
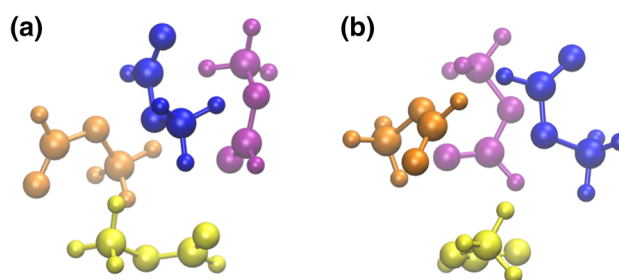

Fig. 6 Optimized tetramer geometry. Again, **b** represents a rotation of the geometry depicted on the left-hand side in **a**

ing monomers cannot easily be described by the aforementioned motifs. Here, the tetramer's dipole moment does not vanish. Moreover, the relationship, i.e., the path between the geometry of the optimal trimer and the optimal tetramer appears to be non-trivial. This might imply that the growth of larger cohesive clusters is hindered and that substantial energetic barriers between stable configurations have to be crossed which is unlikely at low-temperature conditions.

3.2 Molecular modeling of methyl formate

3.2.1 Parametrization of the model

The CHARMM force field CGenFF and its parameters represent a mature framework for the simulation of various molecular systems, including organics such as MF. Therefore, the intramolecular bond, angle and dihedral parameters are considered for the simulation of MF (see Tables 3, 4 and 5 in the appendix). However, because the intermolecular interactions are weak and because the common model compounds which are utilized for fitting CHARMM parameters did not include MF specifically and because the manifest values which are given in Table 2 for comparison did not yield any meaningful results, we here aim at a dedicated molecular model for MF. This however comes at the cost of an elaborate parameter estimation procedure. In this article, an initial step is taken so as to shed light onto spont-electric behavior which is one example of the plethora of interesting phenomena involving molecular ices. There-

Table 1 Description of atoms in the MF molecule, their respective designation in CGenFF, the partial charges q obtained through optimization of the molecular model and, as comparison, Mulliken charges from MP2/ADZ (with the hydrogen charges summed into the heavier atoms)

No.	Description	CGenFF	Mulliken q	Model q
1	Ester O	OG302	-0.739	-0.520
2	Carbonyl O	OG2D1	-0.596	-0.588
3	Methyl C	CG331	$+0.472$	$+0.343$
4	Carbonyl C	CG2O2	$+0.863$	$+0.765$
5,6,7	Methyl H	HGA3	–	–
8	Aldehyde H	HGR52	–	–

q is given in units the elementary charge e. Numbering is according to Fig. 1

Table 2 Description of van-der-Waals parameters: comparison between CGenFF and the actually deployed parameter estimates

No.	CGenFF	ε	R	Model ε	Model R
1	OG302	-0.100	3.30	-0.064	3.275
2	OG2D1	-0.120	3.40	-0.064	3.275
3	CG331	-0.078	4.10	-0.297	3.759
4	CG2O2	-0.098	3.40	-0.297	3.759
5,6,7	HGA3	-0.024	2.68	-0.000	–
8	HGR52	-0.046	1.80	-0.000	–

Well-depth is given in kcal/mol and equilibrium distance is provided in A

fore, in this initial approach we choose a simple molecular model for describing the intermolecular interactions between MF molecules. In the famous TIP3P model for water molecules [40,41], van-der-Waals forces are solely accounted for between the heavy, i.e., oxygen, atoms. Similarly, for MF van-der-Waals parameters are designated regarding the four heavy atoms where moreover we neglect atom-typing, i.e., the discrimination between chemically unlike atoms of the same element, e.g., ester and carbonyl oxygen or methyl and carbonyl carbon. Hence, so far four independent parameters have to be determined, namely van-der-Waals welldepth for oxygen and for carbon as well as equilibrium distance for both elements. Additionally, atomic partial charges have to be quantified. Again, we aim at reducing complexity and restrict ourselves to atomic charges residing on the heavy atoms. Thus, since neutral molecules are treated here, only three additional free parameters have to be estimated which adds up to a total of seven fitting parameters. This intricate multidimensional search problem is tackled with the global optimization scheme described in Sect. 2.3. Prior to estimating the actual parameter values of the molecular model based on the aforementioned *ab initio* results, the meta-parameters are determined in order to obtain accurate results within a reasonable time span. The step size, i.e., the strength for mutating the parameters randomly by an increment represents one meta-parameter. Here, one parameter is changed where the increment is drawn from a normal distribution having the mutation strength as value. The others being the ensemble size S and the batch size. After some testing, the following setup is deployed: as batch size 100 is used, an increment strength of 0.1 and an ensemble size of 25. Each local search consists in 10,000 iterations and a total of 80 global optimization steps are taken.

The derived molecular model leads to a MAD on the independent holdout set of $\Delta_{\text{holdout}} = 0.15$ kcal/mol. For comparison, the MAD from MP2/ADZ energies to other quantum methods has been also evaluated (on a smaller set of about 400 geometries), it yields $\Delta_{\text{holdout}}^{\text{MP2/ATZ}} = 0.02$ and $\Delta_{\text{holdout}}^{\text{APFD/ATZ}} = 0.07$ and $\Delta_{\text{holdout}}^{\text{B97D/ATZ}} = 0.09$ kcal/mol.

The estimated parameters of the classical molecular model are detailed in Tables 1 and 3. Even though the electrical dipole moment $\mu_{\text{monomer}}^{\text{model}}$ has not been explicitly incorporated into the optimization procedure, the estimated partial charges yield $|\mu_{\text{monomer}}^{\text{model}}| = 1.92$ D which is in good agreement with the experimental value of 1.77 D and also its direction is accurate, see Fig. 1, when compared to the known literature result [42]. From the water molecule models [40] which are parametrized in a similar fashion as our molecular model, it has become evident that it is intricate to have accurate models applicable for both potential energy and dipole moment calculations [41].

The partial charges correspond reasonably well to the summed Mulliken charges from the MP2/ADZ calculation. Lennard-Jones radii approximately match CGenFF values. Particularly, the Lennard-Jones welldepth of the simplified model deviates significantly from the all-atom GCenFF values, because the energy contributions from hydrogen are not explicitly modeled.
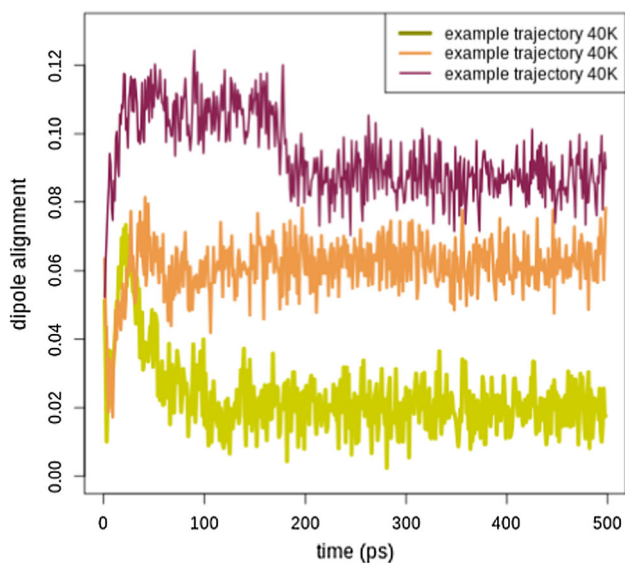
Fig. 7 Illustrative time evolution of the system's dipole alignment. The last 200 ps of each stable trajectory are used to calculate the thermodynamic average alignment, see Fig. 8
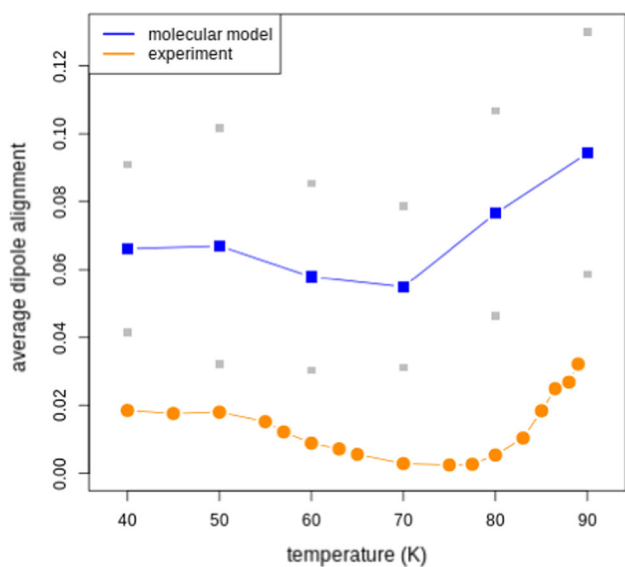

Fig. 8 Simulated average alignment, as defined in Eq. (8), in bulk MF versus temperature. The experimentally observed curve from [21] with the anomalous increase toward higher temperatures is also drawn. The gray squares are error bars indicating one standard deviation and are based on 30 independent trajectories per temperature

3.3 Simulating the spontelectric state

Since the spontelectric effect is thought to be a bulk phenomenon, we validate the molecular model by considering a simulation box with periodic boundary conditions. The box contains $N = 216$ molecules which are randomly oriented at the beginning of the simulation. This corresponds on average to 6 molecular lay-

ers in each spatial direction which roughly represents a minimal extent for observing spontelectricity. Initially the populated simulation box matches a pre-specified density $\rho = 987$ kg/m^3 (being the density of liquid MF). The simulation box is allowed to adapt its side length in the course of the simulation. This differs from a situation when a nanofilm is gradually laid down: the packing of the molecules would be results and not an input to the simulation. The number of atoms and hence the box size presents a compromise: it is large enough to span 2 times the cutoff radius of the van-der-Waals and electrostatic interactions, but small enough to allow collection of some statistics from re-running simulations. Particle mesh Ewald summation is used for treatment of long-range electrostatics. Using periodic boundary conditions, the system is simulated for 500 ps at a given temperature T. A total of 200 independent runs is conducted at each temperature between $T = 40$ and $T = 90$ K. The balance between simulated time and number of runs represents a trade-off which allows for the collection of sufficient statistics on converged simulations. As discussed, intramolecular parameters inferred from the CGenFF are taken. In contrast, intermolecular parameters are determined by our custom global estimation heuristic (see Table 2). NAMD [43] is deployed for the classical calculations.

Into our analysis, we included 30 trajectories for which we verified that they converged and that they capture a stable system. Some of the simulated 200 trajectories at each temperature were short to converge, e.g., in terms of the system's energy. In Fig. 7, the orientational time evolution of individual trajectories is exemplarily depicted. The initialized random orientation spontaneously relaxes to a certain alignment level. Here, dipole alignment is defined as the average net electrical dipole moment of the system considering all molecules. It is normalized relative to the dipole moment of the monomer:

$$\frac{\langle \mu \rangle}{|\mu_{\text{monomer}}|} \quad \text{with} \quad \langle \mu \rangle = \frac{1}{N} \sum_i^N \overrightarrow{\mu}_i. \quad (8)$$

In respective experiments, the z-axis which is perpendicular to substrate as well as film surface, is the natural principal direction of alignment. Because no substrate and no film surface are present in the bulk simulation, the system as a whole will not per se exhibit a preferential alignment in the z-direction. Hence, in our simulation the vector $\langle \mu \rangle$ can point in any direction, not necessarily along the z-axis.

We compare our calculations in Fig. 8 with the experimental observation. The simulated average alignment is the thermodynamic average deduced from the 30 independent runs. Even though rather big uncertainty exists in the calculated curve, it reproduces the anomalous polarization for MF with an increase toward higher temperatures. Overall, the alignment shows higher values than the experimental curve. This fact might be due to the difference in boundary conditions between
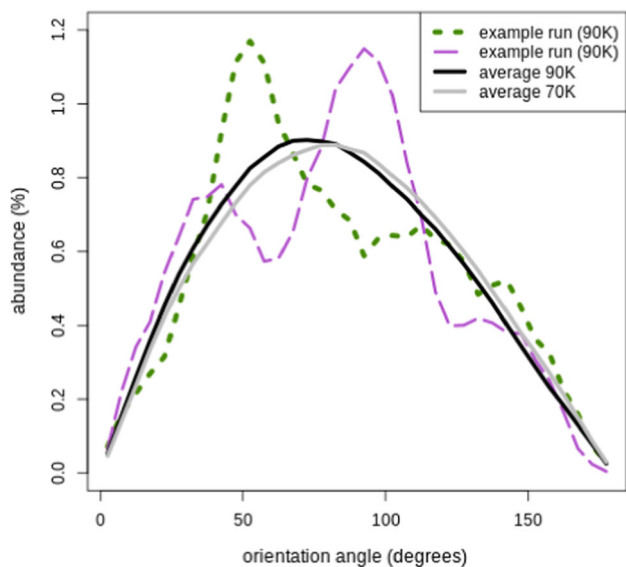
Fig. 9 Statistical distribution $h(\alpha, T)$ of orientation angles relative to the principle alignment direction $\langle \mu \rangle$. Predictions in bulk MF by the developed molecular model. The thermodynamic average shape originates from a superposition of multimodal distributions. The histogram at 90 K (higher overall alignment in Fig. 8) is more skewed than the histogram at 70 K (lower overall alignment in Fig. 8)

simulated bulk system versus deposited thin film. Also, this deviation might be due to lower or higher packing of the dipoles in the periodic-boundary bulk simulation. Furthermore, (i) the mentioned simplifications in the molecular model might lead to discrepancies as well as (ii) the parametrization of the force field which has been based on quantum calculation of dimers that do not adequately reflect the situation in the solid film with an increased number of nearest neighbors, e.g., in terms of the molecule's polarizability. It is important to refer to challenging and long-lasting efforts to derive water molecule models which are applicable to a range of physical/chemical settings [40].

There is great scientific interest regarding microscopic details of the spontelectric state, especially concerning properties which are currently not directly observable through experimental studies. Such a quantity is the temperature-dependent distribution $h(\alpha, T)$ of dipole orientations measured relative to the preferred alignment direction. From experiments solely, the average angle can be extracted which is directly related to the average orientation via

$$\langle \alpha \rangle = \arccos \left(\frac{\langle \mu \rangle}{|\mu_{\mathrm{monomer}}|} \right) \tag{9}$$

We deploy the molecular model to monitor and thereby predict the relative angles in the course of the simulation. Results are given in Fig. 9. This simple and initial model can hereby assist in better interpreting findings related to experiments on the spontelectric effect.

4 Conclusions

Quantum and classical calculations have been reported allowing insights i.a. into the nature of spontaneous polarization of methyl formate (MF). Especially, we predict a broad and slightly skewed distribution of orientation angles suggesting that the appreciable electric surface potentials of nanofilms observed in experiment originate from a subtle alignment tendency on the microscopic level. In principle, classical models can assist in simulating other temperature-dependent phenomena, e.g., the thermal processing of icy solids in astrochemistry.

Additionally, the binding of the MF monomer to aromatic substrate molecules as well as the formation of small MF clusters has been studied by a first-principles approach. The investigation enables a better understanding of the onset of MF nanofilm growth by discussing different scenarios, particularly where the molecular flux onto the surface and the temperatures are as low as in astrochemical settings.

Further research directions could include:

1. The dynamical simulation of the condensation onto a substrate in order to follow the emergence of the spontelectric state and to gain a characterization of the grown nanofilms
2. Explicit account for hydrogen binding which has been omitted in the presented initial model
3. Simulation of an extended system to study possible spatial inhomogeneity
4. Long-term calculation in order to examine temporal fluctuations
5. Investigation of less complex species, i.e., small as well as more symmetric molecules such as CO or N_2O [20], so as to gain fundamental knowledge regarding spontelectricity, because we have tackled an intricate molecule in this article which exhibits a competition between hydrogen bonding and dipole–dipole interaction

Prescinding from spontelectric phenomena, the vibrational absorption infrared spectrum is a common tool in experiments with ices which can also be calculated using the molecular-dynamics method in order to draw further links between both the experimental and theoretical perspective. Lastly, besides the current focus on intermolecular interactions, irradiation-driven molecular-dynamics [4] represents a novel approach to study astrochemical and other systems where intramolecular topology alterations, i.e., bond formation or breakage, are encountered.

Author contributions

CK and AVS designed and discussed the research project. CK carried out the calculation and analyzed the data. CK drafted the article. CK and AVS contributed to the manuscript. AVS supervised the research.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: Data can be shared upon request.]

A Intramolecular parameters

For the molecular-dynamics simulation besides the intermolecular non-bonded parameters also, intramolecular parameters are needed which are however not central to this work. A selection is given in Tables 3, 4 and 5.

Table 3 CGenFF covalent bond parameters

No.	Description	c_b in kcal/mol/A^2	R_b in A
$1 \leftrightarrow 3$	Methyl acetate	340	1.430
$1 \leftrightarrow 4$	Methyl acetate	150	1.334
$2 \leftrightarrow 4$	Methyl acetate	750	1.220
$4 \leftrightarrow 8$	Formic acid	348	1.096
$3 \leftrightarrow 5,6,7$	Alkane	322	1.111

Table 4 CGenFF intramolecular angle parameters

No.	Description	c_a	θ_a in deg
$1 \leftarrow 4 \rightarrow 2$	Acetic acid	90.0	125.9
$1 \leftarrow 4 \rightarrow 8$	Formic acid	47.0	105.0
$2 \leftarrow 4 \rightarrow 8$	Formic acid	39.0	119.0
$3 \leftarrow 1 \rightarrow 4$	Methyl acetate	40.0	109.6
$1 \leftarrow 3 \rightarrow 5,6,7$	Methyl acetate	60.0	109.5
$5 \leftarrow 3 \rightarrow 6,7$	Alkane	35.5	108.4
$6 \leftarrow 3 \rightarrow 7$	Alkane	35.5	108.4

The spring constant is given in units of kcal/mol/deg^2

Table 5 CGenFF dihedral angle parameters

No.	Description	c_d	n	δ
$3 \leftarrow 1 \leftrightarrow 4 \rightarrow 2$	Methyl acetate	0.965	1	180
$3 \leftarrow 1 \leftrightarrow 4 \rightarrow 2$	Methyl acetate	3.850	2	180
$5 \leftarrow 3 \leftrightarrow 1 \rightarrow 4$	Diethylether	0.284	3	0

The spring constant is detailed in units of kcal/mol

References

1. C. Kexel, S. Schramm, A.V. Solov'yov, Eur. Phys. J. B **88**, 221 (2015)
2. C. Kexel, A.V. Verkhovtsev, G.B. Sushko, A.V. Korol, S. Schramm, A.V. Solov'yov, J. Phys. Chem. C **120**, 25043 (2016)
3. G.B. Sushko, I.A. Solov'yov, A.V. Verkhovtsev, S.N. Volkov, A.V. Solov'yov, Eur. Phys. J. D **70**, 12 (2016)
4. G.B. Sushko, I.A. Solov'yov, A.V. Solov'yov, Eur. Phys. J. D **70**, 217 (2016)
5. I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Phys. Rev. A **65**, 053203 (2002)
6. A. Lyalin, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Phys. Rev. A **67**, 063203 (2003)
7. D.J. Burke, W.A. Brown, Phys. Chem. Chem. Phys. **12**, 5947 (2010)
8. A. Tielens, Rev. Mod. Phys. **85**, 1021 (2013)
9. E. Herbst, Phys. Chem. Chem. Phys. **16**, 3344 (2014)
10. S. Grimme, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 211 (2011)
11. T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, K. Ruud, Chem. Rev. **112**, 543 (2012)
12. I.A. Solov'yov, A.V. Yakubovich, P.V. Nikolaev, I. Volkovets, A.V. Solov'yov, J. Comput. Chem. **33**, 2412 (2012)
13. G. Sushko, A. Verkhovtsev, C. Kexel, A.V. Korol, S. Schramm, A.V. Solov'yov, J. Phys. Condens. Matter **28**, 145201 (2016)
14. C.M. Handley, J. Behler, Eur. Phys. J. B **87**, 152 (2014)
15. V. Oiko, T. Mathieu, L. Cao, J. Liu, R. Palmer, J. Chem. Phys. **145**, 166101 (2016)

16. P. Ghesquière, T. Mineva, D. Talbi, P. Theulé, J. Noble, T. Chiavassa, Phys. Chem. Chem. Phys. **17**, 11455 (2015)
17. V. Wakelam, J.C. Loison, E. Herbst, B. Pavone, A. Bergeat, K. Béroff, M. Chabot, A. Faure, D. Galli, W.D. Geppert et al., Astrophys. J. Suppl. Ser. **217**, 20 (2015)
18. R.C. Fortenberry, X. Huang, T.D. Crawford, T.J. Lee, Astrophys. J. **796**, 139 (2014)
19. I.I. Mizus, A.A. Kyuberis, N.F. Zobov, V.Y. Makhnev, O.L. Polyansky, J. Tennyson, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **376**, 20170149 (2018)
20. O. Plekan, A. Rosu-Finsen, A.M. Cassidy, J. Lasne, M.R. McCoustra, D. Field, Eur. Phys. J. D **71**, 162 (2017)
21. O. Plekan, A. Cassidy, R. Balog, N.C. Jones, D. Field, Phys. Chem. Chem. Phys. **14**, 9972 (2012)
22. Y. Nunes, G. Martins, N. Mason, D. Duflot, S.V. Hoffmann, J. Delwiche, M.J. Hubin-Franskin, P. Limao-Vieira, Phys. Chem. Chem. Phys. **12**, 15734 (2010)
23. L. Feketeová, A. Pelc, A. Ribar, S. Huber, S. Denifl, Astron. Astrophys. **617**, A102 (2018)
24. T.S. Thakur, M.T. Kirchner, D. Bläser, R. Boese, G.R. Desiraju, Phys. Chem. Chem. Phys. **13**, 14076 (2011)
25. P. Modica, M. Palumbo, Astron. Astrophys. **519**, A22 (2010)
26. D.J. Burke, F. Puletti, P.M. Woods, S. Viti, B. Slater, W.A. Brown, J. Phys. Chem. A **119**, 6837 (2015)
27. M. Roman, A. Dunn, S. Taj, Z. Keolopile, A. Rosu-Finsen, M. Gutowski, M. McCoustra, A. Cassidy, D. Field, Phys. Chem. Chem. Phys. **20**, 29038 (2018)
28. T.H. DunningJr, J. Chem. Phys. **90**, 1007 (1989)
29. M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson et al., Wallingford. CT **6492** (2009)
30. A.D. MacKerellJr, D. Bashford, M. Bellott, R.L. DunbrackJr, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha et al., J. Phys. Chem. B **102**, 3586 (1998)
31. K. Vanommeslaeghe, A.D. MacKerell Jr., J. Chem. Inf. Model. **52**, 3144 (2012)
32. J. Hernández-Rojas, F. Calvo, Front. Chem. **7**, 573 (2019)
33. D.J. Wales, J.P. Doye, J. Phys. Chem. A **101**, 5111 (1997)
34. Z. Li, H.A. Scheraga, Proc. Natl. Acad. Sci. **84**, 6611 (1987)
35. H.G. Beyer, H.P. Schwefel, Nat. Comput. **1**, 3 (2002)
36. J. Hernández-Rojas, F. Calvo, D. Wales, Phys. Chem. Chem. Phys. **18**, 13736 (2016)
37. S. Tsuzuki, T. Uchimaru, K. Matsumura, M. Mikami, K. Tanabe, J. Chem. Phys. **110**, 11906 (1999)
38. S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, J. Chem. Phys. **124**, 114304 (2006)
39. S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, J. Am. Chem. Soc. **124**, 104 (2002)
40. C. Vega, J.L. Abascal, Phys. Chem. Chem. Phys. **13**, 19663 (2011)
41. C. Vega, Mol. Phys. **113**, 1145 (2015)
42. R. Curl Jr., J. Chem. Phys. **30**, 1529 (1959)
43. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, J. Comput. Chem. **26**, 1781 (2005)