

Ab Initio Optimization of System-specific Dispersion Corrections for Ab Initio Molecular Dynamics Simulations: Acetonitrile as a Test Case

*Nuno Barbosa,¹ Marco Pagliai,² Sourab Sinha,¹ Vincenzo Barone,^{1,3} Dario Alfé^{4,5,6} and Giuseppe
Brancato^{1,3*}*

¹Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy and CSGI.

²Dipartimento di Chimica “Ugo Schiff”, Università degli Studi di Firenze, Via della Lastruccia 3,
50019 Sesto Fiorentino, Italy.

³Istituto Nazionale di Fisica Nucleare (INFN) sezione di Pisa, Largo Bruno Pontecorvo 3, 56127
Pisa, Italy.

⁴Department of Earth Sciences, Thomas Young Center, University College London, 5 Gower
Place, London WC1E 6BS, United Kingdom.

⁵London Centre for Nanotechnology, Thomas Young Centre, University College London, 17-19
Gordon Street, London WC1H 0AH, United Kingdom.

⁶Dipartimento di Fisica Ettore Pancini, Università di Napoli Federico II, Monte S. Angelo, 80126
Napoli, Italy.

*Corresponding author:

GB, E-mail: giuseppe.brancato@sns.it

KEYWORDS: Acetonitrile, Density Functional Theory, Ab Initio Molecular Dynamics,
Dispersion Corrections, DFT-D3

Abstract

Grimme's dispersion-corrected density functional theory (DFT-D) methods have emerged among the most practical approaches to perform accurate quantum mechanical calculations on molecular systems ranging from small clusters to microscopic and mesoscopic samples, i.e., including hundreds or thousands of molecules. Moreover, DFT-D functionals can be easily integrated into popular *ab initio* molecular dynamics (MD) software packages to carry out first-principle condensed phase simulations at an affordable computational cost. Here, starting from the well-established D3 version of the dispersion correction term, we present a simple protocol to improve the accurate description of the intermolecular interactions of molecular clusters of growing size, considering acetonitrile as a test case. Optimization of the interaction energy was performed with reference to diffusion quantum Monte Carlo calculations, successfully reaching the same inherent accuracy of the latter (statistical error of ~ 0.1 kcal/mol per molecule). The refined DFT-D3 model was then used to perform *ab initio* MD simulations of liquid acetonitrile, showing again significant improvements towards available experimental data with respect to the default correction.

1 Introduction

Poor description of van der Waals interactions by standard density functional theory (DFT) approximations has boosted the development of new improved theoretical and empirical models,(1–4) especially for the treatment of extended molecular systems of great interest for life and material sciences (see Ref. (5) for a review). Among others, the so-called dispersion-corrected DFT (DFT-D) methods, as proposed by Grimme and coworkers,(4, 6, 7) have emerged as some of the most versatile, accurate, and computationally efficient approaches for modeling and simulating large molecular systems. The DFT-D3(4) approach, in particular, is a well-established atom

pairwise correction, which has been successfully tested for calculations on various molecular complexes and architectures.(8, 9) The general assumption underlying all the DFT-D variants is that the dispersion energy can be added to the electronic energy obtained from standard Kohn-Sham DFT as a “classical” (i.e. independent from the electronic structure) interatomic potential, including two or more high-order multipole interaction terms (typically, C_6/R^6 , C_8/R^8 , and so on) modulated by further damping functions and scaling factors. Such semi-classical corrections have proved very valuable, if not unavoidable with respect to the uncorrected DFT, for the proper calculations of structural as well as thermodynamic properties of a wide range of chemical systems.(6, 8, 9) In fact, it has been unequivocally shown that well-known and widely used DFT approximations (e.g., B3LYP(10, 11)) can fail badly when used to model simple molecular complexes dominated by van der Waals interactions.(5) However, DFT-D models can deliver satisfactory results only if carefully parametrized by fitting a few adjustable parameters, which are generally dependent on the specific density functional (parameters for more than 80 functionals are available), to accurate reference data. In this context, typical benchmark sets range from high-level coupled cluster calculations of small organic complexes(12) to experimentally determined association energies of large supramolecular systems(13) and to sublimation energies of molecular crystals,(14) though experimental energies always require some a posteriori corrections in order to be compared to single-point energy calculations. At least for medium-sized molecular systems (i.e., few hundred atoms), (local) correlated wave function and quantum Monte Carlo methods are nowadays feasible,(15, 16) with the latter showing a favorable cubic scaling with the number of electrons. A common feature of molecular test sets is the use of isolated equilibrium structures in order to avoid the perturbing effects of thermal fluctuations and of the chemical environment. This is a sensible choice for assessing systematically a large number of different electronic structure

models towards a balanced selection of assorted chemical systems. However, in our opinion, alternative approaches could be more convenient for developing dispersion-corrected DFT models tailored to the accurate description of condensed phase systems, especially liquids. This is an active area of interest, since DFT-based *ab initio* molecular dynamics (AIMD) techniques offer the advantage of an explicit electronic treatment as compared to force-field based simulations, though limited in their time-scale reach. In past studies, for example, dispersion corrections have been successfully tested in AIMD simulations of liquid water,(17, 18) demonstrating the beneficial inclusion of London interactions for reproducing basic liquid properties.

Here, we aimed at assessing the use of DFT-D3 as an accurate computational model for the consistent description of non-covalent interactions of a given molecular system when going from microscopic clusters to the liquid phase, taking acetonitrile as a test case. We adopted the D3 version since it is more popular among DFT and AIMD software packages but the same computational protocol could be easily extended to other updated DFT-D variants, such as DFT-D4.(19, 20) To this end, we devised a computational protocol summarized in **Figure 1**. In contrast to typical test sets that include mostly equilibrium structures and binary systems, we generated multiple non-equilibrium configurations of molecular clusters of growing size, ranging from dimers to hexadecamers, as obtained from corresponding classical molecular dynamics (MD) simulation of liquid acetonitrile. For this task, we think that carefully developed force fields are particularly well suited for sampling liquid-like uncorrelated configurations. Overall, the interaction energy of 500 cluster configurations were tested towards diffusion quantum Monte Carlo (DMC) benchmark data purposely performed for this work. In fact, we believe that validating DFT-D approaches on molecular aggregates of variable dimension may help to emphasize possible flaws in the description of the London dispersion interactions. Then, we

refined the D3 parametrization of two very popular DFT approximations, i.e. BLYP and B3LYP, aiming at improving the agreement with DMC interaction energies. We found that optimization of the dispersion correction term (namely, the S_8 parameter) accounting for the $n > 6$ order multipole interactions at short/medium range distances led to a significant improvement of the results, which showed mean absolute deviations within the accuracy of reference data (0.1 kcal/mol per molecule). Finally, the optimized BLYP-D3 model, as well as the default one and the standard uncorrected BLYP, were used to perform *ab initio* MD simulations of liquid acetonitrile, showing again remarkable improvements towards available experimental data.

2 Methods

1.1 Molecular cluster configurations

Various molecular cluster configurations were obtained by extracting molecular assemblies of different sizes (i.e., from dimers to hexadecamers) from a large number of snapshots issuing from a NpT ensemble MD simulation of liquid acetonitrile carried out at normal conditions ($T = 300$ K, $p = 1$ atm). The recently optimized acetonitrile molecular model of Nikitin and Lyubartsev(21) was adopted for this purpose, since it can reproduce fairly well all main properties of the liquid, such as structure, density and thermodynamics. All bonds were constrained using the LINCS(22) algorithm and simulations were performed using a timestep of 1 fs. As a result, the following cluster configurations were obtained: 150 dimers (i.e., 2-mer), 150 tetramers (i.e., 4-mer), 100 hexamers (i.e., 6-mer), 50 octamers (i.e., 8-mer) and 50 hexadecamers (i.e., 16-mer). The GROMACS(23) software package was used for all MD simulations.

1.2 Diffusion Monte Carlo

Diffusion Monte Carlo calculations were performed with the Casino code(24), using trial wavefunctions of the Slater-Jastrow type:

$$\Psi_T(\mathbf{R}) = D^\uparrow D^\downarrow e^J \quad (1)$$

where D^\uparrow and D^\downarrow are Slater determinants of up- and down-spin single-electron orbitals, and e^J is the so called Jastrow factor, which is the exponential of a sum of one-body (electron-nucleus), two-body (electron-electron) and three-body (electron-electron-nucleus) terms(25). Imaginary time evolution of the Schrödinger equation was performed with the usual short time approximation, using the T-move scheme(26).

We used Dirac-Fock pseudo-potentials (PP) for C, N and H(27). The C and N PPs have a frozen He core and core radii of 0.58 and 0.44 Å, respectively. The H PP has a core radius of 0.26 Å. The single particle orbitals were obtained by DFT plane-wave (PW) calculations using the local density approximation (LDA) and a PW cutoff of 500 Ry, using the pwscf package(28), and re-expanded in terms of B-splines(29), using the natural B-spline grid spacing given by $a = \pi/G_{\max}$, where G_{\max} is the length of the largest vector employed in the PW calculations. The PW calculations were performed by putting the systems in boxes with at least 5 Å of empty space in each direction. The DMC calculations were then performed with no periodic boundary conditions (PBC), using the Ewald interaction to model electron-electron interactions. Note that switching off PBC eliminates finite size effects due to periodic images, which in a many-body technique such as DMC would decay more slowly with the size of the simulation cell. To investigate convergence of the binding energy with respect to time step we repeated the calculations on 10 different dimer configurations using time steps of 0.005 and 0.002 a.u., which showed differences in the binding energies of less than the statistical error of ~ 5 meV/dimer. We therefore decided to use a time step of 0.005 a.u.

1.3 DFT and dispersion corrections

In the next step of this study, single-point energy calculations at the DFT level of theory were carried out on the acetonitrile molecular clusters issuing from MD simulations. DFT calculations were performed with Gaussian09(30) using a combination of Becke's exchange functional(31) with the correlation function LYP(32, 33) and the B3LYP(34) hybrid functional. For the sake of comparison, also the M062X and the double-hybrid B2PLYP functional were considered in a few calculations. 6-31+G(d,p), 6-31+D(2d,2p) and Dunning's correlation (aug-cc-pVTZ)(35) basis sets were used. Single-point energies were corrected using the original Grimme's D3 dispersion correction term. Grimme and coworkers showed that D3 is less empirical than previous D1 and D2 corrections, showing a better overall accuracy, as well as dispersion coefficients computed explicitly(4). Total energy (i.e., including dispersion corrections) can be described as:

$$E_{DFT-D3} = E_{DFT} - E_{dis} \quad (2)$$

where E_{dis} is expressed as (neglecting the three-body or higher terms):

$$E_{dis} = \sum_{a,b}^N \sum_{n=6,8,\dots} S_n \frac{C_n^{ab}}{r_{ab}^n} f_{d,n}(r_{ab}) \quad (3)$$

where C_n^{ab} is the n -th order dispersion coefficient (orders $n = 6, 8, \dots$) defined for any given atom pair (a, b) in the system, r_{ab} is the internuclear atom pair distance, $f_{d,n}(r_{ab})$ is a damping function introduced to avoid singularities at small interatomic distances, and S_n are scaling factors (typically, dependent on the DFT method). The damping function has the form:

$$f_{d,n}(r_{ab}) = \frac{1}{1 + 6 \left(\frac{r_{ab}}{(S_{R,n} R_0^{ab})} \right)^{-\alpha_n}} \quad (4)$$

with $S_{R,n}$ the order-dependent scaling factor of the cutoff radii R_0^{ab} and α_n the steepness parameter. For a detailed discussion of the meaning and definition of all parameters see Ref. (4). Here, we

note that in practical implementations of DFT-D3 the n -th order is usually truncated after $n = 8$ and most of the parameters are computed *ab initio* (C_6^{ab}), derived recursively (C_8^{ab}), or kept fixed (e.g., $S_{R,8}$ and S_6 are set to 1 for all DFT methods, except those accounting for dispersion energy). On the other hand, S_8 and $S_{R,6}$ are empirical parameters, which are adjusted based upon the density functional (e.g., for B3LYP, $S_8 = 1.703$ and $S_{R,6} = 1.261$; for BLYP, $S_8 = 1.682$ and $S_{R,6} = 1.094$). In particular, the S_8 scaling factor is considered to account implicitly for higher multipolar terms beyond the dipole-dipole contribution. For the purpose of further testing, the D3BJ(36) (which includes the Becke and Johnson damping function) and D4(19, 20) variants were also considered.

1.4 Optimization procedure

The optimization procedure of the dispersion correction term (Eq. 3) was tailored to minimize the deviation in the computed cluster interaction energies between DMC and DFT as issuing from calculations on large sets of molecular clusters of growing size ($n = 2-16$). For each cluster configuration, such an interaction energy deviation is defined by subtracting the 1-body energy deviation, ΔE_{1-body}^n , from the total interaction energy difference, ΔE^n , as follows:

$$\Delta E^n - \Delta E_{1-body}^n \quad (5)$$

where:

$$\begin{aligned} \Delta E^n &= (\Delta E_{DFT}^n - \Delta E_{DMC}^n) \\ \Delta E_{1-body}^n &= (\Delta E_{1-body,DFT}^n - \Delta E_{1-body,DMC}^n) \end{aligned} \quad (6)$$

while for the 1-body and total interaction energy of the corresponding electronic structure calculation, we have:

$$\Delta E_X^n = E_X^n - nE_X^{ref} \quad (7)$$

$$\Delta E_{1-body,X}^n = \sum_i^n (E_X^i - E_X^{ref})$$

E_X^n is the total energy of the n -th cluster (with $n = 2-16$) configuration computed at the X (= DMC, DFT, DFT-D3) level of theory, E_X^{ref} is the energy of the isolated molecule at the reference geometry (according to the force field model, geometry is: CN=1.157 Å, CC=1.458 Å, CH=1.090 Å, HCH=109.5° and HCC=110.0°) computed at the same level of theory, and E_X^i is the energy of the isolated i -th (with $i = 1-n$) molecule (possibly distorted) taken from the cluster configuration. Note that in the present work, since the individual molecules of the generated cluster structures had nearly ideal (reference) geometry (i.e., all bonds were constrained), the contribution of the ΔE_{1-Body}^n was found to be negligible (see below) and, therefore, it was ignored during the optimization procedure. Note that for the assessment of the S_8 scaling factor, a convenient optimization procedure could exploit a global fitting towards the results issuing from all cluster systems. In the present case, however, the S_8 optimization performed through a simple grid search on the 8mer system led to a consistent correction readily extended to all other cluster sizes (vide infra).

1.5 *Ab initio* molecular dynamics simulations

Ab initio molecular dynamics (AIMD) simulations with the Born-Oppenheimer method were carried out with the QuickStep(37) module of the CP2K suite of programs(38) to study the structural and thermodynamic properties of liquid acetonitrile. The interaction potential was computed within density functional theory (DFT), employing the BLYP(31, 32) exchange and correlation functional. The TZV2P basis set in conjunction with Goedecker-Teter-Hutter(GTH)(39) pseudopotentials and a plane wave cutoff of 400 Ry was adopted to describe the electronic structure of the systems. The BLYP-D3 method proposed by Grimme *et al.*(4, 36) was

employed to better describe the dispersion forces using both the default and the modified form of D3 (with the S_8 parameter multiplied by a 0.7253 factor, corresponding to 1.22). A further AIMD simulation was carried out without van der Waals corrections, using standard BLYP. All simulations were performed using a fixed periodic cubic box with 17.6982 Å edge containing 64 acetonitrile molecules, thus corresponding to the experimental density of 0.786 g/cm³. All systems were thermalized by rescaling the velocities at ambient temperature ($T = 300$ K), while performing constant-volume simulations with a time step of 0.1 fs for 8 ps.

3 Results and Discussion

1.6 Assessment of the D3 dispersion energy correction

In the present study, a large set of acetonitrile molecular clusters (i.e., 500 configurations) was considered to assess and then refine the effect of Grimme's D3 correction on the interaction energy computed at DFT (i.e., B3LYP and BLYP) level of theory, as compared to high-level DMC calculations. To assess the extent of the correction, we report in [Figure 2a](#) the discrepancy between standard and corrected B3LYP and BLYP calculations with respect to DMC on a number of representative 8-mer structures, using in both cases the 6-31+G(d,p) basis set. While standard B3LYP and BLYP underestimate, on average, the total interaction energy by about 11.5 and 14.4 kcal/mol, the corresponding DFT-D3 calculations appear to overestimate the same energy by 2.4 kcal/mol, thus showing a notable improvement with respect to the uncorrected DFT energies. This result, as expected, demonstrates the capability of DFT-D3 to effectively recover the missing dispersion energy. Also, it is worth noting the decrease on the energy fluctuations upon introduction of the dispersion corrections, the standard deviation being reduced by half (from ~2 kcal/mol to less than ~1 kcal/mol). Interestingly, when DFT-D3 calculations were performed by setting to zero the $n = 8$ order term (i.e., $S_8 = 0$) in [Eq. 3](#), the overall energy correction was reduced

by half (i.e., on average 48%, [Figure S1](#)), a result showing that, at least for acetonitrile, the latter contribution is quantitatively similar to the $n = 6$ term. For the sake of comparison, we carried out further calculations at M062X-D3 and B2PLYP-D3 level, which showed again some deviations with respect to benchmark calculations, though less pronounced for the former functional ([Figure S2](#)). Moreover, we also tested the D3BJ(36) variant, which includes the Becke and Johnson damping function, and the D4(19, 20) variant, which has updated C6 coefficients, BJ damping function and a three-body interaction term. Results are reported in [Figure S3](#). We noticed that the use of the BJ damping function, even though it is generally recommended, did not change significantly the interaction energy of the acetonitrile clusters with respect to the standard “zero damping” formula, in line with what was observed in ref. (36) for non-covalent systems. D4 showed, on the other hand, a sensible improvement in the description of the dispersion interactions, though deviations with respect to reference data were still present (for B3LYP, MAE = 1.16 kcal/mol; for BLYP, MAE = 0.63 kcal/mol).

In previous results, the total interaction energies (ΔE^8), including the one-body term (ΔE_{1-body}^8), were reported. As a matter of fact, this was justified by the observation that energy deviations of individual monomers from test calculations were very small (i.e., average error was 0.05 kcal/mol, [Figure S4](#)), possibly cancelling each other in larger clusters. The role of ΔE_{1-body}^8 in the context of this work was reconsidered in the following.

In order to further assess the impact of the default D3 correction on clusters of variable size and to better estimate the effect of the basis set, we carried out similar calculations at B3LYP-D3 and BLYP-D3 level of theory by considering molecular structures ranging from dimers to hexadecamers and by comparing the 6-31+G(d,p), 6-31+D(2d,2p), and aug-cc-pVTZ basis set. Results are summarized in [Table S1](#) and [Figure 3](#). First of all, it is apparent the steady increase of

the energy deviation between DFT-D3 and DMC with system size: the mean absolute error (MAE) shows a roughly linear trend going from 2mer to 16mer, with a parallel increase of both B3LYP and BLYP results as depicted in [Figure 2b](#). For the largest clusters considered (i.e., 16mer), MAE is over 5 kcal/mol. More specifically, the MAE per molecule does show an increment from 0.1-0.2 kcal/mol (i.e., 2mer) to over 0.3 kcal/mol (16mer) (see inset of [Figure 2b](#)). Though such energy discrepancies seem not too large, the observed increase with cluster size suggests that the overestimation of the dispersion interaction by DFT-D3 may become more relevant when going to larger acetonitrile clusters or to mesoscopic samples. Overall, the same trend was observed by comparing the three basis sets, with minor changes in the energy deviations between DFT-D3 and DMC (note, however, that the most extended 16mer cluster was excluded in the comparison). Therefore, though present, the basis-set superposition error (BSSE) appeared in this case of less importance with respect to the inaccuracy of the combined density functional/dispersion correction model. Incidentally, the smaller basis set reported the lower MAE. In the following optimization procedure, for the sake of convenience, we decided to employ the 6-31+G(d,p) basis set, so to effectively correct energy deviations of DFT as due to poor description of the dispersion interactions. In view of the observed minor differences with respect to the basis set choice (i.e., for B3LYP, MAE per molecule is 0.06 kcal/mol between 6-31+G(d,p) and aug-cc-pVTZ), the optimization procedure is to be considered essentially not dependent on the basis set.

1.7 Optimization of the D3 term

At this point, one can ask if the discrepancy between DFT-D3 and DMC can be significantly reduced by refining one or more parameters of the D3 empirical dispersion term ([Eq. 3](#)) and whether such an optimization can improve consistently results for molecular samples of variable size or be beneficial only for a given cluster dimension. Since the $n = 8$ order term accounts for a

good extent of the overall dispersion energy correction (Figure S1) and the S_8 scaling factor is one of the few empirically adjustable parameters of the Grimme's D3 version (Eq. 3), we decided to refine this parameter by focusing specifically on two popular hybrid and gradient-corrected density functionals, B3LYP and BLYP, respectively, in combination with the 6-31+G(d,p) basis set. Optimization was performed with the purpose to achieve a better agreement between DFT and DMC benchmark calculations. The default values of S_8 for BLYP-D3 and B3LYP-D3 are 1.682 and 1.703, respectively. In preliminary tests, we focused only on the octamer cluster since it represents a good compromise between molecular size and computational cost (i.e., especially for DMC calculations). In particular, S_8 was refined in order to minimize the MAE issuing from calculations on this cluster. Results of MAE obtained from B3LYP-D3 calculations are reported in Table S2, whereas absolute energy deviations for selected S_8 values are depicted in Figure 4. Data presented on Figure 4 are carried out with the same functional and basis set (i.e., B3LYP-D3/6-31+G(d,p)), differing only for the S_8 parameter. The first set of data were generated with the default value ($S_8 = 1.703$) and show the poor agreement with benchmark values as seen above (MAE = 2.4 kcal/mol). Considering the trend reported on Figure S1, it was expected that refinement of S_8 should assume a smaller value than default ($S_8 < 1.7$) to reduce the observed overestimation of the interaction energy. By changing the S_8 parameter from the default value (i.e., 1.703) to ~ 1.20 , a significant decrease of the MAE of about four times, from 2.4 kcal/mol to 0.6 kcal/mol, was obtained for the 8mer cluster (Table S2). Though results for $S_8 = 1.18-1.23$ appeared rather similar, we took $S_8 = 1.22$ as the optimal value of the scaling factor. In this case, the MAE per molecule is only 0.07 kcal/mol, a satisfactory result considering that the estimated accuracy of DMC is about 0.1 kcal/mol. In addition, the standard deviation was also somewhat reduced (from 0.80 to 0.69 kcal/mol), as displayed in Figure 4. Despite the very low MAE, we decided to better

estimate the interaction energy of the five 8mer configurations leading to maximum energy deviations (i.e., $|\Delta E|$ greater than 1 kcal/mol), by subtracting the contribution of the 1-body energy term (ΔE_{1-body}^n). Results are reported in [Figure 5](#), which shows a marked reduction of the energy deviation by about 40%, with $|\Delta E| < 1$ kcal/mol in all cases except one. Hence, while the contribution to the total energy deviation of the 1-body term was generally small for the tested acetonitrile cluster configurations (about 0.05 kcal/mol per monomer) and could be safely ignored in the optimization procedure outlined above, an even better agreement with DMC benchmark calculations was achieved by properly taking into account ΔE_{1-body}^n in the evaluation of the interaction energies. Nevertheless, since results of the optimized DFT-D3 model appeared to be within the limit of accuracy of DMC (MAE/molecule = ~ 0.1 kcal/mol), the effect of ΔE_{1-body}^n was neglected in the following.

Furthermore, B3LYP-D3 calculations with the refined scaling factor ($S_8 = 1.22$) were carried out for all clusters considered in the present work in order to validate the improvement in the computed interaction energy on molecular assemblies of growing size. The obtained results were compared with default S_8 calculations and reported in [Table 1](#) and [Figure 6-7](#). Overall, we noted a remarkable agreement with DMC results at any size, energy deviations being effectively minimized by a factor of 4 in the largest clusters. Moreover, the MAE per molecule never exceeded 0.1 kcal/mol, thus showing a rather flat trend versus system size as depicted in [Figure 7](#). These findings suggest that the observed improvement could be reasonably projected onto even larger clusters and/or liquid phase systems for which high-level quantum mechanical calculations are unfeasible.

The same general approach was also performed to obtain an effective S_8 parameter to be used in combination with the BLYP functional. Upon optimization, a value of 1.18 was set for the

present scaling factor (Table S3). As shown in Table 1, it was found that MAE became as low as 0.69 kcal/mol for the octamer cluster after applying the new refined S8 parameter, meaning a 70% rectification when compared to the default parameter (i.e., from 2.32 to 0.69 kcal/mol). By extending the refined BLYP-D3 model to all acetonitrile configurations, we found again an overall better agreement with respect to DMC (Table 1).

1.8 Liquid acetonitrile: structural properties and pressure

Structural properties of liquid acetonitrile were determined experimentally by means of X-ray and neutron diffractions studies.(40, 41) Previously, such results were adopted as a benchmark for developing effective interaction potentials for molecular simulations.(21, 42–47) In particular, it was observed that the models proposed by Böhm *et al.*(42, 43) and Edwards *et al.*(44) provided structural information in very good agreement with experiments. Some liquid acetonitrile properties, such as density, dielectric constant, and enthalpy of vaporization are accurately reproduced by using the interaction potential developed by Gee *et al.*(47) which shows pair radial distribution functions (RDF) close to those obtained by Edwards *et al.*(44). Hence, the RDFs obtained with the models proposed by Edward *et al.*(44) and Gee *et al.*(47) were taken as reference to assess the accuracy of the AIMD results reported in the present work. Figure 8 shows RDFs for selected intermolecular interactions obtained with AIMD simulations of liquid acetonitrile performed with (i.e., BLYP-D3) and without (i.e., BLYP) van der Waals corrections. Although all the AIMD simulations provided structural results in agreement with the reference models,(44, 47) simulations performed with the addition of the D3 dispersion interactions with the new refined S_8 parameter faithfully reproduced the positions of both maxima and minima.

Then, the effect on the pressure of the dispersion-corrected BLYP was assessed from calculation of the stress tensor on a series of configurations sampled every 0.1 ps during the AIMD

simulations of liquid acetonitrile. Similar attempts were reported in past studies of liquid water.(17, 18) Note that, as usual, pressure calculations of small molecular systems are characterized by very high fluctuations. Hence, according to standard practice, we considered the time evolution of the corresponding running average. In **Figure 9**, the pressure running average is depicted as issuing from all three AIMD simulations. It was observed that a rather high pressure (~ 0.6 GPa) was obtained in the simulation without dispersion corrections, suggesting that BLYP would underestimate the experimental density of acetonitrile. Inclusion of the D3 correction term was expected to decrease the computed pressure as a result of the enhanced attractive interactions. Indeed, the resulting pressure was significantly lower and somewhat negative for BLYP-D3 simulations, showing again that a better agreement with experiments (i.e., ambient pressure) was observed when employing the optimized D3 term accounting for the dispersion energy correction. This also suggests that BLYP-D3 would overestimate the density of liquid acetonitrile at normal conditions, a result somewhat in analogy to what observed for liquid water by Ma *et al.*(18) using Car-Parrinello MD simulations.

4 Conclusions

In the present work, we refined the parametrization of two very popular empirically dispersion-corrected DFT functionals, i.e. BLYP and B3LYP, aiming at improving the description of the interaction energy as occurring in molecular clusters of growing size and, ultimately, in condensed phase systems. Our approach adopted the well tested and computationally efficient Grimme's D3 correction model and used diffusion quantum Monte Carlo calculations as a benchmark, due to the statistically high accuracy of the latter (~ 0.1 kcal/mol). Note, however, that while keeping the same general idea, the present approach can be easily extended to other "versions" of the still growing family of DFT-D methods and/or to other purposely chosen reference calculations. For example,

the more recent D4 correction shares the same two-body dispersion term of D3, as expressed in Eq. 3. In particular, we performed benchmark calculations on a large set of molecular configurations of acetonitrile, ranging from dimers to hexadecamers. We believe this represents a key step of our computational protocol, since, on one hand, it allows to (over-)emphasize the possible flaws affecting both well-established or newly developed DFT-D models (e.g., under- or over-estimation of the London interactions), and, on the other hand, helps to extrapolate results to much larger molecular samples. Results showed that while the default D3 correction works surprisingly well, especially for the cost-effective BLYP functional, there was significant room for improvements on the computed interaction energy of medium to large-sized molecular systems, as compared to DMC data. By optimizing just one parameter of the D3 dispersion energy term, namely the S_8 scaling factor of the $n = 8$ order term, we were able to decrease by a factor of 4 the observed MAE, basically reaching the same accuracy of reference data (0.1 kcal/mol per molecule). Remarkably, the optimized D3 term was easily transferred to AIMD simulations of liquid acetonitrile, where small but notable improvements in the structural (i.e., interatomic pair distribution functions) and thermodynamic (i.e., pressure) properties of the condensed phase system were apparent, thus validating the extrapolation of parameters previously tailored towards molecular clusters of increasing size.

While we are aware that the optimization protocol outlined above is essentially system-dependent, meaning that results are tailored specifically towards a given molecular system, we believe that this is an unavoidable necessity of an empirical correction aiming at high chemical accuracy. Yet, in our opinion, such a computational approach is simple and efficient enough to be applicable to a large number of chemical systems (including, for example, solute-solvent systems,

in which solvent interactions are optimized), whenever high accuracy is required, and the resulting dispersion corrections are readily transferable to widely used QM and AIMD software packages.

AUTHOR INFORMATION

Corresponding Author

*To whom correspondence should be addressed:

Giuseppe Bancato, email: giuseppe.bancato@sns.it

Acknowledgments

This work was funded by MIUR through the PRIN program (contract n° 2012SK7ASN). Technical staff of the HPC Avogadro center is kindly acknowledged for managing the computing facilities at SNS.

Abbreviations

MD: Molecular dynamics, AIMD: ab initio molecular dynamics, DMC: diffusion Monte Carlo

References

1. K. Berland, *et al.*, van der Waals forces in density functional theory: a review of the vdW-DF method. *Rep. Prog. Phys.* **78**, 066501 (2015).
2. A. D. Becke, E. R. Johnson, A density-functional model of the dispersion interaction. *J. Chem. Phys.* **123**, 154101 (2005).
3. A. Tkatchenko, M. Scheffler, Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **102**, 073005 (2009).
4. S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
5. S. Grimme, A. Hansen, J. G. Brandenburg, C. Bannwarth, Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **116**, 5105–5154 (2016).
6. S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **25**, 1463–1473 (2004).

7. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
8. T. Risthaus, S. Grimme, Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **9**, 1580–1591 (2013).
9. S. Spicher, E. Caldeweyher, A. Hansen, S. Grimme, Benchmarking London dispersion corrected density functional theory for noncovalent ion– π interactions. *Phys. Chem. Chem. Phys.* **23**, 11635–11648 (2021).
10. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
11. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).
12. P. Jurečka, J. Šponer, J. Černý, P. Hobza, Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **8**, 1985–1993 (2006).
13. S. Grimme, Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. – Eur. J.* **18**, 9955–9964 (2012).
14. A. Otero-de-la-Roza, E. R. Johnson, A benchmark for non-covalent interactions in solids. *J. Chem. Phys.* **137**, 054103 (2012).
15. A. Ambrosetti, D. Alfè, R. A. DiStasio, A. Tkatchenko, Hard Numbers for Large Molecules: Toward Exact Energetics for Supramolecular Systems. *J. Phys. Chem. Lett.* **5**, 849–855 (2014).
16. J. G. Brandenburg, *et al.*, Physisorption of Water on Graphene: Subchemical Accuracy from Many-Body Electronic Structure Methods. *J. Phys. Chem. Lett.* **10**, 358–368 (2019).
17. J. Schmidt, *et al.*, Isobaric–Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions. *J. Phys. Chem. B* **113**, 11959–11964 (2009).
18. Z. Ma, Y. Zhang, M. E. Tuckerman, Ab initio molecular dynamics study of water at constant pressure using converged basis sets and empirical dispersion corrections. *J. Chem. Phys.* **137**, 044506 (2012).
19. E. Caldeweyher, C. Bannwarth, S. Grimme, Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **147**, 034112 (2017).
20. E. Caldeweyher, *et al.*, A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **150**, 154122 (2019).

21. A. M. Nikitin, A. P. Lyubartsev, New six-site acetonitrile model for simulations of liquid acetonitrile and its aqueous mixtures. *J. Comput. Chem.* **28**, 2020–2026 (2007).
22. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
23. H. J. C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
24. R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys. Condens. Matter* **22**, 023201 (2010).
25. L. Mitáš, E. L. Shirley, D. M. Ceperley, Nonlocal pseudopotentials and diffusion Monte Carlo. *J. Chem. Phys.* **95**, 3467 (1991).
26. M. Casula, Beyond the locality approximation in the standard diffusion Monte Carlo method. *Phys. Rev. B* **74** (2006).
27. J. R. Trail, R. J. Needs, Smooth relativistic Hartree–Fock pseudopotentials for H to Ba and Lu to Hg. *J. Chem. Phys.* **122**, 174109 (2005).
28. S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73**, 515–562 (2001).
29. D. Alfè, M. J. Gillan, Efficient localized basis set for quantum Monte Carlo calculations on condensed matter. *Phys. Rev. B* **70** (2004).
30. A. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09w reference. *Wallingford USA 25p* (2009).
31. A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).
32. C. Lee, W. Yang, R. G. Parr, Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
33. B. Miehlich, A. Savin, H. Stoll, H. Preuss, Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **157**, 200–206 (1989).

34. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648 (1993).
35. T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007 (1989).
36. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
37. J. VandeVondele, *et al.*, Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **167**, 103–128 (2005).
38. J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, cp2k: atomistic simulations of condensed matter systems: cp 2 k Simulation Software. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 15–25 (2014).
39. S. Goedecker, M. Teter, J. Hutter, Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703–1710 (1996).
40. H. Bertagnolli, P. Chieux, M. D. Zeidler, A neutron-diffraction study of liquid acetonitrile: I. CD 3 C 14 N. *Mol. Phys.* **32**, 759–773 (1976).
41. H. Bertagnolli, M. D. Zeidler, Molecular pair-correlation function of liquid acetonitrile from X-ray and neutron-diffraction studies. *Mol. Phys.* **35**, 177–192 (1978).
42. H. J. Böhm, I. R. McDonald, P. A. Madden, An effective pair potential for liquid acetonitrile. *Mol. Phys.* **49**, 347–360 (1983).
43. H. J. Böhm, C. Meissner, R. Ahlrichs, Molecular dynamics simulation of liquid CH₃F, CHF₃, CH₃Cl, CH₃CN, CO₂ and CS₂ with new pair potentials. *Mol. Phys.* **53**, 651–672 (1984).
44. D. M. F. Edwards, P. A. Madden, I. R. McDonald, A computer simulation study of the dielectric properties of a model of methyl cyanide: I. The rigid dipole case. *Mol. Phys.* **51**, 1141–1161 (1984).
45. W. L. Jorgensen, J. M. Briggs, Monte Carlo simulations of liquid acetonitrile with a three-site model. *Mol. Phys.* **63**, 547–558 (1988).
46. E. Guàrdia, R. Pinzón, J. Casulleras, M. Orozco, F. J. Luque, Comparison of Different Three-site Interaction Potentials for Liquid Acetonitrile. *Mol. Simul.* **26**, 287–306 (2001).
47. P. J. Gee, W. F. van Gunsteren, Acetonitrile revisited: a molecular dynamics study of the liquid phase. *Mol. Phys.* **104**, 477–483 (2006).

Table of Contents Graphic