

Hard Numbers for Large Molecules: Towards Exact Energetics for Supramolecular Systems

Alberto Ambrosetti¹, Dario Alfè², Robert A. DiStasio, Jr.³, and Alexandre Tkatchenko^{1*}

¹*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195, Berlin, Germany*

²*Department of Earth Sciences and Department of Physics and Astronomy and London Centre for Nanotechnology and Thomas Young Centre@UCL, University College, London WC1E6BT, United Kingdom*

³*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States of America*

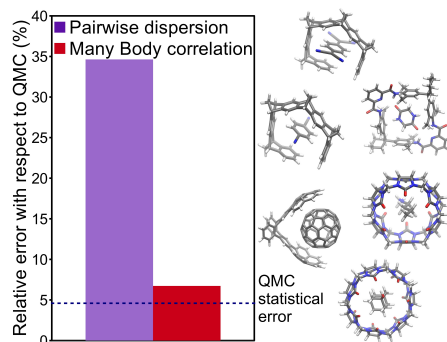
E-mail: tkatchenko@fhi-berlin.mpg.de

Abstract

Non-covalent interactions are ubiquitous in molecular and condensed-phase environments, hence a reliable theoretical description of these fundamental interactions could pave the way towards a more complete understanding of the microscopic underpinnings for a diverse set of systems in chemistry and biology. In this work, we demonstrate that recent algorithmic advances coupled with the availability of large-scale computational resources make the stochastic quantum Monte Carlo approach to solving the Schrödinger equation an optimal contender for attaining "chemical accuracy" (1 kcal/mol) in the binding energies of supramolecular complexes of chemical relevance. To illustrate this point, we considered a select set of seven

*To whom correspondence should be addressed

host-guest complexes, representing the spectrum of non-covalent interactions, including dispersion or van der Waals forces, π - π stacking, hydrogen bonding, hydrophobic interactions, and electrostatic (ion-dipole) attraction. A detailed analysis of the interaction energies reveals that a complete theoretical description necessitates treatment of terms well beyond the standard London and Axilrod-Teller contributions to the van der Waals dispersion energy.



Supramolecular complexes have long been recognized for their remarkable versatility¹⁻⁴ and have therefore become increasingly utilized in a vast array of practical applications, including molecular recognition, self-assembly, template-directed synthesis, and biomimetics.⁴⁻⁸ Countless realizations of supramolecular complexes exist,^{5,9} and typically consist of molecular assemblies stabilized by cooperative binding motifs, with energetic contributions arising from strong covalent and ionic bonds as well as weaker non-bonded intermolecular forces. Therefore a central problem emergent in supramolecular chemistry is characterization and subsequent control of the delicate balance between the different underlying interactions that determine the relative stability of such systems.

Of particular importance in supramolecular chemistry is the class of host-guest complexes, comprised of a host molecule, such as the so-called molecular “tweezers” or “pincers,” and a guest molecule, typically a relatively smaller organic molecule, which are primarily stabilized by non-covalent interactions. Hence, host-guest complexes serve as prototypes for molecular recognition and transient binding events—processes which are primarily dictated by this underlying set of non-covalent interactions. As such, non-covalent interactions play a central role in determin-

ing the functionality of host-guest complexes, with an influence that encompasses conformational energetics, entropic contributions, and solvation effects.

A crucial step in the control and rational design of host-guest complexes is therefore an accurate theoretical description of this underlying set of non-covalent interactions in the absence of complex temperature and environment effects, *i.e.*, “clean-room conditions”, which could provide direct access to the energetics of these supramolecular systems. However, the large size of most functional host-guest complexes of chemical relevance poses enormous challenges for current theoretical methodologies in terms of both accuracy and computational feasibility. In this regard, high-level quantum chemistry methods such as full configuration interaction (FCI) or coupled cluster theory with single, double, and perturbative triple excitations (CCSD(T)), could certainly provide highly accurate binding energies for such systems. In fact, a number of databases of binding energies computed at the CCSD(T) level of theory for small molecular dimers (containing up to a few dozen atoms in size) have recently become available.^{10–12} However, the steep associated computational cost (scaling as N^7 for CCSD(T), where N is a measure of the system size), makes the application of such high-level quantum chemistry methods to large supramolecular systems very challenging, if not impossible, with the computational resources available today.

On the other hand, stochastic methods for solving the Schrödinger equation such as quantum Monte Carlo (QMC) could in principle be utilized to obtain the exact description of molecular ground-state wavefunctions and energies.¹³ In particular, diffusion quantum Monte Carlo (DQMC) represents an optimal approach for treating large systems,^{14,15} since the DQMC method allows for a direct and highly accurate sampling of the ground-state electronic wavefunction, with a more favorable N^3 computational scaling. While the DQMC approach has the ability to describe non-covalent interactions with benchmark (*i.e.*, sub-chemical) accuracy in small molecular dimers,¹⁶ the applicability of DQMC to large supramolecular systems has not been systematically demonstrated to date.

In this work, we considered a select set of six host-guest complexes (see Fig. 1) from the recently proposed S12L database of Grimme,^{17,18} representing the spectrum of non-covalent in-

teractions, including dispersion or vdW forces, π - π stacking, hydrogen bonding, hydrophobic interactions, and electrostatic (ion-dipole) attraction, and ascertained the quality of binding energies obtained via extrapolation of experimental association free energies¹⁷ with respect to benchmark binding energies computed at the DQMC level of theory. Although this comparison revealed a fair degree of overall fidelity between the extrapolated and DQMC binding energies, quantitative differences as large as 3.6 kcal/mol (*e.g.*, for the case of the cucurbit[6]uril–butylammonium cationic complex, **6a** in Fig. 1), which are well above the chemical accuracy benchmark of 1 kcal/mol, persist and are indicative of the inherent limitations in the approximate corrections utilized to extract binding energies from experimentally determined association free energies.

To further investigate the underlying non-covalent interactions determining the stability of host-guest complexes, we performed a many-body decomposition analysis of the long-range correlation energy in the aforementioned systems. Such an analysis is complementary to the DQMC methodology, which provides benchmark energetics for the systems considered herein, yielding detailed physical insight into the fundamental role played by non-covalent interactions in governing supramolecular chemistry. As a result of this analysis, we found that the many-body expansion of the long-range correlation energy is slowly convergent and displays non-trivial behavior, depending on the symmetry and underlying topology of a given host-guest complex, strongly indicating that a chemically accurate theoretical description of supramolecular binding energies requires terms well beyond the standard London (two-) and Axilrod-Teller (three-) body contributions to the dispersion energy. We further investigate this point by extending our analysis to the long-range correlation energy of a double-walled carbon nanotube (DWCNT, see Fig. 2). The marked anisotropy of polarization interactions in this system leads to a reduction of the interwall dispersive binding, which amounts to $\approx 25\%$ with respect to the (isotropic) pairwise vdW energy.

To construct an accurate reference for the energetics in host-guest complexes, we performed DQMC calculations to determine the binding energies for a subset of six complexes from the S12L database¹⁷ (see Fig. 1). This subset (namely **2a**, **2b**, **4a**, **5a**, **6a**, and **7b**, following the original nomenclature of Grimme¹⁷) was selected to represent the broad range of geometries and

non-covalent interactions of primary relevance in supramolecular chemistry, thus preserving the general character of the full S12L database.

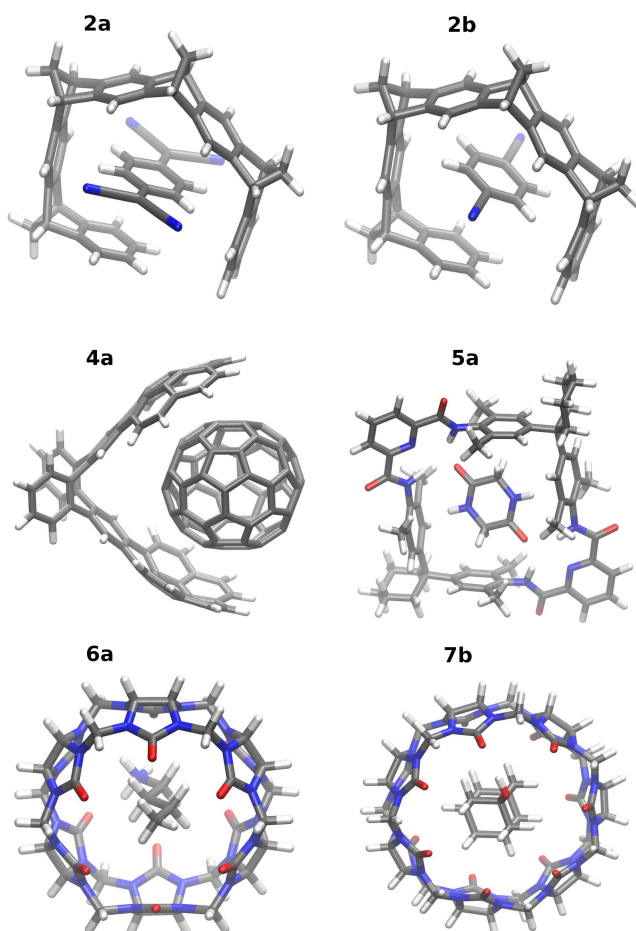


Figure 1: Molecular geometries of the six host-guest complexes studied in this work following the original nomenclature of Grimme in Ref. (17). *From upper-right*, **2a**: tetracyanoquinone–tweezer (TCNQ@tweezer), **2b**: 1,4-dicyanobenzene–tweezer (DCB@tweezer), **4a**: buckyball–catcher (C_{60} @catcher), **5a**: glycine anhydride–macrocycle (GLH@mcyle), **6a**: butylammonium–cucurbit[6]uril cation ($BuNH_4$ @CB6), and **7b**: 1-hydroxyadamantane–cucurbit[7]uril (ADOH@CB7).

The stochastic DQMC electronic structure method is a well-established *ab initio*, or first-principles, approach to solving the Schrödinger equation, and can therefore be utilized in the computation of highly accurate ground-state energies and properties.¹⁴ Since the DQMC methodology intrinsically accounts for dynamical electron correlation effects at all interelectronic separations, DQMC can be considered as a natural benchmark reference for approximate density function-

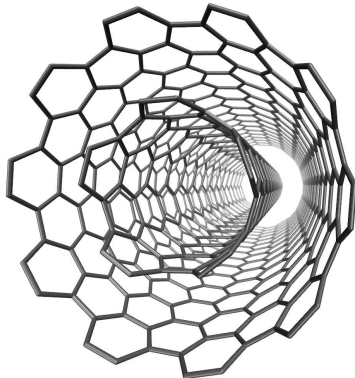


Figure 2: Graphical depiction of a periodic double-walled carbon nanotube, composed of coaxial (10,10) and (5,5) single-walled carbon nanotubes. The supercell shown contains 900 carbon atoms.

als and other perturbative approaches. Exhibiting a favorable computational scaling with system size (N^3), the DQMC method can optimally utilize the computational resources afforded by high-performance massively parallel (super)computer architectures, thereby enabling the challenging large-scale applications carried out in this work.

All DQMC calculations presented herein have been performed utilizing the CASINO suite of programs,¹⁹ employing Slater-Jastrow trial wavefunctions,

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

in which D^\uparrow and D^\downarrow are Slater determinants assembled from single-particle spin orbitals representing the α (up) and β (down) electron spin projections, respectively, and e^J is the so-called Jastrow factor, an exponential comprised of a sum over explicitly correlated one- (electron-nucleus), two- (electron-electron), and three-body (electron-electron-nucleus) terms.

The computed DQMC binding energies for the six host-guest complexes considered in this work are provided in Table 1. The binding energy of the host (H) and guest (G) forming the host-guest complex (H-G), was defined as $\Delta E = E(\text{H-G}) - E(\text{H}) - E(\text{G})$. All DQMC calculations were performed utilizing molecular geometries optimized with dispersion-corrected density functional theory (DFT) in Ref. (18). Convergence tests were performed to verify the dependence of the

computed binding energies on the imaginary time-propagation step. In addition, the reliability of the fixed node approximation was also tested through the use of different trial wavefunctions. As discussed in the Methods section in greater detail, both the time step and nodal errors fall within the statistical uncertainties reported in Table 1 (corresponding to $\pm\sigma$).

While DQMC provides direct and reliable access to benchmark energetics in supramolecular systems, an indirect empirical estimate of the binding energies for the S12L host-guest complexes can also be determined from experimental association free energies, as was recently done by Grimme.¹⁷ Apart from the binding energy, the measured free energies contain many other contributions, such as entropic and solvation effects, which are often of comparable magnitude and opposite sign to the binding energy. For these reasons, the experimental association free energies are roughly one order of magnitude smaller than the corresponding binding energies. Hence, the determination of reliable binding energies from experimentally determined association free energies is a delicate task, which is only further complicated by the need to introduce a number of approximations in the computation of both entropic and solvation contributions, the accuracy of which is often difficult to assess. For example, the solvation effects for the S12L database were computed utilizing a simplified continuum solvent model, while entropic contributions were treated in the rigid rotor-harmonic oscillator (RRHO) approximation, with further approximations required to avoid divergences in the low-frequency regime. Furthermore, the configurational entropy was neglected, *i.e.*, a non-dynamical, single structure approach was used during the computation of binding energies.

To assess the reliability of this approach, the comparison between binding energies computed at the DQMC level of theory and extrapolated utilizing the aforementioned prescription, is also presented in Table 1. From this data set, one can immediately observe absolute differences ranging from 1.4 to 3.6 kcal/mol, with the largest deviations for complexes **2a** (2.4 kcal/mol), **2b** (3.3 kcal/mol), and **6a** (3.6 kcal/mol). In complexes **2a** and **2b**, the host-guest binding is predominantly due to dispersion or vdW forces. Therefore, the host-guest interaction is expected to be responsible for the appearance of soft vibrational modes, with further low-lying vibrational modes appearing

that are related to the host system (*i.e.*, the opening and closing of the “tweezer” moiety). In this regard, anharmonicity, which is neglected in the RRHO approximation, might play an important role in the qualitative and quantitative description of these modes. In addition, the empirical interpolation between the harmonic vibrational and rotational entropy contributions (*i.e.*, to avoid divergences in the low-frequency regime) might also contribute to such large deviations from the DQMC values. For the complex **6a**, the electrostatic cation-dipolar binding is a likely source of inaccuracy for the continuum solvent model—the characterization of the solvent by a macroscopic dielectric function might not be well-suited here due to the polarization effects occurring within such a complex asymmetric system.

We note in passing that the differences between the DQMC and extrapolated binding energies encountered for every host-guest complex considered herein (and in particular, complexes **2a**, **2b** and **6a**) are well above the demanding “chemical accuracy” benchmark of 1 kcal/mol. Despite this fact, the qualitative agreement among these binding energies is certainly remarkable, and constitutes an important “sanity check” between two radically different approaches for determining the gas-phase binding energetics of large supramolecular systems.

The binding energies computed at the DQMC level of theory and presented herein provide the most reliable benchmarks available to date for the energetics of large supramolecular systems of chemical relevance, and can therefore serve as reliable references for the development (and subsequent validation) of computationally efficient approximate electronic structure methods. In fact, a main advantage of utilizing the DQMC method lies in its ability to accurately treat long-range correlation effects—effects that are inherently quantum mechanical, many-body, and non-local in character—which pose quite a challenge for many approximate electronic structure methods. However, the DQMC method alone does not directly allow for a detailed analysis of the various electron correlation contributions to the binding energies in question. Due to the stochastic nature of the sampling of the many-body ground-state wavefunction, the DQMC method computes ground-state energies in a non-perturbative fashion, making a distinction among the different energy components impractical. Therefore, to gain direct physical insight into the role played by the

long-range correlation energy in the stabilization of host-guest complexes, we will make combined use of an alternative and complementary approach, which allows for a detailed analysis of the pairwise and many-body contributions to the long-range correlation energy within the framework of DFT.

Semi-local DFT is a self-consistent quantum-mechanical electronic structure method, which accurately describes electrostatics, induction and hybridization effects, but does not include long-range electron correlation, and therefore fails to account for dispersion or vdW interactions. As discussed in greater detail in the Methods section, we explicitly treat the long-range correlation energy within DFT by utilizing the random-phase approximation (RPA) in the dipole limit²⁰ (obtained through the MBD* method) based on a range-separation of the interelectronic Coulomb potential. The RPA approach seamlessly includes many-body effects in the correlation energy to all orders and is a very accurate theory for the long-range correlation energy, provided that correct polarizabilities are utilized as an input. To efficiently compute the long-range RPA correlation energy, we map the molecular system onto a set of atom-centered quantum harmonic oscillators (QHOs), and utilize an effective oscillator Hamiltonian.²¹ With respect to the recently published MBD method,²¹ MBD* offers an improved description of highly anisotropic systems, primarily due to range-separation of the Coulomb operator and a resultant treatment of the long-range electrodynamic response.

Although the MBD* method is not expected to reach the same degree of accuracy as DQMC, performance beyond chemical accuracy has been demonstrated utilizing this methodology in both small molecules²¹ and extended systems.²² In this regard, the MBD* method can be regarded as complementary to DQMC, which not only provides a detailed many-body decomposition analysis of the long-range correlation energy, but also allows for even more challenging large-scale applications due to its high computational efficiency. Before proceeding any further, a comparison against reliable benchmark data remains essential in order to assess the accuracy of the MBD* method and its predictivity for supramolecular systems.

The computed PBE+MBD* (MBD* coupled with the PBE functional²³) binding energies for

the six host-guest complexes considered in this work are also provided in Table 1 (all DFT calculations were performed with the FHI-aims code²⁴). In general, we observed very good performance across the entire S12L database, with a mean absolute relative error (MARE) of 5.5% computed with respect to the extrapolated experimental values, which is similar to the MARE obtained for smaller gas-phase molecular dimers.²¹ In this regard, it should be emphasized that the PBE+MBD* method, with a corresponding mean absolute error (MAE) of 1.6 kcal/mol over the entire S12L database, provides an accuracy comparable to that of the reference data. In fact, the binding energies computed at the PBE+MBD* level of theory are consistently in closer agreement with the benchmark DQMC results than the extrapolated values: by performing a statistical analysis restricted to the subset of six host-guest complexes considered in this work, we found a MAE of 1.7 kcal/mol with respect to the DQMC results, which should be compared to the MAE of 2.3 kcal/mol obtained when comparing the extrapolated binding energies to the same reference DQMC values. Through this analysis, we further confirm the importance of the long-range correlation energy, the contribution of which can amount to more than 90% of the total binding,²⁵ which is clearly an integral component of the binding energy that is not captured at the underlying DFT level of theory.

The agreement between the PBE+MBD* and DQMC binding energies is essentially comparable to the statistical error of the stochastic DQMC method, except for the complexes **4a** (C₆₀@catcher) and **7b** (ADOH@CB7), which can be explained by the underlying approximations employed in the PBE+MBD* method. For instance, the approximation of localized QHOs might not provide the flexibility to adequately describe the response of the delocalized electrons present in the C₆₀ guest of complex **4a**, leading to a moderate deviation (*i.e.*, 1.0 kcal/mol outside the error bar). Regarding complex **7b**, the elevated number of hydrogens in the guest pointing towards the host causes the combined appearance of weak hydrogen bonds and Pauli exchange-repulsion effects. Here, the approximate treatment of exchange at the semi-local DFT level of theory combined with the non-exact range separation of the Coulomb interaction certainly represent a limitation in this context. A systematic study of the influence of these effects on the binding energies of the

entire S12L database is currently under investigation.

To better analyze the role of the many-body (many-atom) interactions in the binding energies of the S12L complexes, we have carried out a many-body decomposition of the infinite-order MBD* energy into pairwise (two-body), three-body, and higher-order terms. This many-body decomposition is facilitated by the fact that the MBD* correlation energy ($E_c^{\text{MBD}^*}$) can be expanded in powers of the product of the bare response function χ_0 with the interaction v ²⁰ as:

$$E_c^{\text{MBD}^*} = -\frac{1}{2\pi} \int_0^\infty d\omega \sum_{n=2}^\infty \frac{1}{n} \text{Tr}[(\chi_0 v)^n]. \quad (2)$$

The present procedure for the perturbative expansion of the MBD* long-range correlation energy differs from that followed in Ref. (26), which was based on an averaging of the free QHO characteristic frequencies, which were all set to a single value, chosen in order to preserve the total correlation energy. In contrast, the many-body expansion utilized herein is based on a straightforward Taylor series decomposition of the logarithm term that naturally arises in the RPA correlation energy; as such, this expansion allows for a clear diagrammatic interpretation of each perturbative term, now expressed in powers of $\chi_0 v$. In addition, we also note that an alternative range-separation of the Coulomb interaction, with smoother and isotropic short-range behavior, has been employed in this work.

The second-order truncation of the series contained in Eq. (2) leads to the well-known C_6/R^6 London pairwise summation, the third-order term contains the so-called Axilrod-Teller-Muto²⁷ (ATM) energy contribution, while the summation to infinite-order corresponds to the full MBD* energy. As an expected general trend, the terms corresponding to the even powers in Eq. (2) are typically negative (attractive), while odd powers provide contributions of positive sign (repulsive). The alternating sign behavior was observed in all complexes, the only exception being the **2a** host-guest complex (which loses this alternating trend after 5th-order).

An analysis of the data contained in Table 2 confirms that the many-body terms provide substantial decreases in the binding energies. The magnitude of these decreases depends on the par-

ticular system and exceeds 10% for all systems considered, apart from **5a** (7.2%). The reason for this smaller effect (encountered also in the **5b** S12L complex, with a 7.5% decrease) can be attributed to a combination of effects, such as the rather sparse and symmetric conformation, and the relatively low impact of dispersion forces on the overall binding in a complex that is primarily bound by hydrogen bonding. Interestingly, with the mere inclusion of the ATM (three-body) term, the binding decrease is exaggerated, as the higher-order terms provide a further reduction by about a factor of two. Hence, the neglect of higher-order ($n > 3$) terms often amounts to a few kcal/mol and will lead to errors much larger than the highly desired chemical accuracy threshold. As visible in Fig. 3, the progressive decrease in the absolute values of the perturbative contributions with n only converges at relatively higher orders. The rate of convergence again depends on the system, and deviations from MBD* below 1% are usually achieved only after $n \approx 6$. We stress that a deviation of 1% from MBD* can correspond to ≈ 1 kcal/mol for such supramolecular systems, *i.e.*, only slightly below the accuracy of the reference data.

To further illustrate this point, we analyzed the MBD* long-range correlation energy of an infinite, periodic, double walled carbon nanotube (DWCNT). This system is composed of two coaxial single-walled nanotubes, namely of the (10,10) and (5,5) type, as seen in Fig. 2. Despite the metallicity inherent to the DWCNT, this system only possesses two graphene-like crossing linear bands²⁸ at the Fermi surface. Hence, the contribution of these delocalized states to the overall correlation energy will be rather limited. To ensure convergence with respect to finite-size effects, a supercell of 35.7 Å along the DWCNT longitudinal axis was adopted, corresponding to 900 carbon atoms. The MBD* contribution to the binding energy is defined as above, in which the inner and outer nanotubes represent the two fragments of the host-guest complex. The apparent dimension of the system does not represent a limitation for the MBD* method, which can easily be applied to systems containing thousands of atoms. By virtue of the spatial extension and highly anisotropic character of the DWCNT, long-range many-body effects appear to be strongly enhanced in this complex: the long-range MBD* contribution to the binding energy is reduced by 24.7% from second to infinite order, as illustrated in Fig. 3. In addition, a strikingly slow percentage-wise

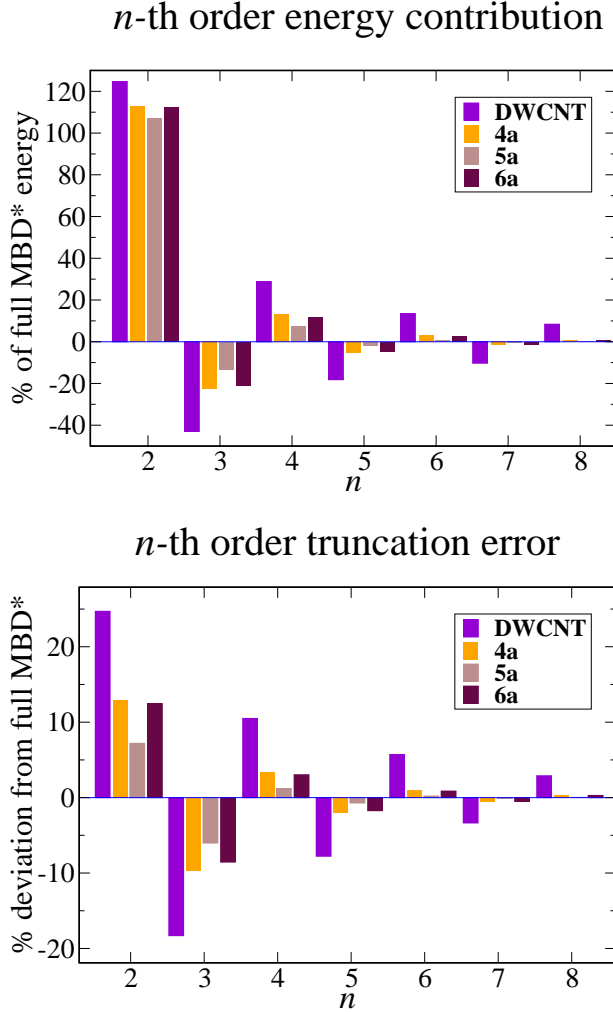


Figure 3: The many-body decomposition of the long-range MBD* correlation energy, as defined in Eq. (2). Results are reported for the complexes showing the slowest and fastest convergence (**4a** and **5a**, respectively), and for an intermediate case (**6a**). The additional case of a periodic double walled carbon nanotube (DWCNT) is also considered for comparison. (Upper Panel) Deviation of the cumulative summation of the many-body contributions up to *n*-th order with respect to the full infinite-order MBD* binding energy. (Lower Panel) Ratio of the single *n*-th order energy contribution with respect to the full infinite-order MBD* binding energy.

convergence of the perturbative MBD* series is also observed, leading to a deviation of 5.7% from full infinite-order MBD* at sixth-order, corresponding to a remarkable ≈ 32 kcal/mol per supercell. A comparable behavior is also expected in finite-extension nanotubes with length scales comparable to the present supercell, although the convergence of the many-body expansion was shown to be very slow in anisotropic low-dimensional systems.²⁹

The sensitive dependence of the many-body effects on the structure of the host-guest complex and the nature of the binding strongly indicates that a simple renormalization of the standard C_6/R^6 pairwise summation is unable to provide the same degree of accuracy as a full many-body treatment of the long-range correlation energy in all types of supramolecular systems. In the same breath, a perturbative approach limited to a fixed finite-order also appears to be inaccurate, due to the relatively slow convergence of the perturbative series. In particular, second-order Moeller-Plesset perturbation theory (MP2) completely neglects many-body dispersion effects, as they arise from higher-order correlation contributions. On the other hand, the *ad hoc* addition of the ATM term to empirical dispersion-corrected DFT approaches^{30,31} is questionable in this regard, further complicated by the fact that unphysical damping functions have to be utilized. In fact, employing different empirical damping functions for two-atom and three-atom interactions is inconsistent, as the attenuation of this fundamental interaction at any perturbative order stems from the same origin. In the MBD* approach, instead, the short-range attenuation of the Coulomb interaction is seamlessly achieved, requiring only the attenuation of the interaction between two atoms at a time, which is consistent with the presence of terms up to two-particle interactions (at most) in the full nucleo-electronic Hamiltonian.

By making use of state-of-the-art DQMC algorithms and large-scale computational resources, we provided benchmark binding energies for a set of six host-guest complexes from the S12L database. The DQMC data represent the first accurate benchmark for large supramolecular systems, with estimated errors not far from chemical accuracy. Very close agreement is found between DQMC and the PBE+MBD* method, and a perturbative many-body decomposition analysis of the long-range correlation energy in these host-guest complexes clearly demonstrated the need for an

accurate description of many-body correlation effects. The influence of these many-body interactions was found to have a significant dependence on the symmetry and underlying topology of the host-guest complexes. As a consequence, these effects cannot be recovered by an effective pairwise approach as high perturbative orders are required in order to converge to the full infinite-order MBD* limit for the long-range correlation energy. Moreover, the mere inclusion of the three-body Axilrod-Teller-Muto energy contribution was shown to provide an overestimated reduction of the binding with respect to the full infinite-order energy and cannot be used to reproduce the infinite-order long-range correlation energy with high fidelity. The relatively successful application of the PBE+MBD* method to host-guest complexes of chemical relevance demonstrates that this is a promising approach for the challenging investigation of large-scale supramolecular systems. The remaining issues to be addressed include a systematic analysis of the underlying semi-local density functional approximation and many-body correlation effects beyond the dipole approximation.

Supporting Information Available

A description of the MBD* computational method along with technical details regarding the DQMC calculations are provided as supplementary material. Further analysis concerning the role of fractional exact exchange and a comparison among pairwise dispersion methods are also given for completeness. The material is available free of charge via the Internet at <http://pubs.acs.org/>.

This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Acknowledgement

A.T. and A.A. received support from the European Research Council (ERC Starting Grant VDW-CMAT). This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy (DOE) under Contract No. DEAC05-00OR22725. R.A.D. received funding from the Department of Energy under Grant No. DE-SC0005180.

References

- (1) Pedersen, C.J. Cyclic Polyethers and Their Complexes with Metal Salts. *J. Am. Chem. Soc.* **1967**, *89*, 7017-7036.
- (2) Pedersen, C.J. The Discovery of Crown Ethers (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1021-1027.
- (3) Cram, D.J. The Design of Molecular Hosts, Guests, and Their Complexes (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 1009-1020.
- (4) Lehn, J.-M. Supramolecular Chemistry—Scope and Perspectives Molecules, Supramolecules, and Molecular Devices (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 89-112.
- (5) Atwood, J.L.; Steed, J.W. *Supramolecular Chemistry*. Wiley **2000**, Weinheim.
- (6) Lehn, J.-M. *Supramolecular Chemistry*. *Science* **1993**, *260*, 1762-1763.
- (7) Zhang, S.G. Fabrication of Novel Biomaterials through Molecular Self-Assembly. *Nature Biotechnology* **2003**, *21*, 1171-1178.
- (8) Balzani, V.; Gomez-Lopez, M.; Stoddart, J.F. Molecular Machines. *Accounts of Chemical Research* **1998**, *31*, 405-414.
- (9) Waller, M.P.; Kruse, H.; Mueck-Lichtenfeld, C.; Grimme, S. Investigating Inclusion Complexes Using Quantum Chemical Methods. *Chem. Soc. Rev.* **2012**, *41*, 3119-3128.
- (10) Rezac, J.; Riley, K.E.; Hobza, P. S66: A Bell-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.
- (11) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. 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.* **2006**, *8*, 1985-1993.

- (12) Rezac, J.; Riley, K.E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285-4292.
- (13) Booth, G.H.; Grüneis, A. Kresse, G. Alavi, A. Towards an Exact Description of Electronic Wavefunctions in Real Solids. *Nature* **2012**, *493*, 365-370.
- (14) Foulkes, W.M.C.; Mitas, L.; Needs, R.J.; Rajagopal, J. Quantum Monte Carlo Simulations of Solids. *Rev. Mod. Phys.* **2001**, *73*, 33-83.
- (15) Diedrich, C.; Lüchow, A.; Grimme, S. Weak Intermolecular Interactions Calculated with Diffusion Monte Carlo. *J. Chem. Phys.* **2005**, *123*, 184106.
- (16) Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo Methods Describe Noncovalent Interactions with Subchemical Accuracy. *J. Chem. Theory Comput.* **2013**, *9*, 4287-4292.
- (17) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955-9964.
- (18) Risthaus, T.; Grimme, S. Benchmarking of London Dispersion-Accounting Density Functional Theory Methods on Very Large Molecular Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 1580-1591.
- (19) Needs, R.J.; Towler, M.D.; Drummond, N.D.; López Ríos, P. Continuum Variational and Diffusion Quantum Monte Carlo Calculations. *J. Phys.: Cond. Matt.* **2010**, *22*, 023201-023215.
- (20) Tkatchenko, A.; Ambrosetti, A.; DiStasio, Jr. R.A. Interatomic Methods for the Dispersion Energy Derived from the Adiabatic Connection Fluctuation-Dissipation Theorem. *J. Chem. Phys.* **2013**, *138*, 074106.
- (21) Tkatchenko, A.; DiStasio, Jr. R.A.; Car R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

- (22) Reilly, A.M.; Tkatchenko, A. Seamless and Accurate Modeling of Organic Molecular Materials. *J. Phys. Chem. Lett.* **2013**, *4*, 1028-1033.
- (23) Perdew, J.P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (24) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab Initio Molecular Simulations with Numeric Atom-Centered Orbitals. *Comp. Phys. Comm.* **2009**, *180*, 2175-2196.
- (25) Tkatchenko, A.; Alfè, D.; Kim, K.S. First-Principles Modeling of Non-Covalent Interactions in Supramolecular Systems: the Role of Many-Body Effects. *J. Chem. Theory Comput.* **2012**, *8*, 4317-4322.
- (26) DiStasio, Jr. R.A.; von Lilienfeld, O.A.; Tkatchenko, A. Collective Many-Body van der Waals Interactions in Molecular Systems. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14791-14795.
- (27) Axilrod, B.M.; Teller, E. Interaction of The van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299-300.
- (28) Kwon, Y.-K.; Tománek, D. Electronic and Structural Properties of Multiwall Carbon Nanotubes. *Phys. Rev. B* **1998**, *58*, 16001(R).
- (29) Gobre, V.V.; Tkatchenko, A. Scaling Laws for van der Waals Interactions in Nanostructured Materials. *Nat. Commun.* **2013**, *4*, 2341.
- (30) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (31) Otero-de-la-Roza, A.; Johnson, E.R. Many-Body Dispersion Interactions from the Exchange-Hole Dipole Moment Model. *J. Chem. Phys.* **2013**, *138*, 054103.

Table 1: Binding energies for each of the considered S12L host-guest complexes in kcal/mol. (Second Column) DQMC computed binding energies with associated statistical sampling uncertainties given in parentheses. (Third Column) Binding energies extrapolated from experimentally measured association free energies via approximate solvation and entropic corrections.¹⁸ (Last Column) Indicated as PBE+MBD*, binding energies for the PBE functional including long-range correlation at the MBD* level (see text for details).

Binding Energies			
complex	DQMC	Extrap. Expt.	PBE+MBD*
2a TCNQ@tweezer	-27.5 (1.2)	-29.9	-29.0
2b DCB@tweezer	-17.2 (1.0)	-20.5	-18.8
4a C ₆₀ @catcher	-25.8 (1.5)	-27.5	-28.3
5a GLH@mcyle	-33.4 (1.0)	-34.8	-33.8
6a BuNH ₄ @CB6	-81.0 (1.6)	-77.4	-82.1
7b ADOH@CB7	-24.1 (1.8)	-22.6	-27.4

Table 2: Truncation errors in the PBE+MBD* long-range correlation energies for the host-guest complexes considered in this work. (Second Column) Differences between the second- and infinite-order MBD* correlation energies (percentage-wise, with respect to the infinite-order value). (Third Column) Differences between the second- and third-order MBD* correlation energies (percentage-wise, with respect to the infinite-order value).

Correlation Energy Analysis		
complex	2 nd - ∞	2 nd -3 rd
2a TCNQ@tweezer	12.2%	20.7%
2b DCB@tweezer	11.9%	19.6%
4a C ₆₀ @catcher	12.9%	22.6%
5a GLH@mcyle	7.2%	13.2%
6a BuNH ₄ @CB6	12.5%	21.0%
7b ADOH@CB7	12.3%	21.3%