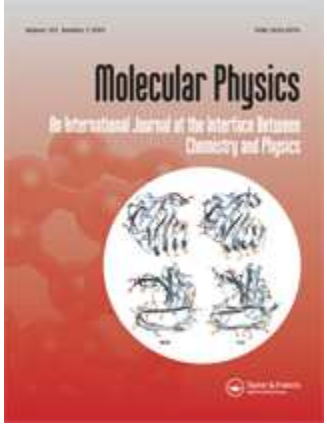
Dispersion energies for small organic molecules: First row atoms

Alston J Misquitta, Anthony J. Stone



Dispersion energies for small organic molecules: First row atoms

Journal:	*Molecular Physics*
Manuscript ID:	TMPH-2008-0125.R1
Manuscript Type:	Invited Article
Date Submitted by the Author:	27-May-2008
Complete List of Authors:	Misquitta, Alston; University of Cambridge, Chemistry; Girton College Stone, Anthony; Cambridge University, University Chemical Laboratory
Keywords:	dispersion, distributed dispersion, WSM, SAPT(DFT), polarizabilities
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Misquitta_Stone.tex natbib.sty	

Dispersion energies for small organic molecules: First row atoms

By Alston J. Misquitta*[†,‡] and Anthony J. Stone[†]

†University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.
‡University College London, 20 Gordon Street, London, WC1H 0AJ, U.K.

May 27, 2008

Accurate atom–atom dispersion coefficients are needed if we are to be able to successfully model the intermolecular interactions between organic molecules. To go beyond the usual C_6 description we have to resort to accurate *ab initio* methods, and we use the Williams–Stone–Misquitta methodology to obtain anisotropic distributed dispersion coefficients to C_{12} on each atomic site. Comparisons with SAPT(DFT) energies show that the resulting descriptions are very accurate, even for molecular contacts, where the asymptotic series is thought to be invalid. The complexity of our most accurate models limits their applicability, so we explore simplifications that are more suitable for use in calculations of the condensed phase, in particular, we describe a means of calculating optimized isotropic C_6 models that can be readily used in conventional programs.

1 Introduction

The dispersion energy plays a fundamental role in the bonding of gases, liquids and many solids. Of particular importance is the role of the dispersion in biological systems and organic crystals, where it is perhaps second in importance only to the electrostatic and polarization interactions found in hydrogen-bonding systems, but of primary importance where π-stacking is important. But it must not be forgotten that while the electrostatic, induction (also called polarization) and dispersion are all long-range interactions, the dispersion is always present between every pair of atoms and always attractive, so its cumulative effect can be very high, particularly in the condensed phase.

It is often argued that the dispersion is fairly isotropic and consequently can be considered to be a *correction* to the more anisotropic electrostatic and induction interactions. This is only partly true. For many systems the dispersion is indeed less sensitive to orientation, but in systems exhibiting π-stacking the anisotropy can be considerable, for here, besides the geometric anisotropy associated with a planar molecule, the atomic polarizabilities exhibit an intrinsic anisotropy that is reflected in strongly anisotropic atom–atom dispersion coefficients. The geometric anisotropy can be described using an isotropic atom–atom description, but the intrinsic anisotropy should require the explicit inclusion of anisotropy in the atom–atom terms.

Anisotropic atom–atom dispersion models are bound to be complex, and they will not always be needed. After all, the missing anisotropy can be at least partially absorbed in the anisotropic short-range part of the potential, as is done in many *ab initio* potentials[1–3]. Even in empirical potentials that include the dispersion only through an isotropic $-C_6/R^6$ term a partial inclusion of the anisotropy is possible through the anisotropy in the electrostatic term, if present. For example, one of the important anisotropic R^{-6} dispersion terms has the same angular dependence as the quadrupole–quadrupole (R^{-5}) electrostatic interaction[4], and the radial behaviour is very similar over a short distance close to contact.

To decide quantitatively on the importance of the anisotropy, we need a means of calculating anisotropic atom–atom dispersion coefficients accurately and in a computationally efficient manner. Empirical methods cannot be used in the construction of anisotropic models due to the large number of parameters involved. Rather we must rely on theoretical methods. Further, if we are to achieve a new level of accuracy in the description of the dispersion energy, we need to go beyond the $-C_6/R^6$ term and include contributions from the C_7, C_8, and higher-order terms. Once again, only theoretical methods can provide us with these terms.

Until recently, there were very few theoretical methods for obtaining atom–atom dispersion coefficients of organic molecules, and none of them were really suitable for small organic molecules. One of the approaches used has been to fit an atom–atom model to dispersion energies calculated either from a perturbation theory like symmetry-adapted perturbation theory (SAPT)[5–7], or the more modern version based on density functional theory (SAPT(DFT))[8–14] or from dispersion energies calculated from a single-centre asymptotic expansion. This method

*Corresponding author: phone +44 1223 763874; fax +44 1223 336362; email am592@cam.ac.uk.

has been used to obtain distributed dispersion coefficients for the water dimer[2] and the benzene dimer[3], but has many serious limitations. First of all, the resulting description may be unphysical, as the parameters are always strongly coupled, and may be ill-defined if the atoms are 'buried' under other atoms. Second, anisotropic dispersion coefficients cannot be obtained using this method because of the large number of parameters involved. Lastly, for a well-defined fit, a large number of dispersion energies need to be calculated; if calculated with SAPT or SAPT(DFT), this can be a computationally expensive step, and if calculated using a single-centre asymptotic expansion, the resulting energies will be nearly useless for even the smallest of organic molecules owing to the large sphere of divergence[15] associated with such an expansion.

The more successful methods for calculating atom–atom dispersion models have been based on a distribution of molecular polarizabilities which have been recently reviewed[16,17]. Methods based on real-space partitioning, such as that proposed by Hattig *et al.*[18] seem to be slowly convergent, and are not ideal for practical use; while those based on fitting distributed polarizability models to point-to-point polarizabilities calculated on a grid around the molecule, such as that proposed by Williams & Stone[19], require considerable effort to avoid the presence of unphysical terms, and this can make the calculation of higher-rank polarizabilities very tedious. Arguably the most successful method for calculating distributed polarizabilities is the Williams–Stone–Misquitta (WSM) method[20,21], which is a combination of the constrained density-fitted distribution of Misquitta & Stone[16] and the Williams & Stone method. There are a number of features of the WSM method that make it ideal for calculating dispersion models:

- It can be applied to frequency-dependent polarizabilities, which is essential for calculating dispersion coefficients.
- There are no known basis set restrictions. This is quite important as fairly large and diffuse basis sets are needed to calculate accurate polarizabilities.
- High ranks are possible. At present, this method has been used to calculate polarizabilities to rank 3.
- The polarizabilities are calculated using linear response DFT using the same functional as used in a standard SAPT(DFT) calculation. Therefore, the polarizabilities, and consequently, the dispersion coefficients are consistent with SAPT(DFT) energies. This is an essential requirement when building an analytic potential.
- The atomic polarizabilities are physically reasonable, and the very few instances of loss of positive-definiteness occur mainly in the high-rank terms.
- The WSM polarizability model can be chosen to suit the problem. This has the advantage of allowing us to create a dispersion model as simple or complex as needed. Further, exact or approximate symmetries can be trivially imposed on the model, which can significantly simplify the resulting dispersion description.
- Additionally, the WSM polarizabilities can be calculated for small organic molecules at relatively modest computational expense.

In this paper we have set out to describe the dispersion models that can be calculated using the WSM polarizabilities and provide quantitative answers to the questions that arise when setting out to perform such a calculation. These include questions of the basis sets that must be used, the order of terms that must be included for accurate energies, and whether we can do without the explicit anisotropy in the dispersion description. These are important questions when high-accuracy benchmarks are needed, but are probably of little concern in practice owing to the relatively large computational cost of evaluating energies using our most complicated description. So we will also explore isotropic models — in particular, the popular isotropic C_6 model. Such models could herald a new generation of accurate atom–atom potentials that are derived from first principles alone, but are no more complex than conventional empirical potentials. These models could also benefit the dispersion-corrected DFT methods such as those of Grimme and colleagues[22], which currently rely on empirical dispersion terms but are otherwise derived *ab initio*.

We will limit our investigations to molecules containing atoms from the first row of the periodic table. Molecules containing atoms from the second and third rows, in particular the halogens, are expected to exhibit a large intrinsic anisotropy. We will explore such systems in a subsequent paper.

This paper is organised as follows. In sec. 2 we present the basic theoretical background of the WSM method, the dispersion models and SAPT(DFT). The numerical details of the calculations are described in sec. 3, and in sec. 4 we describe the methods we have used to display our results. In secs. 5 and 6 we discuss the important topics of damping and the basis sets to be used for calculations of the models. The main body of our results is in sec. 7, where we explore a range of dispersion models of different order and complexity for a variety of systems. Isotropic C_6 models are discussed in sec. 8, and we summarize our results in sec. 9.

2 Theory

2.1 The Williams–Stone–Misquitta (WSM) method

The WSM method for calculating distributed (frequency-dependent) polarizabilities has been described in detail elsewhere[17,20,21], so we will provide only a brief outline of the method here. There are three main steps in this procedure:

- First we obtain a distributed polarizability model that includes non-local polarizabilities and charge-flow terms. This description is based on a partitioning of the density susceptibility function using a constrained density-fitting algorithm[16].
- Next, we use a localization scheme to transform away the non-local and charge-flow terms. We have usually used the localization scheme proposed by LeSueur & Stone[23], but have recently adopted the method proposed by Lillestolen & Wheatley[24].
- The localization step results in a degradation of the convergence properties of the polarizability description. To remedy this, we use the method of Williams & Stone[19] to refine the local polarizability description to reproduce a few million point-to-point polarizabilities computed on a grid around the molecule; typically between the vdW×2 and vdW×4 surfaces (that is, surfaces at twice or 4 times the van der Waals radius from each atom). The standard Williams & Stone method would result in unphysical terms in the description. In order to prevent this, the refinement is done with quadratic constraints imposed with the values calculated in the localization step used as anchors. Since we are dealing with frequency-dependent polarizabilities, we have modified the coefficients of these constraints, the $g_{kk'}$ of eq. 36 in ref. 20, to be

$$g_{kk'}(\omega) = 0 \text{ if } k \neq k'$$
$$g_{kk}(\omega) = \begin{cases} 10^{-5}/(1+|\omega|^2) & \text{if } k \in \{10, 10c, 10s\} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

That is, only the dipole–dipole polarizabilities are constrained to remain close to the anchor values, with weights that decrease with increasing magnitude of the frequency.

2.2 Dispersion models

Having obtained the localized frequency-dependent polarizabilities, we can now write the site-site dispersion energy in the form[25]

$$\begin{aligned} E_{\text{disp}}^{\text{asymp}} &= -\frac{\hbar}{\pi} \sum_{a\in A, b\in B} T_{tu}^{ab} T_{t'u'}^{ab} \int_0^\infty \alpha_{tt'}^a(\mathrm{i}\nu)\alpha_{uu'}^b(\mathrm{i}\nu)\mathrm{d}\nu \\ &= -\sum_{a\in A, b\in B} \left(\frac{C_6^{ab}(\Omega)}{R_{ab}^6} + \frac{C_7^{ab}(\Omega)}{R_{ab}^7} + \frac{C_8^{ab}(\Omega)}{R_{ab}^8} + \cdots \right), \end{aligned} \tag{2}$$

where a and b are sites on molecules A and B respectively, T_{tu}^{ab} are the T-functions that contain the distance and orientation information[15], R_{ab} is the distance between the two sites, and t, u, etc. are the angular momentum labels and take on values $00, 10, 11c, 11s, \cdots$. The dispersion coefficients depend on the relative orientation Ω, and can be described by a series in the orientational S functions[15]. The general formulation in terms of spherical tensors is given by Stone & Tough[4] and specific examples are given in refs. 19 and 26.

The dispersion coefficients depend on integrals of the form

$$\int_0^\infty \alpha_{tt'}^a(\mathrm{i}\nu)\alpha_{uu'}^b(\mathrm{i}\nu)\mathrm{d}\nu, \tag{3}$$

where rank-l_1 – rank-l_1' and rank-l_2 – rank-l_2' polarizability tensors contribute to a C_n coefficient with $n = l_1 + l_1' + l_2 + l_2' + 2$. At present, we are able to calculate WSM polarizabilities to rank 3, so while we can calculate C_{10} and C_{12} terms, they will lack the contributions from the hexadecapole and higher-rank polarizabilities. As we shall see, this is probably not a serious limitation. In any case, the present dispersion description is already cumbersome, and the inclusion of terms of even higher order could make it almost unusable.

The integration in eq. 3 is performed numerically using Gauss–Legendre quadrature with the now standard transformation $\nu = \beta(1+t)/(1-t)$ with $\beta = 0.5$ and a grid of 10 points.

A word about the notation and terminology used: the dispersion energy calculated with an anisotropic model in which C_n is the highest order term will be denoted by $E^{(2)}_{\mathrm{disp,d}}(n)$ and described as a model of 'order n'. Likewise, if the model used includes only isotropic terms, the energy will be denoted by $E^{(2)}_{\mathrm{disp,d}}(n,\mathrm{iso})$. We use 'rank' to refer to the maximum rank of polarizabilities included in the model. The C_{10} and C_{12} models will generally be calculated with a polarizability description of maximum rank 3, but the C_{10} model may also be calculated with polarizabilities of maximum rank 2. The corresponding dispersion energies will be indicated by $E^{(2)}_{\mathrm{disp,d}}(10(\mathrm{L2}))$ ($E^{(2)}_{\mathrm{disp,d}}(10,\mathrm{iso(L2)})$ for the isotropic models).

2.3 SAPT(DFT) dispersion energies

Reference dispersion energies are calculated using SAPT(DFT)[9,27]. These energies are arguably some of the most accurate possible for small organic molecules. Within SAPT(DFT), the dispersion has two components: $E^{(2)}_{\mathrm{disp,pol}}$ is the second-order dispersion energy defined through the polarization expansion[5]. Exchange effects are missing in $E^{(2)}_{\mathrm{disp,pol}}$. These are included in the term $E^{(2)}_{\mathrm{disp,exch}}$, the exchange-dispersion energy. We will denote by $E^{(2)}_{\mathrm{disp,tot}}$ the sum of these two terms, that is

$$E^{(2)}_{\mathrm{disp,tot}} = E^{(2)}_{\mathrm{disp,pol}} + E^{(2)}_{\mathrm{disp,exch}}. \tag{4}$$

The multipole-expanded dispersion energy (eq. 2) is formally obtained from $E^{(2)}_{\mathrm{disp,pol}}$ by expanding the interaction operator using the multipole expansion[4,15], so we should expect $E^{(2)}_{\mathrm{disp,pol}}$ to be well approximated by $E^{\mathrm{asymp}}_{\mathrm{disp}}$ (suitably damped). This is indeed the case. However it is important to bear in mind that because $E^{(2)}_{\mathrm{disp,exch}}$ decays exponentially with intermolecular separation R, it has no multipole expansion, so that $E^{(2)}_{\mathrm{disp,tot}}$ has the same multipole expansion as $E^{(2)}_{\mathrm{disp,pol}}$.

3 Numerical details

It is well known that the dispersion energy is difficult to converge using standard basis sets[28]. Consequently, SAPT(DFT) calculations are usually performed using monomer basis sets augmented with basis functions placed in the bonding region between the two molecules and on the atomic sites of the partner monomer, resulting in the so-called 'monomer-centred-plus', or MC+ basis sets. The basis functions in the bonding region (so-called bond functions) are needed to converge the dispersion energy, while the functions on the partner monomer sites are needed for the induction energy[28]. The latter are not needed in this work, but are retained for consistency. We have used either the aug-cc-pVTZ or the Sadlej-pVTZ basis sets for the monomer parts of the MC+ basis sets, and a $3s2p1d$ basis set for the bond functions, which have been placed at a position determined by a generalization of the weighting scheme described in ref. 29. Dispersion energies calculated in such a basis are very close to convergence with respect to basis set.

All SAPT(DFT) calculations have used the molecular orbitals and energies obtained from the asymptotically-corrected PBE0 exchange-correlation functional. The linear-response DFT calculations needed for the second-order SAPT(DFT) energies and the frequency-dependent polarizability calculations used a hybrid adiabatic LDA and coupled Hartree–Fock kernel. For details see ref. 11. The SAPT(DFT) calculations were performed using the CamCASP[30] and Sapt2006[31] programs.

The vertical ionization potentials (IPs) of the interacting molecules are used in the asymptotic correction and for calculating the damping coefficients (see below). They were obtained either from a ΔDFT calculation using the PBE0 functional[16] or from experimental results reported in ref. 32. For water, urea and benzene we have used IPs of 0.464, 0.367 and 0.340 a.u. respectively, obtained from ref. 32, and for formamide and N-methyl propanamide we have used IPs of 0.375 and 0.345 respectively, obtained using the ΔDFT method[16].

Rather than select dimers at just their minimum energy configurations, we have used the algorithm described in ref. 33 to generate dimers randomly so as to get a uniform coverage of the space of physically important geometries. In brief, this is done by first obtaining a dimer configuration using a Sobol pseudo-random sequence and Shoemake's uniform distribution algorithm[34]. Next the intermolecular separation is adjusted so as to bring the dimers into van der Waals contact, and finally, 2 to 4 dimer separations are chosen randomly to lie around the contact distance. This algorithm has been implemented in the CamCASP program.

4 Displaying the energies

One of the problems we frequently encounter is the analysis of multi-dimensional data. One solution is to simply abandon the dimensionality information and present the data as a scatter plot, with the model energies plotted

against a suitable reference energy. We shall make use of such plots in the paper. While there is a correlation of the intermolecular separation with the energy scale, such a plot contains no information about the nature of the atoms in contact, or the angular configuration of the molecules. At least some of this data can be obtained through an energy map on a suitable surface around the molecule[17,21]. Such a map can be constructed by calculating the required energy, here the dispersion energy, using a suitable probe placed at various points on the chosen surface. At present, we use an energy probe with spherical symmetry, which for the dispersion energy is taken to be a neon atom. This has the advantage of allowing the dispersion energy maps to be made with a reasonable amount of computational effort. However, the small polarizability of the neon atom results in a rather small range of dispersion energies. Typical dispersion energies of organic molecules will be about 5 times larger than those between a neon atom and an organic molecule.

We have used the vdW×2 surface defined in ref. 21 with van der Waals radii prescribed by Bondi[35], except for hydrogen atoms capable of forming hydrogen bonds, which have their radii set to zero. This is sometimes referred to as the 'water accessible' surface.

5 Damping to reproduce $E^{(2)}_{\mathrm{disp,tot}}$ or $E^{(2)}_{\mathrm{disp,pol}}$?

Like any multipole expansion, the dispersion series must be damped at short-range. Damping is particularly important when terms above C_6 are included. Even the C_6/R^6 term will diverge as $R \to 0$, and although its divergence can often be ignored, since it manifests itself at rather small R, it can be a problem for Monte Carlo simulations. We have used the Tang–Toennies[36] damping functions $f_n(\beta R)$, which are based on incomplete Gamma functions. In general, the damping function should probably be anisotropic, but the description would then become very complicated, and in any case the form of any anisotropy is unknown. Besides, the anisotropy, if needed, can probably be absorbed in an anisotropic short-range term in the potential, such as the Born–Mayer term used to represent the exchange-repulsion and penetration terms.

The damping coefficients used in the Tang–Toennies functions will generally have to depend on the atom types. In accurate *ab initio* potentials, the damping parameter is often taken to be the same as the exponent in the Born–Mayer potential, because the repulsion and damping both arise from the overlap of the molecular wavefunctions. Therefore a natural method for obtaining the atom–atom damping parameter would be to fit the exchange-repulsion, penetration and damping terms simultaneously. Instead, we have chosen to use a single damping parameter β that depends only on the interacting molecules. As we shall see, this can work very well. The small residual errors can then be absorbed in the fit to short-range energies.

It then remains to determine what the damped expansion is to reproduce: $E^{(2)}_{\mathrm{disp,tot}}$ or $E^{(2)}_{\mathrm{disp,pol}}$? There are many good reasons for damping $E^{\mathrm{asymp}}_{\mathrm{disp}}$ to match $E^{(2)}_{\mathrm{disp,tot}}$ rather than $E^{(2)}_{\mathrm{disp,pol}}$. First of all, $E^{(2)}_{\mathrm{disp,pol}}$ is evaluated using wavefunctions that are not antisymmetrized with respect to electron exchange between the interacting monomers, so it is unphysical at short-range, where the effects of electron exchange become important. It seems far more satisfactory to consider the total dispersion energy $E^{(2)}_{\mathrm{disp,tot}}$ as the physically important quantity. Taken individually, $E^{(2)}_{\mathrm{disp,pol}}$ and $E^{(2)}_{\mathrm{disp,exch}}$ are artefacts of the symmetrized Raleigh–Schrödinger perturbation theory on which SAPT(DFT) is based. A more pragmatic reason is that for our most accurate dispersion models, the damped asymptotic expansion, when taken to high order, is found to agree with $E^{(2)}_{\mathrm{disp,tot}}$ as well as or better than with $E^{(2)}_{\mathrm{disp,pol}}$ in an energy range from 0 to $-50\,\mathrm{kJ\,mol^{-1}}$, for all dimers considered in this paper. In particular, for the physically important dimer configurations the r.m.s. errors made by either choice are between 0.2 and $0.8\,\mathrm{kJ\,mol^{-1}}$ only. A further reason for comparing with $E^{(2)}_{\mathrm{disp,tot}}$ concerns the damping coefficient (β in $f_n(\beta R)$). When comparing with $E^{(2)}_{\mathrm{disp,pol}}$, β has to be obtained numerically, by a least-squares fit, but when comparing with $E^{(2)}_{\mathrm{disp,tot}}$, we find that for all systems investigated here, we can use the same damping factor as for the induction energy, which is determined solely from the molecular vertical ionization energies I_A and I_B:

$$\beta = \sqrt{2I_A} + \sqrt{2I_B}. \tag{5}$$

As we shall see below, this damping, which is motivated by the asymptotic form of the overlap between the ground-state wavefunctions of the two monomers, is near-optimal for all systems and dispersion models we have considered.

6 Basis sets

It may seem strange to question which basis set should be used to calculate the dispersion coefficients. After all, should we not use the same basis as was used in calculating the reference SAPT(DFT) energies? The problem is that the basis set used for the SAPT(DFT) energies is of the MC+ type and therefore not only depends on the

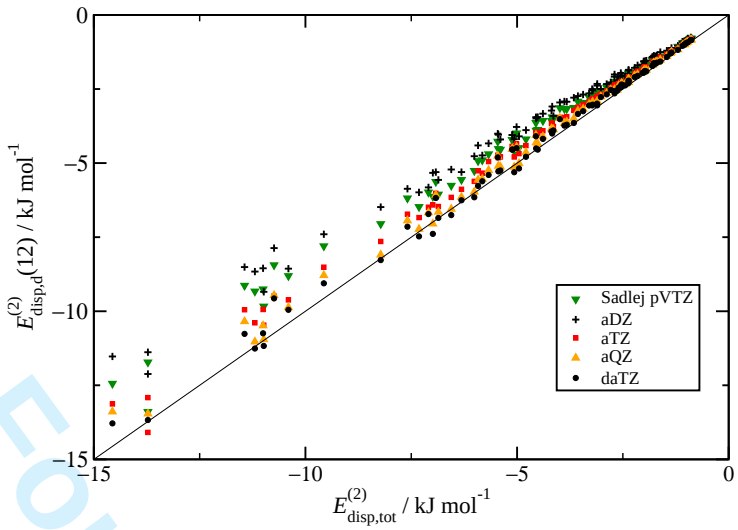


Figure 1: Dispersion energies for the formamide···water dimer. Scatter plot of $E^{(2)}_{\mathrm{disp,tot}}$ and $E^{(2)}_{\mathrm{disp,d}}(12)$ dispersion models calculated using a variety of basis sets.

dimer geometries, but consists of a standard molecular basis *augmented* with functions in the bonding region and elsewhere. Because of the augmentation, the effective size of the MC+ basis is larger than that of the standard molecular basis set that forms the main part of the MC+ basis type. For example, the effective size of the aug-cc-pVTZ basis of the MC+ type is probably that of the aug-cc-pVQZ or larger. For consistency with the SAPT(DFT) energies, the dispersion coefficients need to be calculated in a basis that is the equivalent of the MC+ basis used in the SAPT(DFT) calculations.

This problem was realised in earlier work [10], where it was suggested that the criterion for equivalence should be that the (damped) expanded and non-expanded dispersion energies should match at the dimer minimum energy configuration. This is a reasonable requirement, as it is at the important geometries that we need the highest accuracy. For systems larger than those considered in ref. 10, such as clusters and crystals, it is not only the global minimum energy dimer configuration that is important; rather we need to generalise this requirement to include dimers at and near all of a variety of important intermolecular contacts. This is done using the algorithm described in sec. 3.

Fig. 1 shows energies calculated using the damped asymptotic expansion $E^{(2)}_{\mathrm{disp,d}}(12)$ plotted against the total dispersion energy $E^{(2)}_{\mathrm{disp,tot}}$. The SAPT(DFT) energies were calculated using the aug-cc-pVTZ basis of the MC+ type. The trend is obvious: the larger and more diffuse the basis set, the closer the agreement between the expanded and non-expanded energies. The aug-cc-pVDZ basis (abbreviated to 'aDZ' in Fig. 1) is clearly inadequate, with an r.m.s. error of 1.17 kJ mol^{-1} for energies in the range −15 to −1 kJ mol^{-1}. The r.m.s. error is more than halved, to 0.52 kJ mol^{-1}, with the aug-cc-pVTZ basis, which is the monomer part of the MC+ basis used for the SAPT(DFT) calculations. The aug-cc-pVQZ and d-aug-cc-pVTZ basis sets yield dispersion models with r.m.s. errors of only 0.40 and 0.38 kJ mol^{-1} respectively. This is very good given the large range of energies. Of the last two basis sets, the d-aug-cc-pVTZ basis ('daTZ') not only yields slightly more accurate dispersion models, but is smaller than the aug-cc-pVQZ basis. Therefore, unless otherwise specified, all dispersion models presented in this work have been calculated using this basis set.

In an earlier study of the induction energy [21], we recommended using the Sadlej pVTZ [37,38] basis sets for calculations of molecular moments and polarizabilities. This basis was optimized for molecular properties and despite its small size, was shown to yield induction energies comparable to those from the much larger aug-cc-pVTZ basis. Unfortunately, it fares rather poorly for calculations of the dispersion energies, with an r.m.s. error of 0.86 kJ mol^{-1}, intermediate between the aug-cc-pVDZ and aug-cc-pVTZ bases. This is unfortunate as the small size of this basis set (the same size as aug-cc-pVDZ) makes it very attractive for calculations on organic molecules. The large errors are almost certainly due to the lack of sufficiently high angular momenta in the basis set: for the first row atoms, the Sadlej basis includes functions of at most *d* symmetry. It is possible that the augmentation of this basis with an additional diffuse shell would result in significantly more accurate dispersion models, while still keeping it smaller than the aug-cc-pVTZ basis. We are currently exploring this possibility.

7 Order and anisotropy

The WSM polarizability models allow us to construct dispersion models that are very detailed and accurate, but they can be quite unwieldy, because they may comprise a large number of terms, and the resulting computational costs can limit their applicability severely. Consequently, we need to examine the relationship between complexity and accuracy in some detail. The questions we would like to answer here are, first, at what order can we truncate the expansion in eq. 2? And secondly, how important is the anisotropy in the dispersion coefficients? That is, can we construct isotropic dispersion models of sufficient accuracy? For the purposes of this paper, we will take the target to be an r.m.s. error of a few tenths of a $\mathrm{kJ\,mol^{-1}}$ over an energy range of between 15 to 25 $\mathrm{kJ\,mol^{-1}}$, depending on the complex.

These are not the only questions we are interested in. The WSM method allows great flexibility in the choice of dispersion model. For example, we could retain anisotropic terms for some atom–atom interactions and not others, or we could truncate the expansion at low order for interactions involving hydrogen atoms, and retain the higher-order terms for the heavier atom pairs. Once the model is chosen, the WSM procedure ensures that the resulting polarizability description, and hence the dispersion description, will be the most accurate possible within the constraints of the model. We already use some of this flexibility by limiting the polarizability description on the hydrogen atoms to rank 1, but apart from this, we shall not consider hybrid models in this paper.

Fig. 2 shows the dispersion energy of the formamide⋯neon system, mapped onto the vdW×2 surface of formamide. The SAPT(DFT) energies have been calculated using a d-aug-cc-pVTZ basis without bond functions. Ideally we should have displayed $E^{(2)}_{\mathrm{disp,tot}}$ rather than $E^{(2)}_{\mathrm{disp,pol}}$, but due to a current limitation of the CamCASP program[30], this was not possible. In any case, $E^{(2)}_{\mathrm{disp,exch}}$ is not expected to be large in the absence of bond functions. All dispersion models have been calculated using the d-aug-cc-pVTZ basis.

As mentioned in sec. 4, the energy range for $E^{(2)}_{\mathrm{disp,pol}}$ from SAPT(DFT) is quite narrow (it is only 2 $\mathrm{kJ\,mol^{-1}}$) because of the small polarizability of neon. The strongest dispersion interaction occurs at the hydrogen atoms in the amide group, because of the close contacts involved. The dispersion models are presented as difference maps against the SAPT(DFT) energies. For simplicity of the discussion, none of the models is damped. We see that the C_6 model clearly underestimates the dispersion energy, but the underestimation is not uniform: it is highest near the amide group and smallest near the –CH hydrogen. The C_8 model is substantially better, and the C_{10} model is excellent except at the $-NH_2$ hydrogen atoms, where the dispersion energy is over-estimated. This is due to the lack of damping, and on damping (not shown), the agreement of the C_{10} model and the SAPT(DFT) energies is near perfect. From fig. 3 we see that the same holds true for the benzene and N-methyl propanamide molecules.

These observations are supported by results for the formamide⋯water complex, shown in fig. 4, in which we have plotted the damped model dispersion energies against the SAPT(DFT) total dispersion energies $E^{(2)}_{\mathrm{disp,tot}}$. The damping coefficient $\beta = 1.83$ a.u. was calculated using eq. 5. Ideally we would like the energies from the damped dispersion models to lie close to the line in fig. 4, preferably with only a small scatter. That is, a fit of the form $y = mx$ should have the slope $m = 1$. As expected, the C_6 model underestimates $E^{(2)}_{\mathrm{disp,tot}}$ quite substantially with slope $m = 0.62$ and a rather large r.m.s. error of 2.19 $\mathrm{kJ\,mol^{-1}}$. Interestingly, this model also exhibits a fairly large scatter of energies throughout the energy range, even for small energies where it is often assumed that the C_6 model is reasonably accurate. As the rank of the models used increases, the slope tends to 1 and the scatter decreases. The damped C_8, C_{10} (L2), C_{10} and C_{12} models have slopes of 0.83, 0.87, 0.95 and 0.97 respectively, and reproduce $E^{(2)}_{\mathrm{disp,tot}}$ with r.m.s. errors of 0.98, 0.79, 0.37 and 0.38 $\mathrm{kJ\,mol^{-1}}$ respectively. The C_8 model is clearly a big improvement on the C_6 model. The C_{10} (L2) model, which includes C_{10} coefficients calculated with polarizabilities of maximum rank 2, is not much of an improvement over the C_8 model, but the C_{10} model, which includes polarizabilities of rank 3, is substantially better, being as good as the more complex C_{12} model. Both of these result in dispersion energies that agree with the SAPT(DFT) reference energies across the entire energy range of 15 $\mathrm{kJ\,mol^{-1}}$. This energy range includes dimers with total interaction energies as much as 16 $\mathrm{kJ\,mol^{-1}}$; that is, these models are valid even in the repulsive region. This excellent agreement continues even below $-15\,\mathrm{kJ\,mol^{-1}}$ (not shown), where the damped C_{10} and C_{12} models have r.m.s. errors of 0.54 and 1.72 $\mathrm{kJ\,mol^{-1}}$ respectively, for dimers with total interaction energy as much as 45 $\mathrm{kJ\,mol^{-1}}$.

In figs. 5, 6 and 7 are shown similar comparisons for the formamide, urea and benzene dimers. All SAPT(DFT) energies have been calculated with the Sadlej pVTZ basis[37,38] in the MC+ format. The dispersion models for the formamide and urea dimers have been calculated using the d-aug-cc-pVTZ basis, but due to a technical problem in the DFT calculation, we have used the aug-cc-pVTZ basis for the benzene dimer. Using eq. 5, the damping coefficients for the formamide, urea and benzene dimers are determined to be 1.73, 1.71 and 1.65 a.u. respectively.

The damped C_{10} and C_{12} models for the formamide and urea dimers are in excellent agreement with $E^{(2)}_{\mathrm{disp,tot}}$, with r.m.s. errors of 0.28 & 0.24 $\mathrm{kJ\,mol^{-1}}$ for formamide and 0.33 & 0.32 $\mathrm{kJ\,mol^{-1}}$ for urea. Once again, there is little to choose between the C_{10} and C_{12} models, but since the C_{10} model is simpler, it would be the model of choice for accurate calculations.

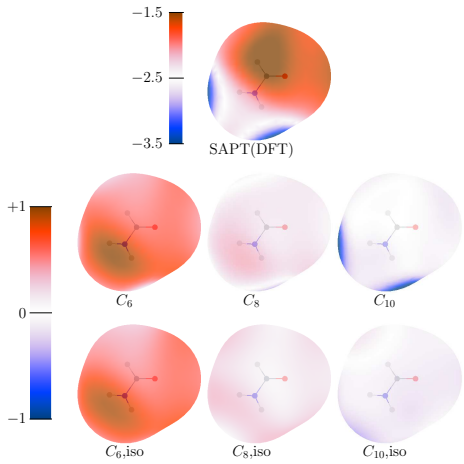


Figure 2: Dispersion energy maps of formamide with neon as a probe. The SAPT(DFT) dispersion energy $E^{(2)}_{\text{disp,pol}}$ is displayed using an absolute scale in kJ mol^{-1}. The model dispersion energies are displayed as *differences* taken against $E^{(2)}_{\text{disp,tot}}$ from SAPT(DFT). The models are not damped. The d-aug-cc-pVTZ basis has been used for all calculations.

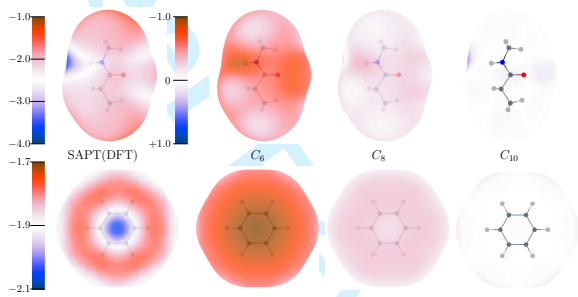


Figure 3: Dispersion energy maps of benzene and N-methyl propanamide with neon as a probe. See the caption of fig. 2 for a description. The description of benzene has been obtained using the aug-cc-pVTZ basis.

While the C_{10} and C_{12} models are certainly quite good for the benzene dimer, the r.m.s. errors are somewhat larger at 0.83 kJ mol^{-1} in each case. These larger errors could be in part because the dispersion models for this system were calculated using the aug-cc-pVTZ basis, which, from the discussion in sec. 6, has been shown to underestimate the total dispersion energy. But there is also the possibility that the large anisotropy of the benzene molecule requires the use of an anisotropic damping. This needs to be investigated. In any case, even for the this system, the errors made by the damped C_{10} and C_{12} models in recovering $E^{(2)}_{\text{disp,tot}}$ are small. This is true even for the particularly difficult and important stacked configurations, which are highlighted in fig. 7.

7.1 Isotropic models

We now turn to the question of anisotropy: is it really needed, or can we construct isotropic models of comparable accuracy? We need to explore this issue, because although the dispersion models with anisotropic terms are accurate, they are too complex to be used in most kinds of calculation.

We have previously constructed isotropic dispersion models from the models with anisotropy by simply retaining only the isotropic terms from the final dispersion coefficient calculation [17,33]. However we can do better by restricting the WSM polarizability models so that the refinement stage includes only isotropic atomic polarizabilities. These isotropic polarizabilities can then adjust to absorb some of the effects of polarizability anisotropy. For example, this method leads to an r.m.s. error of 0.45 kJ mol^{-1} for the formamide···water complex, compared with 0.55 kJ mol^{-1} by retaining the isotropic part of the anisotropic C_{12} model.

In fig. 2 we display the dispersion energy difference maps for the isotropic dispersion models of the formamide molecule. The results are almost indistinguishable from the difference maps obtained with the models that include anisotropy. In fact, the $C_{10,\text{iso}}$ model seems to be better than the C_{10} model, but this is only because these models

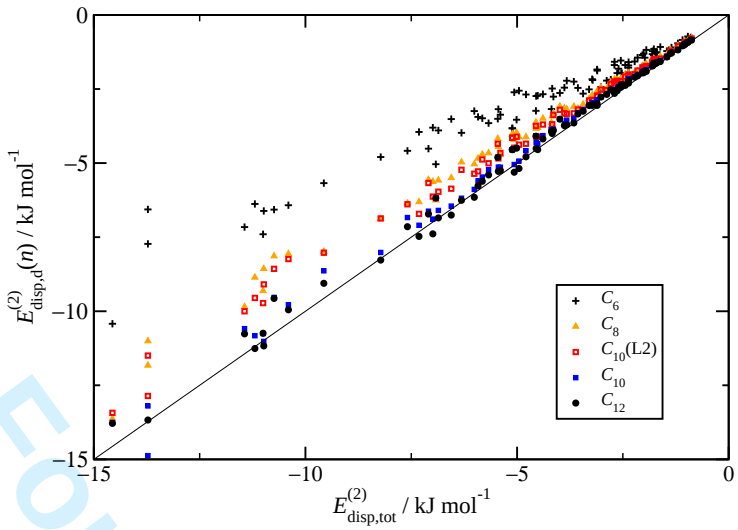


Figure 4: Dispersion energies for the formamide···water dimer. Scatter plot of various anisotropic, damped dispersion models plotted against $E^{(2)}_{\mathrm{disp,tot}}$. All dispersion models have been calculated with the d-aug-cc-pVTZ basis set.

have not been damped. On damping, the isotropic models generally tend to underestimate the dispersion energy slightly, while the agreement is rather good with the anisotropic models (not shown). The same holds true for the N-methyl propanamide and benzene molecules.

The real differences between the anisotropic and isotropic dispersion models becomes visible in the scatter plots of model energies against $E^{(2)}_{\mathrm{disp,tot}}$ from SAPT(DFT). In fig. 8 we present such a scatter plot for the damped isotropic dispersion models for the formamide···water complex. Compare this with fig. 4. One difference that stands out quite dramatically is the increased scatter of the dispersion energies calculated using the damped isotropic models, with the exception of the $C_{6,\mathrm{iso}}$ model. The r.m.s. errors made by the $C_{6,\mathrm{iso}}$, $C_{8,\mathrm{iso}}$, $C_{10,\mathrm{iso}}$ (L2), $C_{10,\mathrm{iso}}$ and $C_{12,\mathrm{iso}}$ models in reproducing $E^{(2)}_{\mathrm{disp,tot}}$ are 2.12, 1.02, 0.85, 0.45 and 0.45 kJ mol^{-1} respectively. It is interesting that the $C_{6,\mathrm{iso}}$ model has a slightly smaller r.m.s. error than the C_6 model. Apart from this anomaly, the errors are larger than those in the corresponding anisotropic models, but they are not very large and could well be acceptable for many situations.

These observations are also true for the formamide and urea dimers (not shown). For the benzene dimer the $C_{10,\mathrm{iso}}$ and $C_{12,\mathrm{iso}}$ models do describe the dispersion energies reasonably well, but errors are non-negligible at the stacked geometries, where these models underestimate the total dispersion energy. For the stacked dimer with molecules separated by 3.8 Å the underestimation with these models is 2.01 and 1.51 kJ mol^{-1}. This error is large compared with the total interaction energy, which is only −7.6 kJ mol^{-1} [21]. However, as we shall see, the errors made by the even simpler scaled $C_{6,\mathrm{iso}}$ model are even larger.

Therefore, the damped $C_{10,\mathrm{iso}}$ model in particular is a good alternative to the more complex C_{10} model. The price to pay is a larger scatter in dispersion energies and an underestimation of the dispersion energies at the stacked configurations of benzene. However, it is possible that at least some of these deficiencies could be compensated by using an anisotropic short-range potential. We are currently investigating this possibility.

8 Isotropic C_6 models

The small scatter of the damped $C_{6,\mathrm{iso}}$ model naturally leads us to the ask whether this model could be scaled to reproduce $E^{(2)}_{\mathrm{disp,tot}}$ better. This scaling could be done during the refinement step of the WSM procedure by biasing the point-to-point polarizabilities to points close to the molecule, for example by choosing the points to lie between the vdW×1.8 and vdW×3.0 surfaces, or by scaling the $C_{6,\mathrm{iso}}$ model to reproduce the SAPT(DFT) energies as well as possible. The former method has the advantage of being unbiased by orientation, but the latter allows us to choose selectively those configurations at which we require higher accuracies. We will adopt the latter method here.

In general, the scaling coefficient would depend on the atom pairs; that is, we would minimize a function of

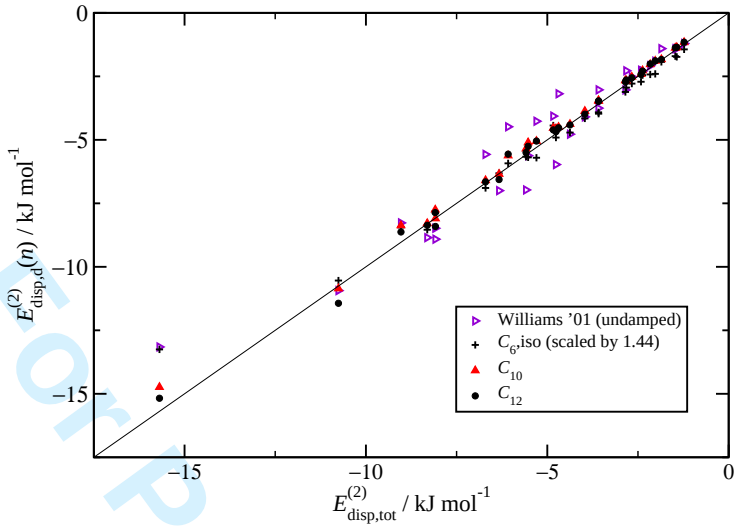


Figure 5: Dispersion energies for the formamide dimer. Scatter plot of dispersion energies calculated using the damped anisotropic C_{10} and C_{12} models and the isotropic scaled C_6 model plotted against $E^{(2)}_{\mathrm{disp,tot}}$ calculated using SAPT(DFT). Also shown are undamped dispersion energies calculated using the C_6 parameters from the Williams '01 potential[39,40]. All dispersion models have been calculated using the d-aug-cc-pVTZ basis set.

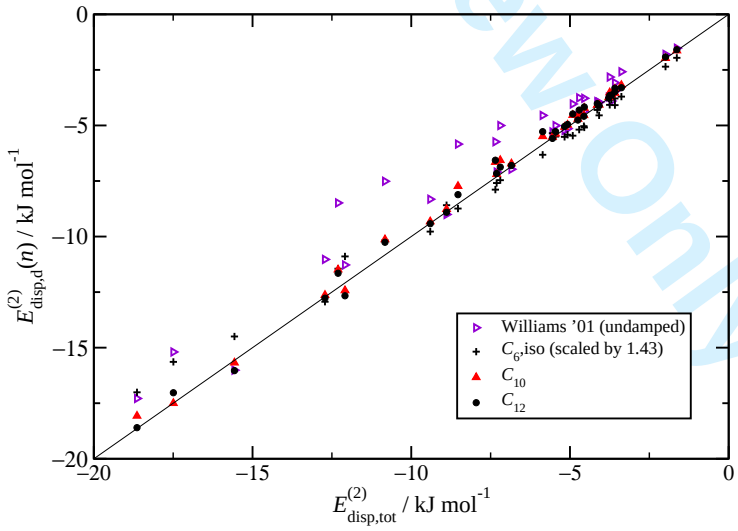


Figure 6: Dispersion energies for the urea dimer. See the caption of fig. 5 for a description. All dispersion models have been calculated using the d-aug-cc-pVTZ basis set.

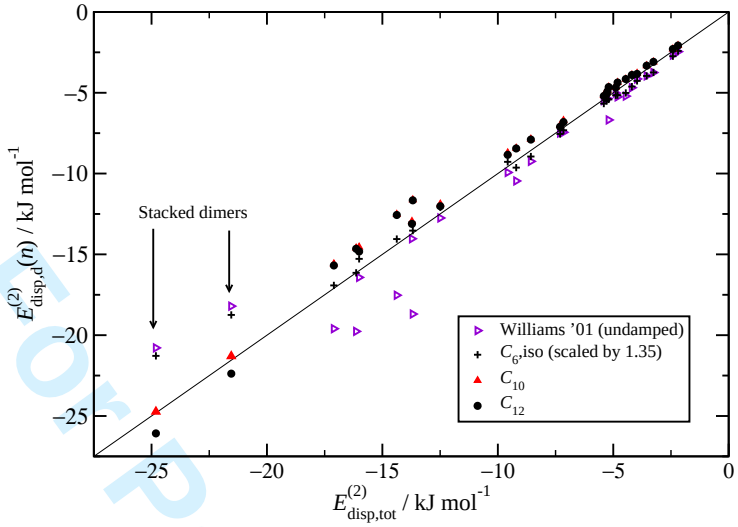


Figure 7: Dispersion energies for the benzene dimer. See the caption of fig. 5 for a description. In contrast to the other dispersion models calculated in this paper, the models for this system have been calculated with the aug-cc-pVTZ basis set.

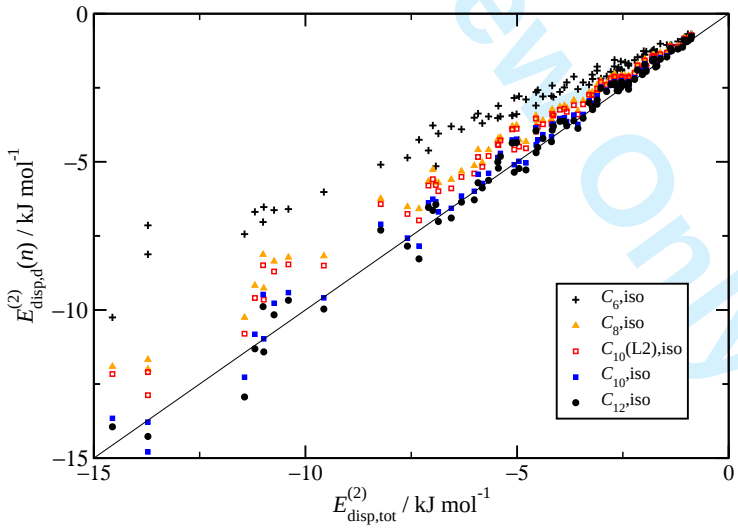


Figure 8: Dispersion energies for the formamide···water dimer. Scatter plot of various isotropic, damped dispersion models plotted against $E^{(2)}_{\mathrm{disp,tot}}$. All dispersion models have been calculated with the d-aug-cc-pVTZ basis set.

the form:

$$\Delta = \sum_i w(i) \left(E^{(2)}_{\mathrm{disp,tot}}(i) + \sum_{a\in A, b\in B} \xi_{ab} \frac{C_6^{ab}, \mathrm{iso}}{R^6_{ab}} \right)^2 \tag{6}$$

where i labels the configurations, the coefficients ξ_{ab} are determined by a least squares fit and w is a weight, which will generally be energy-dependent. Here we have considered the simplest possible fit: all configurations are weighted equally and a single constant of proportionality ξ is used. This constant has been determined to be 1.60, 1.44, 1.43 and 1.35 for the formamide···water complex and formamide, urea and benzene dimers respectively.

The scatter plots of the damped scaled $C_{6,\mathrm{iso}}$ models are shown in figs. 9, 5, 6 and 7. On the whole, the results are very encouraging. The r.m.s. errors for the scaled models are 0.49, 0.52, 0.62 and 0.94 kJ mol^{-1} for the formamide···water complex and the formamide, urea and benzene dimers respectively. These errors are quite tolerable, but there are large errors at configurations with total dispersion energies less than −15 kJ mol^{-1}, where the dispersion is underestimated in magnitude. Likewise, the dispersion energy is overestimated (in magnitude) above −10 kJ mol^{-1}.

These errors are inevitable: the scaled C_6 model is too simple to describe the dispersion energy across the entire range of energies and contacts. What we get is a compromise model that may be adequate for some purposes, but seriously inadequate for others. One such failure occurs at the stacked geometries of the benzene dimer. Because of the importance of this configuration in organic crystals and biological systems, it is important that we get it right. However, the scaled C_6 model under-binds one stacked configuration, with molecules separated by 3.8 Å, by 2.8 kJ mol^{-1}. Compare this to the underestimation by 2.01 and 1.51 kJ mol^{-1} made by the $C_{10,\mathrm{iso}}$ and $C_{12,\mathrm{iso}}$ models (above). In this particular case, the stacked configurations were not included in the data set used to obtain the scaling parameter, so we could have obtained a better fit at this configuration, but it would be at the cost of an even larger over-estimation of the dispersion energies at other geometries.

For the sake of comparison, we have also included energies calculated using the C_6 coefficients taken from the Williams '01 potential parameters[39,40]. These parameters have been obtained by a simultaneous fit of the short-range and long-range parameters to a range of crystal structures and heats of sublimation. Strictly, the long-range parameters cannot be isolated from the short-range parameters, as they will be coupled, but the comparison does give us an idea of how experimentally derived dispersion models compare with ours. The Williams potentials were fitted with undamped dispersion, so damping has not been used. Additionally, the potential specification requires the effective centre of the hydrogen in the C—H and O—H bonds to be foreshortened by 0.1 Å. We have not done this, so the resulting dispersion energies for the formamide and benzene dimers will be slightly overestimated. These results are presented in figs. 5, 6 and 7. The most noticeable feature of these energies is their large scatter, which is much more than any of our models. Additionally, the Williams parameters and our scaled damped C_6 models both fail at configurations with dispersion energies that are large in magnitude. However, for configurations with energies smaller in magnitude than −15 kJ mol^{-1}, our scaled model is considerably better.

Interestingly, for the benzene system our scaled C_6 coefficients and those from the Williams '01 potential differ by only 5 to 8%, yet the dispersion energies calculated using these models differ quite considerably around the −15 kJ mol^{-1} energy. This is due to the lack of damping in the Williams potential. However, damping makes little difference to the scatter in the Williams '01 energies for the urea and formamide dimers.

9 Summary

We have presented a method for calculating dispersion models for small organic molecules that is based on the theoretically well-founded and *ab initio* Williams–Stone–Misquitta method. Using a theoretically motivated damping factor that depends on the vertical ionization potentials of the interacting molecules, we have demonstrated an excellent agreement between the dispersion energies from our most advanced damped models, including anisotropic terms up to C_{12}, and the total dispersion energies from SAPT(DFT). Our best and most complex models reproduce the benchmark SAPT(DFT) energy $E^{(2)}_{\mathrm{disp,tot}}$ with r.m.s. errors of only 0.2 to 0.8 kJ mol^{-1} across an energy range of 15 to 25 kJ mol^{-1}, depending on the complex.

Though accurate, these models are probably too complex to be used in practice, so we have made use of the inherent flexibility of the WSM method to construct isotropic polarizability models from which we have calculated isotropic dispersion models, which have proved only slightly less accurate than their anisotropic counterparts. In particular, using a large sample of configurations of the formamide···water complex and the formamide, urea and benzene dimers, we have shown the isotropic $C_{10,\mathrm{iso}}$ model to result in r.m.s. errors between 0.45 and 1.06 kJ mol^{-1}, making this model ideal for reasonably accurate calculations on a variety of systems. However, we emphasise that the isotropic models can be deficient for molecules like benzene which exhibit a large anisotropy. In particular, there are undoubted shortcomings at the important stacked configurations of benzene. This is the price we may have to pay for simplicity.

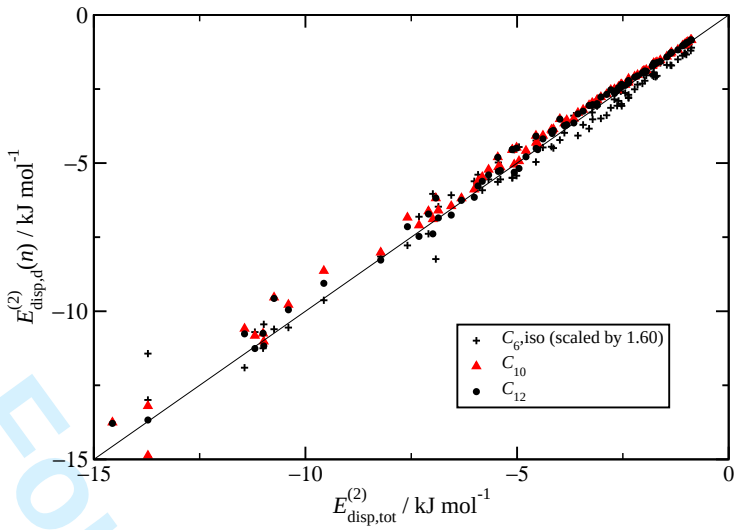


Figure 9: Dispersion energies for the formamide···water dimer. Scatter plot of dispersion energies calculated using the damped anisotropic C_{10} and C_{12} models and the isotropic scaled C_6 model plotted against $E^{(2)}_{\mathrm{disp,tot}}$ calculated using SAPT(DFT). All dispersion models have been calculated using the d-aug-cc-pVTZ basis set.

Simple though the $C_{10,\mathrm{iso}}$ models are, many existing programs are not in a position to use them, as they are limited to isotropic C_6 models only. Therefore we have investigated scaled $C_{6,\mathrm{iso}}$ models, and have shown that they can be constructed to be of reasonable accuracy, but at the cost of larger errors at potentially important dimer configurations, like the stacked geometry of the benzene dimer. Additionally, the scaled $C_{6,\mathrm{iso}}$ models tend to over-bind at configurations with total dispersion energies less than about $-7\,\mathrm{kJ\,mol^{-1}}$ in magnitude. Since the bulk of the dispersion interactions occur in this energy range, these models could over-bind denser and more compact structures over the more open ones. Interestingly, the scaling parameter is not a constant, but varies from 1.35 to 1.60 across systems. This has implications about the transferability of the derived atomic dispersion coefficients and seems to suggest that we will have to look at the transferability of larger groups instead.

All of these models have been constructed using the CamCASP[30] program. Such calculations are now quite routine on a modest workstation. But we should point out that these accurate dispersion models cannot be used with repulsion parameters derived empirically. Each system requires a dedicated set of parameters derived from first principles. We are working on developing an efficient route to obtaining these parameters.

10 Programs

Many of the theoretical methods described in this paper are implemented in programs available for download. Some of these, together with their main uses in the present work, are:

- Sapt2006[31]: $E^{(2)}_{\mathrm{disp,exch}}$ calculations.
- CamCASP 5.1.02[30]: Calculation of WSM polarizabilities, the dispersion models and the SAPT(DFT) dispersion energies.
- Orient 4.6[41]: Localization of the distributed polarizabilities, calculation of dimer energies using the dispersion models and visualization of the energy maps.
- Dalton 2.0[42]: DFT and CKS calculations. A patch[31] is needed to enable Dalton 2.0 to work with CamCASP.

11 Acknowledgements

AJM would like to dedicate this article to Anthony J. Stone on his 70th birthday as a small token in recognition of the sheer fun it has been working with him these last few years. We are grateful to Sarah L. Price for her comments and suggestions. AJM wishes to thank Girton College, Cambridge for a research fellowship. This research was supported by EPSRC grant EP/C539109/1.

References

[1] Millot, C. and Stone, A. J., 1992, *Mol. Phys.*, **77**, 439.

[2] Mas, E. M., Bukowski, R., Szalewicz, K., Groenenboom, G. C., Wormer, P. E. S., and van der Avoird, A., 2000, *J. Chem. Phys.*, **113**, 6687.

[3] Podeszwa, P., Bukowski, R., and Szalewicz, K., 2006, *J. Phys. Chem. A*, **110**, 10345.

[4] Stone, A. J. and Tough, R. J. A., 1984, *Chem. Phys. Lett.*, **110**, 123.

[5] Jeziorski, B., Moszynski, R., and Szalewicz, K., 1994, *Chem. Rev.*, **94**, 1887.

[6] Jeziorski, B. and Szalewicz, K., 1998, in *Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer *et al.*, volume 2, p. 1376, (Wiley, Chichester, UK).

[7] Jeziorski, B. and Szalewicz, K., 2002, in *Handbook of Molecular Physics and Quantum Chemistry*, edited by S. Wilson, volume 8, chapter 8, pp. 37–83, (Wiley).

[8] Misquitta, A. J. and Szalewicz, K., 2002, *Chem. Phys. Lett.*, **357**, 301.

[9] Misquitta, A. J., Jeziorski, B., and Szalewicz, K., 2003, *Phys. Rev. Lett.*, **91**, 33201.

[10] Misquitta, A. J. and Szalewicz, K., 2005, *J. Chem. Phys.*, **122**, 214109.

[11] Misquitta, A. J., Podeszwa, R., Jeziorski, B., and Szalewicz, K., 2005, *J. Chem. Phys.*, **123**, 214103.

[12] Hesselmann, A. and Jansen, G., 2002, *Chem. Phys. Lett.*, **357**, 464.

[13] Hesselmann, A. and Jansen, G., 2002, *Chem. Phys. Lett.*, **362**, 319.

[14] Hesselmann, A. and Jansen, G., 2003, *Chem. Phys. Lett.*, **367**, 778.

[15] Stone, A. J., 1996, *The Theory of Intermolecular Forces*, (Clarendon Press, Oxford).

[16] Misquitta, A. J. and Stone, A. J., 2006, *J. Chem. Phys.*, **124**, 024111.

[17] Stone, A. J. and Misquitta, A. J., 2007, *Int. Revs. Phys. Chem.*, **26**, 193.

[18] Hattig, C., Jansen, G., Hess, B. A., and Angyan, J. G., 1997, *Mol. Phys.*, **91**, 145.

[19] Williams, G. J. and Stone, A. J., 2003, *J. Chem. Phys.*, **119**, 4620.

[20] Misquitta, A. J. and Stone, A. J., 2008, *J. Chem. Theory Comput.*, **4**, 7.

[21] Misquitta, A. J., Stone, A. J., and Price, S. L., 2008, *J. Chem. Theory Comput.*, **4**, 19.

[22] Grimme, S., 2004, *J. Comp. Chem.*, **25**, 1463.

[23] Le Sueur, C. R. and Stone, A. J., 1994, *Mol. Phys.*, **83**, 293.

[24] Lillestolen, T. C. and Wheatley, R. J., 2007, *J. Phys. Chem. A*, **111**, 11141.

[25] Stone, A. J. and Tong, C. S., 1989, *Chem. Phys.*, **137**, 121.

[26] Williams, G. J., 2004, *Molecular distributed polarizabilities*, Ph.D. thesis, Churchill College, University of Cambridge.

[27] Podeszwa, R., Bukowski, R., and Szalewicz, K., 2006, *J. Chem. Theory Comput.*, **2**, 400.

[28] Williams, H. L., Mas, E. M., Szalewicz, K., and Jeziorski, B., 1995, *J. Chem. Phys.*, **103**, 7374.

[29] Akin-Ojo, O., Bukowski, R., and Szalewicz, K., 2003, *J. Chem. Phys.*, **119**, 8379.

[30] Misquitta, A. J. and Stone, A. J., 2007, CamCASP*: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form*, University of Cambridge, http://www-stone.ch.cam.ac.uk/programs.html#CamCASP.

[31] Bukowski, R., Cencek, W., Jankowski, P., Jeziorski, B., Jeziorska, M., Kucharski, S. A., Lotrich, V. F., Misquitta, A. J., Moszynski, R., Patkowski, K., Podeszwa, R., Rybak, S., Szalewicz, K., Williams, H. L., Wheatley, R. J., Wormer, P. E. S., and Zuchowski, P. S., 2006, *SAPT2006: an ab initio program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies*, University of Delaware and University of Warsaw.

[32] Lias, S. G., 2000, in *NIST Chemistry WebBook (NIST Standard Reference Database Number 69)*, edited by W. G. Mallard and P. J. Linstrom, (NIST, Gaithersburg), (http://webbook.nist.gov).

[33] Misquitta, A. J., Welch, G. W. A., Stone, A. J., and Price, S. L., 2008, *Chem. Phys. Lett.*, **456**, 105.

[34] Shoemake, K., 1992, in *Graphics Gems III*, pp. 124–132, (Academic Press).

[35] Bondi, A., 1964, *J. Phys. Chem.*, **68**, 441.

[36] Tang, K. T. and Toennies, J. P., 1984, *J. Chem. Phys.*, **80**, 3726.

[37] Sadlej, A. J., 1988, *Coll. Czech Chem. Commun.*, **53**, 1995.

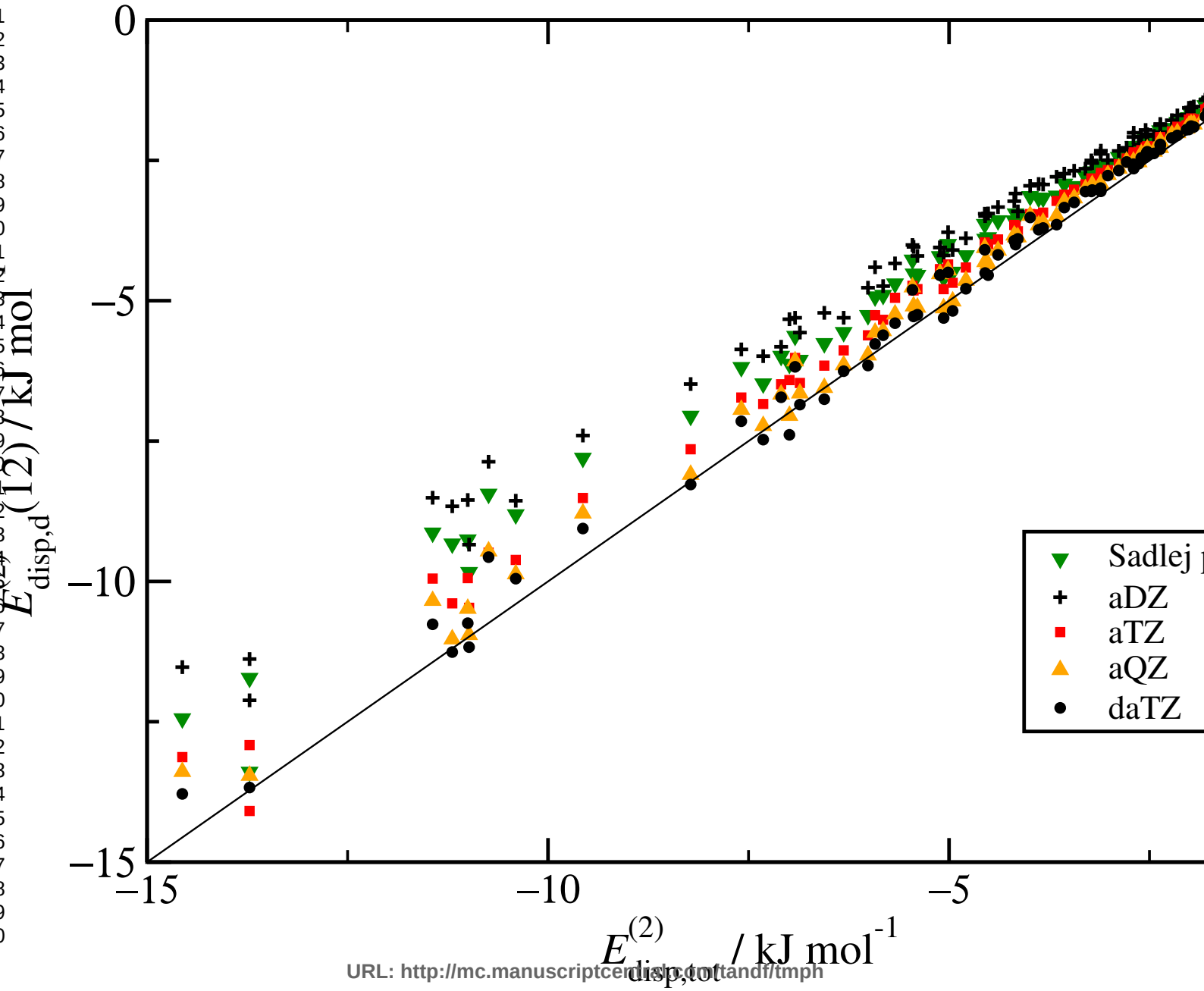
[38] Sadlej, A. J., 1991, *Theor. Chim. Acta*, **79**, 123.

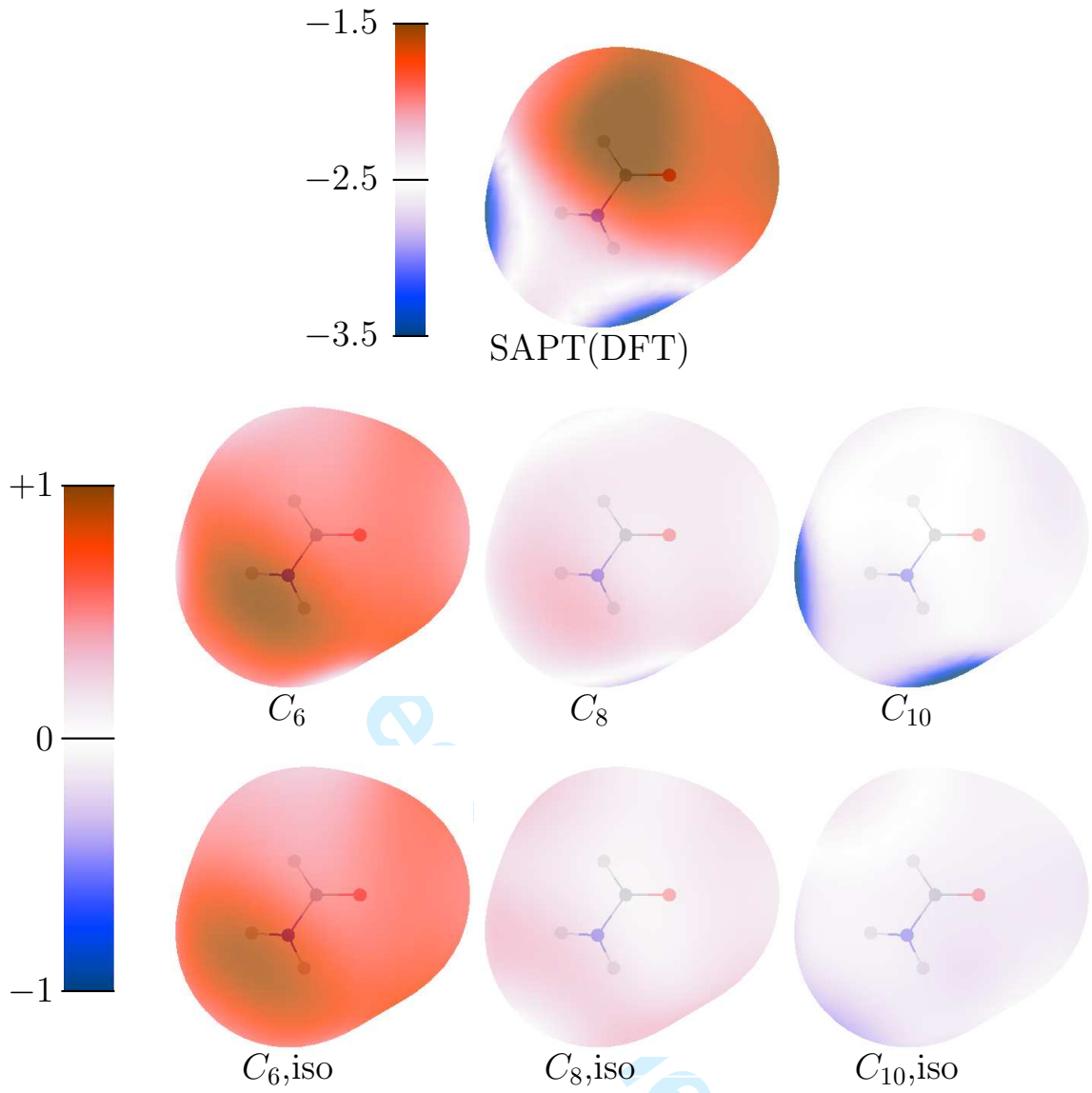
[39] Williams, D. E., 2001, *J. Comp. Chem.*, **22**, 1.

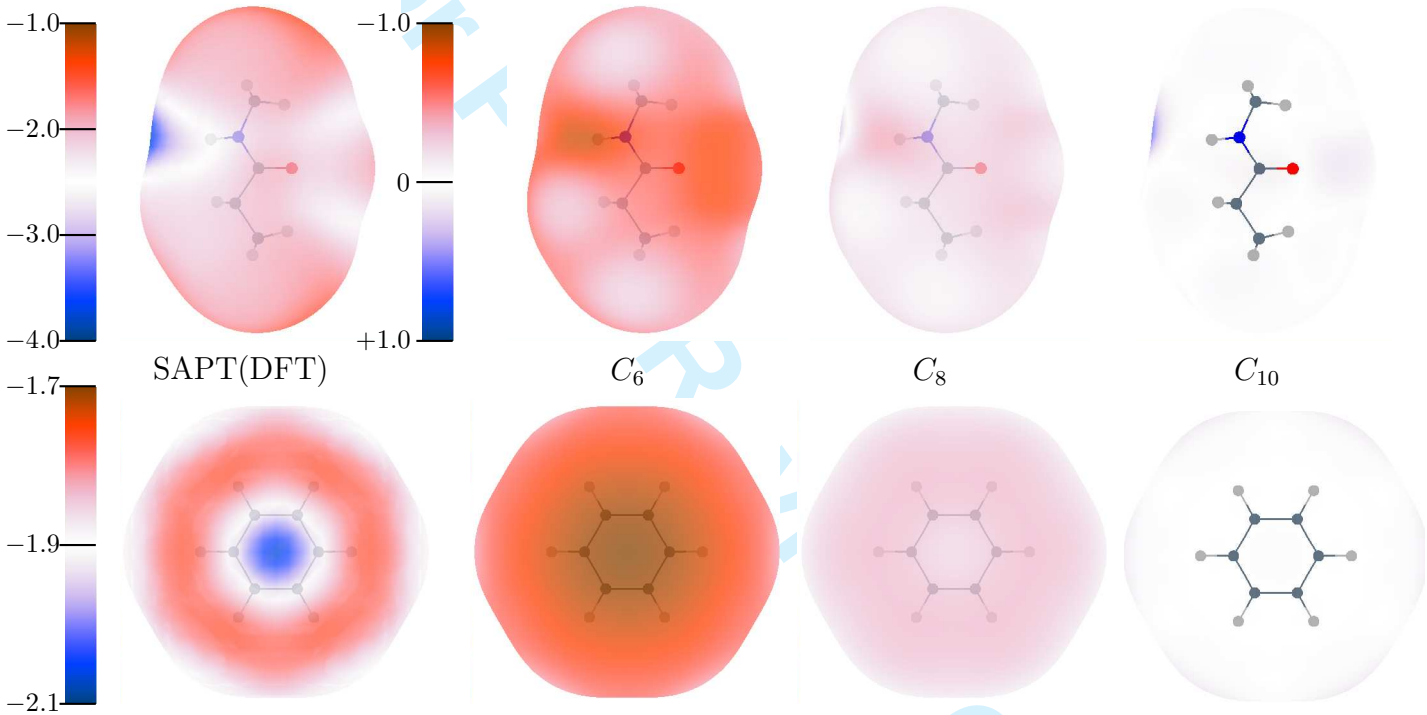
[40] Williams, D. E., 2001, *J. Comp. Chem.*, **22**, 1154.

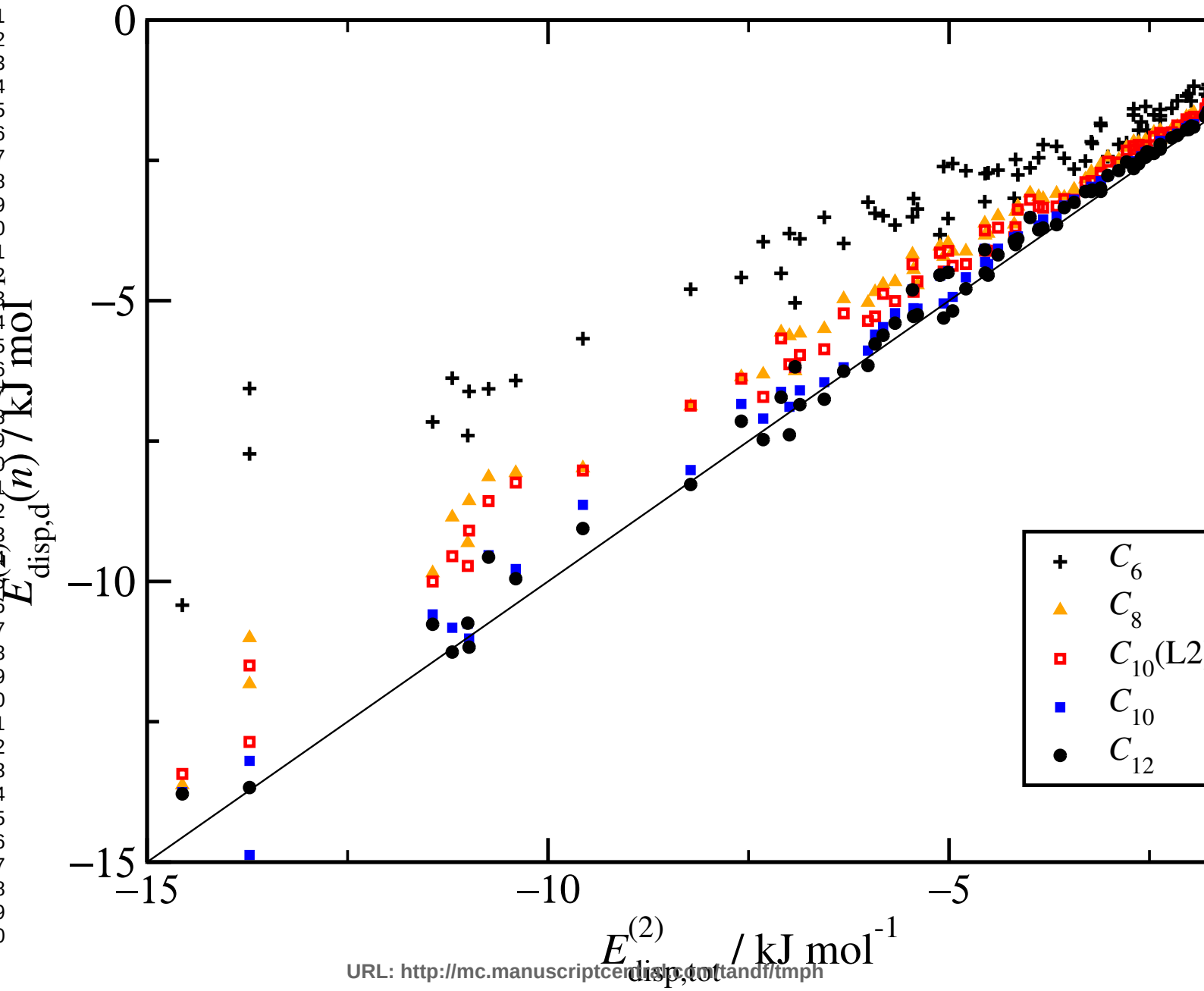
[41] Stone, A. J., Dullweber, A., Engkvist, O., Fraschini, E., Hodges, M. P., Meredith, A. W., Nutt, D. R., Popelier, P. L. A., and Wales, D. J., 2006, Orient*: a program for studying interactions between molecules, version 4.6*, University of Cambridge, http://www-stone.ch.cam.ac.uk/programs.html#Orient.

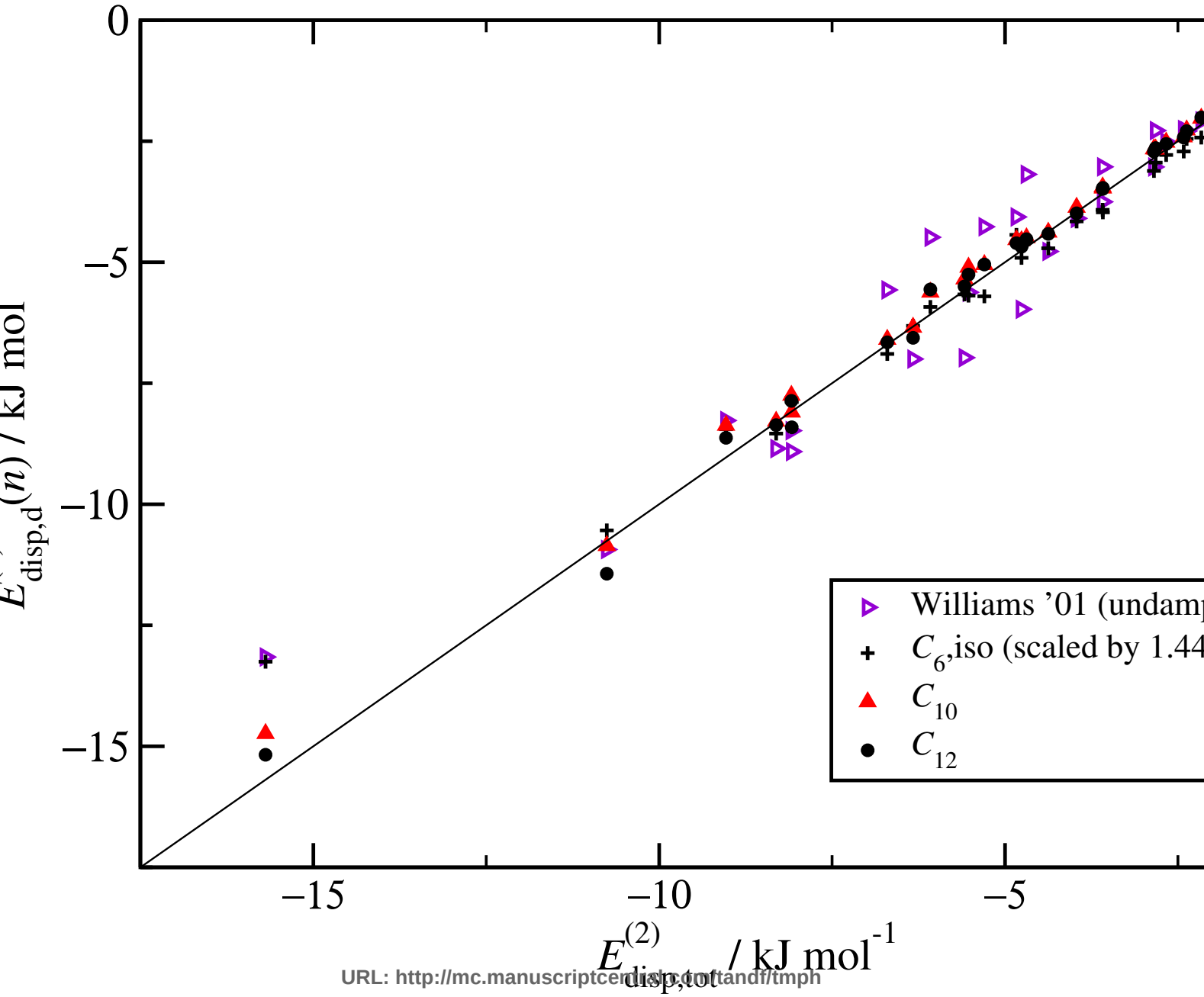
[42] *DALTON, a molecular electronic structure program, Release 2.0*, 2005. See http://www.kjemi.uio.no/software/dalton/dalton.html.

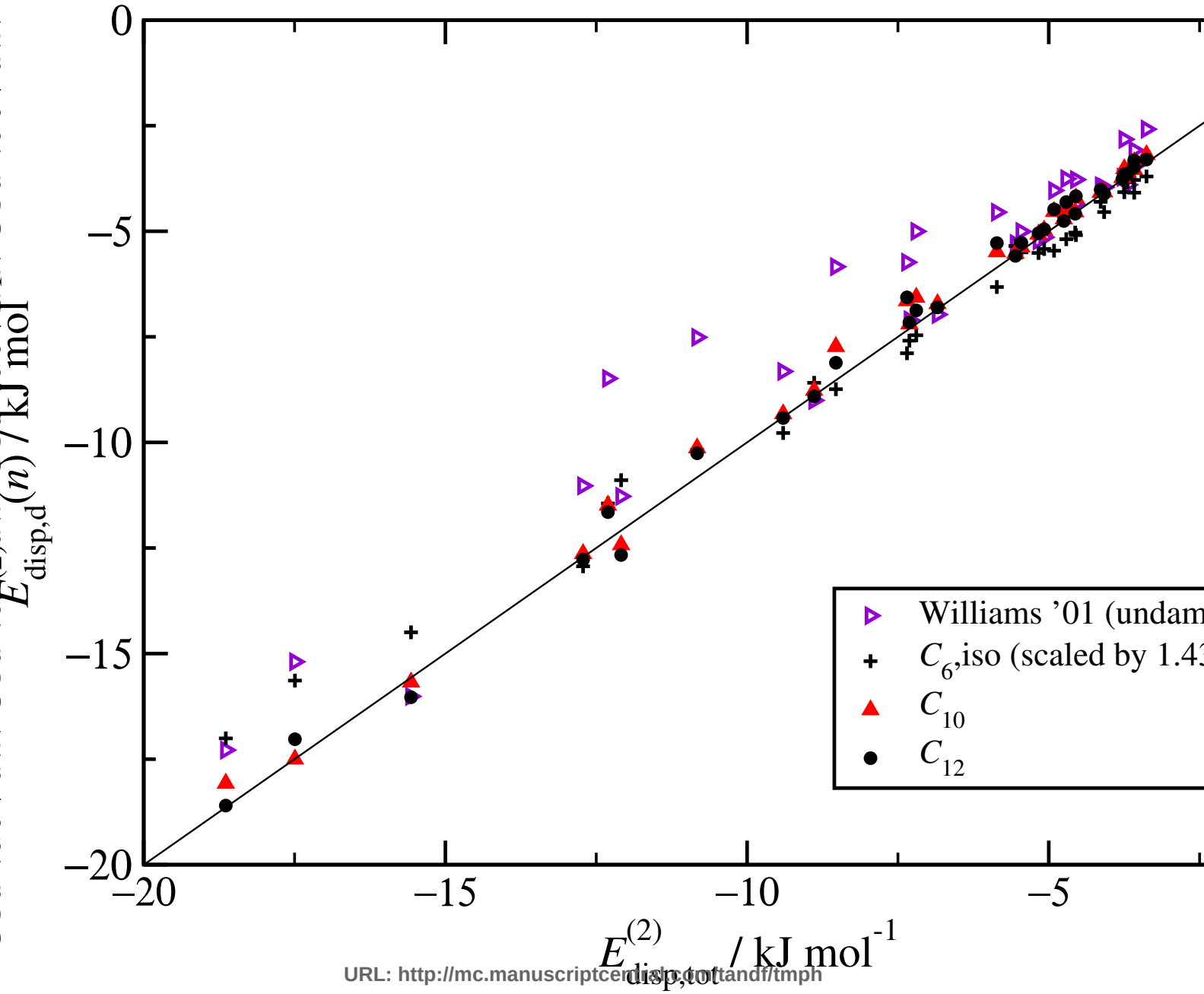

0
−5
−10
−15
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(12)$ / kJ mol
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
Sadlej
aDZ
aTZ
aQZ
daTZ

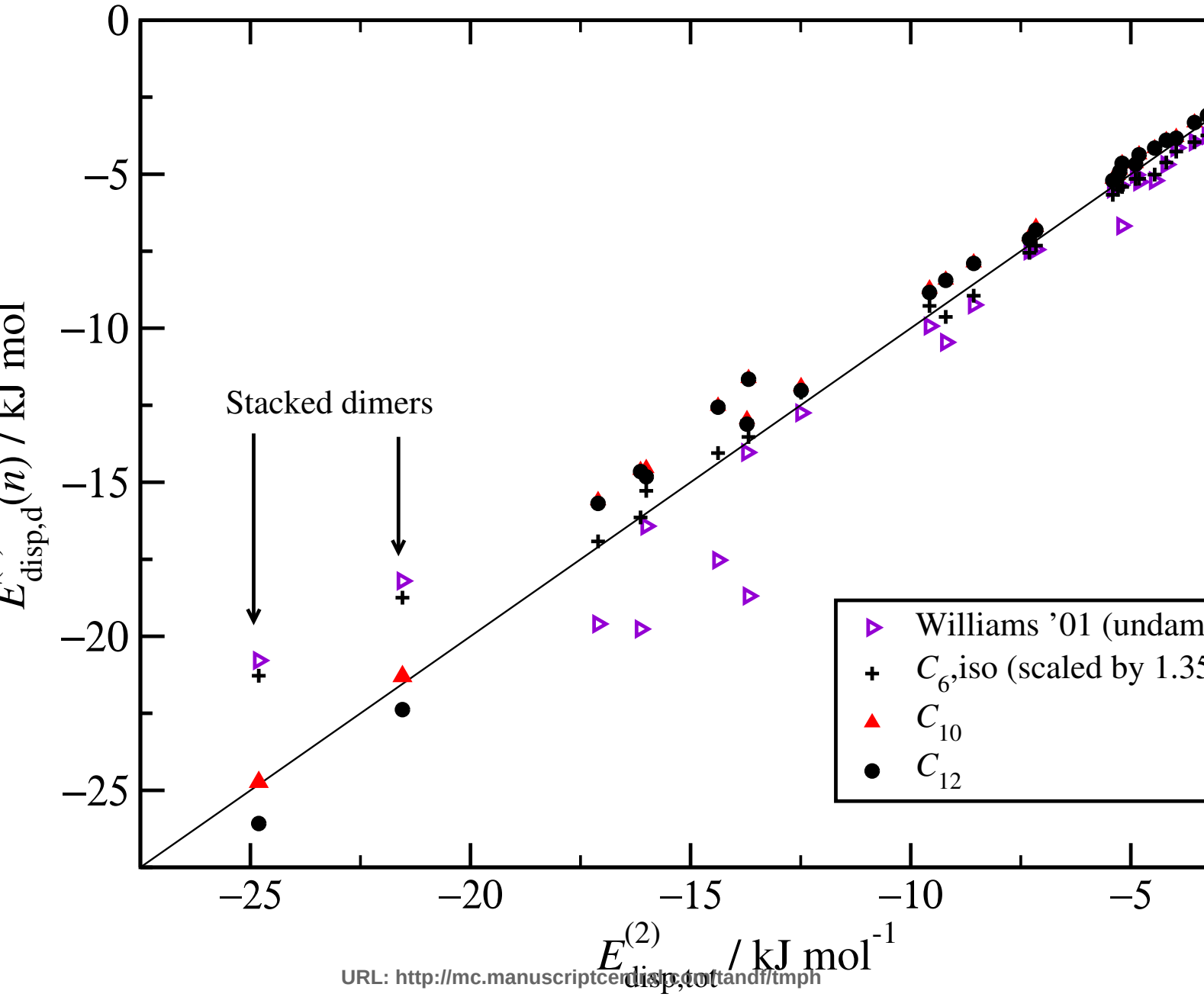


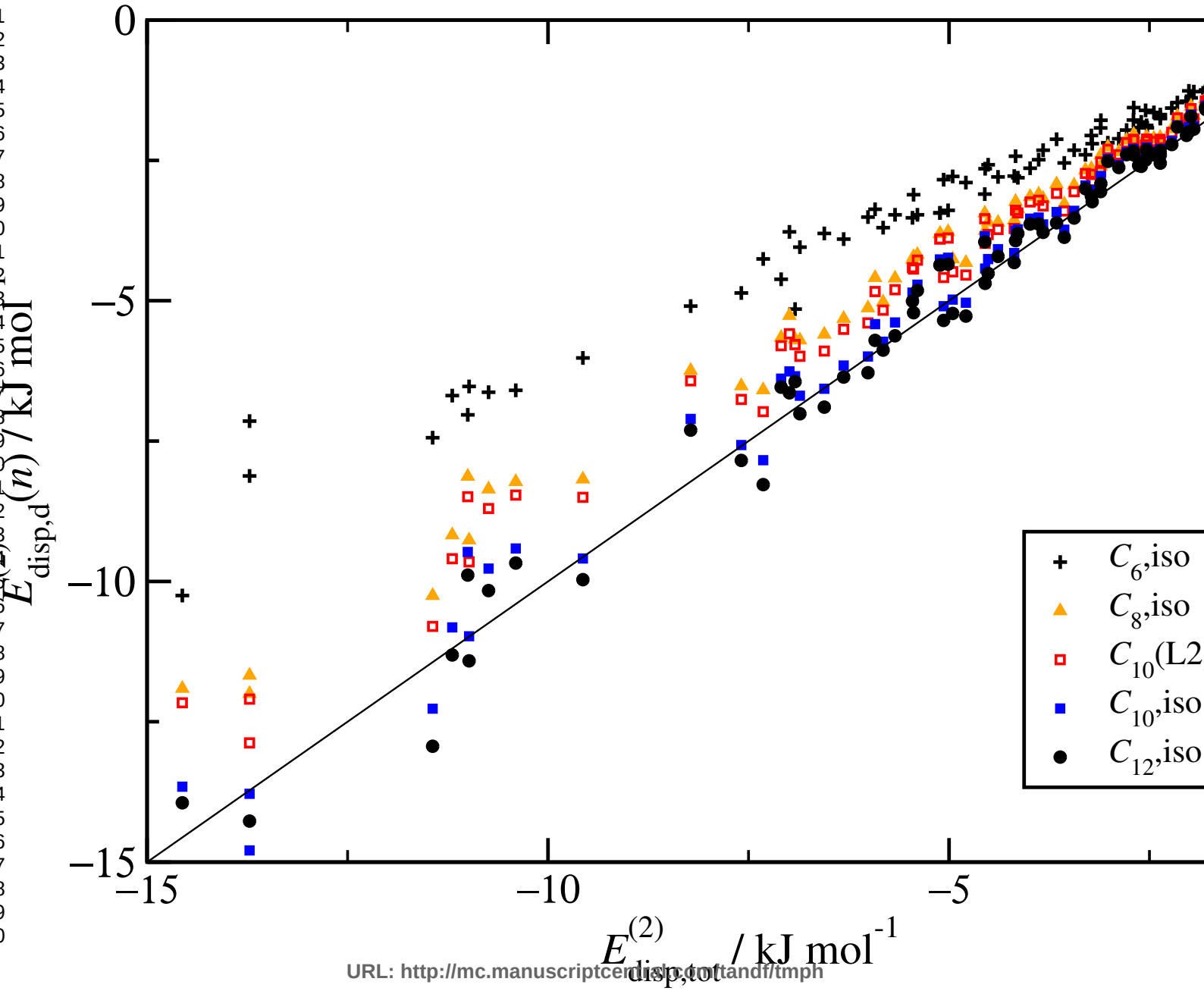


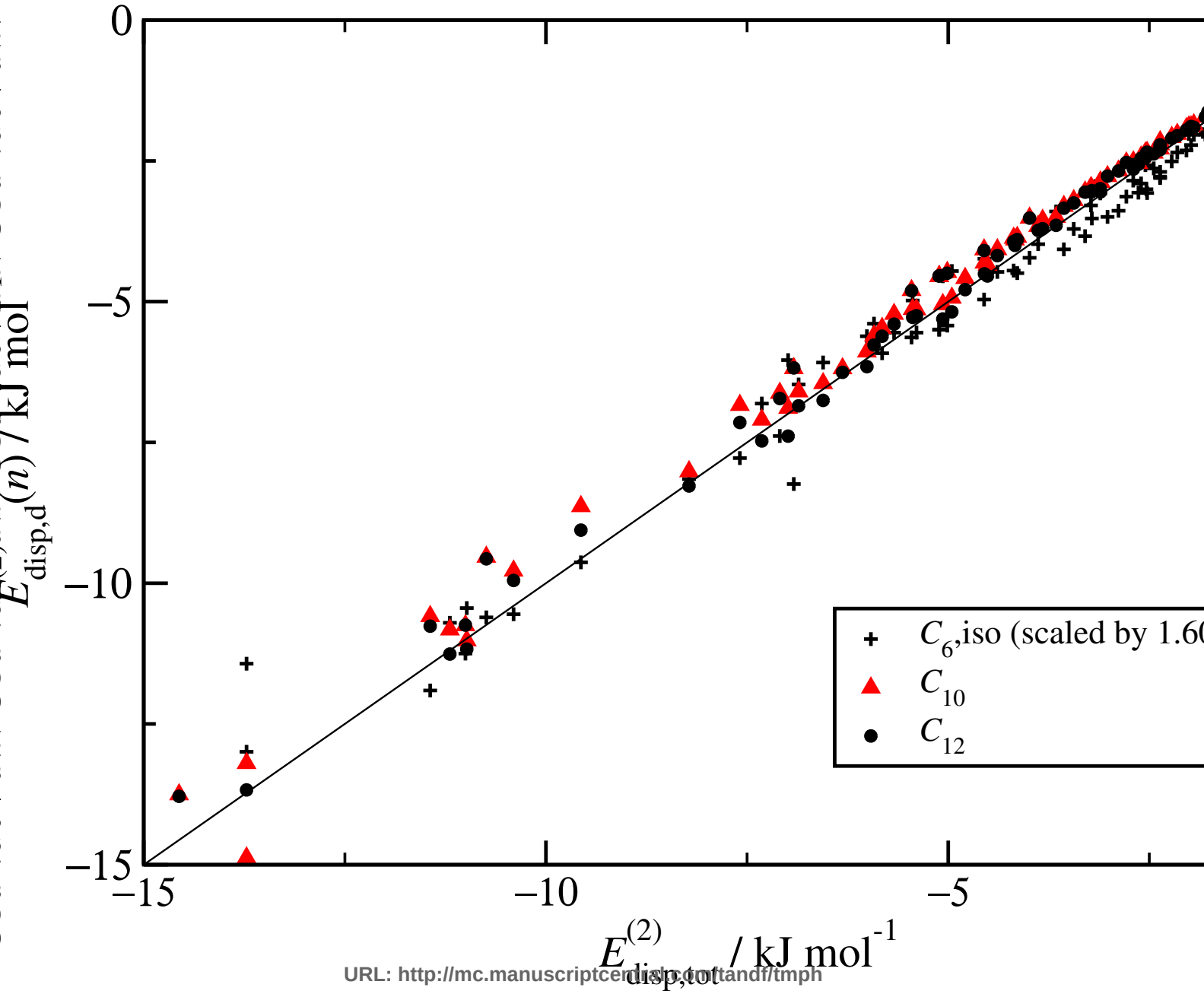

0
−5
−10
−15
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
C_6
C_8
C_{10}(L2
C_{10}
C_{12}


0
−5
−10
−15
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol^{-1}
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
Williams '01 (undamp
C_6,iso (scaled by 1.44
C_{10}
C_{12}


0
−5
−10
−15
−20
−20
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
Williams '01 (undam
C_6,iso (scaled by 1.43
C_{10}
C_{12}


0
−5
−10
−15
−20
−25
Stacked dimers
Williams '01 (undam
C_6,iso (scaled by 1.35
C_{10}
C_{12}
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol^{-1}
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}


0
−5
−10
−15
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol^{-1}
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
C_6,iso
C_8,iso
C_{10}(L2
C_{10},iso
C_{12},iso


0
−5
−10
−15
−15
−10
−5
$E^{(2)}_{\mathrm{disp,d}}(n)$ / kJ mol^{-1}
$E^{(2)}_{\mathrm{disp,tot}}$ / kJ mol^{-1}
C_6,iso (scaled by 1.60
C_{10}
C_{12}