Cambridge Centre for Computational Chemical Engineering

University of Cambridge

Department of Chemical Engineering and Biotechnology

Preprint ISSN 1473 – 4273

A first principles development of a general anisotropic potential for polycyclic aromatic hydrocarbons

Tim S. Totton[1], Alston J. Misquitta[2], Markus Kraft[1]

released: 12 September 2009

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site, Pembroke Street
Cambridge CB2 3RA
United Kingdom

[2] Department of Physics
Cavendish Laboratory
University of Cambridge
J J Thomson Avenue
Cambridge, CB3 0HE
United Kingdom

Preprint No. 81

c4e

Key words and phrases: PAH, anisotropy, intermolecular potential, soot modelling

Edited by
Cambridge Centre for Computational Chemical Engineering
Department of Chemical Engineering and Biotechnology
University of Cambridge
Cambridge CB2 3RA
United Kingdom.

Fax: + 44 (0)1223 334796
E-Mail: c4e@cheng.cam.ac.uk
World Wide Web: http://www.cheng.cam.ac.uk/c4e/

Abstract

Standard empirical atom-atom potentials are shown to be unable to describe the binding of PAH molecules in the variety of configurations seen in clusters. The main reason for this inadequacy is the lack of anisotropy in these potentials. We have constructed an anisotropic atom-atom intermolecular potential for the benzene molecule from first principles using SAPT(DFT) interaction energy calculations and the Williams-Stone-Misquitta method for obtaining molecular properties in distributed form. Using this potential as a starting point we have constructed a transferable anisotropic potential to model intermolecular interactions between polycyclic aromatic hydrocarbons (PAH). This new potential has been shown to accurately predict interaction energies for a variety of dimer configurations for four different PAH molecules including certain configurations which are poorly predicted with current isotropic potentials. It is intended that this potential will form the basis for further work on the aggregation of PAHs.

Contents

1 Introduction

Polycyclic aromatic hydrocarbon (PAH) molecules have often been invoked as intermediates in the chemistry of soot formation and growth [1]. The presence of stacked PAH molecular structures in experimental HRTEM images of soot particles [2–4] has led some to suggest that the intermolecular binding of PAH molecules may be responsible for particle inception. This hypothesis has provoked a large number of theoretical studies on the stability and relative orientation of PAH molecules present in dimers and larger stacks in flame environments [1, 5–9]. Currently, many numerical simulations of soot formation in flames consider the dimerisation of molecules as small as pyrene ($C_{16}H_{10}$) [9, 10] to be the particle inception step, however the validity of this assumption is still debated [11].

Previous studies into the intermolecular chemistry of PAHs have been largely based on computationally convenient model potentials such as isotropic Lennard-Jones 12-6 and exp-6 potentials [1, 12, 13]. In general atom-atom potentials approximate the total interaction energy, U, as sum over all pairwise atomic interactions between molecules:

$$U = \sum_{A} \sum_{A<B} \sum_{a \in A} \sum_{b \in B} U_{ab}(R_{ab}, \Omega_{ab}). \tag{1}$$

Here $U_{ab}(R_{ab}, \Omega_{ab})$ denotes an atom-atom interaction potential. The indices A and B are for molecules, and the indices a and b run over all the atomic sites within these molecules. In general the interaction potential depends upon the atom-atom separation, R_{ab}, and the relative molecular orientation, described in some way by Ω_{ab}. Often, however, orientational dependence is removed as a simplification and such potentials are 'isotropic', *i.e.* the atoms in a molecule are considered to be spherically symmetric. Common functional forms for $U_{ab}(R_{ab}, \Omega_{ab})$ include the 12-6 Lennard-Jones potential (Eq. 2) and the exp-6 potential (Eq. 3). Explicit electrostatic models are often added to these forms, the simplest being based upon partial atom-centred point charges (Eq. 4).

$$U_{\text{LJ}} = 4\epsilon_{ab}\left[\left(\frac{\sigma_{ab}}{R_{ab}}\right)^{12} - \left(\frac{\sigma_{ab}}{R_{ab}}\right)^{6}\right]. \tag{2}$$

$$U_{\text{exp-6}} = B_{ab}\exp(-C_{ab}R_{ab}) - \frac{A_{ab}}{R_{ab}^{6}}. \tag{3}$$

$$U_{\text{elst}} = \frac{q_a q_b}{R_{ab}}. \tag{4}$$

In recent years, with the advance of computational power, the theoretical understanding of intermolecular interactions has developed significantly yet empirical potentials have remained largely unchanged. Current isotropic literature potentials have typically been parameterised to be applicable to a wide range of organic molecules [14–17]. Such potentials are typically parameterised with heats of sublimation and crystallographic data, and whilst transferable, are only accurate for the configurations they were parameterised for and often fail at others. For example, consider two widely used potentials: the Williams W99 potential [15, 16] based on the exp-6 form and a 12-6 Lennard-Jones potential [14] both including point charges (the latter parameterised from an earlier form of the Williams potential). The performance of these isotropic potentials has been examined for naphthalene and anthracene dimer orientations shown in Figs. 1 and 2. In Figs. 3 and 4 we show

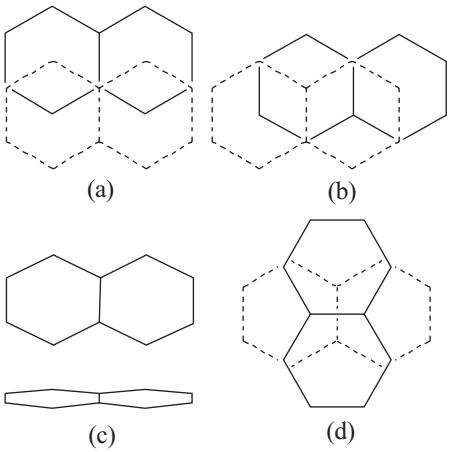


Figure 1: *Naphthalene: (a) Slipped-parallel, symmetry C_{2h} (b) Graphite-type, symmetry C_i (c) T-shape, symmetry C_{2v} (d) Crossed, symmetry D_{2d}.*

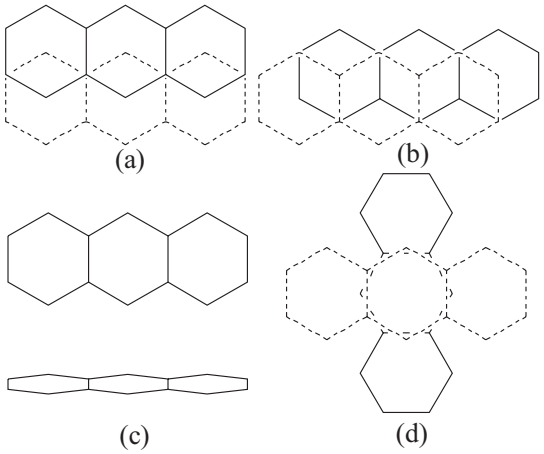


Figure 2: *Anthracene: (a) Slipped-parallel, symmetry C_{2h} (b) Graphite-type, symmetry C_i (c) T-shape, symmetry C_{2v} (d) Crossed, symmetry D_{2d}.*

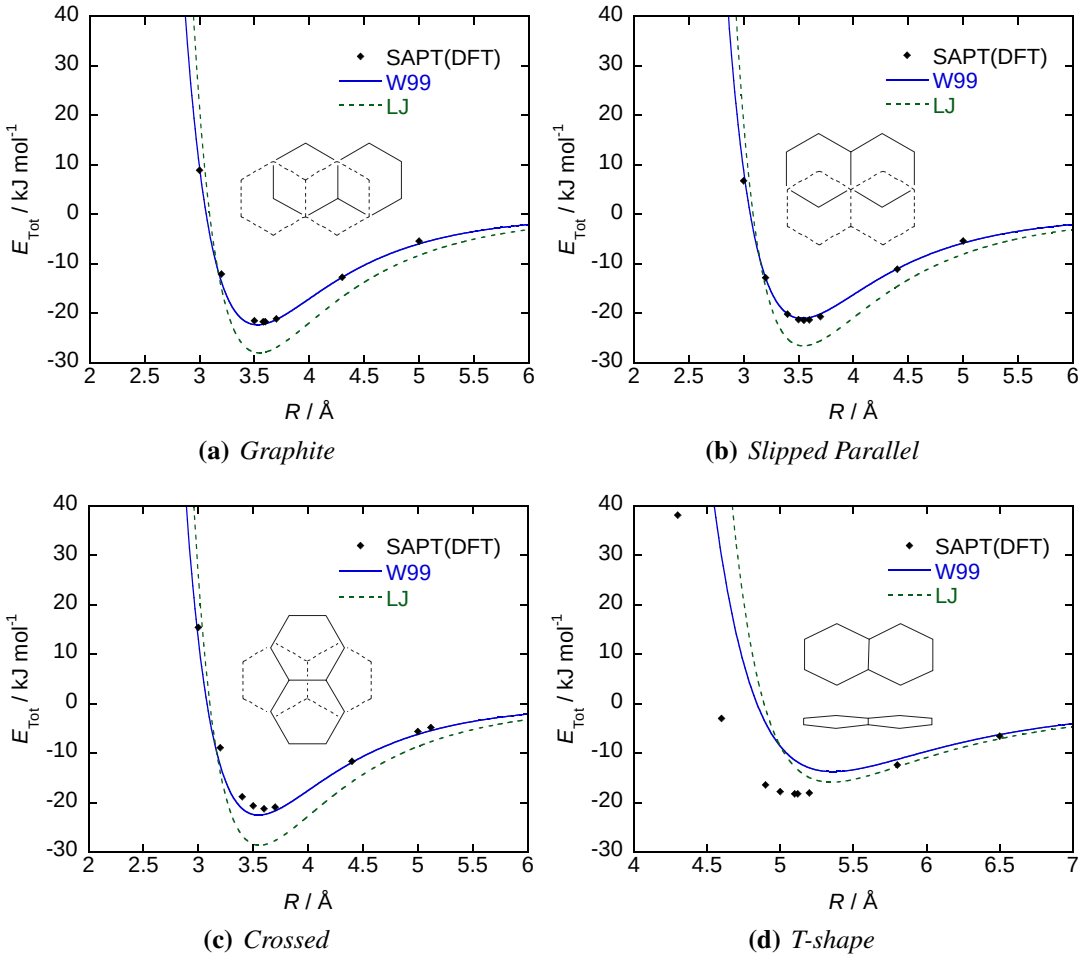
cross sections of the potential energy surface at these orientations where separation, R, is between the centres of mass of the monomers. The reference energies are taken from *ab initio* SAPT(DFT) calculations performed by Podeszwa and Szalewicz [18].

The W99 potential performs remarkably well for stacked PAH geometries whilst the Lennard-Jones plus point charges potential tends to overestimate well depths by 5–10 kJ mol^{-1}. However both potentials show substantial errors for the T-shape dimer with an error in equilibrium separation of 0.3–0.4 Å and an underbinding of as much as 7 kJ mol^{-1} in the case of the W99 potential.

Being isotropic, the potentials cannot accurately model the atom-atom interactions where there is significant anisotropy in the electron distribution around constituent atoms, such as in PAHs where there is significant π-bonding. These potentials also suffer from being required to possess too large a degree of transferability to make them sufficiently accurate for the specific system of interest, and development of anisotropic potentials empirically is precluded due to insufficient data.

To accurately model dimers in all orientations either accurate *ab initio* methods must be used directly (on-the-fly methods) or new anisotropic atom-atom potentials are required, parameterised using *ab initio* results. Currently most *ab initio* methods are prohibited due to high computational expense. Density functional theory is the only method which is computationally feasible but currently there are no practical and quantitative functionals which correctly predict intermolecular dispersion energies. In reality the size of PAH systems and the complexity of the calculations restrict us to using model atom-atom potentials. In the context of PAHs and investigating soot structure at a molecular level there are a number of requirements for a potential:

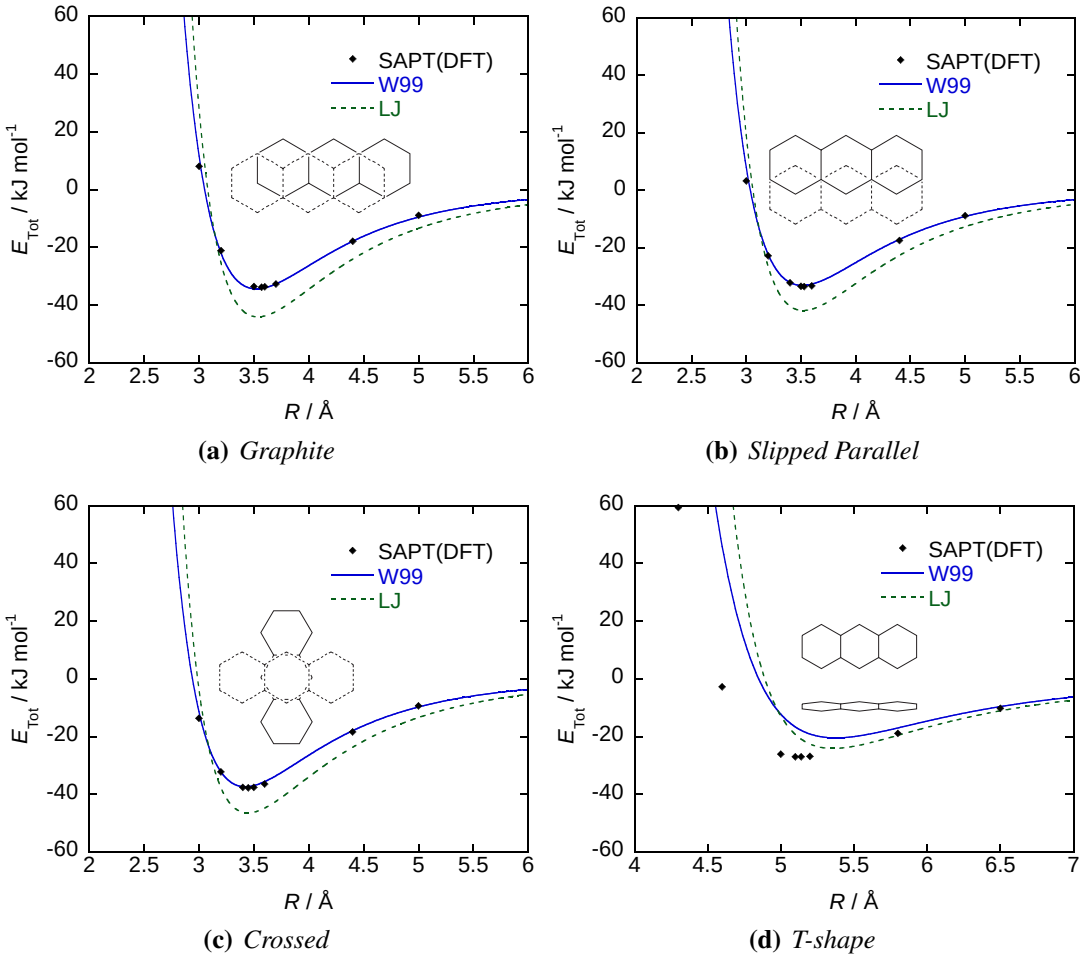
- **Accuracy**: The potential is expected to be accurate for the variety of dimer configurations which are expected to be sampled in a flame environment. In particular the potential must correctly predict barriers on the potential energy surface (PES)



(a) *Graphite* **(b)** *Slipped Parallel*

(c) *Crossed* **(d)** *T-shape*

Figure 3: *Comparison of isotropic model potentials with SAPT(DFT) energies for different naphthalene dimers. A key to the geometries is given in Fig. 1. Model potential energies have been calculated using the* ORIENT *[19] program.*



(a) *Graphite* **(b)** *Slipped Parallel*

(c) *Crossed* **(d)** *T-shape*

Figure 4: *Comparison of isotropic model potentials with SAPT(DFT) energies for different anthracene dimers. A key to the geometries is given in Fig. 2. Model potential energies have been calculated using the* ORIENT *[19] program.*

of the molecular cluster.

- **Transferability**: In a flame environment typically large ensembles of different PAH molecules exist of varying size (C_6 – C_{400}) [20] and consequently it is very important that any potential developed can be easily transferably to different PAHs.

- **Simplicity**: Large PAH clusters need to be studied requiring extensive calculations. This will limit the functional complexity of the potential expression used to model interactions. In later work it is hoped that a coarse-grained potential can be developed which reflects this requirement, but this will require an accurate atom-atom potential as a reference.

In the context of PAHs there have been several studies using *ab initio* methods. For example, coupled cluster calculations at CCSD(T) level have been used to study naphthalene dimers [21] whilst MP2 level calculations have been used to obtain dimer interaction energies for various PAHs [10]. However, Møller-Plesset perturbation theory is inadequate to study intermolecular interactions between systems with a significant amount of π-bonding (such as PAH clusters). Compared to the more reliable CCSD(T) calculations, MP2 calculations have been shown to considerably overestimate attraction between molecules, in some cases by almost a factor of two [21–23] throwing into doubt some earlier studies of soot particle inception [10]. However CCSD(T) is computationally demanding and is not really suitable for potential development due to the inability to decompose the overall interaction energy into physically significant contributions. This makes it hard to parameterise analytic potentials comprised of multiple terms used to describe different interactions, such as dispersion and repulsion.

By contrast, intermolecular perturbation theory provides an ideal framework for the development of model potentials because it provides the interaction energy as a sum of physically significant contributions. This allows the separate parameterisation of different terms representing different interactions within multi-term model potentials. The development of symmetry-adapted perturbation theories has enabled both long and short-range interactions to be accurately calculated and the recent development of SAPT(DFT)[24–31] has made possible highly accurate studies of intermolecular interactions at a level comparable to CCSD(T) [32, 33], with modest computational resources.

This methodology has already been used to develop intermolecular potentials. Misquitta *et al.* [34] have developed an anisotropic potential to predict the crystal structure of the 1,3-dibromo-2-chloro-5-fluorobenzene ($C_6BrClFH_2$) molecule giving results in excellent agreement with experiment. Similarly an analytic potential derived from SAPT(DFT) interaction energies has been used to study the potential energy surface of cyclotrimethylene trinitramine (RDX) dimer [35]. A model benzene potential has been constructed using SAPT(DFT) energy calculations of 491 dimer geometries [33]. However, in addition to the usual atomic sites this potential also contains off-atomic sites and it is difficult to see how the parameters for off-atomic sites can be transferred easily to larger PAHs.

For our purposes we desire a transferably potential which can accurately predict a variety of PAH dimer energies and this necessitates the potential to be based upon interactions between atomic sites only. The potential will have the following functional form

$$U_{ab} = \underbrace{G\exp\Big[-\alpha_{ab}\Big(R_{ab}-\rho_{ab}(\Omega_{ab})\Big)\Big]}_{short-range} \underbrace{-f_6(\beta R_{ab})\frac{C_{6,\mathrm{iso}}}{R_{ab}^6}+\frac{q_a q_b}{R_{ab}}}_{long-range}, \tag{5}$$

where the first term is the Born-Mayer term describing short-range interactions, the second is the dispersion term multiplied by the appropriate Tang-Toennies damping function [36], $f_6(\beta R_{ab})$, and the final term is the ESP point charge model. This potential form differs from exp-6 potetial in Eq. 3 in two important ways: the short-range term includes a shape-function, ρ_{ab}, that models the anisotropy of the interacting sites, and the singularity in the dispersion term is removed by the damping function.

We begin this Article with a description of the methods we have used to parameterise this potential for benzene. The resulting parameter set will act as a starting point for the generalisation of the potential to larger PAHs. This stage will be aided using SAPT(DFT) interaction energies calculated by Podeszwa and Szalewicz [18] for dimers of naphthalene ($C_{10}H_8$), anthracene ($C_{14}H_{10}$), and pyrene, ($C_{16}H_{10}$) in a variety of configurations. Finally, we conclude with possibilities for further work.

2 Constructing the intermolecular potential

The basic strategy for constructing an analytic potential for molecules consisting of more than two atoms has been described in a recent review [37]. The potential is logically separated into long and short-range parts (Eq. 5). The long-range part depends upon molecular properties such as multipole moments, polarisabilities and dispersion coefficients. Long-range polarisation, or induction, is expected to be weak in molecules which do not possess strong multipole moments, such as PAHs. We have therefore neglected an explicit induction term in our model potentials. The short-range energies include the exchange-repulsion, the penetration energies (see below) and the second-order induction effects (which make a small but significant contribution). The short-range energies all decay exponentially with increasing separation and we have fitted the parameters of the exponential terms via the density overlap model in the procedure described below [34].

2.1 Molecular geometry and basis sets

The geometry of benzene was obtained by *in vacuo* optimisation using DFT with the B3LYP functional and a 6-31G* basis set with the GAUSSIAN03 [38] program. The molecule was assumed to be rigid and calculated atom coordinates are provided in the supporting information. The D_{6h} symmetry of benzene allows us to identify just two unique atom types: a carbon and a hydrogen. Symmetry was imposed during the calculation of the distributed properties and the subsequent fitting process for the Born-Mayer parameters.

Interaction energies and molecular properties have been calculated using the CAMCASP [39] program from molecular wavefunctions obtained using the DALTON [40] program. The molecular wavefuctions were calculated with the asymptotically corrected

PBE0 exchange-correlation functional and the Sadlej-pVTZ [41] basis. We have used the Tozer-Handy asymptotic correction [42, 43] with a vertical ionization potential of 0.3397 a.u., obtained from Ref. [44]. The linear-response DFT calculations needed for second-order SAPT(DFT) energies were performed using a hybrid adiabatic LDA and coupled Hartree-Fock kernel [27, 45]. Two different types of basis set have been used for dimer calculations: a 'monomer-centred' (MC) basis which includes basis function on atomic sites only and a 'monomer-centred-plus' (MC+) basis which takes the MC basis and adds basis functions placed in the bonding region between the two molecules and on atomic sites of the partner molecule. This is necessary to ensure the dispersion energy converges, which is difficult with standard basis sets [46]. We have used the Sadlej-pVTZ basis [41] set for the monomer parts of the MC+ basis set, and a $3s2p1d$ basis set for the bond functions, which have been placed at a position determined by a generalisation of the weighting scheme described in Ref. [47]. Molecular polarisabilities and dispersion coefficients were calculated using the Sadlej-pVTZ basis.

We used two kinds of auxiliary basis sets during the course of these calculations: the aug-cc-pVTZ auxiliary basis [48] has been used for the calculations of molecular properties and SAPT(DFT) energies with the MC+ basis type, and the smaller JK-TZVPP basis [49] has been used for the calculation of the density overlap and the first order SAPT(DFT) energies used in the first part of the fitting process.

2.2 SAPT(DFT) dimer energies

SAPT(DFT) benzene dimer interaction energies were calculated for a variety of configurations to model the exchange-repulsion, penetration and induction energies, as well as providing a standard for comparison of our dispersion models. Based upon an earlier study [45, 50], we used the following formulation of the SAPT(DFT) interaction energy:

$$U^{(2)} = E^{(1)}_{\text{elst}}(\text{KS}) + E^{(1)}_{\text{exch}}(\text{KS}) + E^{(2)}_{\text{ind}} + E^{(2)}_{\text{disp}} + E^{(2)}_{\text{ind,exch}} + E^{(2)}_{\text{disp,exch}}. \tag{6}$$

Here, $E^{(1)}_{\text{elst}}(\text{KS})$ and $E^{(1)}_{\text{exch}}(\text{KS})$ are the first-order electrostatic and exchange-repulsion energies recovered using Kohn-Sham density functional theory. $E^{(2)}_{\text{disp}}$ and $E^{(2)}_{\text{disp,exch}}$ are the second-order dispersion and exchange-dispersion energies, and similarly $E^{(2)}_{\text{ind}}$ and $E^{(2)}_{\text{ind,exch}}$ are the second-order induction and exchange-induction energies. These terms are calculated using Kohn-Sham linear response theory. Terms of third and higher-order in the interaction operator have been neglected as these are not expected to be significant for systems without hydrogen bonds [33, 45].

It is important to select the dimers so as to get a uniform coverage of the space of physically important configurations. We have done this by keeping one of the molecules of the dimer fixed and centred at the origin, and by translating and rotating the other using the following algorithm [34]:

- Using a Sobol pseudo-random sequence, generate a random direction vector for the translation and, using Shoemakes uniform distribution algorithm [51], generate a quaternion for rotation.

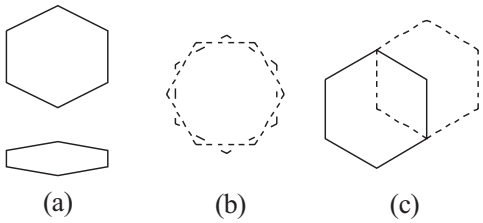


Figure 5: *Benzene: (a) T-shape, symmetry C_{2v} (b) Crossed, symmetry D_{6d} (c) Slipped-parallel, symmetry C_{2v}.*

- Starting with both molecules centred at the origin, rotate one using the quaternion. Using the standard van der Waals radii [52], determine the distance of van der Waals contact R_0 along the direction vector.
- Translate the rotated molecule along the direction vector by a few (1 to 5) randomly selected distances chosen to lie between $R_0 - \Delta R_{\text{min}}$ and $R_0 + \Delta R_{\text{max}}$. We have used ΔR_{min} and ΔR_{max} to be 1.5 and 1.2 a.u., respectively.

We have used a two-tiered approach in the development of our potential. Firstly, we have calculated first-order SAPT(DFT) interaction energies and density overlaps in the MC basis for 500 benzene dimer configurations selected using the algorithm described above. These results have been used to perform an initial fit for the short-range terms in our potential. Secondly, we have calculated first and second-order SAPT(DFT) interaction energies in the MC+ basis for the first 100 dimer configurations of the 500 initially chosen. An additional 27 energies were calculated at specific dimer orientations shown in Fig. 5. Here the slipped parallel and crossed configurations represent stacked dimer configurations and the interaction energy has been calculated at various interplanar spacings. The T-shaped configuration energies have been calculated at various separations of the monomer centres of mass. This second stage has been used to refine both our initial short-range parameterisation and our dispersion model.

2.3 Molecular properties

Whilst the electrostatic interaction could be described in detail by high-rank multipole models, in the interest of simplicity and transferability, a distributed point charge model has been used. This was calculated with the GAUSSIAN03 [38] program using the PBE0 functional and Sadlej-pVTZ basis using the Merz-Singh-Kollman scheme [53] which fits the molecular electrostatic potential to a set of atom-centred point charges.

The frequency-dependent polarisabilities are needed to calculate the dispersion coefficients from which we calculate the second-order dispersion energies at intermolecular separations where orbital overlap effects can be neglected. These polarisabilities need to be distributed in order to ensure rapid convergence with rank of the multipole expansion; in fact, even for a molecule the size of benzene, the single centre multipole expansion will not converge for the physically important dimer configurations. Distributed frequency-dependent polarisabilities of ranks 1, 2 and 3 for carbon and hydrogen atoms have been

obtained using the Williams-Stone-Misquitta (WSM) [45, 50, 54–56] method. This distribution method has been shown to result in models which exhibit very good convergence properties while resulting in a physically meaningful partitioning of the molecular properties.

2.4 Dispersion models

In general the second order dispersion energy between two atoms, $E^{(2)}_{\rm disp}$, can be modelled as an expansion given in the form

$$E^{(2)}_{\rm disp}(\text{model}) = -\frac{C_6}{R^6_{ab}} - \frac{C_7}{R^7_{ab}} - \frac{C_8}{R^8_{ab}} - \ldots, \tag{7}$$

where the C_n are the dispersion coefficients which are generally angular-dependent and odd terms vanish if atoms are spherically symmetric.

The WSM method allows calculation of a variety of dispersion models varying from the simple isotropic C_6 model to the very elaborate anisotropic C_{12} model. The CAMCASP [39] program was used to calculate isotropic and anisotropic C_6, C_{10} and C_{12} dispersion models (a C_n *model* would include all term from C_6 to C_n) for all pairs of interacting atoms in which the $D_{6\rm h}$ symmetry of benzene was taken into account. At present, we are able to calculate WSM polarisabilities to rank 3, so whilst we can calculate C_{10} and C_{12} terms, they lack contributions from the hexadecapole and higher-rank polarisabilities, however this is not thought to be a serious limitation [56].

Like any multipole expansion, the dispersion series must be damped at short range in order to avoid the divergence as $R \to 0$. This requirement is particularly important when higher ranking terms above C_6 are included. Whilst the divergence associated C_6 terms can often be ignored owing to being manifest at rather small R_{ab}, it can still remain a problem for some Monte Carlo simulations. It has previously been argued [50] that the Tang-Toennies damping functions [36] $f_n(\beta R_{ab})$ are suitable for damping the classical induction energy with an isotropic damping parameter of $\beta = 2\sqrt{2I}$, where I is the vertical ionization energy in a.u. It has further been suggested based on numerical evidence [56] that this damping parameter is also suitable for the dispersion energy. The Tang-Toennies damping functions are based upon incomplete Gamma functions having the form,

$$f_n(\beta R_{ab}) = 1 - \exp(-\beta R_{ab}) \sum_{k=0}^{n} \frac{(\beta R_{ab})^k}{k!}. \tag{8}$$

In general, the damping function should probably depend on atom type and be anisotropic, but the description would become very complicated, and the form of the anisotropy is unknown. Any anisotropy needed can probably be satisfactorily accommodated in the anisotropic short-range term in the potential. For benzene, using the value of I presented above, we obtain $\beta = 1.6485$ a.u.

Dispersion energies for 100 benzene dimer configurations were calculated using the ORIENT program [19]. The performance of the damped dispersion models compared to SAPT(DFT) total dispersion energies (taken as the sum of the second-order dispersion,

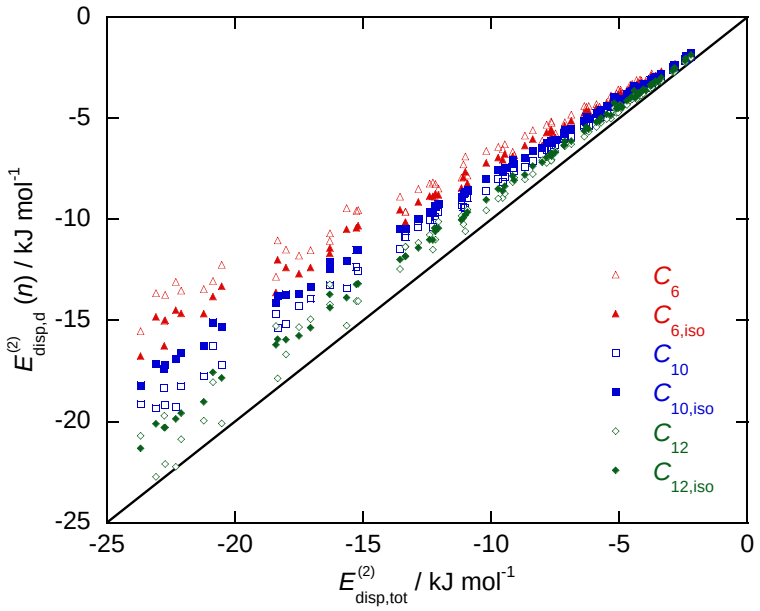


Figure 6: *Dispersion energies for the benzene dimer. Scatter plot of dispersion energies calculated using the damped dispersion models represented by* $E^{(2)}_{disp,d}(n)$ *against* $E^{(2)}_{disp,tot}$ *calculated using SAPT(DFT). The dispersion models presented are anisotropic unless given the suffix 'iso', in which case they are isotropic.*

$E^{(2)}_{\mathrm{disp}}$ and exchange-dispersion energies, $E^{(2)}_{\mathrm{ex\text{-}disp}}$) is demonstrated in Fig. 6. The general trend seen in the dispersion models is that as higher terms in the dispersion energy series are added, energies become more accurate. Adding anisotropy to the dispersion model also generally tended to produce more accurate results with the C_{12} anisotropic model most closely matching the SAPT(DFT) results, although a higher level of scattering was noticed. The scatter is probably due to the use of a single isotropic damping parameter and highlights the need to incorporate anisotropy in this term to produce the most accurate dispersion models.

2.4.1 Refining the isotropic C_6 dispersion model

In practice, elaborate anisotropic dispersion models are difficult to use and consequently in the interests of simplicity and transferability, a damped isotropic C_6 model has been chosen to represent dispersion in our potential. The scatter plot shown in Fig. 6 shows that deviation of the model dispersion energies from the SAPT(DFT) energies is approximately linear for all the dispersion models. It thus becomes possible to introduce a scaling factor by which the $C_{6,\mathrm{iso}}$ model dispersion energies can be scaled to recover more accurately the SAPT(DFT) energies. In order to find the scaling coefficients a function of the following form was minimised:

$$\Lambda = \sum_i w_i \left[E^i_{\mathrm{disp,tot}} + \xi \sum_{a\in A, b\in B} \frac{f_6(R_{ab}) C^{ab}_{6,\mathrm{iso}}}{R^6_{ab}} \right]^2, \tag{9}$$

where i labels the configurations and $E_{\text{disp,tot}}$ is the total SAPT(DFT) dispersion energy. The coefficient, ξ is determined by a least-squares fit and w_i is a weight, which will generally be energy dependent. In the general case the scaling coefficient would depend on the atom pairs, but in this work the simplest possible fit has been considered: all configurations are weighted equally and a single constant of proportionality is used. The appropriate scaling factor for the damped isotropic C_6 model was found to be 1.372 for the physically significant dispersion energies defined by the range from -20 to 0 kJ mol^{-1}. The r.m.s. error for the damped and scaled $C_{6,\text{iso}}$ model over this range is only 0.47 kJ mol^{-1}. There are large errors at configurations with total dispersion energies lower than -20 kJ mol^{-1}, where the dispersion is underestimated in magnitude, but at such short ranges repulsive interactions are expected to dominate.

3 Short-range energies

For our potential the short-range energy is defined as the sum of the exchange-repulsion, penetration and induction energies:

$$E_{\text{sr}} = E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}} + E^{(2)}_{\text{ind,tot}}. \tag{10}$$

Here $E^{(1)}_{\text{pen}}$ is the penetration energy defined as $E^{(1)}_{\text{elst}} - E^{(1)}_{\text{elst}}(\text{ESP})$; $E^{(1)}_{\text{elst}}(\text{ESP})$ being the electrostatic energy calculated using the distributed electrostatic-potential-fitted point charge model. The total induction energy is given as $E^{(2)}_{\text{ind,tot}} = E^{(2)}_{\text{ind}} + E^{(2)}_{\text{exch-ind}}$. The induction is included as a short-range energy because, in the absence of strong permanent multipoles, $E_{\text{ind,tot}}$ is almost all due to orbital overlap effects. Unlike the exchange-repulsion energy, E_{sr} is not always positive, as there will be configurations for which the penetration energies - which are generally negative - and the negative induction energies will be larger in magnitude than the corresponding exchange-repulsion energies. Nevertheless, where negative, E_{sr} is very small in magnitude, and, in any case, for most dimer configurations E_{sr} is positive, so the negative values have been neglected. Furthermore, since the penetration energy and exchange-repulsion energies both arise from the overlap of the molecular wavefunctions, they both exhibit an exponential dependence on intermolecular separation. Assuming the same distance dependence for both terms, and a similar exponential dependence for short-range induction, we have fitted the positive values of E_{sr} to an atom-atom form of the Born-Mayer expansion:

$$G \exp[-\alpha_{ab}(R_{ab} - \rho_{ab}(\Omega_{ab}))]. \tag{11}$$

Here $\rho_{ab}(\Omega_{ab})$ describes the shape of the interaction between atoms a and b as a function of their relative orientation Ω_{ab} and thus describes anisotropy in the overall interaction. The hardness of the interaction is described by α_{ab} and G is a constant energy unit taken to be 10^{-3} a.u. The shape functions have been assumed to be the sum of the shape functions of the individual sites (see Stone [57]):

$$\rho_{ab}(\Omega_{ab}) = \rho^a(\theta_a, \phi_a) + \rho^b(\theta_b, \phi_b), \tag{12}$$

where

$$\rho^a(\theta_a, \phi_a) = \sum_{l\kappa} \rho^a_{l\kappa} C_{l,\kappa}(\theta_a, \phi_a), \tag{13}$$

with a similar expression for $\rho^b(\theta_b, \phi_b)$. Here the angles θ_a and ϕ_a describe the site-site vector from a to b in the local axis system of site a. We have used an approximate axial symmetry at each atomic site with the z-axis in the local axis system pointing radially outward from the carbon to the attached hydrogen. Consequently, the shape functions $\rho_{ab}(\Omega_{ab})$ can be expanded in Legendre polynomials giving renormalised spherical harmonics $C_{l,\kappa}(\theta_a, \phi_b)$ (in the Racah definition) multiplied by a constant $\rho^a_{l\kappa}$ to model anisotropy. In this work the expansion was terminated at rank 2, but for the hydrogen atoms, where anisotropy is less important, only terms up to rank 1 were found to be necessary.

Fitting the short-ranges energies was split into two stages because a direct fit to all the short-range terms of the type given by Eq. 11 would result in an unphysical parameter set due to the highly coupled nature of the parameters. In the first stage, using the density overlap model [58], we fitted the first-order energies only, *i.e.* $E_{\text{sr}} = E^{(1)}_{\text{exch}} + E^{(1)}_{\text{pen}}$. This was done using SAPT(DFT) energies and density overlaps calculated at 500 benzene configurations using the CAMCASP [39] program. These calculations were performed using the MC Sadlej-pVTZ basis and the JK-TZVPP auxiliary basis. In the second stage, we refined the parameters obtained from the first step against SAPT(DFT) energies calculated at the first 100 configurations used in step one and a further 27 specific dimers at orientations shown in Fig. 5. This time we used the much larger MC+ basis type, and the short-range energy was defined to include the second-order induction energy as in Eq. 10.

The density overlap model postulates that the short-range energy is nearly proportional to the overlap between the molecular electron densities. The short-range energy is generally taken to be the exchange-repulsion energy, but here we additionally include the penetration and, in step two of the fitting process, the induction energies. We fit the sum, E_{sr}, to the total density overlap:

$$E_{\text{sr}} \approx E_{\text{fit}} = K_0 S_\rho^\gamma, \tag{14}$$

where, if ρ_{e}^X is the electron density of molecule X, the density-overlap is defined as $S_\rho = \int \rho_{\text{e}}^A(\mathbf{r})\rho_{\text{e}}^B(\mathbf{r})\mathrm{d}^3\mathbf{r}$. For the asymptotically correct densities we have used here, the exponent γ has been shown [59] to be exactly 1, so K_0 is the only free parameter in this model and can be determined trivially by minimising

$$\chi^2 = \sum_i w_i \left[\frac{E^i_{\text{fit}}}{E^i_{\text{sr}}} - 1 \right]^2, \tag{15}$$

where w_i is the weight and E_{fit} is the energy calculated from the fit. We have used the Gaussian/Log weighting scheme [60] which is suitable when fitting a data set spanning many orders of magnitude. In this scheme the weight, w_i, is given by

$$w_i = \exp\left[-\alpha \left[\ln\left(\frac{E^i_{\text{sr}}}{E_0} \right) \right]^2 \right], \tag{16}$$

where $\alpha = 1/\ln 10$ and $E_0 = 25$ kJ mol^{-1} in the first stage and 8 kJ mol^{-1} in the second.

The fit to the total density overlap function is usually poor and a better fit is obtained if we partition the electron density into atomic contributions, so that $\rho_{\text{e}}^A(\mathbf{r}) = \sum_{a\in A} \rho_{\text{e}}^a(\mathbf{r})$,

Table 1: *Parameters of benzene anisotropic atom-atom potential in a.u.*

Atom Pair	$l_a\kappa_a$	$l_b\kappa_b$	ρ	α	C_6
C C	00	00	4.1780	1.8683	30.452
	10	00	0.2535		
	20	00	−2.0390		
C H	00	00	5.4242	1.7370	12.490
	00	10	−0.4663		
	10	00	0.1472		
	20	00	−0.1422		
H H	00	00	3.4400	1.5263	5.092
	10	00	0.3611		

and we fit E_{sr} to the distributed overlap model [60, 61],

$$E_{\text{sr}} \approx E_{\text{fit}} = \sum_{a\in A, b\in B} K^{ab} S_\rho^{ab}, \tag{17}$$

where a and b are the atomic sites of A and B, respectively, and $S_\rho^{ab} = \int \rho_{\text{e}}^a(\mathbf{r})\rho_{\text{e}}^b(\mathbf{r})\mathrm{d}^3\mathbf{r}$ is the site-site density overlap. The density-partitioning is not unique and can be achieved in a variety of ways. We have used density-fitting to achieve the partitioning, which is analogous to the Gaussian multipole method of Wheatley [62]. Unfortunately, because the electron density of some atoms are not sampled sufficiently, a straightforward fit to the distributed overlap model results in unphysical values of K^{ab} for some pairs of sites. This problem can be avoided by using the constant K_0 obtained from the fit to the total density-overlap model as a constraint on the parameters K^{ab} [34]. So we minimise the function

$$\Delta = \frac{\chi^2}{\sum_i w_i} + \lambda \sum_{ab} (K^{ab} - K_0)^2, \tag{18}$$

where λ is a parameter which could be dependent on the sites a and b, but has been chosen to be a constant, and the χ^2 has been normalised by the sum of the weights to avoid a dependence on the number of points in the grid. We have used $\lambda = 10^{-8}$ to give reasonably good fits without the occurrence of unphysical values for K^{ab}.

Having obtained the distributed overlap model (Eq. 17), it is now relatively easy to fit the contributions of individual pairs of sites to a single Born-Mayer term of the form Eq. 11, this time with physically sensible values for the parameters. In this way each fit involves only a small number of adjustable parameters. The D_{6h} symmetry of benzene was taken into account at this stage of the fitting process. These stages of fitting were performed using the CAMCASP [39] and the ORIENT [19] programs and the overall weighted r.m.s. error for the fitted energies was 0.82 kJ mol^{-1}.

In the second stage, the fitting was performed using penalty functions of the form $(p_i - p_i^0)^2$ where p_i^0 are the anchor values obtained from the first stage of the fitting process. In this way the fit was refined whilst preventing the parameters from taking non-physical values. The final choice of parameters to be relaxed and the weights given to the harmonic

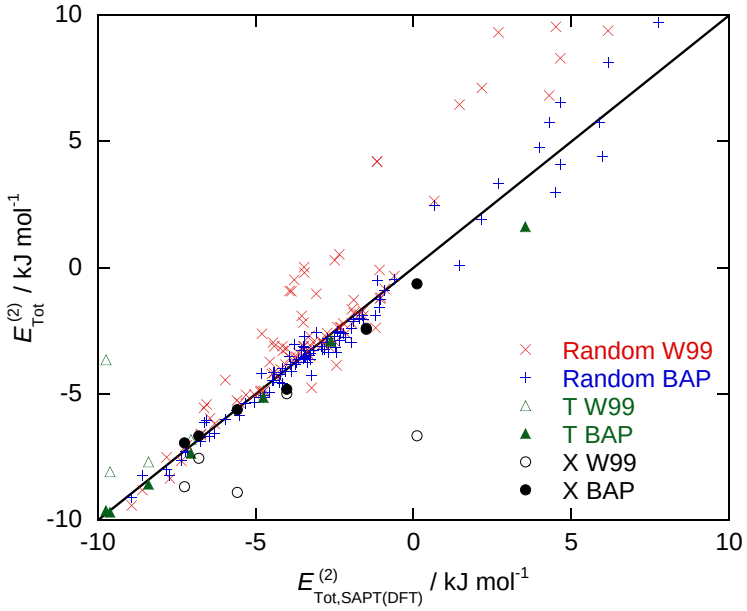


Figure 7: *Comparison of the benzene anisotropic potential with SAPT(DFT) energies calculated for 100 random benzene dimer orientations.*

constraints were chosen with an element of empiricism, although values of α_{ab} and $\rho^{ab}_{00,00}$ were constrained more tightly than other parameters. The final fit had a weighted r.m.s. residual error of 0.75 kJ mol^{-1}.

3.1 Overall benzene potential

The final fitting procedure described above was repeated to relax the Born-Mayer parameters and the damped $C_{6,\mathrm{iso}}$ dispersion coefficients within constraints to fit SAPT(DFT) energies calculated in the MC+ basis ($E^{(1)}_{\mathrm{exch}}$, $E^{(1)}_{\mathrm{pen}}$, $E^{(2)}_{\mathrm{ind,tot}}$ and $E^{(2)}_{\mathrm{disp,tot}}$) for the 100 random benzene geometries and 27 geometries relating to specific configurations. The SAPT(DFT) energies were weighted so as to favour more negative energies to ensure the potential well was accurately fitted. The weighting scheme used has been adapted from that used in Ref. [63] and is given as

$$w_i = \Theta(E^i_{\mathrm{tot}} - E_0)\left(\frac{E_0}{E^i_{\mathrm{tot}}}\right)^2 + \left[1 - \Theta(E^i_{\mathrm{tot}} - E_0)\right]\exp\left[\gamma(E_0 - E^i_{\mathrm{tot}})\right], \qquad (19)$$

where $\Theta(x)$ is the Heaviside step function. E^i_{tot} are the SAPT(DFT) energies and the parameters E_0 and γ were set to 3 kJ mol^{-1} and 0.1 mol kJ^{-1} respectively.

The parameters for the benzene anisotropic potential, based on the ESP point charge model, are given in Table 1. The point charges in a.u. used for carbon and hydrogen atoms are -0.1111 and 0.1111 respectively. It should be noted that for this potential the definition of the shape function given in Eq. 12 was relaxed and instead shape functions for individual atoms were allowed to vary depending on the specific atom pair considered. This allowed a better fit for the benzene dimers but for the development of a general

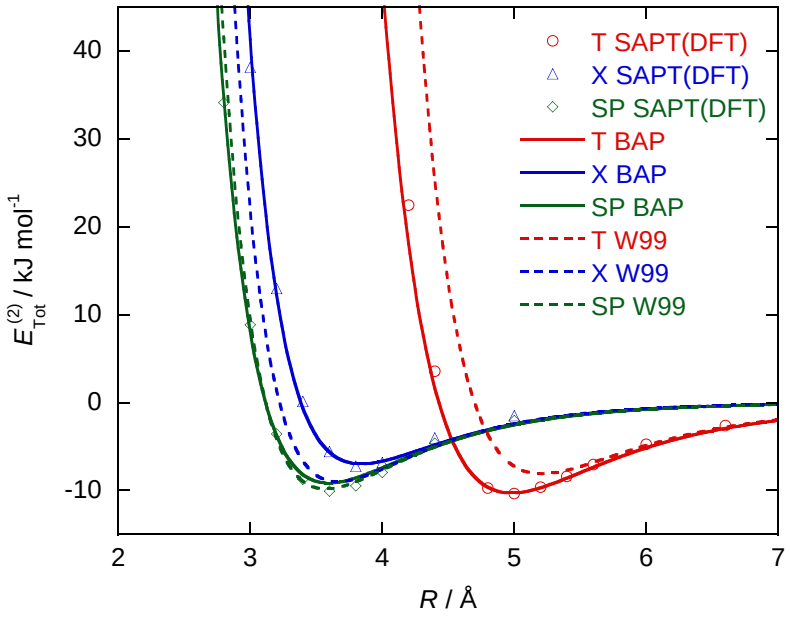


Figure 8: *Comparison of benzene anisotropic potential with SAPT(DFT)energies and W99 potential for T-shaped (T), crossed (X) and slipped parallel (SP) dimer configurations.*

anisotropic potential for PAHs (discussed below) we have retained this definition for the sake of transferability.

The weighted r.m.s. residual energy ($E^i_{\rm tot} - E^i_{\rm fit}$) for the benzene anisotropic potential when compared with the 127 SAPT(DFT) energies was found to be 0.49 kJ mol^{-1}. Figure 7 shows the scatter of energies of the new potential compared to SAPT(DFT) energies for the random benzene dimer configurations and some of the specific configurations chosen. For comparison the W99 potential has been included and the plot shows energies calculated with the new potential are noticeably less scattered. The scatter which remains for the new potential results is likely to be due to the damped C_6 isotropic dispersion model and for which there cannot be further improvement without going to a more detailed dispersion model. Figure 8 shows a comparison of the new benzene potential with SAPT(DFT) energies and the W99 potential for the orientations shown in Fig. 5. The new potential matches the SAPT(DFT) results in all configurations, especially the T-shaped configuration where the W99 potential is notably poor.

4 Generalising to larger PAH molecules

The shape function shown in Eq. 12 imposes certain constraints on the parameters of the potential. For example, $\rho^{\rm HH}_{10,00} = \rho^{\rm CH}_{00,10}$. These conditions were not imposed during the construction of the benzene anisotropic potential, but they can be, and if we do impose them, the reduced flexibility of the functional form results in a poorer fit with a weighted r.m.s. residual of 0.96 kJ mol^{-1}. These shape constraints are probably inconsequential if the potential is restricted to a single system, but, since they impose the idea of transferabil-

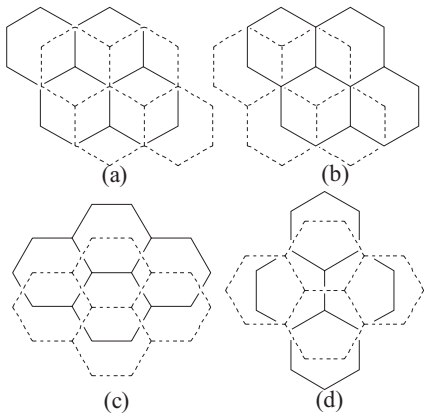


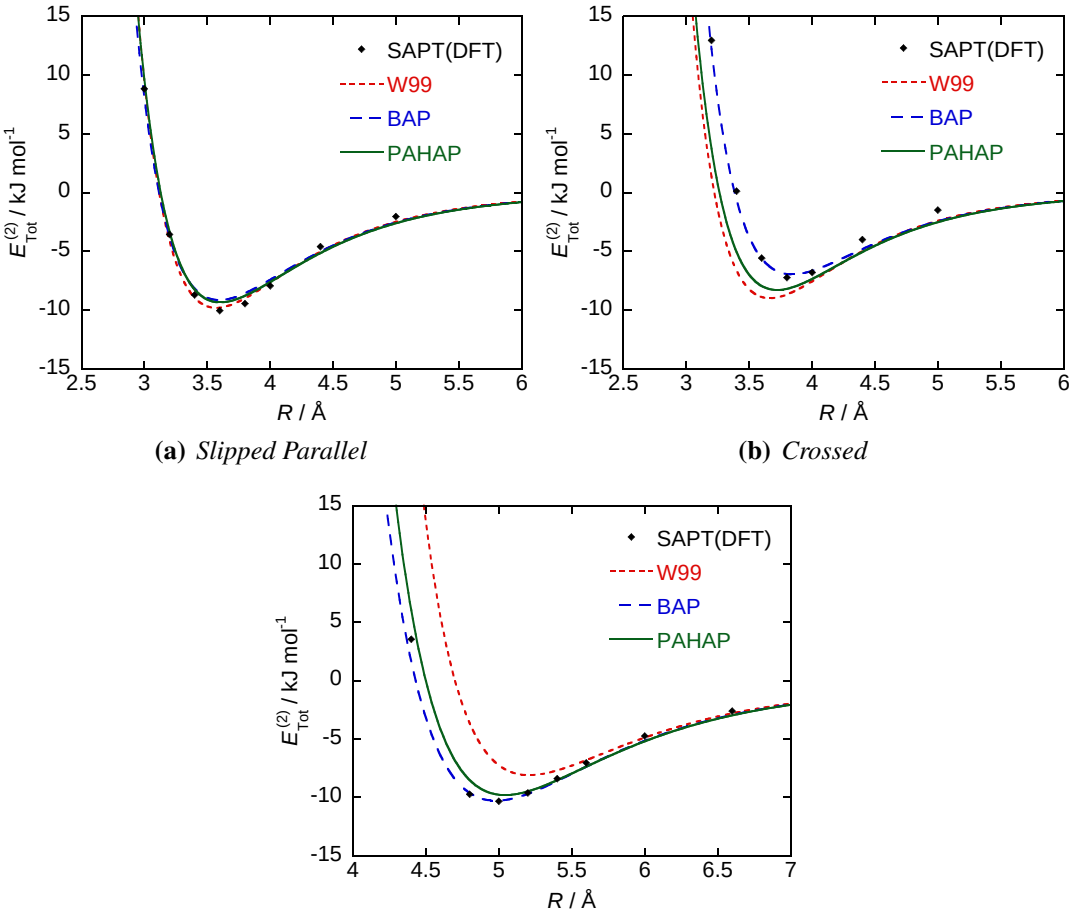
Figure 9: *Pyrene: (a) Slipped-parallel L, symmetry C_{2h} (b) Graphite-type, symmetry C_i (c) Slipped-parallel S, symmetry C_{2h} (d) Crossed, symmetry D_{2d}.*

ity [57] are needed if we wish to use the potential parameters on other, related, systems as well. So when generalising the benzene potential to larger PAH molecules we have imposed these constraints. This results in a potential with fewer parameters, but as shall be shown, appears to perform remarkably well for the larger PAH dimers.

The benzene anisotropic potential (BAP) parameters serve as the starting parameterisation for the transferable anisotropic PAH potential. As can be seen from Figs. 10-13 the parameters from this potential describe the interactions between the larger PAH molecules reasonably well, demonstrating that they are a good starting point for a more accurate, generalised potential. We have tuned the potential parameters against 111 SAPT(DFT) dimer energies calculated by Podeszwa and Szalewicz [18] for the naphthalene, anthracene and pyrene dimers at the orientations shown in Figs. 1– 9). Molecular geometries were taken from [18] and ESP point charge models were calculated for each molecule with the GAUSSIAN03 [38] program in the same way as described for benzene. The geometries and partial atomic charges are given in the supporting information.

The ORIENT program cannot simultaneously fit parameters to multiple types of molecular dimers, so an iterative scheme has been adopted. In this scheme the initial benzene parameterisation is used as the starting point for fitting the parameters for the naphthalene dimer; having obtained a the new set of parameters, these now become the starting point for fitting to the anthracene dimer energies; this process is then continued, cycling through each set of dimer energies for each of the four PAH molecules. In order to converge to a parameter set, the harmonic constraints used in the fitting procedure were tightened after each iteration. Eventually the parameters are so tightly constrained that they could not be varied; this gave us the final parameter set. Whilst this procedure is by no means optimum, it has proved adequate and has resulted in a generalized parameter set that is not only able to model the interactions of the larger PAH molecules, but also the 127 benzene dimer geometries.

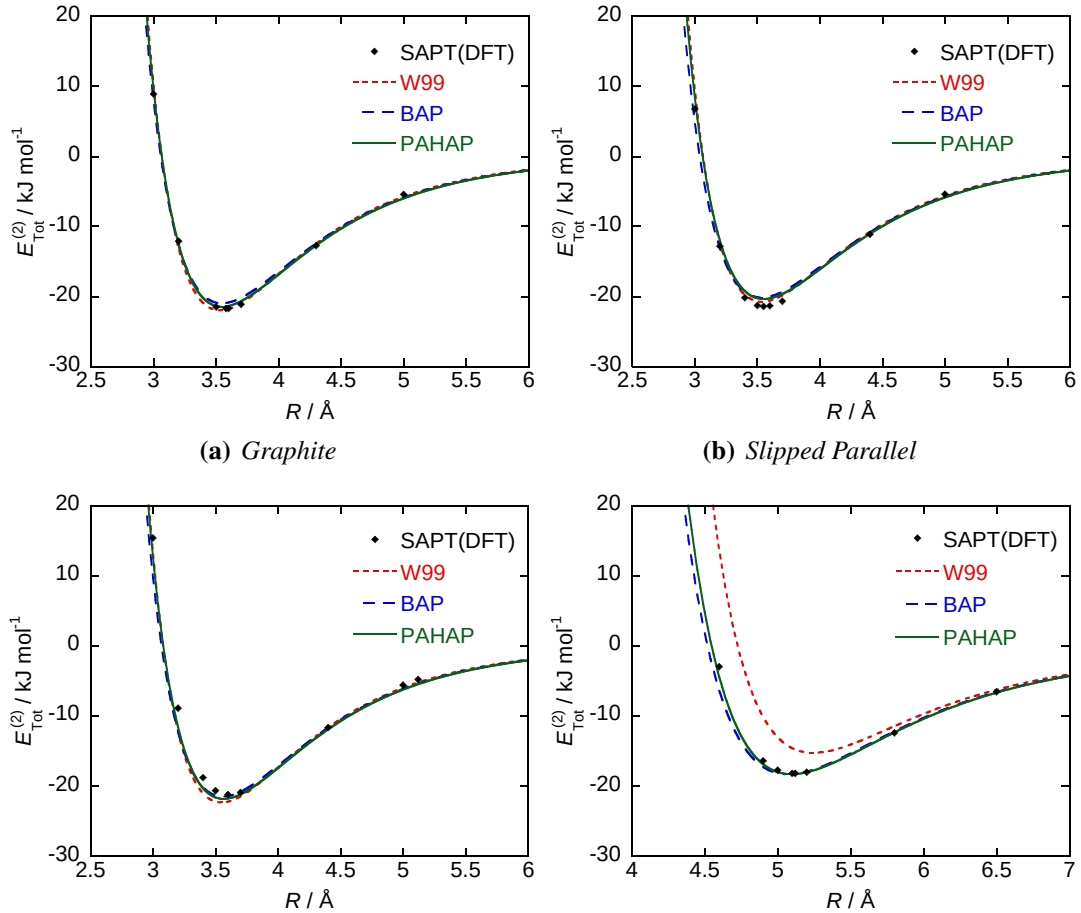
Table 2 shows the set of parameters obtained for the general PAH anisotropic potential (PAHAP) which, unlike the initial benzene parameterisation, satisfy the constraints imposed by the shape function given in Eq. 12. From Figs. 10-13 we see that the PAHAP parameterisation is also a slight improvement over the benzene parameterisation for the



(a) *Slipped Parallel* **(b)** *Crossed*

(c) *T-shape*

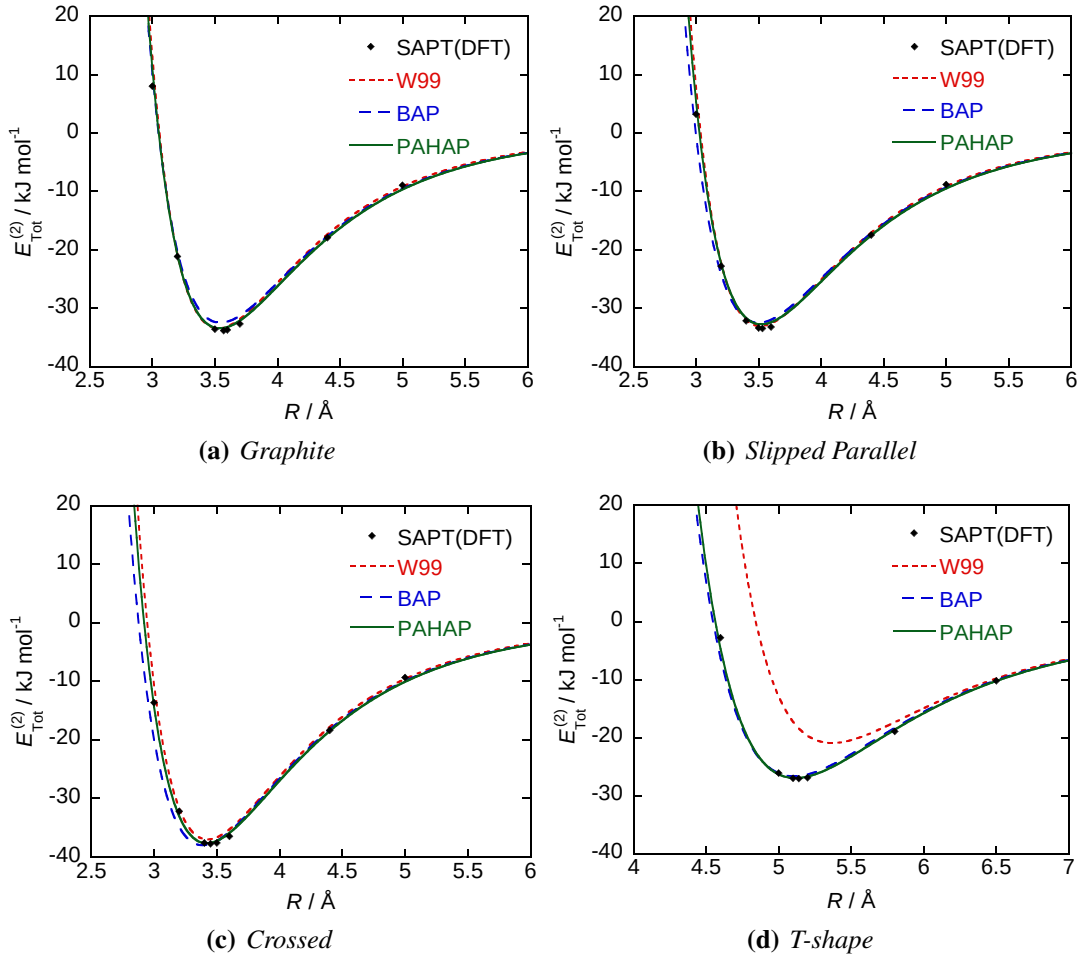
Figure 10: *Comparison of the W99 potential, the benzene anisotropic potential (BAP) and the PAH anisotropic potential (PAHAP) with SAPT(DFT) energies for benzene dimers. Model potential energies have been calculated using the* ORIENT *[19] program.*



(a) *Graphite* **(b)** *Slipped Parallel*

(c) *Crossed* **(d)** *T-shape*

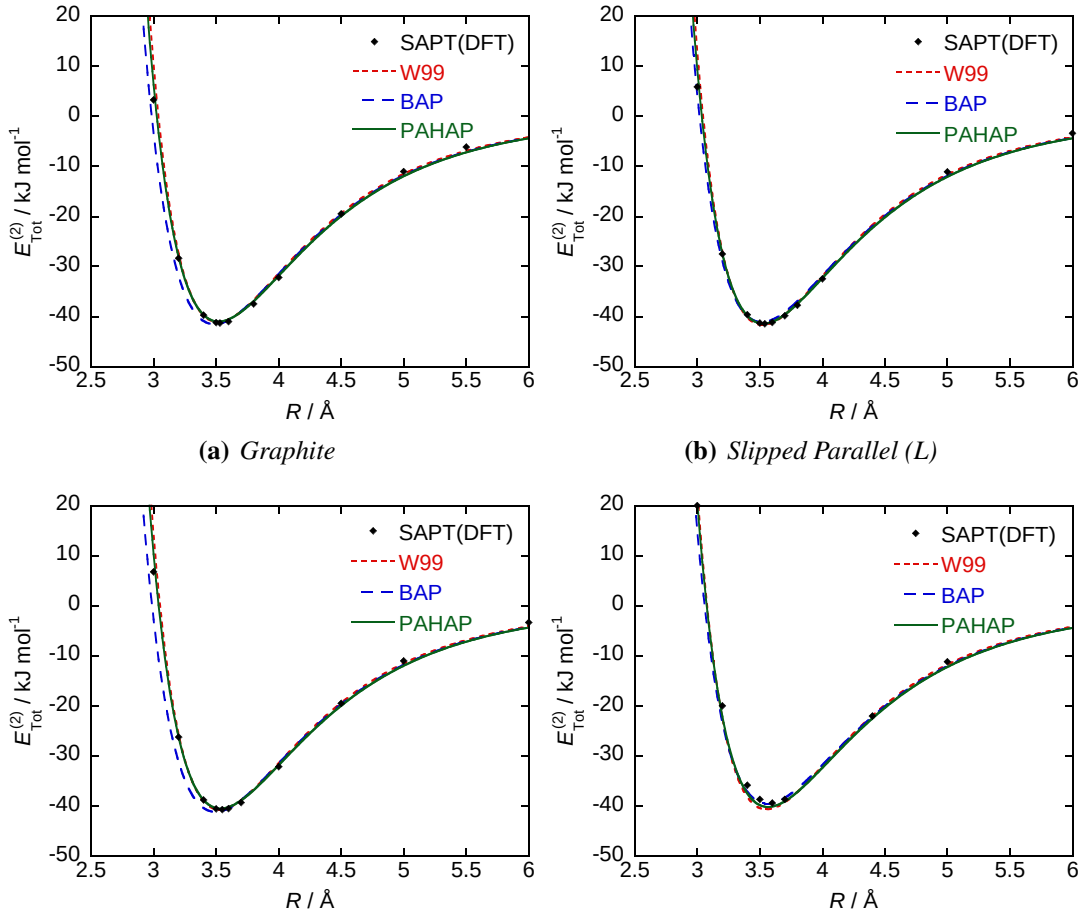
Figure 11: *Comparison of the W99 potential, the benzene anisotropic potential (BAP) and the PAH anisotropic potential (PAHAP) with SAPT(DFT) energies for naphthalene dimers. Model potential energies have been calculated using the* ORIENT *[19] program.*



(a) *Graphite* **(b)** *Slipped Parallel*

(c) *Crossed* **(d)** *T-shape*

Figure 12: *Comparison of the W99 potential, the benzene anisotropic potential (BAP) and the PAH anisotropic potential (PAHAP) with SAPT(DFT) energies for anthracene dimers. Model potential energies have been calculated using the* ORIENT *[19] program.*



(a) *Graphite* **(b)** *Slipped Parallel (L)*

(c) *Slipped Parallel (S)* **(d)** *Crossed*

Figure 13: *Comparison of the W99 potential, the benzene anisotropic potential (BAP) and the PAH anisotropic potential (PAHAP) with SAPT(DFT) energies for pyrene dimers. Model potential energies have been calculated using the* ORIENT *[19] program.*

Table 2: *Parameters of PAH anisotropic atom-atom potential in a.u.*

Atom Pair	$l_a\kappa_a$	$l_b\kappa_b$	ρ	α	C_6
C C	00	00	5.8147	1.8615	30.469
	10	00	0.0217		
	20	00	−0.2208		
C H	00	00	5.1505	1.7756	12.840
	00	10	−0.2718		
	10	00	0.0217		
	20	00	−0.2208		
H H	00	00	4.4862	1.4312	5.359
	10	00	−0.2718		

larger PAHs, without a significant loss of accuracy for the benzene dimer energies. In particular, in contrast to the W99 potential, the PAHAP potential correctly models the PAH interaction energies at both the stacked as well as the T-shaped configurations. The overall weighted r.m.s. residual error over the 238 dimer configurations considered was found to be 0.73 kJ mol^{-1}, which is more than three times less than the error of 2.54 kJ mol^{-1} incurred by the W99 potential.

5 Discussion

Using *ab initio* calculations we have developed a transferable anisotropic potential for polycyclic aromatic hydrocarbons that surpasses some of the best empirically derived isotropic potentials in accuracy. In particular, this potential accurately predicts intermolecular interactions for both stacked and non-stacked dimer configurations, such as the T-shape dimer. This is important as whilst stacked configurations are generally energetically more favourable for most PAH dimers, when modelling clusters of PAHs as in the context of nascent soot particles, non-stacked configurations are also present [13, 64].

To assess the overall accuracy of our potential we have to consider both the accuracy of the fit and the accuracy of the SAPT(DFT) energies. The accuracy of the former is shown by the weighted r.m.s. residual error which was calculated to be 0.73 kJ mol^{-1} over the 238 dimer configurations considered. To determine the accuracy of the latter comparison must be made to other *ab initio* results. Highly accurate benzene dimer energies have been obtained at CCSD(T) and QCISD(T) level by Janowski and Pulay [65]. In this work the largest calculation at QCISD(T)/aug-cc-pVQZ level involved 30 correlated orbitals and 1512 basis functions. Of the three benzene dimers configurations considered, only energy calculations for the T-shape dimer are directly comparable to our SAPT(DFT) calculations. At a separation of 4.989 Å the QCISD(T) binding energy extrapolated to infinite basis is 11.23 kJ mol^{-1}, whereas the corresponding SAPT(DFT) binding energy calculated at a separation of 5.0 Å in our MC+ basis is 10.33 kJ mol^{-1}. This error in our SAPT(DFT) calculations is probably due to the difference in basis sets. The MC+ basis comprises the Sadlej-pVTZ basis for the monomer centred functions and

extra $3s2p1d$ midbond functions, but this combined basis is considerably smaller than that used in QCISD(T) calculations and is the likely cause for the underestimation of the binding energy. Using larger basis sets would reduce the error but would increase computational demands prohibitively. Thus at our chosen level of theory the error is unavoidable and given the transferable nature of our PAH potential we believe this error to be quite acceptable.

The new transferable PAH anisotropic potential represents a first step in our planned investigation of the intermolecular chemistry involved in the clustering of PAHs which is thought to be an important step in the formation of nascent soot particles. Whilst this potential can be used in its own right, it is hoped that it will also provide an accurate reference against which we can produce a general coarse-grained PAH potential, necessary for the study of large molecular clusters. The potential may also find applications in other fields were the effects of anisotropy could be important, such as organic crystal structure prediction [66].

Acknowledgements

T.S.T. gratefully acknowledges financial support from the EPSRC, Shell Research Ltd. and Churchill College, Cambridge. The authors thank R. Podeszwa and K. Szalewicz for data provided relating to work reported in Ref. [18].

References

[1] J. D. Herdman and J. H. Miller. Intermolecular potential calculations for polynuclear aromatic hydrocarbon clusters. *Journal of Physical Chemistry A*, 112(28):62496256, 2008.

[2] H. X. Chen and R. A. Dobbins. Crystallogenesis of particles formed in hydrocarbon combustion. *Combustion Science and Technology*, 159(1):109–128, 2000.

[3] T. Ishiguro, Y. Takatori, and K. Akihama. Microstructure of Diesel Soot Particles Probed by Electron Microscopy: First Observation of Inner Core and Outer Shell. *Combustion and Flame*, 108(1-2):231–234, 1997.

[4] R. L. Vander Wal, A. Yezerets, N. W. Currier, D. H. Kim, and C. M. Wang. HRTEM Study of diesel soot collected from diesel particulate filters. *Carbon*, 45:70–77, 2007.

[5] J. T. McKinnon and J. B. Howard. The roles of PAH and acetylene in soot nucleation and growth. *Symposium (International) on Combustion*, 24(1):965–971, 1992.

[6] J. H. Miller. The kinetics of polynuclear aromatic hydrocarbon agglomeration in flames. *Symposium (International) on Combustion*, 23(1):91–98, 1990.

[7] J. H. Miller. Aromatic excimers: evidence for polynuclear aromatic hydrocarbon condensation in flames. *Proceedings of the Combustion Institute*, 30(1):1381–1388, 2005.

[8] J. H. Miller, K. C. Smyth, and W. G. Mallard. Calculations of the dimerization of aromatic hydrocarbons: Implications for soot formation. *Symposium (International) on Combustion*, 20(1):1139–1147, 1985.

[9] C. A. Schuetz and M. Frenklach. Nucleation of soot: Molecular dynamics simulations of pyrene dimerization. *Proceedings of the Combustion Institute*, 29(2):2307–2314, 2002.

[10] J. Appel, H. Bockhorn, and M. Wulkow. A detailed numerical study of the evolution of soot particle size distributions in laminar premixed flames. *Chemosphere*, 42(5-7):635–645, 2001.

[11] J. Happold, H. Grotheer, and M. Aigner. Distinction of gaseous soot precursor molecules and soot precursor particles through photoionization mass spectrometry. *Rapid Communications in Mass Spectrometry*, 21(7):1247–1254, 2007.

[12] J. H. Miller, W. G. Mallard, and K. C. Smyth. Intermolecular potential calculations for polycyclic aromatic hydrocarbons. *Journal of Physical Chemistry*, 88(21):4963–4910, 1984.

[13] M. Rapacioli, F. Calvo, F. Spiegelman, C. Joblin, and D. J. Wales. Stacked clusters of polycyclic aromatic hydrocarbon molecules. *Journal of Physical Chemistry A*, 109(11):2487–2497, 2005.

[14] B. W. van de Waal. Calculated ground-state structures of 13-molecule clusters of carbon dioxide, methane, benzene, cyclohexane, and napthalene. *Journal of Chemical Physics*, 79(8):3948–3961, 1983.

[15] D. E. Williams. Improved intermolecular force field for crystalline hydrocarbons containing four- or three-coordinated carbon. *Journal of Molecular Structure*, 485-486:321–347, 1999.

[16] D. E. Williams. Improved intermolecular force field for crystalline oxohydrocarbons including O–H...O hydrogen bonding. *Journal of Computational Chemistry*, 22(1):1–20, 2001.

[17] D. E. Williams. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry*, 22(11):11541166, 2001.

[18] R. Podeszwa and K. Szalewicz. Physical origins of interactions in dimers of polycyclic aromatic hydrocarbons. *Physical Chemistry Chemical Physics*, 10(19):2735–2746, 2008.

[19] A. J. Stone, A. Dullweber, O. Engkvist, E. Fraschini, M. P. Hodges, A. W. Meredith, D. R. Nutt, P. L. A. Popelier, and D. J. Wales. ORIENT: a program for studying interactions between molecules, version 4.6, University of Cambridge, 2002. http://www-stone.ch.cam.ac.uk/programs.html.

[20] P. Weilmünster, A. Keller, and K. H. Homann. Large Molecules, Radicals, Ions, and Small Soot Particles in Fuel-Rich Hydrocarbon Flames Part I: Positive Ions of Polycyclic Aromatic Hydrocarbons (PAH) in Low-Pressure Premixed Flames of Acetylene and Oxygen. *Combustion and Flame*, 116:62–83, 1999.

[21] S. Tsuzuki, K. Honda, T. Uchimaru, and M. Mikami. High-level *ab initio* computations of structures and interaction energies of naphthalene dimers: Origin of attraction and its directionality. *The Journal of Chemical Physics*, 120(2):647–659, 2004.

[22] S. Tsuzuki, T. Uchimaru, K. Matsumura, M. Mikami, and K. Tanabe. Effects of the higher electron correlation correction on the calculated intermolecular interaction energies of benzene and naphthalene dimers: comparison between MP2 and CCSD(T) calculations. *Chemical Physics Letters*, 319(5-6):547–554, 2000.

[23] M. Gerenkamp and S. Grimme. Spin-component scaled second-order Møller-Plesset perturbation theory for the calculation of molecular geometries and harmonic vibrational frequencies. *Chemical Physics Letters*, 392(1-3):229–235, 2004.

[24] A. J. Misquitta and K. Szalewicz. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chemical Physics Letters*, 357(3-4):301–306, 2002.

[25] A. J. Misquitta, B. Jeziorski, and K. Szalewicz. Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.*, 91(3):033201, 2003.

[26] A. J. Misquitta and K. Szalewicz. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *The Journal of Chemical Physics*, 122(21):214109, 2005.

[27] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *The Journal of Chemical Physics*, 123(21):214103, 2005.

[28] A. Heßelmann and G. Jansen. First-order intermolecular interaction energies from Kohn-Sham orbitals. *Chemical Physics Letters*, 357(5-6):464–470, 2002.

[29] A. Heßelmann and G. Jansen. Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn-Sham density functional theory. *Chemical Physics Letters*, 362(3-4):319–325, 2002.

[30] A. Heßelmann and G. Jansen. Intermolecular dispersion energies from time-dependent density functional theory. *Chemical Physics Letters*, 367(5-6):778–784, 2003.

[31] A. Heßelmann, G. Jansen, and M. Schütz. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *The Journal of Chemical Physics*, 122(1):014103, 2005.

[32] R. A. DiStasio, Jr., G. von Helden, R. P. Steele, and M. Head-Gordon. On the T-shaped structures of the benzene dimer. *Chemical Physics Letters*, 437(4-6):277–283, 2007.

[33] R. Podeszwa, R. Bukowski, and K. Szalewicz. Potential energy surface for the benzene dimer and perturbational analysis of π–π interactions. *The Journal of Physical Chemistry A*, 110(34):10345–10354, 2006.

[34] A. J. Misquitta, G. W. A. Welch, A. J. Stone, and S. L. Price. A first principles prediction of the crystal structure of $C_6BrClFH_2$. *Chemical Physics Letters*, 456(1-3):105–109, 2008.

[35] R. Podeszwa, R. Bukowski, B. M. Rice, and K. Szalewicz. Potential energy surface for cyclotrimethylene trinitramine. *Physical Chemistry Chemical Physics*, 9:5561–5569, 2007.

[36] K. T. Tang and J. P. Toennies. An improved simple model for the van der waals potential based on universal damping functions for the dispersion coefficients. *The Journal of Chemical Physics*, 80(8):3726–3741, 1984.

[37] A. J. Stone and A. J. Misquitta. Atom-atom potentials from *ab initio* calculations. *International Reviews in Physical Chemistry*, 26(1):193222, 2007.

[38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02, 2003. Gaussian, Inc., Wallingford, CT, 2004.

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

[40] DALTON, a molecular electronic structure program, release 2.0, 2005. see http://www.kjemi.uio.no/software/dalton/dalton.html.

[41] A. J. Sadlej. *Collection of Czechoslovak Chemical Communications*, 53:1995, 1988.

[42] D. J. Tozer and N. C. Handy. Improving virtual Kohn–Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *The Journal of Chemical Physics*, 109(23):10180–10189, 1998.

[43] D. J. Tozer. The asymptotic exchange potential in Kohn–Sham theory. *The Journal of Chemical Physics*, 112(8):3507–3515, 2000.

[44] S.G. Lias and J.F. Liebman. Ion energetics data in NIST Chemistry WebBook, NIST standard reference database number 69.

[45] A. J. Misquitta and A. J. Stone. Accurate induction energies for small organic molecules: 1. Theory. *Journal of Chemical Theory and Computation*, 4(1):7–18, 2008.

[46] H. L. Williams, E. M. Mas, K. Szalewicz, and B. Jeziorski. On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. *The Journal of Chemical Physics*, 103(17):7374–7391, 1995.

[47] O. Akin-Ojo, R. Bukowski, and K. Szalewicz. Ab initio studies of He–HCCCN interaction. *The Journal of Chemical Physics*, 119(16):8379–8396, 2003.

[48] F. Weigend, A. Köhn, and C. Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics*, 116(8):3175–3183, 2002.

[49] F. Weigend, A. Köhn, and C. Hättig. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics*, 116(8):3175–3183, 2002.

[50] A. J. Misquitta, A. J. Stone, and S. L. Price. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *Journal of Chemical Theory and Computation*, 4(1):19–32, 2008.

[51] K. Shoemake. Uniform random rotations. pages 124–132, 1992.

[52] A. Bondi. van der Waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964.

[53] U. C. Singh and P. A. Kollman. An approach to computing electrostatic charges for molecules. *The Journal of Chemical Physics*, 5(2):129–145, 1984.

[54] J. Williams, G and A. J. Stone. Distributed dispersion: A new approach. *The Journal of Chemical Physics*, 119(9):4620–4628, 2003.

[55] A. J. Misquitta and A. J. Stone. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *The Journal of Chemical Physics*, 124(2):024111, 2006.

[56] A. J. Misquitta and A. J. Stone. Dispersion energies for small organic molecules: first row atoms. *Molecular Physics*, 106(1213):16311643, 2008.

[57] A. J. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, Oxford, 1996.

[58] Y. S. Kim, S. K. Kim, and Lee W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chemical Physics Letters*, 80:574, 1981.

[59] A. J. Misquitta and A. J. Stone. *Ab initio* atom–atom potentials using CAMCASP: pyridine as an example. in preparation.

[60] M. P. Hodges and R. J. Wheatley. Application of the overlap model to calculating correlated exchange energies. *Chemical Physics Letters*, 326(3-4):263–268, 2000.

[61] I. Nobeli and S. L. Price. A non-empirical intermolecular potential for oxalic acid crystal structures. *The Journal of Physical Chemistry A*, 103(32):6448–6457, 2000.

[62] R. J. Wheatley. Gaussian multipole functions for describing molecular charge distributions. *Molecular Physics*, 79(3):597–610, 1993.

[63] R. Bukowski, J. Sadlej, B. Jeziorski, P. Jankowski, K. Szalewicz, S. A. Kucharski, H. L. Williams, and B. M. Rice. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *The Journal of Chemical Physics*, 110(8):3785–3803, 1999.

[64] T. S. Totton, A. J. Misquitta, D. Chakrabarti, D. J. Wales, and M. Kraft. Modelling the internal structure of nascent soot particles, 2009. Submitted for publication.

[65] T. Janowski and P. Pulay. High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chemical Physics Letters*, 447(1-3):27–32, 2007.

[66] S. L. Price and L. S. Price. Modelling intermolecular forces for organic crystal structure prediction. In D.M.P. Mingos and D. J. Wales, editors, *Intermolecular Forces and Clusters I: Structure and Bonding*, volume 115. Springer-Verlag, Berlin Heidelberg, 2nd edition, 1999.

Supporting Information

Table 1: *Benzene monomer coordinates and point charges (a.u.)*

Atom label	x	y	z	Charge
C1	-2.63923	0.00000	0.00000	-0.11114
C2	2.63923	0.00000	0.00000	-0.11114
C3	-1.31961	-2.28564	0.00000	-0.11114
C4	1.31961	-2.28564	0.00000	-0.11114
C5	-1.31961	2.28564	0.00000	-0.11114
C6	1.31961	2.28564	0.00000	-0.11114
H1	-4.69339	0.00000	0.00000	0.11114
H2	4.69339	0.00000	0.00000	0.11114
H3	2.34669	4.06460	0.00000	0.11114
H4	-2.34669	4.06460	0.00000	0.11114
H5	2.34669	-4.06460	0.00000	0.11114
H6	-2.34669	-4.06460	0.00000	0.11114

Table 2: *Naphthalene monomer coordinates and point charges (a.u.)*

Atom label	x	y	z	Charge
C1	-1.33862	-4.59918	0.00000	-0.10048
C2	-2.65019	-2.35249	0.00000	-0.29796
C3	-1.35523	0.00000	0.00000	0.24018
C4	1.35523	0.00000	0.00000	0.24018
C5	2.65019	-2.35249	0.00000	-0.29796
C6	1.33862	-4.59918	0.00000	-0.10048
C7	-2.65019	2.35249	0.00000	-0.29796
C8	2.65019	2.35249	0.00000	-0.29796
C9	1.33862	4.59918	0.00000	-0.10048
C10	-1.33862	4.59918	0.00000	-0.10048
H1	-4.70575	2.34799	0.00000	0.15530
H2	-2.35493	-6.38388	0.00000	0.12304
H3	-4.70575	-2.34799	0.00000	0.15530
H4	4.70575	-2.34799	0.00000	0.15530
H5	2.35493	-6.38388	0.00000	0.12304
H6	4.70575	2.34799	0.00000	0.15530
H7	2.35493	6.38388	0.00000	0.12304
H8	-2.35493	6.38388	0.00000	0.12304

Table 3: *Anthracene monomer coordinates and point charges (a.u.)*

Atom label	x	y	z	Charge
C1	1.36540	2.31298	0.00000	0.23448
C2	1.36540	2.31298	0.00000	0.23448
C3	1.36540	-2.31298	0.00000	0.23448
C4	-1.36540	-2.31298	0.00000	0.23448
C5	-2.65253	0.00000	0.00000	-0.47174
C6	2.65253	0.00000	0.00000	-0.47174
C7	2.65927	4.68538	0.00000	-0.25252
C8	-2.65927	4.68538	0.00000	-0.25252
C9	2.65927	-4.68538	0.00000	-0.25252
C10	-2.65927	-4.68538	0.00000	-0.25252
C11	1.34762	-6.91760	0.00000	-0.11389
C12	-1.34762	-6.91760	0.00000	-0.11389
C13	1.34762	6.91760	0.00000	-0.11389
C14	-1.34762	6.91760	0.00000	-0.11389
H1	4.71450	-4.67888	0.00000	0.14291
H2	2.35428	-8.70751	0.00000	0.12531
H3	-2.35428	-8.70751	0.00000	0.12531
H4	-4.71450	-4.67888	0.00000	0.14291
H5	4.71450	4.67888	0.00000	0.14291
H6	2.35428	8.70751	0.00000	0.12531
H7	-2.35428	8.70751	0.00000	0.12531
H8	-4.71450	4.67888	0.00000	0.14291
H9	-4.70918	0.00000	0.00000	0.19915
H10	4.70918	0.00000	0.00000	0.19915

Table 4: *Pyrene monomer coordinates and point charges (a.u.)*

Atom label	x	y	z	Charge
C1	-1.34794	0.00000	0.00000	-0.04275
C2	1.34794	0.00000	0.00000	-0.04275
C3	2.70059	2.33625	0.00000	0.22339
C4	2.70059	-2.33625	0.00000	0.22339
C5	-2.70059	-2.33625	0.00000	0.22339
C6	-2.70059	2.33625	0.00000	0.22339
C7	1.28651	4.65603	0.00000	-0.24782
C8	5.35355	2.28771	0.00000	-0.29542
C9	1.28651	-4.65603	0.00000	-0.24782
C10	5.35355	-2.28771	0.00000	-0.29542
C11	-1.28651	-4.65603	0.00000	-0.24782
C12	-5.35355	-2.28771	0.00000	-0.29542
C13	-1.28651	4.65603	0.00000	-0.24782
C14	-5.35355	2.28771	0.00000	-0.29542
C15	6.65929	0.00000	0.00000	-0.05466
C16	-6.65929	0.00000	0.00000	-0.05466
H1	2.32543	6.42907	0.00000	0.15533
H2	6.38694	4.06382	0.00000	0.15109
H3	2.32543	-6.42907	0.00000	0.15533
H4	6.38694	-4.06382	0.00000	0.15109
H5	-2.32543	-6.42907	0.00000	0.15533
H6	-6.38694	-4.06382	0.00000	0.15109
H7	-2.32543	6.42907	0.00000	0.15533
H8	-6.38694	4.06382	0.00000	0.15109
H9	8.71284	0.00000	0.00000	0.12425
H10	-8.71284	0.00000	0.00000	0.12425