

# **Energy decomposition analysis based on absolutely localised molecular orbitals for large-scale density functional theory calculations in drug design**

M. J. S. Phipps,<sup>†</sup> T. Fox,<sup>‡</sup> C. S. Tautermann,<sup>‡</sup> and C.-K. Skylaris<sup>\*,†</sup>

*School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK, and  
Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG,  
88397 Biberach, Germany*

E-mail: C.Skylaris@soton.ac.uk

---

\*To whom correspondence should be addressed

<sup>†</sup>School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

<sup>‡</sup>Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany

## Abstract

We report the development and implementation of an energy decomposition analysis (EDA) scheme in the ONETEP linear-scaling electronic structure package. Our approach is hybrid as it combines the localised molecular orbital EDA (Su, P. and Li, H., *J. Chem. Phys.*, **2009**, *131*, 014102) and the absolutely localised molecular orbital EDA (Khaliullin, R. Z. et al. *J. Phys. Chem. A*, **2007**, *111*, 8753-8765) to partition the intermolecular interaction energy into chemically distinct components (electrostatic, exchange, correlation, Pauli-repulsion, polarisation and charge transfer). Limitations shared in EDA approaches such as the issue of basis set dependence in polarisation and charge transfer are discussed and a remedy to this problem is proposed that exploits the strictly localised property of the ONETEP orbitals. Our method is validated on a range of complexes with interactions relevant to drug design. We demonstrate the capabilities for large-scale calculations with our approach on complexes of thrombin with an inhibitor comprised of up to 4975 atoms. Given the capability of ONETEP for large-scale calculations, such as on entire proteins, we expect that our EDA scheme can be applied in a large range of biomolecular problems, especially in the context of drug design.

## 1 Introduction.

Intermolecular interactions are of key importance in determining the physical and chemical properties of molecular systems. For example, the intermolecular forces that govern a hydrogen bond are of great interest to chemists due to the role they play in determining polymeric structures and macroscopic properties of structures.<sup>1,2</sup> The effect of  $\sigma$  holes as observed in a number of halogen-bonding complexes<sup>3</sup> has implications for drug-host binding and therefore this interaction is of key pharmaceutical interest. Scientific fields that involve the study of systems containing large numbers of intermolecular interactions, such as the disciplines of biomolecular<sup>4,5</sup> and supramolecular chemistry<sup>6,7</sup> and condensed matter physics,<sup>8,9</sup> benefit from the insights gained from studies of intermolecular bonding.

The interaction energy is defined in the supermolecular approach as the difference between

the ground-state energies of a number of monomers in isolation and the ground-state energy of the supermolecular complex that these monomers form, and the energy value obtained from this approach can provide for highly detailed and accurate studies of systems of interest. However, the chemical effects such as electrostatics, Pauli-repulsion, polarisation and charge transfer that each contribute to the overall intermolecular interaction are hidden within this value. Such bonding components are important from a chemical stance as they provide valuable insight into the synergy and nature of bonding. Additionally, these components are also valuable pedagogically as they provide a link between the world of *ab-initio* quantum calculations and the familiar picture of intermolecular forces known to chemists.

In recent times many developments have been made in gaining insight into these often complex interactions. As there is no unique decomposition, a large number of perturbative and variational energy decomposition analysis methods exist<sup>10–20</sup> that seek to decompose the interaction energy  $\Delta E$  into its constituent chemical components. These include the symmetry-adapted perturbation theory (SAPT) scheme<sup>19,20</sup> based on the perturbative treatment of fragment wavefunctions in which chemically useful terms such as electrostatic, exchange-repulsion, induction, and dispersion are obtained. Many variational schemes have been developed from the pioneering Kitaura-Morokuma (KM) scheme.<sup>10,21,22</sup> These include for example the reduced variational space (RVS)<sup>11</sup> approach, the constrained space orbital variations (CSOV)<sup>12,23</sup> energy decomposition, and the natural energy decomposition analysis (NEDA) scheme<sup>13,14,24,25</sup> based on natural bond orbital theory.<sup>15,26</sup> Additionally, schemes have been proposed and implemented for analyses of periodic extended systems including the EDA of Raupach and Tonner,<sup>27</sup> the CSOV EDA implementation in the CRYSTAL package<sup>28</sup> and the EDA of Philipsen and Baerends<sup>29</sup> within the BAND code.<sup>30</sup>

These methods have demonstrated varying degrees of success in the study of protein-ligand interactions,<sup>31</sup> and show that the problem of separating polarisation and charge transfer effects is approached in a variety of ways. For example, schemes such as the localised molecular orbital (LMO)<sup>16</sup> and extended transition state (ETS)<sup>32–35</sup> EDAs simply do not seek to separate polarisation and charge transfer effects at all. Other schemes that attempt this partitioning consider the

problem as involving the optimisation of molecular orbitals (MOs) subject to conditions that impose localisation to particular fragments. Schemes that belong to this category include the KM, RVS, CSOV, absolutely localised molecular orbital (ALMO)<sup>17</sup> and block-localised wavefunction (BLW)<sup>36,37</sup> EDAs. The NEDA scheme that involves a highly localised picture of bonding also considers this to be an orbital localisation problem, however its solution does not involve variational optimisation of a polarised state. The EDA of Wu et al.<sup>18</sup> considers charge transfer to be more appropriately partitioned by charge density using the constrained density functional theory (CDFT)<sup>38</sup> approach.

The use of these methods has generally been limited to studies involving no more than a few tens of atoms.<sup>31</sup> Exceptions to this include use of the fragment molecular orbital (FMO) framework<sup>39-41</sup> in schemes such as the PIEDA.<sup>42</sup> Recently, work has been published<sup>43</sup> investigating 18 class A GPCR-ligand crystal structures using the FMO-MP2 (second order Møller-Plesset perturbation theory)<sup>44</sup> PIEDA method with the 6-31G\* basis set, demonstrating the components of the interactions and hence the high suitability of EDA to the field of drug design especially in the case of large drug-protein systems. With the developments made in the field of linear-scaling density functional theory (DFT) methods<sup>45,46</sup> a new opportunity to remove this size limitation arises. The ALMO EDA scheme has been shown to give reasonable polarisation and charge transfer energies<sup>31</sup> and benefits from the use of a properly antisymmetrized state in performing this partitioning. Notably, use of SCF-MI in separating polarisation and charge transfer offers improvement upon the scaling of SCF calculations due to partitioning of the diagonalisation bottleneck.<sup>47</sup> This has been demonstrated by Mo and Gao<sup>48</sup> in their quantum mechanical and molecular mechanical coupled BLW EDA (of relation to ALMO EDA) study of ionic system solvation with up to 32 water molecules being treated quantum mechanically. The frozen density energy component of the ALMO EDA decomposition provides insights into the interaction of fragment densities, however also contained within this term are effects originating from electrostatic, exchange, Pauli-repulsion and correlation interactions. A rigorous treatment of the frozen density component is possible that further decomposes the term into its constituents via a modified LMO EDA<sup>16</sup> analysis.

Within this paper, we provide a brief summary of the ALMO EDA scheme of Head-Gordon et al.<sup>17</sup> and LMO EDA scheme of Su and Li<sup>16</sup> and describe our implementation of a hybrid EDA scheme in the ONETEP<sup>49</sup> linear-scaling DFT package. The limitations of using the ALMO approach to partition polarisation from charge transfer effects is explored and a solution proposed that exploits the strictly-localised property of the ONETEP orbitals. An assessment of the method is provided through its application to a test set of small complexes that display biomolecule-like bonding characteristics. We then demonstrate the ability of our approach to be applied on entire protein-ligand systems by performing EDA calculations on a 4975-atom complex between thrombin and one of its inhibitors.

## 2 Theory.

In the following discussion of the theory, we have adopted the Einstein convention of implicit summation over repeated Greek indices only. We will use the word ‘fragment’ to describe any molecule which is part of a complex held together by intermolecular interactions (*i.e.* no chemical bonds). All our fragments in the following examples are closed-shell, but it is a straightforward task to extend the implementation to open-shell systems.

### 2.1 The localised and absolutely localised energy decomposition analyses.

In order to develop the equations used in the EDA approaches, it is first necessary to formally express the supermolecule interaction energy. This is given by,

$$\begin{aligned}\Delta E &= E[\Psi_X] - \sum_{A \in X}^{N_{\text{frag}}} [\Psi_A^{0,A}] + \Delta E_{\text{nuc}} + \Delta E_{\text{BSSE}} \\ \Delta E_{\text{nuc}} &= E_{\text{nuc}}(X) - \sum_{A \in X}^{N_{\text{frag}}} E_{\text{nuc}}(A)\end{aligned}\tag{1}$$

where  $\Psi$  represents an electronic wavefunction with the subscripts  $A$  and  $X$  used to represent the fragments  $A$  that compose the supermolecule  $X$ , with the superscript ‘0’ used to denote states

optimised in isolation from the field of the remaining fragments of the supermolecule and the superscript ‘A’ used to indicate optimisation only in the fragment A basis set, and where  $N_{\text{frag}}$  is the number of fragments comprising the supermolecule X.  $E[\Psi_Y]$  is the energy of system Y and  $E_{\text{nuc}}(Y)$  represents its nuclei–nuclei electrostatic energy.  $\Delta E_{\text{BSSE}}$  is included as a correction for the basis set superposition error (BSSE).

### 2.1.1 The absolutely localised energy decomposition analysis.

The ALMO EDA developed by Head-Gordon et al.<sup>17</sup> decomposes the interaction energy  $\Delta E$  into the frozen density  $\Delta E_{\text{FRZ}}$ , polarisation  $\Delta E_{\text{POL}}$ , and charge transfer  $\Delta E_{\text{CT}}$  interaction terms as,

$$\Delta E = \Delta E_{\text{FRZ}} + \Delta E_{\text{POL}} + \Delta E_{\text{CT}} . \quad (2)$$

We define a number of key wavefunctions that are necessary to allow clear description of the EDA theory in this paper. The electronic wavefunction  $\Psi_A^{0,A}$  is the wavefunction of the arbitrary fragment A, and  $\Psi_X^{0,\text{orth}}$  is the properly antisymmetrized wavefunction of the interacting X complex constructed from the nonorthogonal frozen occupied MOs of the fragments.  $\Psi_X^{\text{ALMO}}$  is defined as the wavefunction of the polarised state produced by relaxation of the nonorthogonal frozen MOs while still ensuring localisation of the MOs within their corresponding fragments. This localisation is imposed through the constraint that the MOs are only expanded using basis functions on a particular fragment and therefore prevents charge delocalisation between fragments.  $\Psi_X$  is defined as the wavefunction of fully antisymmetrized and fully relaxed state. Each energy component may be described in terms of energy functionals of these intermediate wavefunctions as,

$$\Delta E_{\text{FRZ}} = E[\Psi_X^{0,\text{orth}}] - \sum_{A \in X}^{N_{\text{frag}}} E[\Psi_A^{0,A}] + \Delta E_{\text{nuc}} \quad (3)$$

$$\Delta E_{\text{POL}} = E[\Psi_X^{\text{ALMO}}] - E[\Psi_X^{0,\text{orth}}] \quad (4)$$

$$\Delta E_{\text{CT}} = E[\Psi_X] - E[\Psi_X^{\text{ALMO}}] + \Delta E_{\text{BSSE}} \quad (5)$$

where  $N_{\text{frag}}$  is the number of fragments comprising the supermolecule X and  $\Delta E_{\text{BSSE}}$  is calculated using the Boys–Bernardi<sup>50</sup> counterpoise correction (CP) approach. The geometries of fragments A are defined as found in the (geometry optimised) X complex (*i.e.* only the electronic density is optimised in this calculation).

The ALMO EDA achieves the separation of polarisation and charge transfer effects using the ALMO state,  $\Psi_{\text{X}}^{\text{ALMO}}$ . This state describes a system which is polarised but that has the restriction of no charge transfer between the fragments. The optimisation of the ALMO state proceeds via a procedure termed locally projected SCF for molecular interactions (SCF-MI).<sup>47,51</sup> The SCF-MI equations we use are those of Stoll et al.<sup>52</sup> and are further described in Section SI.1 of the Supporting Information. We note that this approach to separating out polarisation and charge transfer effects is the same as that described within the BLW<sup>36,37</sup> EDA procedure. Specifically, the BLW state that is identical to the ALMO state is optimised either by Jacobi transformation<sup>53</sup> or by the algorithm of Gianinetti and Raimondi et al.<sup>54,55</sup>

### 2.1.2 The frozen density analysis.

In this section we describe a prescription for a frozen density analysis produced via the LMO EDA theory of Su and Li,<sup>16</sup> in which  $\Delta E_{\text{FRZ}}$  is further decomposed into individual electrostatic ( $\Delta E_{\text{ES}}$ ), exchange ( $\Delta E_{\text{EX}}$ ), Pauli–repulsion ( $\Delta E_{\text{REP}}$ ) and correlation ( $\Delta E_{\text{CORR}}$ ) terms as,

$$\Delta E_{\text{FRZ}} = \Delta E_{\text{ES}} + \Delta E_{\text{EX}} + \Delta E_{\text{REP}} + \Delta E_{\text{CORR}} \quad . \quad (6)$$

Combining the classical-like terms of the Kohn Sham (KS) energy (by excluding the exchange and correlation contributions) as,

$$E_{\text{cl}} = 2 \sum_{i,j}^{N_{\psi(\text{occ})}} h_{ij} \sigma^{ij} + \sum_{i,j,k,l}^{N_{\psi(\text{occ})}} \langle ij|kl \rangle \times \sigma^{ij} \sigma^{kl} + E_{\text{nuc}} \quad (7)$$

where  $N_{\psi(\text{occ})}$  is the number of occupied MOs, kinetic energy effects are included by presence of the full core Hamiltonian matrix  $h_{ij}$ , and where the contravariant metric tensor  $\sigma^{ij}$  is the inverse

of the MO overlap matrix  $\sigma_{ij}$ ,

$$\sigma_{ij} = \int d\mathbf{r} \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) = \langle \psi_i | \psi_j \rangle \quad (8)$$

$$\sigma^{ij} \equiv (\sigma^{-1})_{ij} \quad (9)$$

the components of the frozen density interaction may be expressed using intermediate wavefunctions as,

$$\Delta E_{\text{ES}} = E_{\text{cl}}[\Psi_{\text{X}}^0] - \sum_{A \in \text{X}}^{N_{\text{frag}}} E_{\text{cl}}[\Psi_A^{0,A}] \quad (10)$$

$$\Delta E_{\text{EX}} = E_{\text{X}}[\Psi_{\text{X}}^0] - \sum_{A \in \text{X}}^{N_{\text{frag}}} E_{\text{X}}[\Psi_A^{0,A}] \quad (11)$$

$$\begin{aligned} \Delta E_{\text{REP}} = & E_{\text{cl}}[\Psi_{\text{X}}^{0,\text{orth}}] - E_{\text{cl}}[\Psi_{\text{X}}^0] \\ & + E_{\text{X}}[\Psi_{\text{X}}^{0,\text{orth}}] - E_{\text{X}}[\Psi_{\text{X}}^0] \end{aligned} \quad (12)$$

$$\begin{aligned} \Delta E_{\text{CORR}} = & E_{\text{C}}[\Psi_{\text{X}}^{0,\text{orth}}] - \sum_{A \in \text{X}}^{N_{\text{frag}}} E_{\text{C}}[\Psi_A^{0,A}] \\ \equiv & \Delta E_{\text{FRZ}} - (\Delta E_{\text{ES}} + \Delta E_{\text{EX}} + \Delta E_{\text{REP}}) \end{aligned} \quad (13)$$

where  $E_{\text{X}}[n(\mathbf{r})]$  and  $E_{\text{C}}[n(\mathbf{r})]$  are the DFT exchange and correlation energy functionals of the electronic density  $n(\mathbf{r})$  associated with the wavefunction  $\Psi(\mathbf{r})$ , and where we note that the kinetic energy contribution to  $\Delta E_{\text{ES}}$  is zero due to the kinetic energy of the fragment states  $\Psi_A^{0,A}$  cancelling with the kinetic energy of the frozen supermolecule state  $\Psi_{\text{X}}^0$  as  $\sigma^{ij}$  is the identity matrix for  $E_{\text{cl}}[\Psi_{\text{X}}^0]$ , while it is as in eqs. 8 and 9 for  $E_{\text{cl}}[\Psi_{\text{X}}^{0,\text{orth}}]$ .

The components of this frozen energy decomposition are formally similar to the components described in the LMO EDA<sup>16</sup> but with inclusion of a new component  $\Delta E_{\text{CORR}}$  that isolates the correlation energy change on orthogonalisation contained within the LMO EDA ‘polarisation’ term. It is also noted that the electrostatic term of this decomposition is formally similar to that within the BLW<sup>36,37</sup> scheme. Additionally, it is important to note the development of a number of alternative



approaches to the above theory such as the scheme of Horn et al.<sup>56</sup> for investigating (decomposing) the frozen interaction in KS DFT calculations and the density-based energy decomposition analysis (DEDA) of Wu<sup>57</sup> that calculates the frozen density interaction in a variational manner.

## 2.2 The combined ONETEP approach.

Our aim has been to implement within the ONETEP<sup>49</sup> package a combination of the ALMO EDA approach<sup>17</sup> and the frozen density component analysis based upon the LMO EDA theory of Su and Li<sup>16</sup> above. Our prescription includes the electrostatic, exchange, Pauli-repulsion and modified correlation terms of the LMO EDA of Su and Li<sup>16</sup> in the frozen density component of the ALMO EDA, with the ALMO EDA polarisation and charge transfer components remaining as in their original form. By substitution of eq. 6 into eq. 2 we obtain the fully-decomposed  $\Delta E$  as,

$$\begin{aligned} \Delta E = & \Delta E_{\text{ES}} + \Delta E_{\text{EX}} + \Delta E_{\text{REP}} + \Delta E_{\text{CORR}} \\ & + \Delta E_{\text{POL}} + \Delta E_{\text{CT}} \quad . \end{aligned} \quad (14)$$

The implementation of this EDA scheme within a linear-scaling DFT code such as ONETEP that aims to achieve large basis set accuracy in routine calculations is an important capability. By decomposing the frozen density component into its substituents and expressing the interaction energy in terms of 6 rigorously defined energy terms, our EDA scheme provides a detailed analysis of the interaction energy that allows understanding of the key chemical components of binding. In what follows we describe the ONETEP EDA implementation and provide examples of calculations using this.

### 3 Implementation of EDA in ONETEP.

#### 3.1 The ONETEP formalism.

In the ONETEP formalism, the set of KS orbitals  $\{|\psi_i\rangle\}$  can be considered as linear combinations of a set of localised orbitals  $\{|\phi_\alpha\rangle\}$  known as nonorthogonal generalised Wannier functions (NGWFs)<sup>58</sup> as,

$$\psi_i(\mathbf{r}) = \phi_\alpha(\mathbf{r})M_i^\alpha \quad . \quad (15)$$

where  $\{M_i^\alpha\}$  are the KS MO expansion coefficients. The NGWFs  $\{|\phi_\alpha\rangle\}$  are expanded in a basis of  $N_p$  periodic cardinal sine (psinc) functions<sup>59,60</sup>  $\{D_p\}$  as

$$\phi_\alpha(\mathbf{r}) = \sum_{p=1}^{N_p} D_p(\mathbf{r})c_{p\alpha} \quad (16)$$

where  $c_{p\alpha}$  are the expansion coefficients. The psinc functions are centred on a regular grid that spans the whole simulation cell. The grid point each psinc function is centred on is given by the index  $p$ . The information of the KS states is contained within the density matrix given by

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i^{N_{\psi(\text{occ})}} \psi_i(\mathbf{r})\psi_i^*(\mathbf{r}') \quad . \quad (17)$$

and the electronic density is obtained from  $\rho(\mathbf{r}, \mathbf{r}')$  as,

$$n(\mathbf{r}) = 2\rho(\mathbf{r}, \mathbf{r}) \quad . \quad (18)$$

Orthonormality of  $\{|\psi_i\rangle\}$  is a necessary physical constraint for the Pauli principle to be respected, which can be expressed as,

$$\sigma_{ij} = \langle \psi_i | \psi_j \rangle = (M^\dagger)_i^\alpha S_{\alpha\beta} M_j^\beta = \delta_{ij} \quad (19)$$

where the NGWF overlap matrix  $S_{\alpha\beta}$  is given by,

$$S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle \quad (20)$$

By substitution of eq. 15 into eq. 17, we introduce the density kernel  $K^{\alpha\beta}$  as the representation of the density operator in terms of NGWFs,

$$\rho(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta^*(\mathbf{r}') \quad (21)$$

where the density kernel  $K^{\alpha\beta}$  is given by,

$$K^{\alpha\beta} = \sum_i^{N_\phi} M_{fi}^\alpha f_i (M^\dagger)_i^\beta \quad (22)$$

where  $f_i$  are the MO occupancies which are equal to 1 for the doubly-occupied MOs  $\psi_i(\mathbf{r})$ .

### 3.2 Construction of the frozen density kernels.

In the frozen density analysis, we seek expressions of the kernel and set of NGWFs that represent the two intermediate wavefunctions  $\Psi_X^0$  and  $\Psi_X^{0,\text{orth}}$ . The NGWF set of these wavefunctions,  $\{\phi_{A\alpha}(\mathbf{r})\}$ , is constructed by superposition of the NGWFs of the isolated fragments  $A$  onto the supermolecule  $X$ . We now define the two density kernels  $K_0^{A\alpha,B\beta}$  and  $K_{\text{orth}}^{A\alpha,B\beta}$  that are used to represent  $\Psi_X^0$  and  $\Psi_X^{0,\text{orth}}$  in this ‘frozen’ NGWF set.

The density kernel of the frozen state,  $K_0^{A\alpha,B\beta}$ , is constructed by superposition of the occupied states of the fragments onto the supermolecule. This is constructed from the fragment density kernels  $(K_A)^{\alpha\beta}$  as,

$$\begin{aligned} K_0^{A\alpha,A\beta} &= (K_A)^{\alpha\beta} \\ K_0^{A\alpha,B\beta} &= 0|_{A \neq B} \quad . \end{aligned} \quad (23)$$

The density kernel  $K_{\text{orth}}^{A\alpha,B\beta}$  is the properly antisymmetrized form of  $K_0^{A\alpha,B\beta}$ . The construction of this kernel requires the coefficients  $(M_A)_i^\alpha$  of the isolated fragments. From eq. 22 for the density kernel we can write a generalised eigenvalue problem as,

$$K^{\alpha\beta} S_{\beta\gamma} M_i^\gamma = M_i^\alpha f_i \quad . \quad (24)$$

This eigenvalue equation may be expressed for the fragment problems as,

$$(K_A)^{\alpha\beta} (S_A)_{\beta\gamma} (M_A)^\gamma_i = (M_A)^\alpha_i (f_A)_i \quad (25)$$

where the subscript  $A$  is used to denote a fragment quantity. The vectors of the fragment matrices  $M_A$  that are obtained for unoccupied MOs are discarded at this point. This is achieved by ordering the MOs by decreasing orbital occupancy and retaining only  $(M_A)_{\beta i} |_{i \leq N_{\psi(\text{occ},A)}}$  where  $N_{\psi(\text{occ},A)}$  is the number of occupied MOs on fragment  $A$ .

The block-diagonal MO coefficient matrix  $M_{Bi}^{A\alpha}$  of the frozen supermolecule is constructed from the fragment coefficient matrices  $(M_A)_i^\alpha$  as,

$$\begin{aligned} M_{Ai}^{A\alpha} &= (M_A)_i^\alpha \\ M_{Bi}^{A\alpha} &= 0 |_{A \neq B} \quad . \end{aligned} \quad (26)$$

Finally, the density kernel  $K_{\text{orth}}^{A\alpha,B\beta}$  of the properly antisymmetrized state is expressed using these MO coefficients as,

$$\begin{aligned} K_{\text{orth}}^{A\alpha,B\beta} &= \sum_{C,D \in X}^{N_{\text{frag}}} \sum_i^{N_{\psi(\text{occ},C)}} \sum_j^{N_{\psi(\text{occ},D)}} M_{Ci}^{A\alpha} \sigma^{Ci,Dj} (M^\dagger)_{Dj}^{B\beta} \\ &= \sum_i^{N_{\psi(\text{occ},A)}} \sum_j^{N_{\psi(\text{occ},B)}} M_{Ai}^{A\alpha} \sigma^{Ai,Bj} (M^\dagger)_{Bj}^{B\beta} \end{aligned} \quad (27)$$

where we have exploited the block-diagonal structure of  $M$  described in eq. 26 and where  $i$  and  $j$  run over the occupied MOs of the fragments and where the inverse MO overlap matrix  $\sigma^{Ai,Bj}$  is

constructed from the quantity  $\sigma_{Ai,Bj}$ ,

$$\sigma_{Ai,Bj} = \sum_{C,D \in X}^{N_{\text{frag}}} (M^\dagger)_{Ai}^{C\alpha} S_{C\alpha,D\beta} M_{Bj}^{D\beta} \quad (28)$$

where  $S_{C\alpha,D\beta}$  is the full NGWF set overlap matrix for the supermolecule.

For clarity, the structures of the quantities  $M_{Bi}^{A\alpha}$  and  $\sigma_{Ai,Bj}$  involved in constructing  $K_{\text{orth}}^{A\alpha,B\beta}$  are shown diagrammatically in Figure 1.

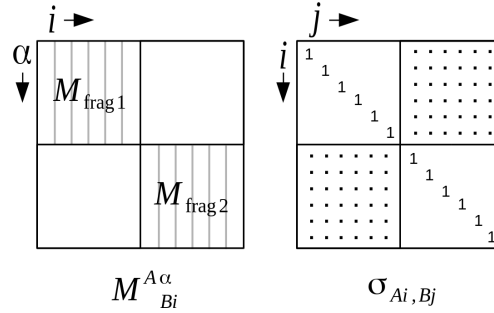


Figure 1: A diagrammatic representation of the quantities involved in the construction of the frozen density kernel  $K_{\text{orth}}^{A\alpha,B\beta}$  for a system comprising of two fragments ‘frag1’ and ‘frag2’. Empty matrix sub-blocks represent zero blocks, and sub-blocks containing dots represent non-zero blocks. The diagonal intrafragmental MO overlap sub-blocks of  $\sigma_{Ai,Bj}$  are identity due to orthonormality of the KS states within fragments, and non-zero for the sub-blocks representing interfragment MO overlaps due to lack of orthonormality.

The energies of the two intermediate wavefunctions  $\Psi_X^0$  and  $\Psi_X^{0,\text{orth}}$  are expressed in terms of energy functionals of the quantities  $K_0^{A\alpha,B\beta}$ ,  $K_{\text{orth}}^{A\alpha,B\beta}$  and  $\{\phi_{A\alpha}(\mathbf{r})\}$  as,

$$E[\Psi_X^0] \equiv E[\{K_0^{A\alpha,B\beta}\}, \{\phi_{A\alpha}(\mathbf{r})\}] \quad (29)$$

$$E[\Psi_X^{0,\text{orth}}] \equiv E[\{K_{\text{orth}}^{A\alpha,B\beta}\}, \{\phi_{A\alpha}(\mathbf{r})\}] \quad (30)$$

### 3.3 Calculation of the polarised density kernel.

A brief overview of the ALMO polarisation procedure as implemented within the ONETEP package is now described. As discussed in section 3.1, the problem of obtaining the ground state of a system in ONETEP involves the optimisation of the NGWFs and the optimisation of the density

kernel for each current set of NGWFs. In ONETEP, the Stoll eigenvalue problem is not solved directly by diagonalisation, but the ground state is obtained by a conjugate gradient minimisation of the energy. An overview of the Stoll equations on which this procedure is based in terms of the quantities used in ONETEP is provided in Section SI.2 of the Supporting Information.

To describe the optimisation of the density kernel in the SCF-MI equations within ONETEP, we begin by introducing the density kernel corresponding to the non-antisymmetrized wavefunction at SCF-MI iteration  $m$  as  $K_{0(m)}^{A\alpha,B\beta}$ . At  $m = 0$ , this kernel is simply set to the kernel  $K_0^{A\alpha,B\beta}$  constructed in the previous frozen density analysis by direct summation of the fragment density kernels. Having obtained this kernel, SCF-MI loops are performed in order to optimise the total energy with respect to the density kernel. The electronic energy is evaluated at each iteration using the antisymmetrized representation of this kernel,  $K_{\text{orth}}^{A\alpha,B\beta}$ . This electronic energy is expressed in terms of the quantities used in ONETEP as,

$$\begin{aligned} E_{\text{orth}} &= \sum_{A,B \in X}^{N_{\text{frag}}} \sum_i^{N_{\psi(\text{occ},A)}} \sum_j^{N_{\psi(\text{occ},B)}} \sigma^{Ai,Bj} \langle \psi_{Bj} | \hat{H} | \psi_{Ai} \rangle - E_{\text{DC}} \\ &= \sum_{A,B \in X}^{N_{\text{frag}}} K_{\text{orth}}^{A\alpha,B\beta} H_{B\beta,A\alpha} - E_{\text{DC}} \end{aligned} \quad (31)$$

where  $E_{\text{DC}}$  is the double-counting correction,

$$E_{\text{DC}} = E_{\text{H}}[n(\mathbf{r})] - E_{\text{XC}}[n(\mathbf{r})] + \int \frac{\delta E_{\text{XC}}[n(\mathbf{r})]}{\delta n(\mathbf{r})} n(\mathbf{r}) d\mathbf{r} \quad (32)$$

where  $E_{\text{H}}$  and  $E_{\text{XC}}$  are the Hartree and exchange and correlation energies, and where,

$$n(\mathbf{r}) = 2 \sum_{A,B \in X}^{N_{\text{frag}}} \phi_{A\alpha}(\mathbf{r}) K_{\text{orth}}^{A\alpha,B\beta} \phi_{B\beta}(\mathbf{r}) \quad . \quad (33)$$

A single iteration of the SCF-MI kernel loop contains the following steps:

1. The fragment eigenvalue problems described in eq. 25 are solved to obtain the isolated fragment MO coefficients  $(M_A)^\alpha_i$  and subsequently  $M_{Bi}^{A\alpha}$ .

2. The MO overlap matrix  $\sigma_{Ai,Bj}$  is constructed using eq. 28 and inverted to obtain  $\sigma^{Ai,Bj}$ .
3. The kernel  $K_{\text{orth}}^{A\alpha,B\beta}$  is constructed from  $M_{Bi}^{A\alpha}$  using eq. 27 and  $\sigma^{Ai,Bj}$ .
4. The unprojected Hamiltonian  $H_{\alpha\beta}$  is calculated and the energy of the ALMOs evaluated as,

$$E_{\text{orth}} = \sum_{A,B \in X}^{N_{\text{frag}}} K_{\text{orth}}^{A\alpha,B\beta} H_{B\beta,A\alpha} - E_{\text{DC}} \quad . \quad (34)$$

If variations in the density kernel are less than a specified tolerance then exit.

5. The locally projected Hamiltonian  $(H_S^A)_{A\alpha,A\beta}$  is constructed from  $H_{A\alpha,B\beta}$ ,  $S_{A\alpha,B\beta}$ ,  $M_{Bi}^{A\alpha}$  and  $\sigma^{Ai,Bj}$  using eq. S9 provided in the Supporting Information.
6.  $(H_S^A)_{A\alpha,A\beta}$ ,  $K_{0(m)}^{A\alpha,B\beta}$  are used in a conjugate gradient minimisation of the Stoll locally projected energy of the fragments  $E_S$ ,

$$E_S = \sum_{A \in X}^{N_{\text{frag}}} K_{0(m)}^{A\alpha,A\beta} (H_S^A)_{A\beta,A\alpha} - E_{\text{DC}} \quad (35)$$

where  $E_{\text{DC}}$  is evaluated using the associated fragment densities of  $K_{0(m)}^{A\alpha,B\beta}$  (eq. 33), to obtain the new density kernel  $K_{0(m+1)}^{A\alpha,B\beta}$ . The minimisation approach used<sup>61</sup> is the method of Li, Nunes and Vanderbilt,<sup>62,63</sup> wherein the density kernel  $K_0^{A\alpha,B\beta}$  is defined in terms of an auxiliary density kernel  $L_0^{A\alpha,B\beta}$  by the McWeeny purifying transform,<sup>64</sup>

$$K_0^{A\alpha,B\beta} = 3(L_0 S_0 L_0)^{A\alpha,B\beta} - 2(L_0 S_0 L_0 S_0 L_0)^{A\alpha,B\beta} \quad (36)$$

where  $S_0$  is the fragment overlap matrix constructed from the fragment overlap matrices  $S_A$  as,

$$\begin{aligned} (S_0)_{A\alpha,A\beta} &= (S_A)_{A\alpha,A\beta} \\ (S_0)_{A\alpha,B\beta} &= 0|_{A \neq B} \quad . \end{aligned} \quad (37)$$

This purification transformation is iterated as required within the LNV subroutine to yield a fully idempotent  $K_0^{A\alpha,B\beta}$ , equivalent to orthonormalisation of the KS states within the fragments. The gradient of the total energy with respect to the auxiliary density kernel  $L_0^{A\alpha,A\beta}$  is calculated as,

$$\begin{aligned} \frac{\partial E_S}{\partial L_0^{A\alpha,A\beta}} &= 6(S_0 L_0 H_S^A + H_S^A L_0 S_0)_{A\beta,A\alpha} \\ &\quad - 4(S_0 L_0 S_0 L_0 H_S^A + S_0 L_0 H_S^A L_0 S_0 + H_S^A L_0 S_0 L_0 S_0)_{A\beta,A\alpha} \\ \frac{\partial E_S}{\partial L_0^{A\alpha,B\beta}} &= 0|_{A \neq B} \end{aligned} \quad (38)$$

and minimisation of  $E_S$  is equivalent to minimisation of  $E_{\text{orth}}$  with the ALMO constraint.

Having obtained the SCF-MI optimised density kernel  $K_{\text{orth}}^{A\alpha,B\beta}$ , the optimisation of the NGWFs proceeds as described in Section SI.3 of the Supporting Information. Once the solution to the minimisation of both the NGWFs and density kernel is found and the final polarised ALMO state energy obtained, the polarisation energy is then calculated using eq. 4. Charge transfer is calculated using eq. 5 but with omission of the  $\Delta E_{\text{BSSE}}$  contribution, *i.e.*

$$\Delta E_{\text{CT}} = E[\Psi_X] - E[\Psi_X^{\text{ALMO}}] \quad (39)$$

where,

$$E[\Psi_X] \equiv E[\{K^{A\alpha,B\beta}\}, \{\phi_\alpha(\mathbf{r})\}] \quad (40)$$

$$E[\Psi_X^{\text{ALMO}}] \equiv E[\{K_{\text{orth}}^{A\alpha,B\beta}\}, \{(\phi_{\text{ALMO}})_{A\alpha}(\mathbf{r})\}] \quad (41)$$

where  $\{K^{A\alpha,B\beta}\}$  and  $\{\phi_\alpha(\mathbf{r})\}$  are respectively the fully optimised density kernel and NGWFs of the supermolecule without ALMO constraint (*i.e.* the quantities obtained from a standard singlepoint calculation),  $\{(\phi_{\text{ALMO}})_{A\alpha}(\mathbf{r})\}$  are the NGWFs determined by the minimisation of eq. S10 provided in the Supporting Information, and where  $\{K_{\text{orth}}^{A\alpha,B\beta}\}$  is the optimised ALMO-constrained kernel



(not equal to  $\{K_{\text{orth}}^{A\alpha,B\beta}\}$  of eq. 27). The exclusion of  $\Delta E_{\text{BSSE}}$  in the ONETEP implementation is justified as it is zero due to the fact that the psinc basis set is independent of atomic positions.

## 4 Results and discussion.

### 4.1 Validation of the method.

We have applied our EDA to a number of small test systems which model interactions relevant within the field of drug design in order to validate the accuracy of the ONETEP EDA against its ALMO and LMO EDA counterparts in the Q-Chem<sup>65</sup> and GAMESS-US<sup>66</sup> *ab initio* packages respectively. These systems have been selected as they provide examples of important interactions such as hydrogen bonding,  $\pi$ - $\pi$ , and dispersion.<sup>67,68</sup> The test systems are derived from those of our earlier work<sup>31</sup> in which we present a benchmark study of a selection of EDA methodologies using 6 test sets of small interacting molecules of biomolecular relevance.

#### 4.1.1 Calculation set-up.

Geometry optimisation was performed at the PBE-D2/aug-cc-pVTZ level of theory<sup>69,70</sup> on all structures using the Q-Chem *ab initio* package.<sup>65</sup> The -D2 correction for dispersion of Grimme et al.<sup>70</sup> was used in order to properly model the dispersion interactions especially observed in the case of the  $\pi$ - $\pi$  interacting systems. The optimised geometries of the systems studied are shown within Figure 2 and Figure 3.

EDA was performed on the geometry optimised structures at the PBE-D2/aug-cc-pVTZ level of theory in the case of the LMO and ALMO EDAs, and at the PBE-D2 level with a 1200 eV

Figure 2: The test set 1 of PBE-D2/aug-cc-pVTZ geometry optimised systems for EDA (inter-molecular distances are given in Å)

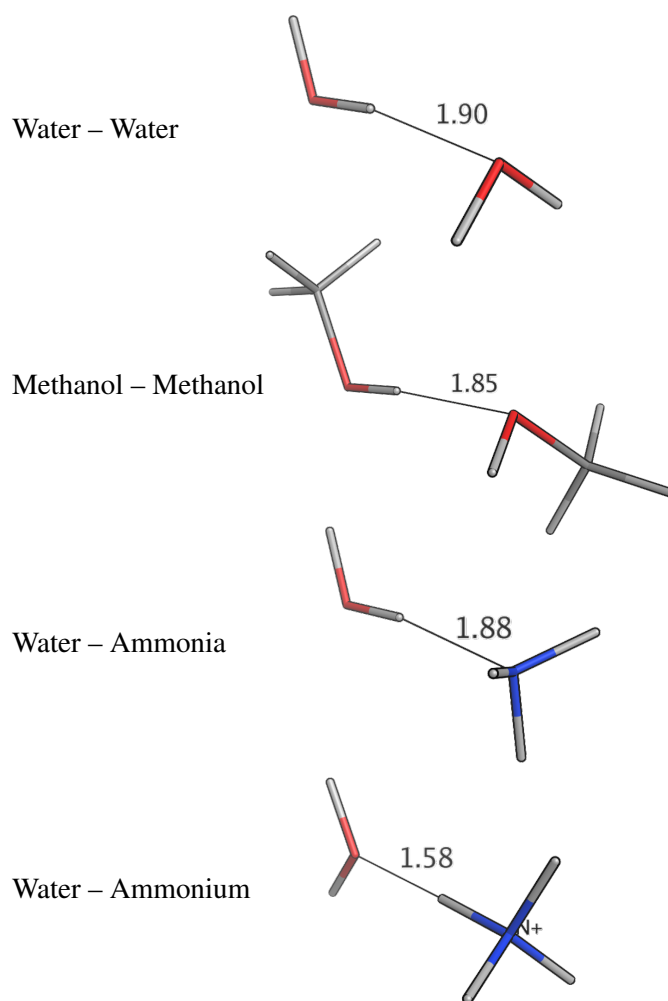
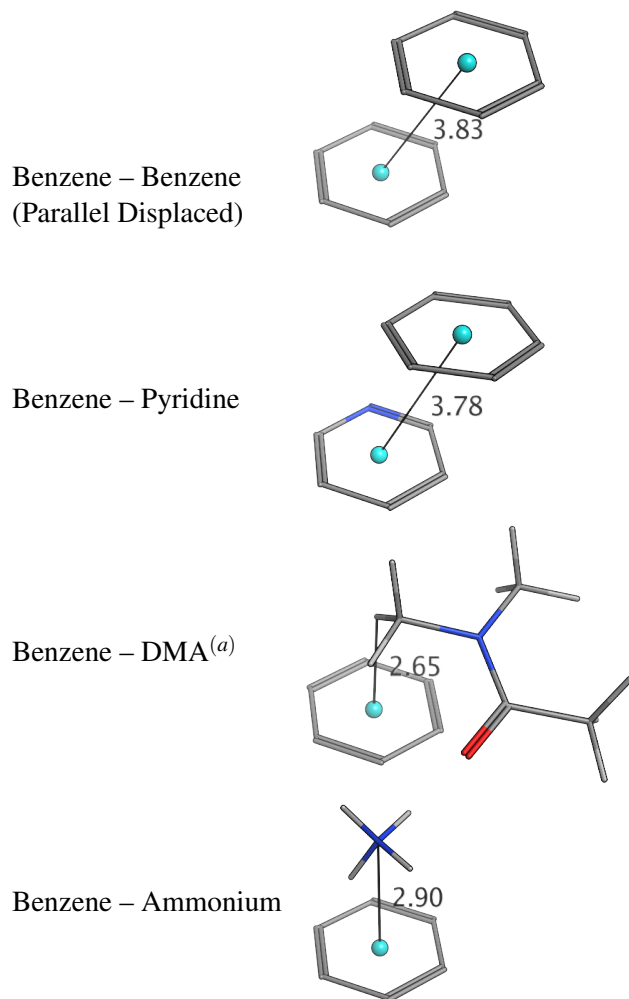


Figure 3: The test set 2 of PBE-D2/aug-cc-pVTZ geometry optimised systems for EDA (inter-molecular distances are given in Å)



<sup>(a)</sup> DMA is dimethylacetamide.

psinc basis set cutoff energy and 8 Bohr NGWF radii in the case of the ONETEP EDA<sup>1,2</sup>. Additionally, the Martyna-Tuckerman approach was used<sup>71,72</sup> in the ONETEP calculations to prevent interactions with periodic system images. Evidence of convergence of the calculations with respect to psinc basis set kinetic energy cutoff is provided in the Supporting Information section of this paper.

#### 4.1.2 Test Set 1: Hydrogen-bonding interactions.

In this test set, we assess the bonding components of important hydrogen bonded systems such as the water dimer, methanol – methanol, water – ammonia and water – ammonium (geometries shown in Figure 2). Studies were published that investigate the covalency of hydrogen-bonding in water,<sup>73–80</sup> and this covalency may be further analysed by considering the strength of the charge transfer component through EDA.

The ALMO EDA frozen density component  $\Delta E_{\text{FRZ}}$ , further partitioned into its constituent electrostatic, exchange and Pauli-repulsion parts, is shown in Figure 4. These constituent components follow the same definitions as in the LMO EDA and are in good agreement throughout the set. Additionally included in the ONETEP EDA is the correlation component that is required to form  $\Delta E_{\text{FRZ}}$  as shown in eq. 6.

We observe that the net effect of these frozen density components is almost negligible for all systems excluding the ammonium interacting case as shown by the small magnitudes of the  $\Delta E_{\text{FRZ}}$  component values ( $+0.1 \text{ kcal/mol} > \Delta E_{\text{FRZ}} > -0.9 \text{ kcal/mol}$ ). Intuitively, a more favourable  $\Delta E_{\text{FRZ}}$  component is expected for the water – ammonium system due to the strong electrostatic interaction of the charged ammonium molecule with the dipole of the water molecule. Our expectations are confirmed by assessment of the electrostatic energy: for this system we observe electrostatic contributions of  $-25.3 \text{ kcal/mol}$  and  $-25.5 \text{ kcal/mol}$  through the ONETEP and LMO EDAs respec-

---

<sup>1</sup>For all calculations excluding the benzene – ammonium system the polarisation stage NGWFs were initialised to the converged fragment NGWFs. For the benzene – ammonium system, re-initialisation to guess NGWFs at the polarisation stage was required in order to prevent under-convergence of the polarisation energy component.

<sup>2</sup>Grimme -D3 correction results are also included for comparison in Section SI.4 of the Supporting Information. The -D3 contributions were calculated using the DFT-D3 program as only the -D2 correction is currently implemented in ONETEP.

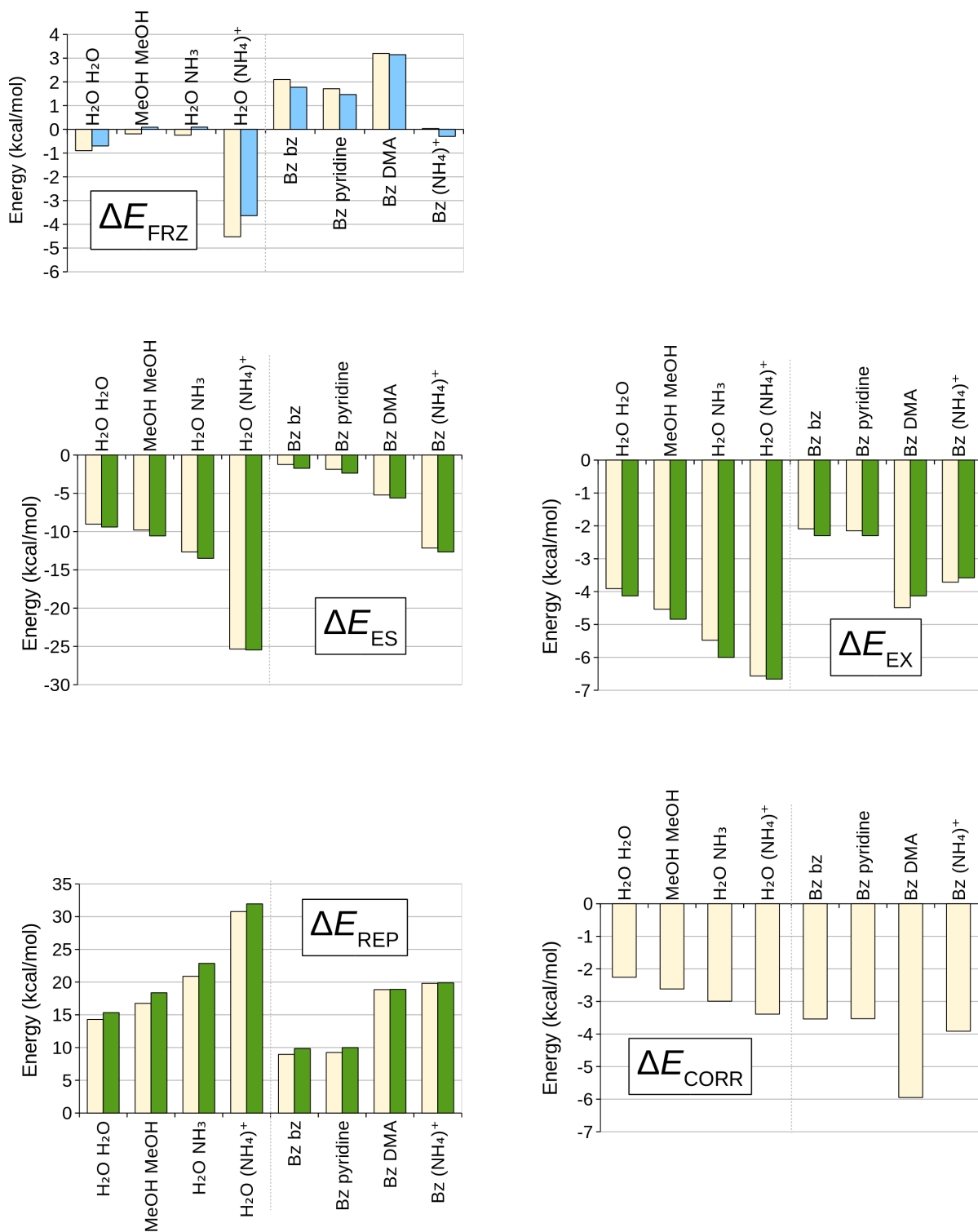


Figure 4: Frozen density analysis energy components for test sets 1 and 2. EDA component values are given in kcal/mol and were calculated with the PBE-D2 functional. ONETEP EDA calculations (yellow) were performed with a psinc basis set with a kinetic energy cutoff of 1200 eV. Q-Chem ALMO EDA calculations (blue) and GAMESS-US LMO EDA calculations (green) were performed with an aug-cc-pVTZ Gaussian basis set. The ALMO frozen density component is the sum of the electrostatic, exchange, Pauli-repulsion and correlation energy components as shown in eq. 6

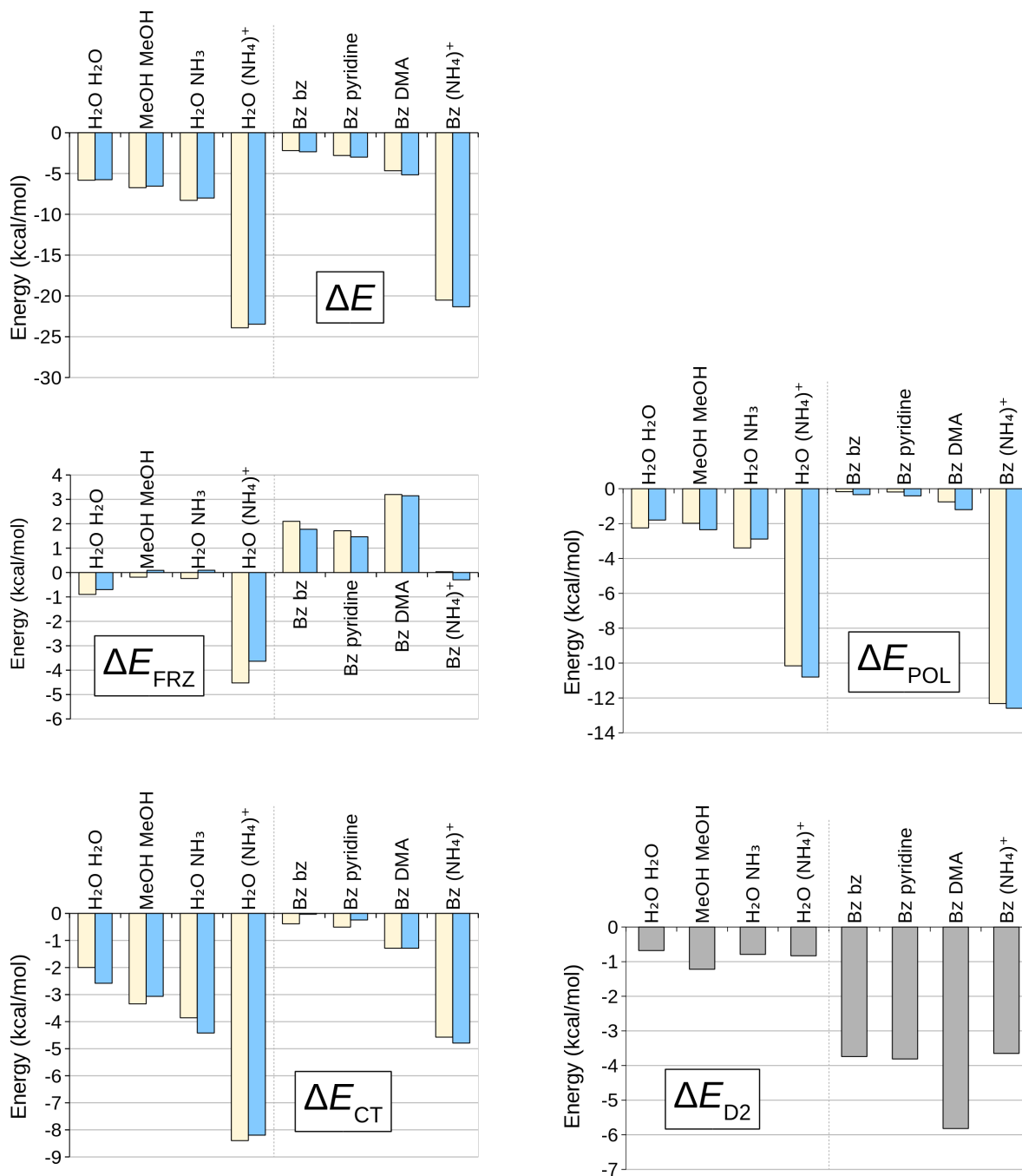


Figure 5: EDA energy components for test sets 1 and 2. EDA component values are given in kcal/mol and were calculated with the PBE-D2 functional. ONETEP EDA calculations (yellow) were performed with a psinc basis set with a kinetic energy cutoff of 1200 eV. Q-Chem ALMO EDA calculations (blue) were performed with an aug-cc-pVTZ Gaussian basis set. Values of Grimme's -D2 correction for dispersion are also provided (grey).

tively, whereas in the case of the uncharged systems we observe  $\Delta E_{\text{ES}} \leq -13.5$  kcal/mol. The increased effect of electrostatics in the water – ammonium system is also moderated significantly by a 9.9 kcal/mol (ONETEP) and 9.1 kcal/mol (LMO) increase in the Pauli-repulsion term compared to the water – ammonia system.

Exchange in the ONETEP EDA is observed to increase through the test set from -3.9 kcal/mol for the water dimer to -6.6 kcal/mol for the ammonium interacting system, in good agreement with the LMO EDA results. The correlation term is also observed to increase in magnitude through the test set from -2.3 kcal/mol for the water dimer to -3.4 kcal/mol for the ammonium interacting system. This observation is rationalised through the increasing electron counts of the systems through the set and decreasing intermolecular separations.

Polarisation and charge transfer energies obtained through the ALMO and ONETEP EDAs are shown in Figure 5. The LMO EDA does not attempt to distinguish polarisation from charge transfer effects and instead simply describes both as an inter- and intramolecular polarisation term. Again, the ALMO and ONETEP schemes show generally good agreement of the polarisation term, with the strength of polarisation increasing through the set to the ammonium interacting system (-10.2 kcal/mol (ONETEP) and -10.8 kcal/mol (ALMO)). Polarisation is shown to be similar in magnitude to the charge transfer contribution in the ONETEP EDA of the water dimer system, with a polarisation energy of -2.2 kcal/mol and a charge transfer energy of -2.0 kcal/mol. In the case of the ALMO EDA, charge transfer is observed to be much more significant (-2.6 kcal/mol) than polarisation effects (-1.8 kcal/mol). Despite these differences which are expected due to the different basis sets used in the ONETEP and ALMO EDAs, the values observed and inferences drawn for these components are generally in good agreement between the schemes.

#### **4.1.3 Test Set 2: $\pi$ -bonding systems.**

In this test set, we assess the energy components of a number of molecules interacting with the  $\pi$ -electron cloud of a benzene molecule. Specifically, we have investigated the interactions of benzene, pyridine, dimethylacetamide (DMA) and ammonium with benzene (geometries shown in

Figure 3).

We note overall good agreement between the ONETEP and LMO EDAs through the electrostatic, exchange and Pauli-repulsion energy components as displayed in Figure 4. We also observe good agreement for the ALMO EDA frozen density component that is partially constructed from these terms. Through the set, the contributions of electrostatics (-1.2 kcal/mol to -12.1 kcal/mol (ONETEP)), Pauli-repulsion (9.0 kcal/mol to 19.8 kcal/mol (ONETEP)), polarisation (-0.2 kcal/mol to -12.3 kcal/mol (ONETEP)) and charge transfer (-0.4 kcal/mol to -4.6 kcal/mol (ONETEP)) generally increase as expected with decreasing intermolecular separation and the presence of charged or larger molecules.

The largest correlation effects are observed for the benzene – DMA system (-5.9 kcal/mol). This compares to correlation energies of -3.5 kcal/mol for the benzene – benzene and benzene – pyridine systems, and -3.9 kcal/mol for the benzene – ammonium interacting system. This difference likely arises due to the small intermolecular separation (2.65 Å) of the DMA molecule compared to the other systems within the set (2.90 Å to 3.83 Å), and the larger number of interacting atoms. A similar trend is observed in the contribution of exchange and this can also be rationalised through similar considerations.

ALMO and ONETEP polarisation and charge transfer contributions for test set 2 are shown in Figure 5 and also display good agreement. The ability of DMA to delocalise charge through its structure appears to enhance polarisation to a degree, as shown by a polarisation energy of -0.8 kcal/mol for the ONETEP EDA and -1.2 kcal/mol for the ALMO EDA. Interestingly, polarisation is observed to increase markedly to -12.3 kcal/mol (ONETEP) and -12.6 kcal/mol (ALMO) in the case of the ammonium interacting system. This is likely the result of the placement of a positively charged molecule directly within the polarisable  $\pi$ -cloud of the partner benzene molecule.

Charge transfer is most significant in the DMA (-1.3 kcal/mol (ONETEP and ALMO)) and ammonium (-4.6 kcal/mol (ONETEP) and -4.8 kcal/mol (ALMO)) interacting systems. This is expected due to the ability of the DMA molecule to delocalise and stabilise incumbent charge, and due to the high favourability of the positively charged ammonium molecule to accept electron



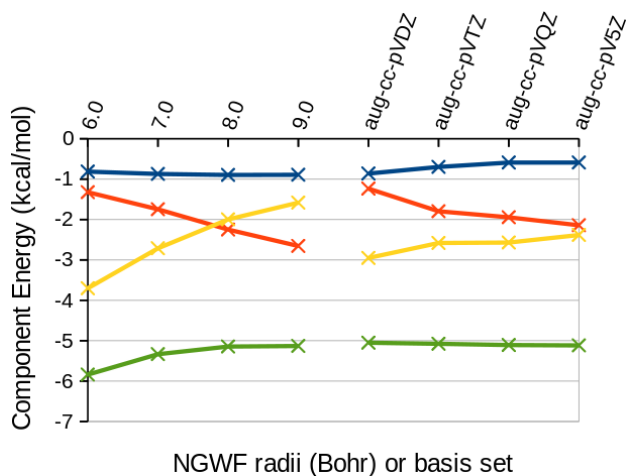
density. Charge transfer is especially strong in the ammonium interacting system, reflecting the consequence of placement of a charged molecule directly within the  $\pi$ -cloud of the partner benzene molecule as was also observed in the case of polarisation.

#### 4.1.4 Basis set dependence.

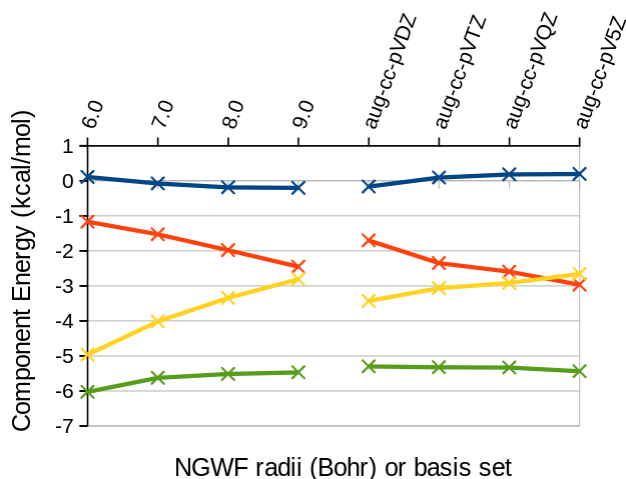
A notable weakness of using the SCF-MI approach to separate polarisation and charge transfer effects is its significant basis set dependency. As the basis set size increases, the polarised state increasingly becomes described using basis functions that extend into the space of the partner fragment and therefore this component increasingly includes charge transfer contributions. This ill-definition of the energy components is similar to the sensitivity<sup>81,82</sup> of Mulliken charges<sup>83</sup> to basis set size. Rigorous separation of the polarisation and charge transfer terms is of key importance for obtaining an accurate chemical description, and the presence of this ill-definition presents an issue. Plots demonstrating the basis set dependence of the frozen density, polarisation, and charge transfer components are provided in Figure 6.

This effect may be countered, and chemically accurate contributions can be obtained by adopting a consistent basis set choice of carefully considered spatial extension. In previous work the use of the aug-cc-pVTZ basis set for ALMO EDA calculations has been shown to offer a fair balance between accuracy and stability of terms.<sup>80,84</sup> Figure 6 shows that adopting between 7.5 Bohr to 8.5 Bohr NGWF radii is approximately equal to using the aug-cc-pVTZ basis set in the Gaussian code ALMO EDA. We suggest the use of 8.0 Bohr NGWF radii NGWFs in order to fairly balance the accuracy of  $\Delta E$  with accuracy in partitioning polarisation and charge transfer effects.

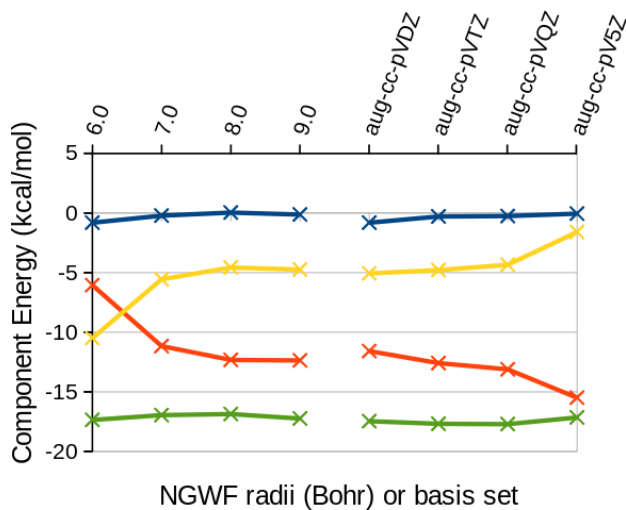
Ascribing polarisation and charge transfer contributions using approaches based on the ALMO methodology is not without challenge. Methods that seek to avoid this ambiguity include the CDFT approach of Wu that exploit population analysis calculations to partition polarisation and charge transfer by real space definitions,<sup>18</sup> and also the recent fragment electric-field response functions (FERF) approach of Horn and Head-Gordon<sup>85</sup> that adopts the use of fragment polarisation subspaces for variationally defining the polarisation contribution. These approaches have



(a)



(b)



(c)

Figure 6: The ONETEP (6 Bohr to 9 Bohr NGWF radii) and ALMO (aug-cc-pVDZ to aug-cc-pV5Z) EDA of PBE-D2/aug-cc-pVTZ geometry optimised (a) water dimer, (b) methanol - methanol, and (c) benzene - ammonium systems. The PBE/1200 eV level  $\Delta E$  (green) is partitioned into the frozen density (blue), polarisation (red), and charge transfer (yellow) energy components as described in eq. 2. The -D2 correction for dispersion of Grimme et al.<sup>70</sup> is independent of basis set size and so was not included for clarity.

shown success in isolating energy contributions with minimal basis set dependence, presenting the opportunity for the development of new EDA schemes with ever more chemically relatable energy components.

#### 4.1.5 Exchange-correlation functional dependence.

It is interesting to also compare the dependence of the EDA with respect to exchange-correlation functional. We have compared the results of the water – ammonium and benzene – DMA systems using the PBE<sup>69</sup> and BLYP<sup>86,87</sup> GGA exchange-correlation functionals and LDA (with VWN correlation<sup>88</sup>) functional using ONETEP with a 1200eV basis set kinetic energy cutoff. As Grimme’s correction for dispersion depends upon the choice of exchange-correlation functional by a simple applied scaling factor, we have omitted this correction from our discussion. The results of this comparison are displayed in Table 1 and Table 2.

Table 1: The water - ammonium ONETEP EDA energy components using a 1200eV basis set kinetic energy cutoff.

Energy term (kcal/mol)	Exchange-correlation functional		
	PBE	BLYP	LDA (VWN)
FRZ	-4.5	-3.4	-8.3
ES	-25.3	-25.7	-25.7
EX	-6.6	-4.9	-12.9
REP	30.8	31.4	32.4
CORR	-3.4	-4.2	-2.1
POL	-10.2	-10.2	-10.4
CT	-8.4	-8.0	-8.6
$\Delta E$	-23.1	-21.5	-27.4

As we may expect, the LDA results display generally poor agreement of  $\Delta E$  values with the GGA functionals. Analysing the EDA components, we observe that this is primarily due to differences in the  $\Delta E_{EX}$  components, for example this component is 11.1 kcal/mol lower for the LDA functional than the PBE functional in the case of the benzene – DMA system. This disagreement is partially cancelled by the  $\Delta E_{REP}$  and  $\Delta E_{CORR}$  components, giving a  $\Delta E_{FRZ}$  component energy 6.2 kcal/mol lower for the LDA functional with this system. Polarisation and charge transfer ef-

Table 2: The benzene - dimethylacetamide (DMA) ONETEP EDA energy components using a 1200eV basis set kinetic energy cutoff.

Energy term (kcal/mol)	Exchange-correlation functional		
	PBE	BLYP	LDA (VWN)
FRZ	3.2	6.8	-3.0
ES	-5.2	-5.4	-5.5
EX	-4.5	-0.6	-15.6
REP	18.9	19.2	21.2
CORR	-5.9	-6.5	-3.1
POL	-0.8	-0.9	-0.8
CT	-1.3	-1.3	-1.8
$\Delta E$	1.2	4.7	-5.6

fects are captured with surprising accuracy by the LDA functional for the two systems studied, with  $\Delta E_{POL}$  within 0.2 kcal/mol of the GGA functionals' values and  $\Delta E_{CT}$  within 0.6 kcal/mol.

There is good agreement between the GGA results, indicating that energy components within this class of functionals could be expected to be comparable. As with the LDA functional, differences using the PBE and BLYP functionals are apparent within the  $\Delta E_{FRZ}$  components, with the  $\Delta E_{FRZ}$  term 1.1 kcal/mol lower for PBE than BLYP in the case of the water – ammonium system and 3.6 kcal/mol lower in the case of the benzene – DMA system. Polarisation and charge transfer are in good agreement (0.4 kcal/mol) between the GGA functionals.

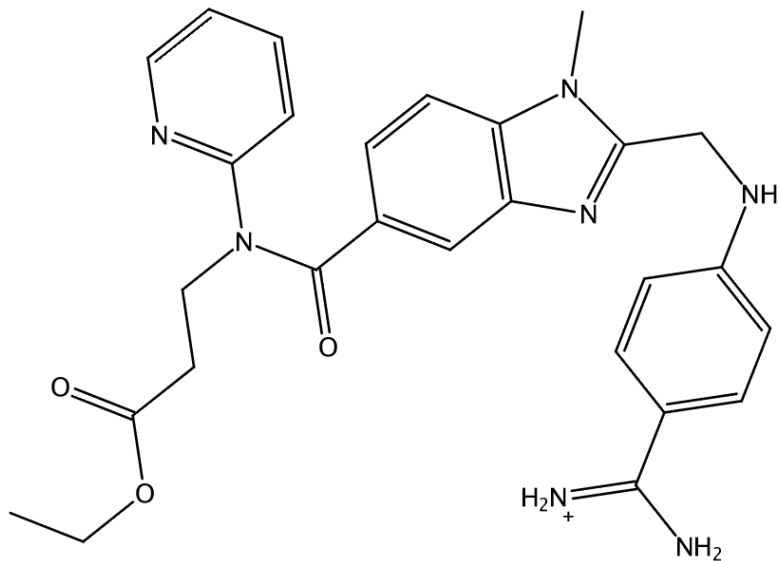
## 4.2 Demonstration of the method on an entire protein with 4975 atoms.

### 4.2.1 Calculation set-up.

Thrombin is a serine protease which plays a central role in the blood coagulation cascade.<sup>89</sup> Therefore, inhibition of thrombin is an approach towards the treatment and prevention of thrombosis. We have analysed the bonding components of thrombin-inhibitor complex models with varying degrees of protein truncation using the ONETEP EDA. A brief description of these systems' preparation follows.

Four model systems were prepared from the structure of the ligand C24 (shown in Figure 7) bound to thrombin (pdb:1KTS).<sup>90,91</sup> The initial thrombin-inhibitor crystal structure was capped

Figure 7: The C24 ligand (pdb:1KTS).



using  $\text{COCH}_3$  and  $\text{NHCH}_3$  groups<sup>3</sup> and protonated using the Protonate3d software.<sup>92</sup> The protein side chains, caps and hydrogens were then optimised using the MMFF94x forcefield with the generalised Born implicit solvent (GBIS) model. The hirudin and its waters were subsequently removed from the system. This structure (referred to as the untruncated system) was then truncated to remove residues beyond various distances (3Å, 9Å, 15Å) from the ligand. Caps ( $\text{COCH}_3$  and  $\text{NHCH}_3$  groups) were then added to the resulting protein fragments. The structures were then re-protonated and any bond clashes introduced by capping removed by further geometry optimisation (MMFF94x, GBIS) of the caps and the hydrogens. ONETEP EDA was performed on the resulting thrombin-C24 systems using a 800 eV psinc basis set cutoff energy and 8 Bohr NGWF radii at the PBE-D2 level of theory. A summary of the protein models and EDA fragment definitions is given in Table 3 and visualisation of the protein binding pocket in Figure 8.

#### 4.2.2 Results and discussion.

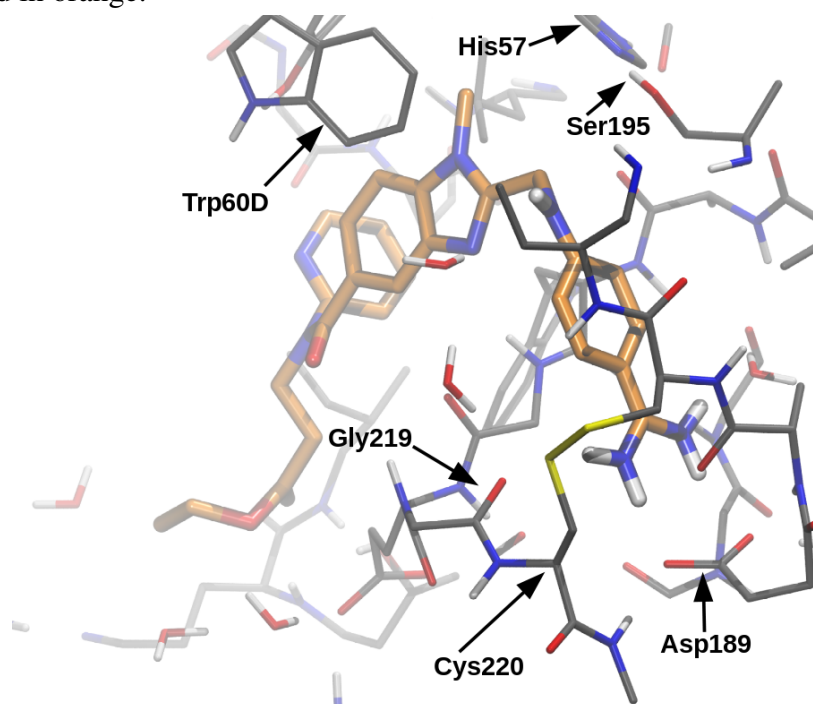
Values of the ONETEP EDA components and their convergence with respect to the untruncated protein system are shown in Figure 9 and Figure 10. The EDA components show a significant

<sup>3</sup>Caps were placed at the Glu1C, Tyr14J, Thr147, Gly150 and Leu264 residues.

Table 3: A summary of the fragments of the thrombin-C24 complexes used in the ONETEP EDA, where ‘∞’ refers to the full, untruncated protein (charges given in parenthesis).

Protein Truncation	Fragment atom count		
	Protein and solvent	Ligand	Supramolecule
3Å	374 (-1)	67 (+1)	441 (0)
9Å	1328 (-1)	67 (+1)	1395 (0)
15Å	2621 (+1)	67 (+1)	2688 (+2)
∞	4908 (+2)	67 (+1)	4975 (+3)

Figure 8: The binding pocket of the untruncated (4975 atom) protein system. The C24 ligand has been highlighted in orange.



electrostatic contribution (-150.3 kcal/mol) to  $\Delta E$ , reflecting the strong influence of the charged Asp189 residue interacting with the positively charged amidine of the ligand. The total contribution of exchange (-53.7 kcal/mol) and correlation (-63.6 kcal/mol) is shown to provide almost as much stability to binding as electrostatics. Stabilisation from electrostatics, exchange and correlation is however largely countered by steric Pauli-repulsion effects (+228.7 kcal/mol). Polarisation (-33.1 kcal/mol) is of similar magnitude to charge transfer (-31.2 kcal/mol), reflecting the balanced contribution of these effects to binding. A significant contribution to binding is provided by Grimme's -D2 correction for dispersion (-79.9 kcal/mol) which arises due to the many long-range

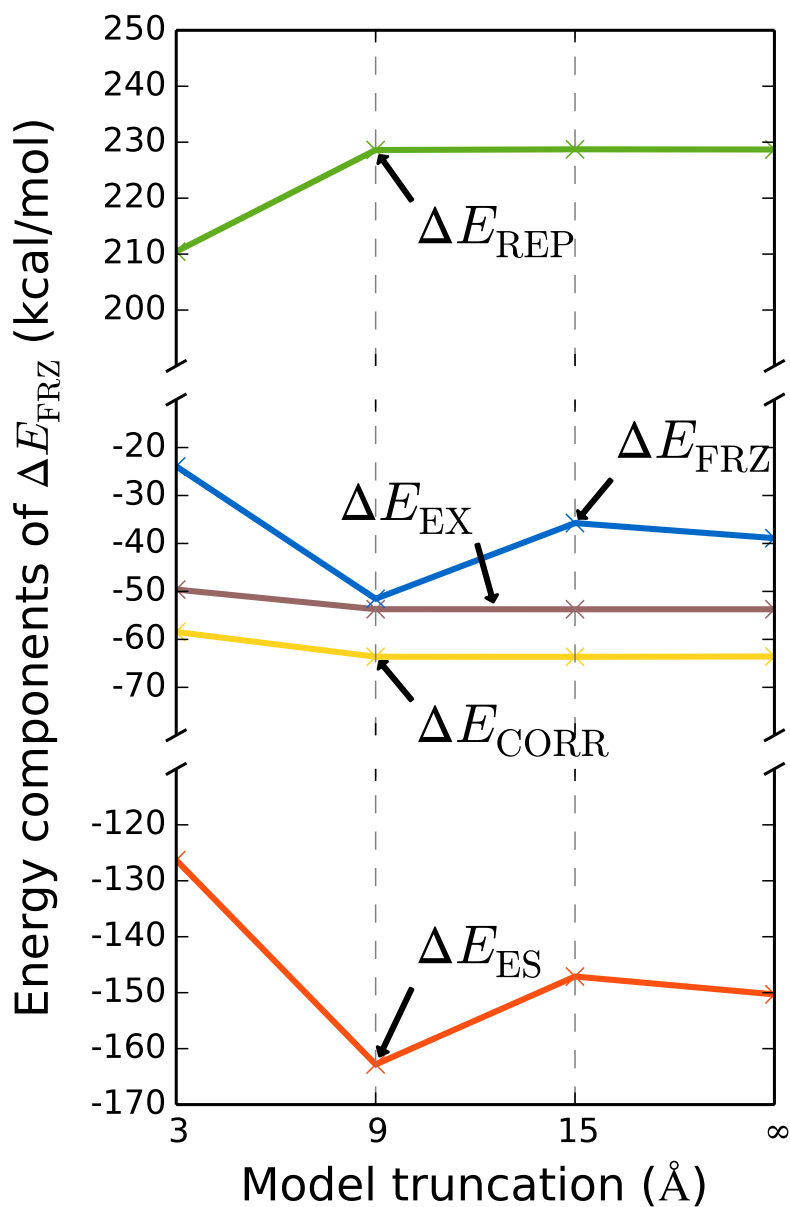


Figure 9: Frozen density analysis of the thrombin-C24 complexes calculated at the PBE-D2/800 eV level of theory (component values are given in kcal/mol). Error of the energy components is shown with respect to the full, untruncated (4975 atom) protein system. The frozen density component is formally equivalent to adding the electrostatic, exchange, Pauli-repulsion and correlation energy components as shown in eq. 6

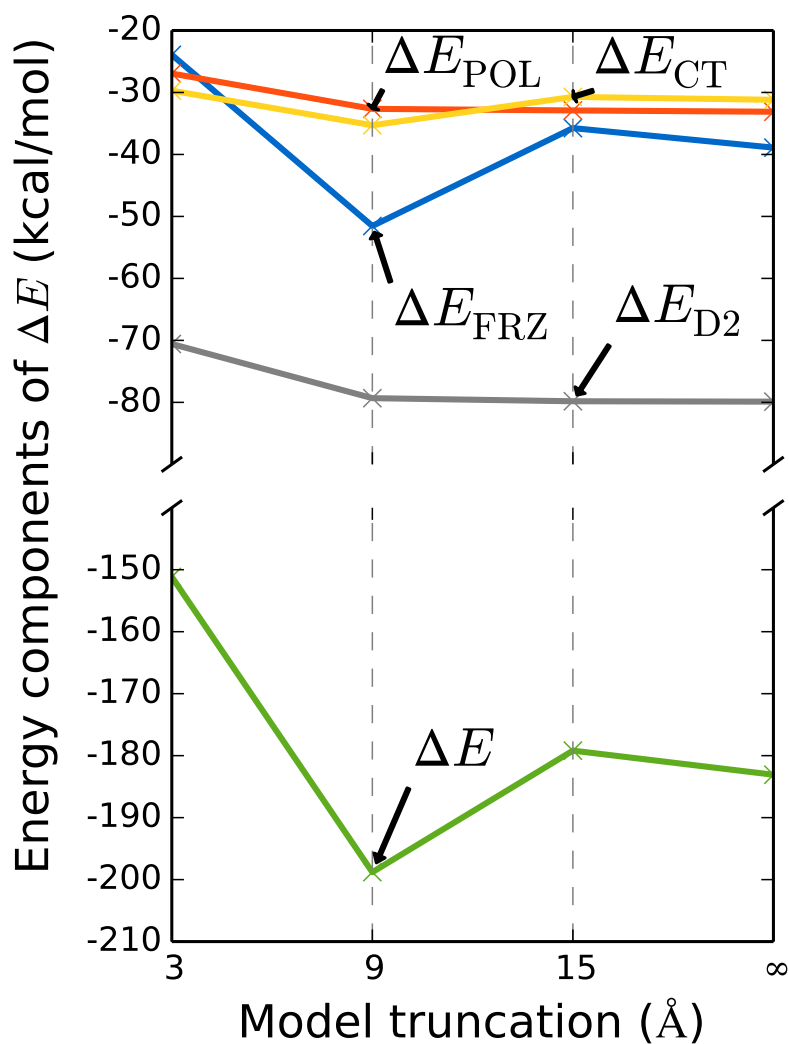


Figure 10: EDA components of the thrombin-C24 complexes calculated at the PBE-D2/800 eV level of theory (component values are given in kcal/mol). Error of the energy components is shown with respect to the full, untruncated (4975 atom) protein system. The full interaction energy  $\Delta E$  is equivalent to adding the frozen density component, polarisation, charge transfer, and Grimme -D2 dispersion energy components.



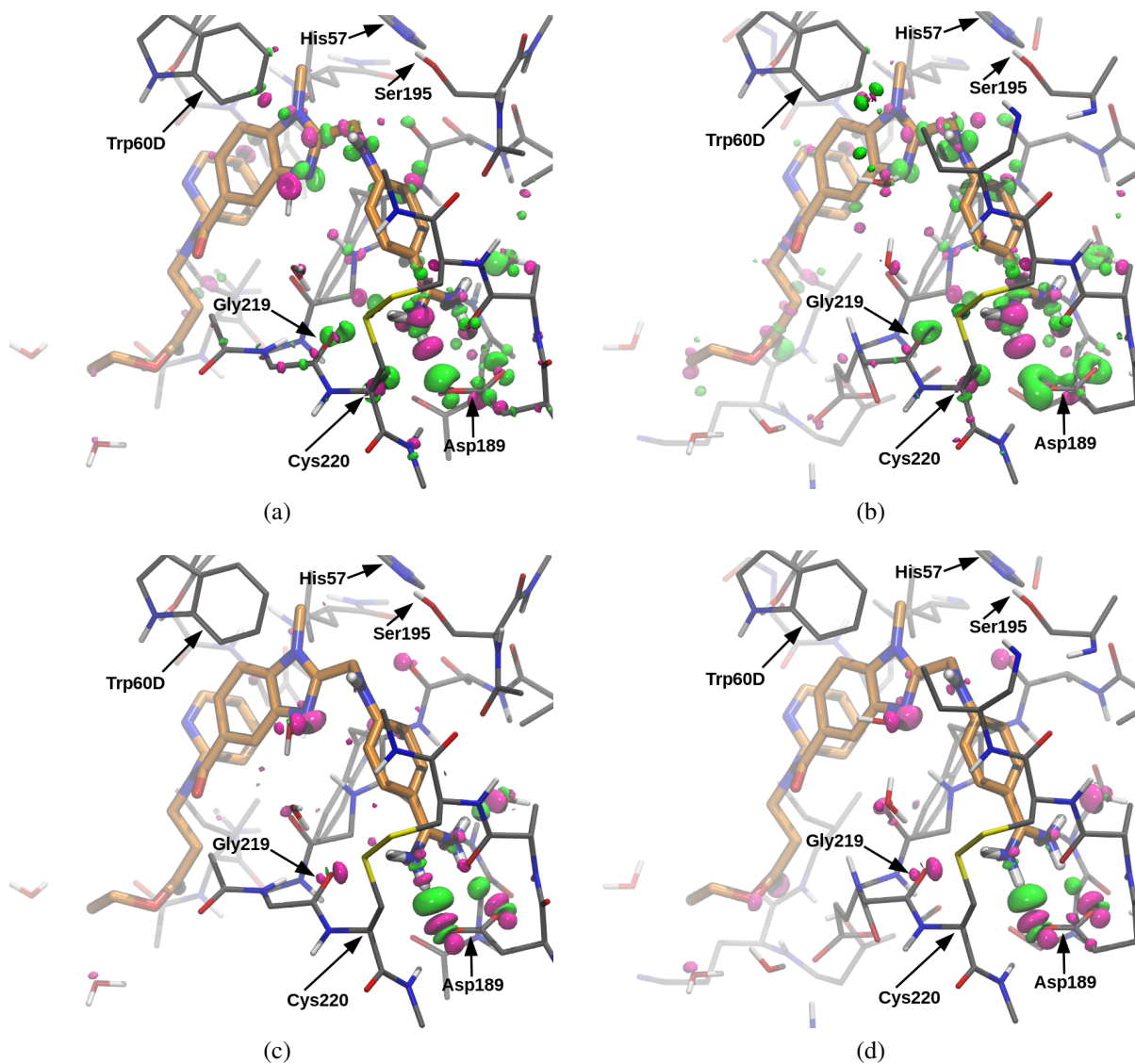
contributions over the large number of atoms in the protein.

Convergence of the exchange, Pauli-repulsion and correlation components is reached between 3Å and 9Å truncation, with full convergence of electrostatics observed between 15Å truncation and the full protein. The large truncation radius required for convergence of electrostatics is likely due to differences in the charge distributions within the truncated proteins and the long-range nature of electrostatics, and is interesting to observe as in molecular mechanics descriptions electrostatics are often truncated at around 9-10Å.<sup>93,94</sup> Polarisation and dispersion are shown to converge smoothly with reduced protein truncation, with charge transfer displaying a similar convergence profile to the frozen density component and electrostatics. Again, the similar convergence profile of charge transfer may be an indication of differences in the protein charge distributions at the different truncation levels. Overall, our results indicate heavy protein truncation to have the potential to adversely affect the outcome of the EDA values. Minimal truncation of protein models is therefore shown to be necessary to ensure convergence of the EDA components.

Qualitative descriptions of intra- and inter-fragment delocalisations are provided by electron density difference (EDD) plots<sup>37</sup> (Figure 11). EDD plots are constructed as electron density differences between intermediate states of the EDA procedure. The polarisation EDD is calculated by subtracting the electron density corresponding to the frozen density state from the polarised state electron density. Similarly, the charge transfer EDD is calculated by subtracting the electron density corresponding to the polarised state from that of the fully (electronically) relaxed state. Earlier work has demonstrated the high value of these plots in drug design, for example by application to HIV-1 protease inhibitors by Hensen et al.<sup>95</sup> and to  $\delta$ -opioid models by Mo et al.<sup>96</sup> using the BLW EDA.

The EDD plots of the 3Å truncated and untruncated systems both display the key polarisation and charge transfer interactions. Overall, polarisation is demonstrated as an important effect with charge transfer interactions involved in the formation of the many hydrogen bonds observed. Electron density is observed to gather in the carboxylate of Asp189 through polarisation effects, with bond formation to the amidine group of the C24 ligand being observed in the charge transfer EDD

Figure 11: EDD plots of polarisation for the (a) 3Å system and (b) untruncated system, and of charge transfer for the (c) 3Å system, and (d) untruncated system. The isosurface contour levels are displayed at 0.0175 electrons per cubic Angstrom with green surfaces representing electron gain and magenta surfaces representing electron loss. The C24 ligand has been highlighted in orange. The EDD plots show the contributions of functional groups to polarisation and charge transfer.



plot. The Gly219 and Cys220 residues of the protein are also shown to interact with the amidine group of the ligand. For these two residues minor differences are observed between the 3Å truncated and untruncated systems in the nature of the charge redistributions through polarisation effects. In addition to interactions of the protein with the ligand, notable interactions involving water molecules are also observed, specifically with the ligand’s amidine and 1-methyl-benzimidazole functional groups. These interactions highlight the importance of solvent participation in ligand-protein binding. Overall, the 3Å and infinite cutoff EDD plots (for both polarisation and charge transfer) appear qualitatively very similar. Thus the 3Å description is sufficient for qualitative and visual interpretations. For accurate energy components however, the 15Å description is needed as we saw in the previous section.

EDD plots are shown to be an important tool for visualising the key interactions that are responsible for the polarisation and charge transfer energies and allow us to associate these energies with particular functional groups. For example it can be seen that during the polarisation stage charge is redistributed in Asp189. Then during charge transfer, charge delocalises to the amidine group of the ligand, clearly showing the formation of bonding in this ion pair. Thus the value of such plots can be very high for assisting in the fine-tuning of drug-protein interactions.

## 5 Conclusions.

In this paper we have presented a new energy decomposition analysis (EDA) scheme based on the ALMO<sup>17</sup> and LMO<sup>16</sup> approaches that decomposes  $\Delta E$  into chemically meaningful electrostatic, exchange, correlation, Pauli-repulsion, polarisation and charge transfer energy terms. This combined method, implemented in the ONETEP linear-scaling electronic structure package,<sup>49</sup> is a fully additive scheme with the advantage of decomposing the ALMO EDA frozen density component into chemically distinct constituents, and is the first implementation of an EDA scheme in a psinc basis set code with strictly localised orbitals and plane wave basis set accuracy. With this new capability we were able to present EDA calculations performed on systems several thousands

of atoms in size, such as an entire protein-ligand complex.

We have validated the method in tests with hydrogen bonding and  $\pi$ -bonding systems and have shown good agreement with the equivalent Q-Chem and GAMESS-US energy components. The recognised sensitivity of partitioning polarisation and charge transfer with respect to basis set size has been shown to be present in our EDA implementation within ONETEP due to use of ALMOs. We suggest use of 8.0 Bohr NGWF radii when using the EDA within ONETEP to mitigate this effect, which we show to be approximately equivalent to use of the aug-cc-pVTZ basis set within the ALMO EDA. At this radius, convergence of the interaction energy  $\Delta E$  with respect to basis set size is achieved whilst also ensuring reasonable separation of charge transfer from polarisation effects.

The ONETEP EDA is capable of handling large systems many thousands of atoms in size, and we have demonstrated application of the scheme to a whole biomolecular protein system of thrombin with the C24 ligand with 4975 atoms. This work has also shown that in order to ensure proper convergence of the EDA values it is a necessary requirement to have only minimal truncation of protein-ligand systems. We found that truncation of the protein system has to be at a radius of 15 Å or larger for converged EDA values. Truncation at this cutoff distance leads to a system with 2688 atoms. Smaller truncations are only useful for visual qualitative studies as we show by EDD plots.

EDD plots are a powerful tool for identifying and visualising the key interactions that are responsible for the polarisation and charge transfer energies and allow us to associate these energies with particular functional groups. The combined EDA and EDD analysis is able to provide accurate quantitative and qualitative insights into the intermolecular bonding driving forces within systems of very large scale, and therefore the method would be highly suitable for drug design applications.

Future extensions of the code for intramolecular energy component partitioning may also be made possible through the appropriate manipulation of fragment electron densities. Isolating energy components in this manner has the potential to provide individual functional group contributions to energy components, in addition to the contributions of intermolecular origin as described

in this paper. Such an approach would be particularly suited to fragment-based techniques for drug design.

## Acknowledgement

M.P. thanks the BBSRC and Boehringer Ingelheim for an industrial CASE PhD studentship (BBSRC grant reference: BB/I015922/1). The authors are grateful for helpful discussions with Prof. Martin Head-Gordon. The calculations in this work were carried out on the Iridis4 Supercomputer of the University of Southampton, and the ARCHER supercomputer via the UKCP consortium (EPSRC grant reference: EP/K013556/1).

## Supporting Information Available

- Overview of the Stoll SCF-MI equations (SI.1).
- The Stoll SCF-MI equations in ONETEP (SI.2).
- The outer loop NGWF optimisation (SI.3).
- The calculated EDA component values (SI.4): PBE-D2/aug-cc-pVTZ (ALMO and LMO), PBE-D2/800 eV and PBE-D2/1200 eV (ONETEP) EDA values of test set 1 (Table S1); PBE-D2/aug-cc-pVTZ (ALMO and LMO), PBE-D2/800 eV and PBE-D2/1200 eV (ONETEP) EDA values of test set 2 (Table S2); PBE-D2/800 eV (ONETEP) thrombin-C24 EDA values (Table S3).

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

## References

- (1) Schwierz, N.; Frost, C. V.; Geissler, P. L.; Zacharias, M. *J. Am. Chem. Soc.* **2016**, *138*, 527–539.

- (2) Acquah, C.; Cagnetta, M.; Achenie, L. E. K.; Suib, S. L.; Karunanithi, A. T. *Ind. Eng. Chem. Res.* **2015**, *54*, 12108–12113.
- (3) Clark, T.; Hennemann, M.; Murray, J.; Politzer, P. *J. Mol. Model.* **2007**, *13*, 291–296.
- (4) Barik, A.; C, N.; Pilla, S. P.; Bahadur, R. P. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2738–2751.
- (5) Wu, E.; Qi, Y.; Park, S.; Mallajosyula, S.; Jr., A. M.; Klauda, J.; Im, W. *Biophys. J.* **2015**, *109*, 2090–2100.
- (6) Ye, R.-P.; Yang, J.-X.; Zhang, X.; Zhang, L.; Yao, Y.-G. *J. Mol. Struct.* **2016**, *1106*, 192–199.
- (7) Liu, X.; Dong, H.; Hu, D.; Shen, H.; Zheng, B.; Ling, B.; Bi, J. *Mol. Cryst. Liq. Cryst.* **2015**, *623*, 319–332.
- (8) Tafreshi, S. S.; Roldan, A.; de Leeuw, N. H. *Surf. Sci.* **2015**, *637-638*, 140–148.
- (9) Jasen, P. V.; Brizuela, G.; Padin, Z.; Gonzalez, E. A.; Juan, A. *Appl. Surf. Sci.* **2004**, *236*, 394–405.
- (10) Morokuma, K. *J. Chem. Phys.* **1971**, *55*, 1236–1244.
- (11) Stevens, W. J.; Fink, W. H. *Chem. Phys. Lett.* **1987**, *139*, 15–22.
- (12) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W. *J. Chem. Phys.* **1984**, *80*, 4378–4386.
- (13) Glendening, E. D.; Streitwieser, A. *J. Chem. Phys.* **1994**, *100*, 2900–2909.
- (14) Glendening, E. D. *J. Phys. Chem. A.* **2005**, *109*, 11936–11940.
- (15) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
- (16) Su, P.; Li, H. *J. Chem. Phys.* **2009**, *131*, 014102.
- (17) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 8753–8765.

- (18) Wu, Q.; Ayers, P. W.; Zhang, Y. *J. Chem. Phys.* **2009**, *131*, 164112.
- (19) Jeziorski, B.; Moszynski, R.; Ratkiewicz, A.; Rybak, S.; Szalewicz, K.; Williams, H. L. In *Methods and Techniques in Computational Chemistry: METECC-94*; Clementi, E., Ed.; STEF: Cagliari, 1993; Vol. B, Chapter 13, pp 79–129.
- (20) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (21) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.
- (22) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294–300.
- (23) Bagus, P. S.; Illas, F. *J. Chem. Phys.* **1992**, *96*, 8962–8970.
- (24) Glendening, E. D. *J. Am. Chem. Soc.* **1996**, *118*, 2473–2482.
- (25) Schenter, G. K.; Glendening, E. D. *J. Phys. Chem.* **1996**, *100*, 17152–17156.
- (26) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (27) Raupach, M.; Tonner, R. *J. Chem. Phys.* **2015**, *142*, 194105.
- (28) Cruz Hernández, N.; Zicovich-Wilson, C. M.; Fdez. Sanz, J. *J. Chem. Phys.* **2006**, *124*, 194105.
- (29) Philipsen, P. H. T.; Baerends, E. J. *J. Phys. Chem. B* **2006**, *110*, 12470–12479.
- (30) te Velde, G.; Baerends, E. J. *Phys. Rev. B* **1991**, *44*, 7888–7903.
- (31) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. *Chem. Soc. Rev.* **2015**, *44*, 3177–3211.
- (32) Ziegler, T.; Rauk, A. *Theor. Chem. Acc.* **1977**, *46*, 1–10.
- (33) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1558–1565.
- (34) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755–1759.

- (35) Mitoraj, M. P.; Michalak, A.; Ziegler, T. *J. Chem. Theory Comput.* **2009**, *5*, 962–975.
- (36) Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530–5538.
- (37) Mo, Y.; Bao, P.; Gao, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760–6775.
- (38) Wu, Q.; Van Voorhis, T. *Phys. Rev. A* **2005**, *72*, 024502.
- (39) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (40) Fedorov, D. G.; Nagata, T.; Kitaura, K. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562–7577.
- (41) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayashi, M. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- (42) Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 222–237.
- (43) Heifetz, A.; Chudyk, E. I.; Gleave, L.; Aldeghi, M.; Cherezov, V.; Fedorov, D. G.; Biggin, P. C.; Bodkin, M. J. *J. Chem. Inf. Model.* **2016**, *56*, 159–172.
- (44) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483–2490.
- (45) Bowler, D. R.; Miyazaki, T. *Rep. Prog. Phys.* **2012**, *75*, 036503.
- (46) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (47) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. *J. Chem. Phys.* **2006**, *124*, 204105.
- (48) Mo, Y.; Gao, J. *J. Phys. Chem. B* **2006**, *110*, 2976–2980.
- (49) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. *J. Chem. Phys.* **2005**, *122*, 084119.
- (50) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (51) Nagata, T.; Takahashi, O.; Saito, K.; Iwata, S. *J. Chem. Phys.* **2001**, *115*, 3553–3560.



- (52) Stoll, H.; Wagenblast, G.; Preuß, H. *Theor. Chim. Acta* **1980**, *57*, 169–178.
- (53) Mo, Y.; Peyerimhoff, S. D. *J. Chem. Phys.* **1998**, *109*, 1687–1697.
- (54) Gianinetti, E.; Raimondi, M.; Tornaghi, E. *Int. J. Quantum Chem.* **1996**, *60*, 157–166.
- (55) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251–266.
- (56) Horn, P. R.; Mao, Y.; Head-Gordon, M. *J. Chem. Phys.* **2016**, *144*, 114107.
- (57) Wu, Q. *J. Chem. Phys.* **2014**, *140*, 244109.
- (58) Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Diéguez, O.; Payne, M. C. *Phys. Rev. B* **2002**, *66*, 035119.
- (59) Mostofi, A. A.; Skylaris, C.-K.; Haynes, P. D.; Payne, M. C. *Comput. Phys. Commun.* **2002**, *147*, 788–802.
- (60) Mostofi, A. A.; Haynes, P. D.; Skylaris, C.-K.; Payne, M. C. *J. Chem. Phys.* **2003**, *119*, 8842–8848.
- (61) Haynes, P. D.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. *J. Phys.: Condens. Matter* **2008**, *20*, 294207.
- (62) Li, X.-P.; Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1993**, *47*, 10891–10894.
- (63) Nunes, R. W.; Vanderbilt, D. *Phys. Rev. B* **1994**, *50*, 17611–17614.
- (64) McWeeny, R. *Rev. Mod. Phys.* **1960**, *32*, 335–369.
- (65) Shao, Y. et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (66) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

- (67) Gabius, H.-J. *Pharm. Res.* **1998**, *15*, 23–30.
- (68) Řezáč, J.; Hobza, P. *J. Chem. Theory Comp.* **2012**, *8*, 141–151.
- (69) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (70) Grimme, S. *J. Comput. Chem* **2006**, *27*, 1787–1799.
- (71) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810–2821.
- (72) Hine, N. D. M.; Dziedzic, J.; Haynes, P. D.; Skylaris, C.-K. *J. Chem. Phys.* **2011**, *135*, 204103.
- (73) Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. *Phys. Rev. Lett.* **1999**, *82*, 600–603.
- (74) Hellemans, A. *Science* **1999**, *283*, 614–615.
- (75) Martin, T. W.; Derewenda, Z. S. *Nat. Struct. Mol. Biol.* **1999**, *6*, 403–406.
- (76) Ghanty, T. K.; Staroverov, V. N.; Koren, P. R.; Davidson, E. R. *J. Am. Chem. Soc.* **2000**, *122*, 1210–1214.
- (77) Rashin, A.; Topol, I.; Tawa, G.; Burt, S. *Chem. Phys. Lett.* **2001**, *335*, 327–333.
- (78) Barbiellini, B.; Shukla, A. *Phys. Rev. B* **2002**, *66*, 235101.
- (79) Beck, J. F.; Mo, Y. *J. Comput. Chem.* **2007**, *28*, 455–466.
- (80) Khaliullin, R.; Bell, A.; Head-Gordon, M. *Chem. Eur. J.* **2009**, *15*, 851–855.
- (81) Åstrand, P.-O.; Ruud, K.; Mikkelsen, K. V.; Helgaker, T. *J. Phys. Chem. A* **1998**, *102*, 7686–7691.
- (82) Cioslowski, J. *J. Am. Chem. Soc.* **1989**, *111*, 8333–8336.
- (83) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

- (84) Azar, R. J.; Horn, P. R.; Sundstrom, E. J.; Head-Gordon, M. *J. Chem. Phys.* **2013**, *138*, 084102.
- (85) Horn, P. R.; Head-Gordon, M. *J. Chem. Phys.* **2015**, *143*, 114111.
- (86) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (87) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (88) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (89) Stubbs, M. T.; Bode, W. *Thromb. Res.* **1993**, *69*, 1–58.
- (90) Huel, N. H.; Nar, H.; Priepke, H.; Ries, U.; Stassen, J.-M.; ; Wienen, W. *J. Med. Chem.* **2002**, *45*, 1757–1766.
- (91) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (92) Labute, P. *Proteins* **2009**, *75*, 187–205.
- (93) Izadi, S.; Aguilar, B.; Onufriev, A. V. *J. Chem. Theory Comput.* **2015**, *11*, 4450–4459.
- (94) Karkov, H. S.; Krogh, B. O.; Woo, J.; Parimal, S.; Ahmadian, H.; Cramer, S. M. *Biotechnol. Bioeng.* **2015**, *112*, 2305–2315.
- (95) Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Hölte, H.-D. *J. Med. Chem.* **2004**, *47*, 6673–6680.
- (96) Mo, Y.; Subramanian, G.; Gao, J.; Ferguson, D. M. *J. Am. Chem. Soc.* **2002**, *124*, 4832–4837.

## Graphical TOC Entry

