Comparison of overlap-based models for approximating the exchange-repulsion energy

Söderhjelm, Pär; Karlström, Gunnar; Ryde, Ulf

Published in:
Journal of Chemical Physics

DOI:
10.1063/1.2206182

2006

Document Version:
Peer reviewed version (aka post-print)

Comparison of overlap-based models for approximating the exchange–repulsion energy

Pär Söderhjelm, Gunnar Karlström, and Ulf Ryde

Department of Theoretical Chemistry,

Chemical Center, Lund University

P.O. Box 124,

S-22100 Lund

Sweden

(Dated: March 1, 2006)

Abstract

Different ways of approximating the exchange–repulsion energy with a classical potential function have been investigated by fitting various expressions to the exact exchange–repulsion energy for a large set of molecular dimers. The expressions involve either the orbital overlap or the electron-density overlap. For comparison, the parameter-free exchange–repulsion model of the Effective Fragment Potential (EFP) is evaluated.

The results show that exchange–repulsion energy is nearly proportional to both the orbital overlap and the density overlap. For accurate results, a distance-dependent correction is needed in both cases. If few parameters are desired, orbital overlap is superior to density overlap, but the fit to density overlap can be significantly improved by introducing more parameters. The EFP performs well, except for delocalised π systems. However, an overlap expression with a few parameters seems to be slightly more accurate and considerably easier to approximate.

PACS numbers:

I. INTRODUCTION

Nonbonded interactions play a central role in chemistry and biology. Among other things, they determine the properties of molecular solids and liquids, govern how proteins fold, and are responsible for biomolecular recognition. Accurate theoretical modelling of nonbonded interactions, preferably based on first principles, is therefore of uttermost importance.

The natural starting point when discussing nonbonded interactions is the quantum-mechanical description. In principle, an exact interaction energy can be obtained by solving the Schrödinger equation for the composite system and subtracting the energies of the isolated monomer molecules. There are two major problems with such a procedure. First, unless extremely large basis sets and a high level of theory is used, the obtained energies are not very accurate. The reason is, like in most applications of quantum chemistry, that the electron correlation energy converges very slowly with both basis set size and excitation level, and thus the dispersion interaction is very difficult to describe quantitatively. Only for very small systems may experimental results be reproduced by ab initio calculations [1].

This does not mean that quantum-chemical calculations are useless to model intermolecular interactions. In some cases, the accuracy is good enough for the purpose and in other cases, the quantum-mechanical model can be used as a good starting point, to which corrections may be added. But then the second problem remains. Even with a moderate basis set and the lowest possible level of theory, quantum chemistry is limited to systems with at most a few hundred atoms. If sampling of the dynamics of the system is needed, the number of atoms that may be modeled decreases even further.

The only solution to this problem is to use a classical force field. Such force fields do not even attempt to calculate the quantum-mechanical wavefunction explicitly, but provide approximate energies and forces, which ideally describe the real physical system. The gain in computational cost when discarding the wavefunction optimisation is so substantial that both handling of large systems and vast sampling is feasible.

The problem with most current force fields are that they are either general but rather inaccurate, or they are accurate only for the types of systems and properties for which they were designed. Whereas the latter type of potentials may be extremely useful in many applications, there are cases when a more general potential is desired, e.g. when predicting energies for systems that have not been examined experimentally. The consensus view

is that with the simple nonbonded potential used in most general-purpose force fields (a Coloumb term and a Lennard–Jones term for each atom–atom pair), we start to reach the best possible accuracy for general systems. Therefore, a more accurate general force field must necessarily contain more elaborate expressions.

The total intermolecular interaction energy of a molecular dimer is often divided into four main contributions [2]: electrostatic energy, induction energy, dispersion energy, and exchange–repulsion energy. The long-range electrostatic contribution can be systematically improved by including higher levels of distributed multipoles (atomic charges, dipoles, quadrupoles, etc.). Similarly, the long-range induction and dispersion energies can in principle be improved by increasing the accuracy of the distributed dipole polarisabilities, although one is in those cases limited by approximations in the models.

It is much more difficult to improve the short-range interaction, by which one usually refers to the additional interaction terms that arise when the overlap between the wavefunctions of the two molecules is significant. It includes overlap corrections to the induction and dispersion energies, but the two primary sources of short-range interaction are the exchange–repulsion energy and the charge penetration energy. The former is a pure quantum-mechanical effect which will be described below, whereas the latter is simply the part of the classical electrostatic energy that cannot be described by interacting distributed multipole moments. It is well known that both these terms formally scale as the square of the overlap. Therefore, they have usually been collected into a rest term, defined as the difference between the Hartree–Fock supermolecular energy and the sum of electrostatic and induction energies, to which some expression with exponential functions of the atom–atom distances is fitted [3]. This may certainly be a good procedure in order to minimise the error in a classical force field, but it does not provide a solid base for understanding the origin of the error. Another problem is that the use of simple exponential functions, in view of the increasing computational resources available, may be a crude approximation that is not necessarily needed.

In this study, we address both of these issues. First, we use the exact exchange–repulsion energy at the Hartree–Fock level as the reference. Second, we do not assume an exponential form of the potential surface, but instead make use of quantities calculated explicitly for each interaction geometry from the monomer wavefunctions, namely the orbital overlap and the electron-density overlap. If good fits can be found, using these quantities as independent

variables and preferably with few fitted parameters, this is a way to quantitatively predict the exchange–repulsion energy when the monomer wavefunctions are known, without having to perform computationally expensive supermolecular calculations.

Several such exchange–repulsion models have been used before [4–8], among which one deserves extra attention [7], because it is, to our knowledge, the only exchange–repulsion model in use that does not contain any fitted parameters. We here investigate the performance of some of these exchange–repulsion models as well as some new suggested models for a wide range of intermolecular interactions. Our aim is to provide an unbiased comparison and to determine the expected accuracy of overlap models when applied to new systems.

II. THEORY AND METHODS

A. Exchange–repulsion energy

The Hartree–Fock exchange–repulsion energy is often divided into an exchange contribution and a repulsion contribution [9]. The former is always attractive and can be derived from Hartree–Fock theory as the sum of exchange contributions from orbitals on different molecules. The latter is repulsive and usually larger in absolute value, thus causing a net repulsion. Its origin can be qualitatively understood [10] by considering two approaching one-electron systems A and B, with like spin, described by wavefunctions ψ_A and ψ_B. The total wavefunction must be antisymmetric with respect to electron exchange and is therefore written as a Slater determinant:

$$\Psi(\mathbf{r_1}, \mathbf{r_2}) = \frac{1}{\sqrt{2-2S^2}}[\psi_A(\mathbf{r_1})\psi_B(\mathbf{r_2}) - \psi_A(\mathbf{r_2})\psi_B(\mathbf{r_1})] \tag{1}$$

where S is the overlap integral $< \psi_A|\psi_B >$. Taking the square and integrating over $\mathbf{r_2}$ gives the electron density:

$$\rho(\mathbf{r_1}) = \frac{1}{1-S^2}[|\psi_A(\mathbf{r_1})|^2 + |\psi_B(\mathbf{r_1})|^2 - 2S\psi_A(\mathbf{r_1})\psi_B(\mathbf{r_1})] \tag{2}$$

The third term indicates a depletion of electron density in the overlap region, which causes the repulsion. Thus, the repulsion is mainly electrostatic in origin, although there are also contributions from changes in the exchange energy upon perturbing the density (which makes the division of exchange–repulsion into exchange and repulsion somewhat misleading).

It is easiliy observed from Eq. 2 that if ψ_A and ψ_B are orthogonal, there is no change in the electron density and the repulsion vanishes. This carries over to many-electron systems and provides a more convenient definition of the repulsion energy as the cost (in Hartree–Fock terms) of orthogonalising the molecular orbitals of one molecule against the molecular orbitals of the other.

Computationally, the most practical approach to obtaining the zeroth-order exchange–repulsion energy between molecules A and B is by subtracting the monomer energies and the electrostatic interaction energy from the Heitler–London energy in a supermolecular Hartree–Fock calculation of the A–B dimer:

$$E_{exch} = \frac{< \Psi_A\Psi_B|\mathcal{A}(H_A + H_B + V)|\Psi_A\Psi_B >}{< \Psi_A\Psi_B|\mathcal{A}|\Psi_A\Psi_B >} - < \Psi_A|H_A|\Psi_A > \\ - < \Psi_B|H_B|\Psi_B > - < \Psi_A\Psi_A|V|\Psi_B\Psi_B > \quad (3)$$

Here, the supermolecular Hamiltonian has been split into the Hamiltonians for the monomers, H_A and H_B, and an interaction term V. The Hartree–Fock wavefunctions of the isolated monomers are denoted by Ψ_A and Ψ_B, and $\mathcal{A}$ is the antisymmetriser.

The Heitler–London energy is obtained in the first Hartree–Fock iteration if one uses the product of the two isolated wavefuntions as starting guess. The electrostatic energy can be obtained in the same type of calculation but skipping orthogonalisation and not including the exchange terms. This is equivalent to a Kitaura–Morokuma decomposition [11]. The exchange–repulsion energy is much less sensitive to basis-set superposition error (BSSE) than the polarisation and charge-transfer terms, because only the occupied orbitals are allowed to mix in the orthogonalisation of the monomer wavefunctions. Nevertheless, the effect has been studied [12] and it has been proposed to use the dimer-centered basis set in the calculations of the isolated monomer wavefunctions. However, the BSSE was considered a minor effect in this study, and, except for one set of calculations where the dimer-centered basis set is specifically mentioned, the monomer-centered basis sets were used in the monomer calculations.

B. Modeling

Approximative methods for calculating the exchange–repulsion energy, given the Hartree–Fock wavefunctions of the isolated molecules, are usually based on calculating either the

squared orbital overlap (which will be refered to simply as orbital overlap)

$$S^2 = \sum_i^A \sum_j^B < \psi_i | \psi_j >^2 \tag{4}$$

where ψ_i and ψ_j are occupied molecular orbitals (MOs) of the isolated monomers A and B, respectively, or the density overlap

$$\Omega = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r} \tag{5}$$

where ρ_A and ρ_B are the electron densities of the isolated monomers. The orbital overlap has greater physical significance, becauses it arises naturally in any (approximate) derivation of exchange–repulsion energy (see for example Ref. 13 and 14). It has been used for approximating the exchange–repulsion [4, 5], sometimes modified by introducing an explicit distance dependence [6, 15]. Another model [4] introduces the energy-weighted orbital overlap

$$S^2_{en} = \sum_i^A \sum_j^B -\frac{\epsilon_i \epsilon_j}{\epsilon_i + \epsilon_j} < \psi_i | \psi_j >^2 \tag{6}$$

where ϵ_i is the orbital energy of ψ_i. This expression is appealing because it takes into account the greater ability of more diffuse orbitals (with higher orbital energy) to redistribute their electrons in response to interaction with another molecule.

Orbital overlap is also the base of a more elaborate exchange–repulsion model [7] included in the Effective Fragment Potential (EFP) method [16]. This model contains no fitted parameters, but starts from an exact expansion of the exchange–repulsion energy in terms of wavefunction overlap truncated after the quadratic term. Then, further approximations are introduced, some of which are exact with an infinite basis set and others that are not. For example, the spherical gaussian overlap approximation [17] assumes certain properties of the molecular orbitals. The EFP model has been carefully tested, but only for a few small model systems [18].

To use the density overlap must be considered as a further approximation. Nevertheless, it has been succesfully applied in many studies ever since its ability to model atomic interactions was demonstrated [19]. It has been used to construct force fields [8, 20], but has also been tested thoroughly for reproducing the exchange–repulsion energy in hydrogen-bonded dimers [21]. There, it was shown that a simple proportionality between the exchange–repulsion energy and the density overlap gave reasonable results, and that weighting differently the contributions from heavy atoms and hydrogen atoms did not improve the results

significantly, although the coefficient for hydrogen was always larger. A series of anisotropic terms was also tested, but was shown to converge very slowly if more terms than the isotropic one was included.

Computationally, the two quantities (S^2 and Ω) can be obtained with a similar effort. The orbital overlap is easier to compute analytically, because relatively few integrals over Gaussian basis functions need to be evaluated, once the monomer wavefunctions are known. Calculating the density overlap involves a larger number of integrals, which makes the exact analytical expression intractable, except for small molecules. One way to reduce the amount of work is to collect contributions to the molecular density from several pairs of basis functions into single Gaussian functions[22], which is conceptually similar to density fitting [23]. Another way, which we have chosen in this study, is to use numerical integration.

The main advantage with using density overlap is that it can easily be further approximated, because it involves a simple scalar function of space, the electron density. One such approximation, which will be outlined below, leads to a simple sum of atom–atom exponential terms, which can be used directly for repulsion (Buckingham potential) or further approximated by a r^{-12} potential, which is one of the most common functional forms in force fields. However, as a side product, this approximation provides a way to divide the total density overlap into atom–atom contributions, in order to allow for different proportionality constants for certain element pairs. This is the only use of the approximation in this study.

To derive the contributions, one starts by approximating the electron density of molecule A as

$$\rho_A(\mathbf{r}) \approx \sum_i^A \kappa_i \exp(-\alpha_i |\mathbf{r} - \mathbf{R_i}|) \tag{7}$$

where i is an atom. In this expression, κ_i, α_i, and $\mathbf{R_i}$ are considered as parameters for each individual atom and are fitted to the density obtained in the monomer calculation. Using Eq. 7, the density overlap can be written as a sum of integrals over pairs of Slater functions. Each such integral can be reasonably well approximated by an exponential function of the interatomic distance. Thus, the overlap can be written as a sum of atom–atom exponential terms:

$$\Omega \approx \sum_i^A \sum_j^B \Omega_{ij} \tag{8}$$

$$\Omega_{ij} = \kappa_i \kappa_j K_{ij} \exp(-A_{ij} |\mathbf{R_i} - \mathbf{R_j}|)$$

where A_{ij} and K_{ij} are functions[31] of α_i and α_j, obtained by fitting each exponential to the numerically evaluated integral for relevant distances and α values. The contribution to the total density from a pair of elements a and b can then be written as

$$x_{ab} = \frac{\sum_i^a \sum_j^b \Omega_{ij}}{\Omega} \tag{9}$$

where the sums are over all atoms of those elements.

C. Selection of molecular dimers

The 22 molecules selected for this study represent all the functional groups occuring in proteins, as well as some other common functional groups in organic chemistry. The molecules were geometry-optimised and then combined into molecular dimers to form a large data set. In order to simplify the interpretation of the results, the dimers were divided into six groups, for which the errors will be reported separately. A summary of which molecular dimers that were included in these groups is found in Table I. All the geometries in the data sets are available from the authors upon request. We encourage other researchers to use these as a test set for other methods.

Dimer geometries for the various groups were obtained in different ways which are briefly described in the following.

- Molecule–water dimers. For each organic molecule, several interactions with a single water molecule were manually constructed and then geomtry-optimised. All selected interactions are attractive. The majority contain a normal hydrogen bond, but O–H..π and C–H..O interactions were also included. For each of these dimers, 11 structures were created by changing the distance between the two main interacting atoms while keeping all other degrees of freedom fixed at the optimised values. In order to safely cover the attractive part of the intermolecular potential, the structures had intermolecular distances between 0.4 Å inside and 2.0 Å outside of the optimised distance.

- Molecule–molecule dimers, in which none of the molecules are water. Apart from this, they were obtained in the same way as the molecule–water dimers. Every interaction contained at least one hydrogen bond.

- Water–water dimers. A standard molecular dynamics (MD) simulation of water, using the TIP3P model for water, was performed and 200 water dimer structures were picked out from an equilibrated snapshot. Only dimers having one atom–atom distance smaller than 2 Å were selected. To avoid biasing from the TIP3P model, the intermolecular distance was varied, producing seven geometries for every dimer structure.

- Ethanethiole–water, toluene–water, and propionamide–water dimers. For each of these groups, a standard MD simulation of one rigid molecule in explicit solvent were performed, using the TIP3P model for water and electrostatic potential derived point charges from a HF/6-31G* calulation for the single molecule. From snapshots of the simulation, molecule–water structures were picked out. The procedure for picking out structures was biased so that, in the case of ethanethiol and propionamide, only interactions with the polar part of the molecule were taken, whereas in the toluene case, only interactions with a hydrogen bond to the π cloud of the aromatic ring were selected. The latter choice was because these interactions have proven difficult to model accurately.

D. Fitting

All expressions investigated in this study, except for the EFP repulsion, contains unknown parameters that need to be fitted in order to reproduce the exact exchange–repulsion energy. To ensure that the results are comparable, all fits were performed in the same way. All the data groups described above were used in the fits (except in one fit that was used to verify transferability). Each group contributed equally in the fitting, irrespectively of the number of geometries it included. All geometries with an exchange–repulsion energy larger than 40 kJ/mol were discarded from the fit. Apart from that, all geometries in a group were given the same weight. The mean absolute error was used as error measure and it was minimised without constraints on the parameter values. Because most of the expressions were non-linear, the Powell method [24] was used for fitting. The algorithm was started with several initial values in order to make sure that the global minimum was found. Local minima were only observed in the fit using Eq. 21, and because they seemed to have approximately the same errors and the same qualitative features of the parameters, one arbitrary minimum

was selected.

E. Computational details

All calculations for obtaining energies and wavefunctions were performed at the Hartree–Fock level of theory. The cc-pVTZ basis set [25, 26] was used, except in the calculations involving the effective fragment potential method, for which aug-cc-pVTZ had to be used instead (see below). The reason for using a correlation-consistent basis set in calculations without correlation is that we wanted to use the same basis set as in a related study concerned with correlated calculations (in preparation). We have confirmed that the same qualitative results are obtained with Pople basis sets. All geometry optimisations were performed using the B3LYP/6-311+G(2d,2p) method.

The orbital overlap was calculated using the expansion of the monomer molecular orbitals into atomic basis functions already applied in the SCF calculations of the monomers. Eq. 4 has the advantage of being invariant to orthogonal transformations of the molecular orbitals. This is not the case with Eq. 6. The results presented here were obtained by using canonical Hartree–Fock orbitals in Eq. 6, but the effect of using localised molecular orbitals (LMOs) was tested for the toluene–water complexes and was found to be minor.

The density overlap was calculated by numerical integration with simple rectangular quadrature. The electron density of the isolated molecule was evaluated on a grid of approximate size 100x100x100 for all considered molecules. Tests with larger grids showed that the error in the numerical integration was less than 2%.

The Effective Fragment Potential (EFP) estimates of the exchange–repulsion energy were obtained by using the default implementation in Gamess. As suggested in the derivation [7], localised molecular orbitals were used. The EFP method was found to be more sensitive to a limited basis set than the orbital overlap itself. In particular, we found that addition of more diffuse functions (as in the aug-cc-pVTZ basis set) were needed to obtain converged EFP results. Because the improvement was substantial (error reduced by 60%), only aug-cc-pVTZ results are considered here. Interestingly, the same change in basis set did not influence the fits involving overlaps (provided of course that the same basis set was used in the calculation of the exchange–repulsion energy).

Several quantum-chemical softwares were used for the various calculations. Turbomole

5.6 [27] was used for geometry optimisations. A modified version of MOLCAS 6 [28] was used for calculating the exchange–repulsion energies and orbital overlaps, whereas Gaussian 98 [29] was used for obtaining the monomer electron densities. All calculations involving the EFP method were performed using Gamess (2004) [30].

III. RESULTS AND DISCUSSION

In order to understand the basic behaviour of exchange–repulsion energy, we started to consider the simplest possible case, namely the helium dimer. The He...He interaction energy at the HF/cc-pVTZ level partitioned into three components is shown in Figure 1, together with the density overlap (Ω) and the orbital overlap (S^2), linearly scaled in order to reproduce the exchange–repulsion energy at large separations. Clearly, both Ω and S^2 are proportional to the exchange–repulsion energy at large separations. At smaller separations, Ω is too repulsive, whereas S^2 is not sufficiently repulsive.

One way of capturing this behaviour is to introduce a multiplicative factor $1+ke^{-R}$, where R is the interatomic distance divided by a reference distance that was arbitrarily chosen to be equal to the Bohr radius (0.529 Å). Choosing another reference distance changes the parameter k but does not influence the resulting fit significantly. The distance-dependent factor goes to unity for long distances, so that proportionality is ensured. In all other respects the choice of an exponential function was arbitrary. Fitting the exchange–repulsion energy curve in Figure 1 to the orbital overlap and the density overlap, respectively, gave the following result:

$$E_{exch} \approx \begin{cases} 0.85S^2(1 + 10.2e^{-R}) \\ 11.0\Omega(1 - 3.1e^{-R}) \end{cases} \tag{10}$$

The mean absolute errors (over 20 interaction distances between 1.6 Å and 2.5 Å) with these expressions were 0.01 kJ/mol and 0.05 kJ/mol, respectively. This excellent agreement could be expected because only one type of interaction was considered. However, these trivial observations raises the question wheather one can, using similar expressions, find parameters that are generally applicable to a large range of intermolecular interactions. To answer this question, we have performed fitting for the large data set of molecular interactions described in section II.

First, fitting to the orbital overlap was considered. Assuming proportionality gave

$$E_{exch} \approx 0.92 S^2 \tag{11}$$

The errors for this and the following fits are collected in Table II. The largest mean absolute error is observed for the water–water dimers, for which the deviation is systematic and similar to the He...He case.

Using the energy-weighted orbital overlap of Eq. 6 we instead obtained

$$E_{exch} \approx 2.70 S_{en}^2 \tag{12}$$

The errors have been reduced for most groups of complexes, but it has increased significantly for the toluene–water complexes. Regarding the energy-weight as a correction to the orbital overlap and optimising the exponent of this correction, we saw that for each group we can actually find an exponent that gives an error under 1 kJ/mol, but the value of the optimal exponent varies considerably. In particular, it was 0.01 for the toluene–water complexes but close to 1.0 for the other groups. Therefore, we did not use the energy-weighted orbital overlap further.

Using a distance-dependent correction, as in the He...He case, we obtained

$$E_{exch} \approx 0.80 S^2 (1 + 20 e^{-R}) \tag{13}$$

where R is an effective interaction distance for the entire dimer, somewhat arbitrary defined as $R = S^2 / (\sum_i^A \sum_j^B < \psi_i | \psi_j >^2 / R_{ij})$ where R_{ij} is the distance between the centers of LMOs i and j, again divided by the Bohr radius. The distance-dependent expression gave significantly improved results, showing that the observations in the He...He case are indeed quite general. The fitted parameters differ from those of Eq. 10, which can be expected because of the limited amount of data in the He...He case. We also tried the expression that is the basis for the model used in the SIBFA force field [6]. Only the functional form was adopted from SIBFA (in a much simplified form without distinguishing between lone pairs and bonds, etc.), whereas the parameters were fitted by us. The result was

$$E_{exch} \approx \sum_i^A \sum_j^B < \psi_i | \psi_j >^2 \left(\frac{5.6}{R_{ij}} - \frac{5.2}{R_{ij}^2} \right) \tag{14}$$

with errors similar to those with Eq. 13. Apparently, the number of fitted parameters (two in these cases) matters more than the exact functional form.

Repeating the two first fits with density overlap instead, we obtained

$$E_{exch} \approx 8.6\Omega \tag{15}$$

$$E_{exch} \approx 9.3\Omega(1 - 8.6e^{-R}) \tag{16}$$

The distance-dependent fit was again an improvement, although not as significant.

Comparing the results of Eq. 16 to those obtained with Eq. 13, we see that density overlap gives slightly larger errors than the orbital overlap, but in the ethanethiol–water complexes the effect is much more pronounced. The error is not reduced if we fit only to these complexes. To further investigate the problem, two qualitatively different ethanethiol–water complexes were selected, shown in Figure 2, each represented by 11 geometries along the interaction axis. Separate fits were performed for each complex, in which the distance-dependence was kept fixed and only the proportionality constant was optimised.

$$\text{dw1: } E_{exch} \approx \begin{cases} 10.2\Omega(1 - 8.6e^{-R}) \\ 0.75S^2(1 + 20e^{-R}) \end{cases} \tag{17}$$

$$\text{dw2: } E_{exch} \approx \begin{cases} 8.3\Omega(1 - 8.6e^{-R}) \\ 0.76S^2(1 + 20e^{-R}) \end{cases} \tag{18}$$

Apparently, the two complexes need significantly different constants if density overlap is used, whereas the same constants can describe both complexes if orbital overlap is used. This means that some effect in the orbital overlap cannot be described by density overlap. Rewriting Eq. 4 explicitly,

$$S^2 = \sum_i^A \sum_j^B \left[\int \psi_i(\mathbf{r}) \psi_j(\mathbf{r}) d\mathbf{r} \right]^2 \tag{19}$$

and expanding Eq. 5 into MO–MO contributions,

$$\Omega = \int \rho_A(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r} = \sum_i^A \sum_j^B \int |\psi_i(\mathbf{r})|^2 |\psi_j(\mathbf{r})|^2 d\mathbf{r} \tag{20}$$

we see that the main difference is that the integrands in Eq. 19 may change sign and possibly cause a sort of destructive interference, whereas the integrands in Eq. 20 are always positive.

If these interference effects are important, the two sets of integrals in Eq. 19 and Eq. 20 should not be well correlated. Therefore, we have plotted each integral of Eq. 19 versus

the corresponding integral of Eq. 20 for the two ethanethiol–water complexes. The plot is shown in Figure 3 together with straight lines indicating the ratio between the total orbital overlap and the total density overlap for the two representative complexes. The intermolecular distances were chosen rather large in order to reduce the importance of the distance-dependent factor. In *dw1* the S–H distance was 2.96 Å and in *dw2* the H–O distance was 2.85 Å.

Several interesting features can be seen in Figure 3. Most evident is that the ratio between the orbital overlap and the density overlap is much more stable in *dw2* than in *dw1*, for which it varies considerably between the various MO–MO pairs. Because *dw1* is an interaction directly with the sulphur atom, this variation probably means that the node structure is more complicated around the sulphur atom than around the oxygen atom. This can be one reason why interactions with second-row elements are more difficult to model than those involving only first-row elements. However, even in the *dw2* case, almost every data point lies sufficiently far from the average line to destroy the nice relationships between orbital overlap and density overlap (and thus between exchange–repulsion and density overlap), had it been the only contribution. Clearly, at least for sulphur-containing complexes, one has to rely on cancellation of error in order to use the density overlap. The results show, however, that in most cases, this cancellation occurs, so that the extra error introduced when neglecting interference effects is smaller than the errors inherent in any model based on the overlap.

We have examined the two data points marked with arrows in Figure 3 more thoroughly. They correspond to two MO–MO pairs in the *dw1* complex that has a striking difference in the S^2/Ω ratio: it equals 30 for the $11-3$ pair and 0.57 for the $17-5$ pair (the HOMO–HOMO pair). Thus, they provide a good example of how different ratios arise. The two pairs of involved orbitals are depicted in Figure 4. In the first case, the large ratio is a result of no destructive interference at all, because the product of the two orbitals has the same sign in the overlap region. In the second case, on the other hand, one of the interacting orbitals (and thereby the product) changes sign in the overlap region, leading to an almost complete cancellation for this particular geometry.

A pragmatic solution to the problem that the same coefficient cannot model all complexes using density overlap is to introduce different coefficients for different atom–atom interactions. The approximation given in eqs. 7-9 was used to partition the total density overlap

into element–element contributions. Fitting to the whole data set gave

$$E_{exch} \approx 8.8\Omega(1 - 6.0e^{-R})(1 + \sum_{a,b} k_{ab}x_{ab}) \tag{21}$$

where a and b are element labels, x_{ab} are relative contributions from a–b interactions (given by Eq. 9), and k_{ab} are fitted coefficients. To reduce the number of parameters, only six element pairs were selected to have nonzero k_{ab}, namely H–H (+0.44), H–C (+0.09), H–F (-0.18), H–S (+0.43), H–Cl (+0.40), and O–S (-0.16). The resulting errors are shown in Table II. Clearly, the problem with ethanethiol–water complexes has disappeared, and most of the other errors have also been reduced. Fitting to a small data set, consisting of the Molecule–water and Toluene–water sets, with relative weights 10 and 1, respectively, did not produce significantly larger errors when the parameters were applied to the other groups. This demonstrates that parameters obtained for these complexes are transferable to the other sets, which is very important in order to use the relation for other types of interactions.

Finally, we have investigated how well the non-empirical model of exchange–repulsion in the Effective Fragment Potential (EFP) method performs for the same set of dimers. The errors are shown in the last column of Table II. In general, EFP performs well, considering that no parameter fitting is involved. It is interesting to see that the error is fairly constant through the various groups (except for the toluene–water complexes which are discussed below). In particular, the complexes not containing water (the molecule–molecule group) do not show a higher error, in contrast to what is seen with the orbital overlap models. However, comparison to the results obtained with Eq. 13 or 21 shows that slightly better results can be obtained with a fitted overlap model with only a few parameters. The density overlap model also has the distinct advantage that further approximations may be introduced.

The EFP error is largest in the toluene–water complexes, which contain mainly the O–H...π interaction. The calculations were repeated using the cc-pVTZ *dimer-centered* basis set in the monomer calculations as well as in the dimer calculation. The resulting energies were almost identical to the monomer-centered aug-cc-pVTZ results, indicating that the error stems from some approximation within the EFP model other than the limited basis set. For example, it is known that the Spherical Gaussian Overlap approximation is rather crude when the molecular orbitals are very delocalised, as the π orbitals are in toluene.

IV. CONCLUSIONS

We have investigated how well the exchange–repulsion can be described in terms of orbital overlap and density overlap. Such descriptions are useful for two reasons. First, calculation of the overlaps only requires quantum-chemical calculations on the isolated monomers, which is a large gain in efficiency compared to performing supermolecular calculations of the whole complex. Second, and more important, overlap expressions provide a starting point for developing parameterised expressions that can be used in molecular mechanics force fields for simulations etc. The advantage of taking the "detour" via overlap expressions is that one can quantitatively separate the error caused by the approximation of exchange–repulsion using monomer wavefunctions from the error caused by more practically motivated approximations, such as a specific functional form and a limited number of fitted parameters.

The results show that, for a large and diverse set of interactions, exchange–repulsion energy is close to being proportional to either the orbital overlap or the density overlap. At short distances, a significant deviation from proportionality was observed in either case. Loosely speaking, the density overlap is too steep whereas the orbital overlap is too flat, and this conclusion holds for all interactions included in this study. Thus, for accurate results, a distance-dependent correction is needed regardless of whether orbital overlap or density overlap is used. This correction may take several forms, but one extra parameter is sufficient to obtain a good model when orbital overlap is used.

When a similar model is applied using density overlap instead, the results are almost as good as with orbital overlap for most types of dimers. However, the performance is less stable, as is shown by the much larger errors obtained in the ethanethiol–water group. The difference comes from the approximation inherent in the density overlap description, namely that no interference effects are present. Because this approximation is the base of almost all molecular mechanics force fields currently in use, it could be important to be aware of the effect. On the other hand, it seems to be reasonably well cancelled out in all cases except for interaction with sulphur, and even in those cases, the problem can be avoided by introducing element-dependent parameters. It should be noted that this is not an attempt to capture the interference effects themselves, but rather to capture one of the most serious consequences of the effects, namely that the effective S^2/Ω ratio becomes different depending on which elements that take part in the interaction. The advantages with the density overlap

model are the simplicity of the model and the ease with which one can introduce further approximations.

EFP performs well, except for the toluene–water comlexes, in which the Spherical Gaussian Approximation seems to be inaccurate. As expected, the absence of fitted parameters makes the model unbiased, so that the error is approximately equal for all types of interactions. This is a great advantage when applying the method for completely unknown systems. However, the error is actually higher than for the fits using overlap expressions and containing few parameters, even when the toluene–water group is disregarded. Another drawback of the EFP method is that it requires very diffuse basis sets to provide reasonable results.

In view of the results presented here, we must conclude that different exchange–repulsion models are preferable in different contexts. If one wants to be sure to have an unbiased model that gives results close to those from a quantum-chemical calculation, the EFP model is the method of choice. If one knows what types of interactions will be studied, one can instead fit a distance-corrected orbital overlap expression to exchange–repulsion energies for some typical dimers. The same parameters apply for a wide range of interactions. For example, Eq. 13 can be expected to be good for most types of interactions encountered in biological systems (metal interactions excluded). Finally, if one wants a more computationally efficient exchange–repulsion model, the density overlap expressions provide a good starting point for approximations.

V. ACKNOWLEDGMENTS

The authors wish to thank Prof. Jan Jensen for help with EFP and fruitful discussion. Ingemar Nilsson, Lars Sandberg, and Ola Engkvist are also acknowledged for interesting discussions and valuable comments. Moreover, we thank LUNARC for generously providing computational resources and AstraZeneca and the Swedish Science Research Council for financial support.

---

[1] G. Chalasinski and M. M. Szczesniak, Chem. Rev. **100**, 4227 (2000).

[2] O. Engkvist, P.-O. Åstrand, and G. Karlström, Chem. Rev. **100**, 4087 (2000).

[3] S. Brdarski and G. Karlström, J. Phys. Chem. A **102**, 8182 (1998).

[4] A. Wallqvist and G. Karlström, Chem. Scripta **29A**, 131 (1989).

[5] N. Gresh, P. Claverie, and A. Pullman, Int. J. Quantum Chem. **29**, 101 (1986).

[6] N. Gresh, J.-P. Piquemal, and M. Krauss, J. Comput. Chem. **26**, 1113 (2005).

[7] J. Jensen and M. S. Gordon, Mol. Phys. **89**, 1313 (1996).

[8] R. J. Wheatley and S. L. Price, Mol. Phys. **71**, 1381 (1990).

[9] E. Fraschini and A. J. Stone, J. Comput. Chem. **19**, 847 (1998).

[10] E. J. Baerends, *Cluster Models for Surface and Bulk Phenomena* (Plenum, New York, 1992).

[11] K. Kitaura and K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976).

[12] M. Gutowski, G. Chalasinski, and J. van Duijneveldt-van de Rijdt, Int. J. Quantum Chem. **26**, 983 (1984).

[13] H. Margenau and N. R. Kestner, Theory of Intermolecular Forces, Pergamon, Oxford (1969).

[14] B. Jeziorski, M. Bulski, and L. Piela, Int. J. Quantum Chem. **10**, 281 (1976).

[15] J. N. Murrell and J. J. N. Teixeiras-Dias, Mol. Phys. **19**, 521 (1970).

[16] M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, J. Phys. Chem. A **105**, 293 (2001).

[17] J. H. Jensen, J. Chem. Phys. **104**, 7795 (1996).

[18] J. Jensen and M. S. Gordon, J. Chem. Phys. **108**, 4772 (1998).

[19] Y. S. Kim, S. K. Kim, and W. D. Lee, Chem. Phys. Letters **80**, 574 (1981).

[20] J. B. O. Mitchell and S. L. Price, J. Phys. Chem. A **104**, 10958 (2000).

[21] I. Nobeli, S. L. Price, and R. J. Wheatley, Mol. Phys. **95**, 525 (1998).

[22] R. J. Wheatley and J. B. O. Mitchell, J. Comput. Chem. **15**, 1187 (1994).

[23] E. J. Baerends, D. E. Ellis, and P. Ros, Chem. Phys. **2**, 41 (1973).

[24] M. J. D. Powell, Computer J. **7**, 155 (1964).

[25] T. H. Dunning Jr, J. Chem. Phys. **90**, 1007 (1989).

[26] D. E. Woon and T. H. Dunning Jr, J. Chem. Phys. **98**, 1358 (1993).

[27] O. Treutler and R. J. Ahlrichs, J. Chem. Phys. **102**, 346 (1995).

[28] G. Karlström, R. Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, et al., Computational Materials Science **28**, 222 (2003).

[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. J. A. Montgomery, R. E. Stratmann, J. C. Burant, et al., *Gaussian 98,*

Revision A.9, Gaussian, Inc., Pittsburgh, PA, 1998.

[30] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, et al., J. Comput. Chem. **14**, 1347 (1993).

[31] $A_{ij} = 0.58(\alpha_i + \alpha_j) + 0.076(\alpha_i - \alpha_j) + 0.68(\alpha_i + \alpha_j)^2 + 0.0048$

$K_{ij} = 0.14 \exp[3.28(\alpha_i + \alpha_j)] - 7.7(\alpha_i - \alpha_j)$

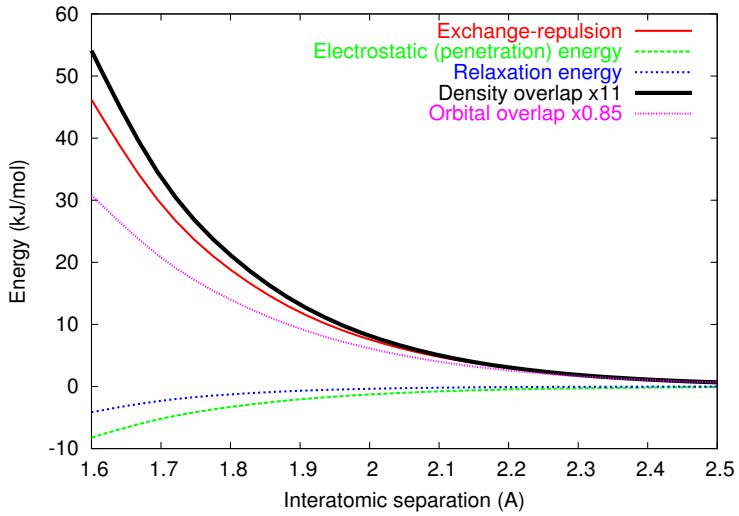

FIG. 1: Contributions to the He...He interaction energy and overlap expressions

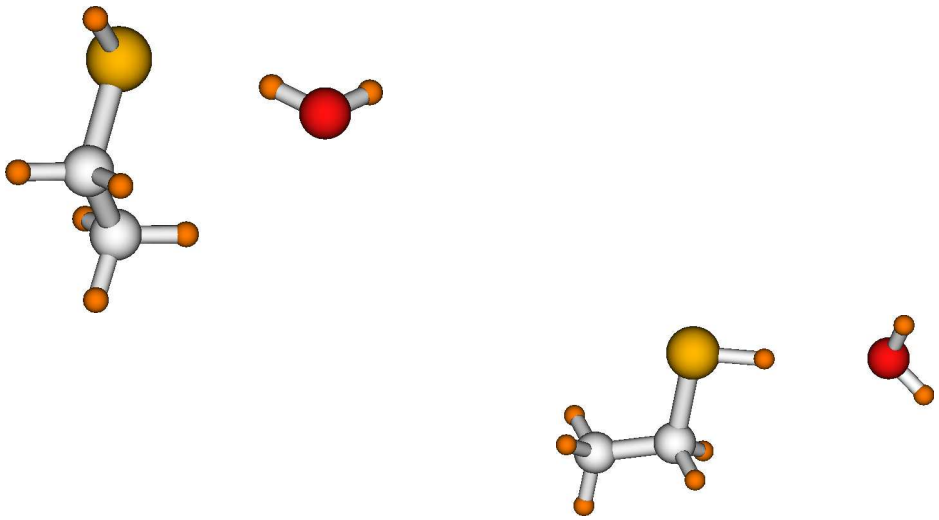

FIG. 2: Selected ethanethiol–water complexes, denoted *dw1* (left) and *dw2* (right)

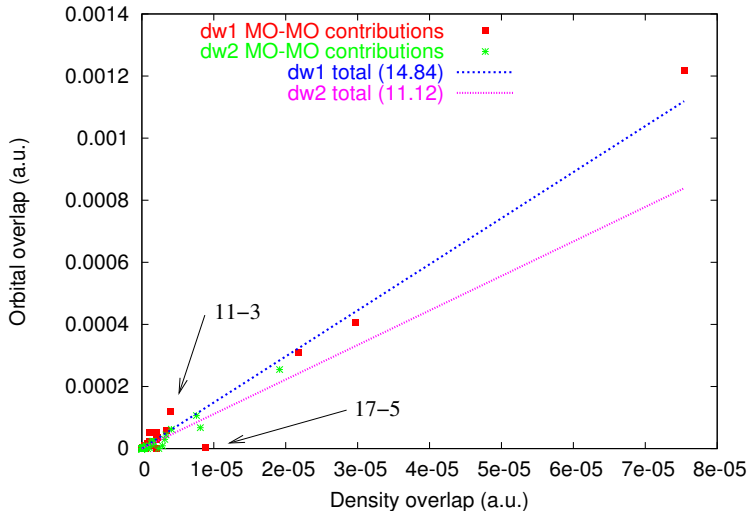

FIG. 3: MO–MO contributions to the orbital overlap and density overlap for the selected ethanethiol–water complexes. The two MO–MO pairs discussed in the text are marked with arrows.

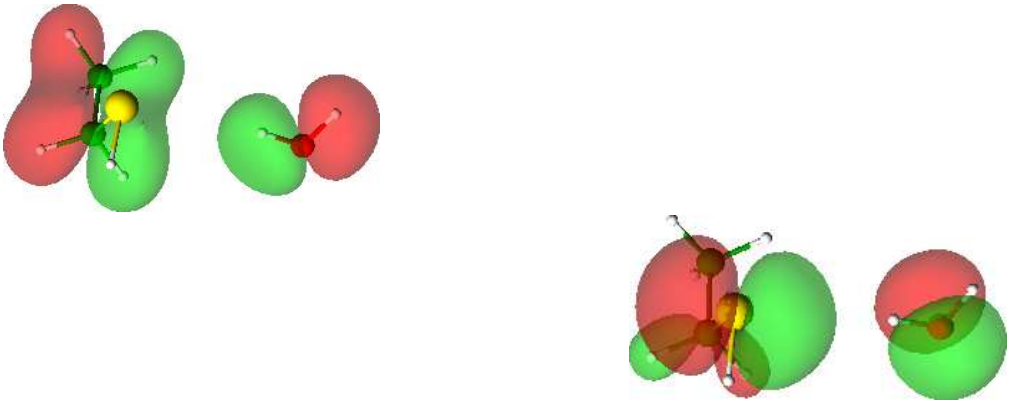

FIG. 4: Space plots of molecular orbitals for two extreme cases of MO–MO interaction in the *dw1* complex. To the left, MO 11 of ethanethiol with MO 3 of water. To the right, MO 17 of ethanethiol with MO 5 of water. The two colors denote different phases (positive and negative). Note that the depicted MOs are the undisturbed monomer orbitals, which have only been inserted in the same figure.

TABLE I: Molecular dimers included in the various data groups. The origin of the geometries (optimised or simulated) is shown after each group label. The acronym in paranthesis after the molecule name indicates for which amino acid the molecule is a model.

Group	**Interacting molecules**	**Number of geometries**
Mol–Wat (opt)	N-acetylglycine methyl amide (backbone) – water	6 x 11
	isobutane (Val) – water	2 x 11
	ethanol (Ser) – water	3 x 11
	ethanethiol (Cys) – water	3 x 11
	methyl ethyl thioether (Met) – water	2 x 11
	propionate ion (Glu) – water	2 x 11
	toluene (Phe) – water	4 x 11
	ethylammonium ion (Lys) – water	1 x 11
	methylguanidinium ion (Arg) – water	4 x 11
	4-methylimidazole (Hie)– water	4 x 11
	4-methylimidazolium ion (Hip) – water	4 x 11
	2-methylimidazole (Hid) – water	4 x 11
	p-cresol (Tyr) – water	5 x 11
	propionamide (Gln) – water	4 x 11
	3-methylindole (Trp) – water	8 x 11
	N-acetylpyrrolidine (Pro) – water	3 x 11
	dimethyl ether – water	2 x 11
	methyl acetate – water	3 x 11
	2,4-difluorotoluene – water	3 x 11
	2,4-dichlorotoluene – water	3 x 11
	1,1,1-trifluoroethane – water	1 x 11
Mol–Mol (opt)	N-acetylglycine methyl amide (backbone) – ethanol (Ser)	3 x 11
	ethanol (Ser) – ethanol (Ser)	1 x 11
	ethanol (Ser) – ethanethiol (Cys) – water	2 x 11
	ethanol (Ser) – 4-methylimidazole (Hie)	2 x 11
	ethanethiol (Cys) – ethanethiol (Cys)	1 x 11
	4-methylimidazole (Hie) – 4-methylimidazole (Hie)	1 x 11
	propionamide (Gln) – propionamide (Gln)	2 x 11
Wat–Wat (sim)	water – water	200 x 7
Eth–Wat (sim)	ethanethiol (Cys) – water	58
Tol–Wat (sim)	toluene (Phe) – water	50
Pro–Wat (sim)	propionamide (Gln) – water	200

TABLE II: Mean absolute errors in kJ/mol for various fits divided into contributions from various dimer sets

	Eq. 11	Eq. 12	Eq. 13	Eq. 14	Eq. 15	Eq. 16	Eq. 21	Refitted Eq. 21[a]	EFP
Mol–Wat	0.78	0.59	0.49	0.46	0.62	0.56	0.45	0.43	0.90
Mol–Mol	1.23	0.65	1.11	1.07	1.01	0.94	0.59	0.58	1.10
Wat–Wat	2.20	1.13	0.74	0.69	0.96	0.61	0.59	0.59	1.16
Eth–Wat	0.65	0.73	0.73	0.84	1.75	1.83	0.57	0.76	1.12
Tol–Wat	1.01	2.68	0.60	0.74	1.43	1.05	0.70	0.62	4.52
Pro–Wat	1.37	0.83	0.70	0.65	1.13	0.81	0.72	0.76	1.26
All sets	**1.21**	**1.10**	**0.73**	**0.74**	**1.15**	**0.97**	**0.60**	**0.64**	**1.68**

[a]same functional form as in Eq. 21 but parameters reoptimised using only the Mol–Wat and Tol–Wat groups, with relative weights of 10 and 1, respectively