A Systematic Nonempirical Method of Deriving Model Intermolecular Potentials for Organic Molecules: Application To Amides

John B. O. Mitchell[†] and Sarah L. Price*

Centre for Theoretical and Computational Chemistry, Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Received: July 5, 2000; In Final Form: September 1, 2000

A systematic method of deriving atom−atom intermolecular potentials from the monomer wave functions has been developed for formamide, acetamide and *trans-N*-methylacetamide (NMA) and tested for its ability to reproduce the crystal structures. The total intermolecular potentials comprised an accurate distributed multipole analysis representation of the multipolar electrostatic interaction energy, an atom−atom C_6 dispersion model, and a short-range repulsion model derived from the overlap of the monomer charge densities. The short-range model has been assessed and validated by comparison with ab initio intermolecular perturbation theory (IMPT) calculations of the exchange−repulsion, penetration, and charge-transfer energies of test sets of around 20 dimer conformations. A range of models has been developed. The simplest version of the overlap model need not require any IMPT calculations (though in this example they are used to calibrate and validate the potential) and can be used to estimate atom−atom repulsive parameters. Removal of various simplifying assumptions in the overlap model gives better reproductions of the IMPT data and the crystal structures, and provides a route to specific potentials for organic molecules. The resulting model potentials, as assessed by crystal structure reproduction, are comparable with the best empirical potentials for amides and superior to some commonly used potential energy functions. An advantage of the method is that transferability of parameters can be tested, rather than assumed. There is an encouraging degree of transferability as the potential generated for NMA reproduces the crystal structures of formamide and both polymorphs of acetamide very well.

1. Introduction

Computational chemistry is widely applied to systems involving interacting molecules, from gas phase van der Waals dimers to protein−drug binding in solution to macroscopic crystalline materials. In all these situations, the accuracy of the calculations is dependent upon being able to evaluate the interaction energy of an assembly of molecules to an appropriate level of precision. In general, the smaller and simpler the molecules, the greater the accuracy with which the interaction energy can be evaluated. For interactions between noble gas atoms, which being spherical have no long-range electrostatic or polarization interactions and no orientation dependence, it is possible to obtain, both from experiment and theory, extremely accurate values of the interaction energy.[1,2] At the other end of the size scale, modeling of protein−drug interactions may rely on either approximate and empirical force-fields[3] or on knowledge-based approaches.[4−7] The prospect of ab initio quality potentials being developed directly for such large systems still seems remote; the use of nonempirical model atom−atom potentials derived on smaller model molecules may provide a more promising route. In recent years there has been a marked increase in the size of systems for which high quality nonempirical potentials have been developed, with accurate intermolecular potentials being published for specific molecules such as water,[8] methanol,[9,10] oxalic acid,[11] and acetonitrile,[12] as well as formamide.[13] We seek here to extend this work to amides, which are of course of interest as models of the protein main-chain. Since amides have been heavily studied, with considerable work on empirical parametrizations, we do not expect that a systematic potential will necessarily do better than the best empirical ones for these molecules. To obtain reasonable results here would, however, validate the approach for use on other atom types, where empirical parametrization is not possible for lack of experimental data, as has already been successful for the dioxaborole functional group.[14]

Our aim is to develop the use of the overlap model (OM) to derive nonempirical atom−atom intermolecular potential energy functions representing the short-range terms from the ab initio charge distributions of the monomers. A key advantage of the overlap model over other methods[15,16] of estimating the repulsive interaction is that it is possible to partition the interaction into atom−atom terms prior to the fitting process. This becomes necessary for organic molecules, where important configurations, including hydrogen bonds, involve several different types of atom−atom contacts, and so the separation into atomic contributions by empirical fitting is ill-determined. The initial aim was to develop a methodology that could be readily used to derive parameters for the many functional groups in industrially important organic molecules that do not have well validated empirical potential parameters. In addition, following careful comparison of the overlap model to a small quantity of Hayes−Stone intermolecular perturbation theory (IMPT)[17] data, we can reduce the number of assumptions made, and produce completely nonempirical nonbonded parameters. These model repulsion potentials are suitable for molecular modeling of condensed phases of neutral organic molecules in conjunction with distributed multipole analysis (DMA)[18] electrostatic and simple dispersion models. Another advantage of the overlap

* Author for correspondence. Fax: (+44)-20-7679-7463. E-mail: s.l.price@ucl.ac.uk.

[†] Address from September 2000: Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

10.1021/jp002400e CCC: $19.00
Published on Web 10/21/2000

model approach is that no assumptions about transferability (i.e., classification of atoms into a limited number of types) need be made. The overlap data for each atom−atom pair can be examined individually and then an assessment of which atoms can be treated as equivalent made. Of course, a high degree of transferability is expected for atoms that are chemically very similar, and the parameters should reflect this in order to be used for extrapolating to larger molecules. Hence, this derivation of model potentials for three closely related amides concludes by testing the transferability of the parameters between the systems.

Our potential consists of the following terms:

$$E_{tot} = E_{es} + E_{rep} + E_{disp} \quad (1)$$

where E_{tot}, the total interaction energy, is subdivided into the electrostatic E_{es} and dispersion E_{disp} long-range terms, and a short-range term E_{rep}. Our potential does not attempt to model the intermolecular polarization (induction) energy,[17,19−21] a many-body term which pair potentials are incapable of describing accurately. Since the polarization energy is always attractive, its omission should in principle leave our models systematically underestimating the attractive part of the interaction energy.

The electrostatic term E_{es} is evaluated from sets of atomic multipoles obtained by distributed multipole analysis (DMA)[18] of the monomer wave function. This representation retains essentially all of the accuracy of the ab initio wave functions of the molecules and hence the accuracy of the long-range electrostatic interaction energies is limited only by the quality of the wave function. The DMA electrostatic model does not include the short-range penetration[22] correction to the electrostatic energy. The dispersion energy is modeled using isotropic atom−atom dispersion coefficients, obtained from monomer wave functions, as a sum of C_6R^{-6} terms. However, we know from studies on small polyatomics that there are also significant R^{-8} and R^{-10} contributions to the dispersion energy.[23] Our model also neglects the effect of interpenetration of the charge clouds on the dispersion energy at short range. Thus, the current nonempirical model neglects the net result of these partially canceling effects on the dispersion energy. For atoms in molecules, all dispersion coefficients are also anisotropic,[24] which is also neglected in the model.

The exponentially decaying short-range term models the exchange−repulsion energy. For empirically fitted model potentials, the other short-range contributions, such as the penetration and charge-transfer energy terms, are absorbed in the parametrization. In a nonempirical potential, such contributions are either neglected or explicitly modeled. In this work, we use IMPT calculations of the penetration and charge-transfer terms to investigate whether they can be effectively incorporated into E_{rep} in a nonempirical model potential.

The atom−atom exponential form for the short-range repulsion E_{rep} is derived using the overlap model.[25,26,27] This is based on the assumed relationship

$$E_{rep} = KS_\rho \quad (2)$$

where K is a proportionality constant and S_ρ is the charge overlap of unperturbed monomer charge distributions of molecules A and B, viz.

$$S_\rho = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r})\, d\mathbf{r} \quad (3)$$

The relationship in eq 2 has been tested for a variety of systems.[11,14,26,27] Earlier, more detailed investigations of the relationship between noble gas atoms[25] showed that a more accurate description of the relationship between overlap and repulsion is

$$E_{rep} = KS_\rho^{\,y} \quad (4)$$

where the power y is usually slightly smaller than 1. The basic relationship (eq 2) is particularly useful for organic molecules as it leaves only one parameter, the proportionality constant, to be derived by fitting to either experimental[9] or ab initio[11,14] data. However, the proportionality constant K for different inert gas pairs does vary slightly,[25] implying that when the model is applied to molecules, a different value of K might be required for contacts between each pair of atom types. This previously made a significant improvement to the representation of the pyridine−methanol exchange−repulsion energy in the hydrogen bonding region.[27]

To exploit its relationship with the short-range energy (eq 2), we model the intermolecular charge overlap, which we hereafter refer to simply as overlap, as the sum of atom−atom exponential functions. These are multiplied by a proportionality constant to obtain atom−atom contributions to E_{rep}. In the basic model, denoted OM-ER, the proportionality constant between the model overlap and the model exchange−repulsion energy K_{er} is fitted to the exchange−repulsion energy calculated by IMPT.

However, since the penetration energy is also known to be essentially proportional to overlap,[28] we can fit another proportionality constant K_{erp} to the sum of the exchange−repulsion and penetration energies (E_{erp}) to obtain the OM-ERP model for E_{rep}. Thus, the OM-ERP model corresponds to E_{erp}, the quantity previously modeled using a probe atom approach by Fraschini and Stone[29] for the methane dimer. E_{erp} also corresponds to the repulsive term fitted to IMPT results by Cabaleiro-Lago and Rios for acetonitrile.[12]

As our results (Table 1) will show, the charge-transfer energy is also approximately proportional to overlap. Hence we can also incorporate an estimated charge transfer into the model potential (eq 1), by fitting K_{erpc} to the sum of the exchange−repulsion, penetration, and charge-transfer energies (the OM-ERPC model). E_{erpc} corresponds to the repulsive term in the formamide potential of Cabaleiro-Lago and Rios.[13] (The major difference in their approach is that they fitted all their repulsion parameters to total IMPT E_{erpc} energies at more than 400 points,

TABLE 1: Details of the Studies on the Three Amides

	formamide	acetamide	*N*-methyl-acetamide
wave function for DMA	6-31G** MP2	6-31G** MP2	6-31G** MP2
wave function for S_ρ and IMPT	6-31G** SCF	6-31G** SCF	6-31G** SCF
atoms	6	9	12
atom types	6	6	7
2nd phase S_ρ geometries	360	309	309
3rd phase IMPT points	21	19	19
Correlation of Total Overlaps S_ρ with IMPT Short-Range Energy Contributions			
correlation overlap vs ER	0.9926	0.9955	0.9703
correlation overlap vs Pen	−0.9973	−0.9904	−0.9934
correlation overlap vs CT	−0.9433	−0.9838	−0.9619
correlation overlap vs ERP	0.9909	0.9927	0.9396
correlation overlap vs ERPC	0.9839	0.9906	0.9321
Optimum Powers for Representing IMPT Short-Range Energy Contributions			
best y for ER = $(S_\rho)^y$	0.925	0.931	0.898
best y for ERP = $(S_\rho)^y$	0.898	0.905	0.850
best y for ERPC = $(S_\rho)^y$	0.933	0.926	0.842
best y for KP models	0.93	0.93	0.95

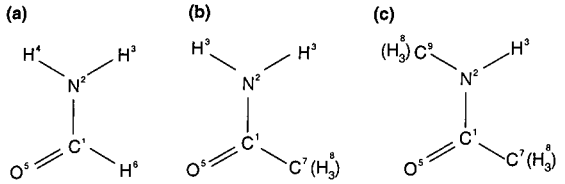

Figure 1. Chemical structures of (a) formamide, (b) acetamide, and (c) *trans-N*-methylacetamide (NMA). The superscripts show the definitions of atom types for each molecule. The parameters were derived independently for each molecule except in the transferability test with the NM-ERP model.

rather than using the overlap model to obtain a more rigorous partitioning into atom−atom contributions.) Although, in principle, including the charge-transfer energy improves the potential, we note that an approximate estimate of a small term need not necessarily bring the calculated energies closer to the experimental results.

When discussing the proportionality constants in terms applicable to all three models, we will simply use the symbol *"K"*. If the *K* values, for the same level of model, vary little between the different amides, then these *K* values could be considered transferable between related molecules. This would avoid the need for the relatively expensive IMPT calculations or the fitting of K to experimental data. However, in the testing of our approach, we have calculated the IMPT interaction energy terms at sufficient points on each potential energy surface to be able to validate the method and to derive improved models by reducing the errors implicit in the overlap model assumption.

Our model intermolecular potentials are systematic[30] in the sense that the interaction energy is considered as the sum of models for the electrostatic, exchange−repulsion, penetration, charge-transfer, and dispersion contributions. This is an exacting approach, which does not present the opportunity to absorb errors by offsetting deficiencies in one term with changes to another; nor can the absence of many-body terms be compensated for in the pair potential. In contrast, empirical potentials are generally designed to give accurate predictions of total energies and various molecular properties, but are not expected to give a reliable term-by-term partitioning of the energy. Our method has the advantage that replacing any term by a more theoretically rigorous model of that component is an improvement to the potential, which should ultimately converge toward the actual intermolecular pair potential.

A major part of the validation of our potentials is the reproduction of experimental crystal structures for the three simple amides (Figure 1a−c) formamide, acetamide, and *trans-N*-methylacetamide (NMA). A model potential should produce a minimum in the lattice energy reasonably close the experimental crystal structure, this being an essential criterion[31] for the model potential to be considered accurate enough for use in crystal structure prediction or other modeling of condensed phases.

2. Methods

The validation of the model potentials against experimental data used the four distinct crystal structures, and the three heats of sublimation,[32−34] available for the three amides. The crystal structures used were the lowest temperature determinations in the Cambridge Structural Database,[35] in order to minimize the errors due to the neglect of temperature effects in static lattice energy minimization. These are a 90 K $P2_1/n$ structure (FORMAM02) of formamide,[36] an $R3c$ structure (ACEMID05) determined at 23 K[37] and a room temperature $Pccn$ structure (ACEMID)[38] of acetamide, and a 110 K $Pna2_1$ structure[39] of NMA (METACM02). The rigid molecular structures used in the crystal structure modeling are taken from the original crystal structure determinations, except that bond lengths to hydrogens are adjusted to standard values (1.009 Å for H−N and 1.083 Å for H−C).

The lattice energy minimization calculations started at the experimental crystal structures and were performed using DMAREL.[40] The ability of our nonempirical potentials to reproduce the crystal structures is compared with that of the empirical FIT repulsion−dispersion parameter set,[41] with both potentials using the same DMA electrostatic model. The FIT parameters set uses the carbon, nitrogen, and nonpolar hydrogen potentials that were empirically fitted to azahydrocarbons by Williams and Cox,[42] and also a polar H−(N) potential[41,43] obtained by fitting to hydrogen bonded crystal structures, including amides. These repulsion-dispersion parameters have been used in a number of studies,[44−48] in conjunction with point charge or DMA electrostatic models, to carry out predictions, lattice energy calculations, and minimizations on organic crystal structures.

The electrostatic model for the crystal structure modeling is derived from the 6-31G** MP2 charge density of the monomer within the crystal structure, calculated using CADPAC,[49] to obtain the distributed multipoles up to and including hexadecapole on each atom. The lattice energy includes all terms in the atom−atom multipole expansion of the electrostatic energy up to R_{ik}^{-5}, with the charge−charge, charge−dipole, and dipole−dipole contributions evaluated by Ewald summation, and all other electrostatic terms by direct summation over all entire molecules with a center of mass distance of less than 15 Å. The repulsion and dispersion contributions to the lattice energy are evaluated by direct summation of all atoms within 15 Å.

We model the dispersion energy as a sum of atom−atom terms where atom i in molecule A and atom k in molecule B are of types ι and κ, respectively

$$E_{\mathrm{disp}} = \Sigma_{i\in A,k\in B} C_{6,\iota\kappa}/R_{ik}^{\ 6} \qquad (5)$$

using the ab initio atomic C_6 parameters[50] derived by Ioannou and Amos. These are strictly C_6 parameters (*not* effective C_6 parameters absorbing other effects), calculated ab initio from the monomer wave functions by integrating over the polarizabilities at imaginary frequencies. Ioannou and Amos have created a database of C_6 parameters for all pairs of atom types in a number of model molecules, including several amino acids. We chose to use those coefficients that they considered most reliable, namely, the set obtained from density functional calculations using the large basis set of Sadlej.[51] They estimate that their grouping of atoms according to type (e.g., one C_6 coefficient for interactions between any amide N and any methyl H atoms) gives an error of ∼5%, relative to calculating C_6 explicitly for each atom pair. This compares with an error of ∼20% for the common assumption of "one atom type per element".[50]

These long-range electrostatic and dispersion terms are constant for all the model potentials derived for a given molecule. They are combined with various short-range models for E_{rep}, derived nonempirically from the intermolecular overlap. All the calculations to derive the short-range potentials use 6-31G** SCF wave functions obtained using CADPAC[49] with molecular structures optimized at the 3-21G SCF level.

Thus, the overlaps and the IMPT interaction energies are calculated with the same molecular wave functions. Hence the repulsion potentials do not incorporate the minor effects due to distortions of the molecular structure by the crystalline environment.

2.1. The Overlap Model. We use the program GMUL[22] both to calculate the intermolecular overlap S_ρ and to partition it (and hence E_{rep}) into atom−atom terms using the positions and exponents of the Gaussian basis functions such that, (i,k) being atoms in molecules A and B,

$$S_\rho = \Sigma_{ik} S_\rho(ik) \tag{6}$$

and applying eq 2 gives

$$E_{rep} = K\Sigma_{ik} S_\rho(ik) \tag{7}$$

For most applications of intermolecular potentials, it is necessary to use relatively simple and rapidly calculable functional forms. Thus, we use an exponential model for the overlap and assume each atom−atom contribution to be isotropic, which is probably adequate for the repulsion parameters of carbon, hydrogen, nitrogen and oxygen. (If the isotropic exponential model did not adequately represent the set of S_ρ values for a given atom pair, then the representation could be improved by the incorporation of atom−atom anisotropy into the S_m(ik) via the use of S-functions,[52] in a similar way to Wheatley and Price[30] and to Nobeli et al.[27]) The model overlap S_m is thus an approximation to the accurately calculated overlap S_ρ,

$$S_m = \Sigma_{ik} S_m(ik) = \Sigma_{ik} A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik}) \tag{8}$$

By applying eq 2 to our *model* overlap S_m, we obtain

$$E_{rep} = K\Sigma_{i\in A, k\in B} A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik}) \tag{9}$$

with $A_{\iota\kappa}$ and $\alpha_{\iota\kappa}$ being constants, to be determined by fitting, for the interaction between atoms of types ι and κ (atoms i and k being of types ι and κ respectively). K will differ slightly when model overlaps are used instead of accurately calculated overlaps, to absorb the errors in representing the overlap.

Our description of the repulsion is validated, and the proportionality constant determined, by comparison with IMPT calculations. Separate proportionality constants are determined for the models where the repulsion term is fitted to the exchange−repulsion energy (OM-ER), to the exchange−repulsion plus penetration energy (OM-ERP), and to the exchange−repulsion plus penetration plus charge-transfer energy (OM-ERPC). Our methodology for obtaining the repulsion model from the overlap involves three phases of analysis, which are described below.

2.1.1. First Phase: Exponential Decay Constants ($\alpha_{\iota\kappa}$). A significant problem with all methods of deriving exponential model repulsion potentials is the strong correlation between the preexponential ($A_{\iota\kappa}$) and exponential decay ($\alpha_{\iota\kappa}$) parameters. This can be avoided by first calculating $\alpha_{\iota\kappa}$ from the variation of the isotropic atom−atom overlaps with distance. Thus, we calculate the isotropic coefficient ($C_{\iota\kappa}^{iso}$) of the expansion of the atom−atom overlaps in terms of *S*-functions[52] at a number of different distances for at least one atom pair (ik) representative of each pair of atom types ($\iota\kappa$). We assume the relationship

$$C_{\iota\kappa}^{iso} = A_{\iota\kappa}^{iso} \exp(-\alpha_{\iota\kappa} R_{ik}) \tag{10}$$

and plot $\ln(C_{\iota\kappa}^{iso})$ against R_{ik} to obtain a value for $\alpha_{\iota\kappa}$. We include all data with

$$(0.005 \geq C_{iso} \geq 0.00005)$$
(X−X pairs, where X is any element but hydrogen)
$$0.005 \geq C_{iso} \geq 0.00001 \text{ (X−H pairs)}$$
$$0.005 \geq C_{iso} \geq 0.000005 \text{ (H−H pairs)}$$

in atomic units. This roughly corresponds to the typical distance ranges 4.0 to 6.0 Å (X−X), 3.0 to 5.8 Å (X−H), and 2.0−5.2 Å (H−H). We carry out the fitting for each atom pair, except that we may choose to group some atoms together into atom types if they are sufficiently similar to share parameters in the final potential.

The $\alpha_{\iota\kappa}$ parameters are retained for the second phase of the fitting. Previous work[11,27] has shown that, although $A_{\iota\kappa}^{iso}$ does give the genuine isotropic part of the overlap, it is not the best effective isotropic coefficient for the intermolecular potential, and thus it is discarded. This is explained by the observation that the anisotropic expansion (defining $C_{\iota\kappa}^{iso}$) models the atomic charge density in the uninteresting intramolecular region as well as the required intermolecular region.

2.1.2. Second Phase: Derivation of the Preexponential Parameters ($A_{\iota\kappa}$). We generate about 300 representative dimer geometries, with the molecules required to be in van der Waals contact, using software based on random configuration generation. This produces a set of geometries in which different pairs of atoms from the two monomers are in contact. One may also add specifically chosen geometries to the set. The program uses radii of C, 1.7 Å; N, 1.5 Å; O, 1.4 Å; and H, 1.0 Å, and has a built-in tolerance of 0.3 Å above and 1.0 Å below the sum of the two atomic radii. The lower limit allows short contacts with repulsion energies large enough to give accurate sampling of the repulsive wall. For each of the 300 or so geometries, we calculate all the intermolecular atom−atom distances and atom−atom overlaps. These data are then arranged according to atom pairs, so that for each atom pair we have 300 distances and 300 overlaps.

We then exclude all data except those points for which

$$0.005 \geq S_\rho(ik) \geq 0.00005 \text{ (X−X pairs)}$$
$$0.005 \geq S_\rho(ik) \geq 0.00001 \text{ (X−H pairs)}$$
$$0.005 \geq S_\rho(ik) \geq 0.000005 \text{ (H−H pairs)}$$

with S_ρ(ik) being the atom−atom overlap in atomic units. The large number of dimer geometries are required because each configuration usually samples overlap contributions within the quoted range for only a few atom pairs, and it is necessary to discard the many uninteresting longer separations to prevent them from dominating the fitting. The atom−atom distance ranges corresponding to these overlaps are similar to those quoted earlier for C_{iso}. Given the $\alpha_{\iota\kappa}$ value from the first phase, the preexponential factor $A_{\iota\kappa}$ is obtained as the intercept of the line of best fit of $\ln[S_\rho(ik)]$ against R_{ik} in a plot combining data from all atom pairs of types ι and κ.

This procedure is carried out for each pair of atom types. For some relatively inaccessible atom pairs there is often a paucity of data with which to carry out the fit, despite the number of dimer geometries used. This is because the atoms are sufficiently buried within the molecules that they are almost never found in van der Waals contact. Since the parameters for such contacts are therefore unimportant from a modeling point

of view, we can confidently transfer parameters from smaller homologous molecules in which the corresponding atoms are more exposed. For instance, we found insufficient data for the interaction between the two amide carbon atoms in the NMA dimer and consequently used the carbon−carbon parameters from formamide to help generate the NMA potential.

Once the parameters have been determined, the fitted $A_{\iota\kappa}$ and $\alpha_{\iota\kappa}$ parameters allow us to calculate the model overlap $S_m(ik)$ as a function of distance only for each atom pair

$$S_m(ik) = A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik}) \tag{11}$$

2.1.3. Third Phase: Derivation of the Proportionality Constant K, and Testing of the Model for the Net Repulsion. The short-range contributions to the intermolecular energy are calculated using IMPT for about 20 dimer geometries. These are chosen to be representative both of a range of total exchange−repulsion energies (spanning 2 orders of magnitude between about 1 and 100 kJ/mol) and also of different atom pairs making major contributions to the total overlap. For each of these geometries, the exchange−repulsion and charge-transfer energies are obtained separately, and the penetration energy is calculated by subtracting the multipolar (DMA) electrostatic energy from IMPT's electrostatic term. The IMPT charge-transfer term is calculated free of basis set superposition error, using Stone's adaptation[53] of the earlier[17] methodology.

For each of the dimer geometries on the IMPT short-range potential surface, the atom−atom distances are calculated and total model overlap for all atom pairs is obtained as

$$S_m = \Sigma_{ik} S_m(ik) = \Sigma_{i\in A, k\in B} A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik}) \tag{12}$$

Since the underlying assumption in our model for the exchange−repulsion energy is that

$$E_{er} = K_{er} S_m \tag{13}$$

we plot E_{er} against the total model overlap S_m, constraining the line to pass through the origin. The gradient of the resulting line of best fit, minimizing the root mean-square (RMS) deviation of the points from linearity, is K_{er}. Then the total overlap model exchange−repulsion (OM-ER) energy for each test conformation is simply $K_{er}S_m$. These model repulsion energies are compared with the corresponding IMPT repulsion energies in the fitting of K_{er} and measures of the goodness-of-fit (principally RMS % error and correlation coefficient) provide indications of the accuracy of the model repulsion potential.

We can trivially repeat this process to generate an overlap model for the exchange−repulsion plus penetration energy (OM-ERP), using the assumed relationship

$$E_{erp} = K_{erp} S_m \tag{14}$$

between the overlap and the IMPT exchange−repulsion plus penetration energy. The analogous procedure for the exchange−repulsion plus penetration plus charge-transfer energy (OM-ERPC model) uses

$$E_{erpc} = K_{erpc} S_m \tag{15}$$

Since, for the dimer configurations considered, both the penetration and charge-transfer energies are attractive,

$$K_{er} > K_{erp} > K_{erpc} \tag{16}$$

2.2. Models with Two *K* Values (KO-Models). The overall quality of the fit between the model overlap and IMPT exchange−repulsion energy could be improved by allowing different atom−atom interactions to have different K values,[27] and would be justified in terms of the small variations observed for the noble gases.[25] A study of the plot of exchange−repulsion energy against model overlap for 21 configurations of the formamide dimer (Figure 2a) showed a simple trend. Those configurations where atom pairs involving oxygen made a significant contribution to the overlap tended to lie below the line (implying that their IMPT exchange−repulsion energies were smaller than those predicted by the overlap model) and those without oxygen contributions tended to lie above the line. A similar pattern was found when the penetration (Figure 2b) and charge-transfer energies (Figure 2c) were considered, and also for acetamide (Figure 3) and NMA (Figure 4). As an

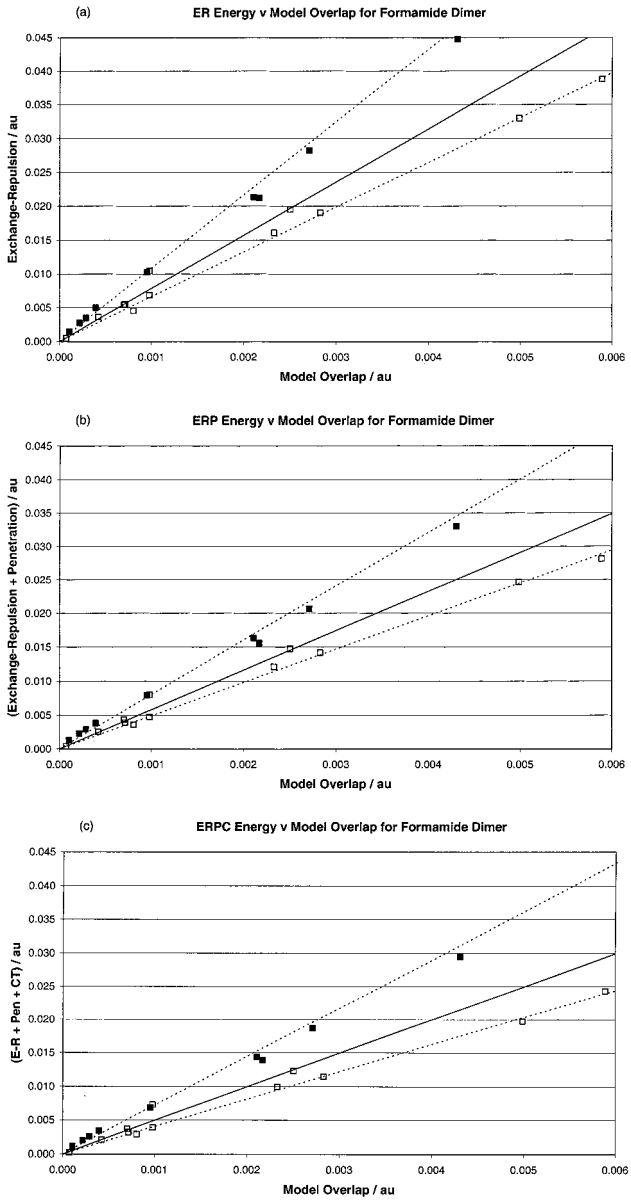

Figure 2. Various IMPT short-range energies plotted against the total model overlap for 21 configurations of the formamide dimer. (a) Exchange−repulsion energy (E_{er}); (b) exchange−repulsion plus penetration energy (E_{erp}); (c) exchange−repulsion plus penetration plus charge transfer (E_{erpc}). The solid line represents the K values used in the OM-models. Points for which more than half the predicted overlap involves oxygen are shown as open squares, other points as filled squares; this distinction is made binary for visualization purposes only. The upper and lower dashed lines represent K^X and K^O respectively from the KO-models.

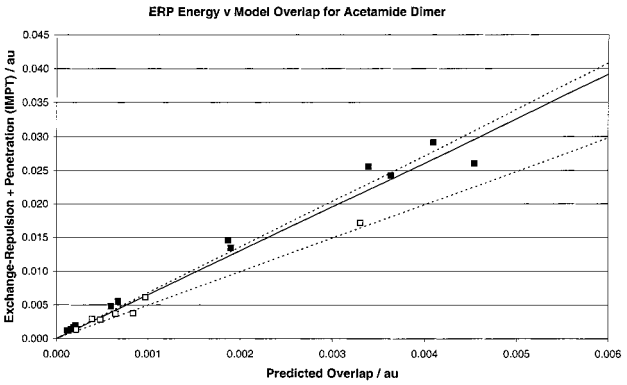

Figure 3. IMPT exchange−repulsion plus penetration energy (E_{erp}) plotted against the total model overlap for 19 configurations of the acetamide dimer. The solid line represents the K_{erp} value used in the OM-ERP model for acetamide. The upper and lower dashed lines represent K^{X}_{erp} and K^{O}_{erp}, respectively, from the KO-ERP model. Points for which more than half the predicted overlap involves oxygen are shown as open squares, other points as filled squares; this distinction is made binary for visualization purposes only.

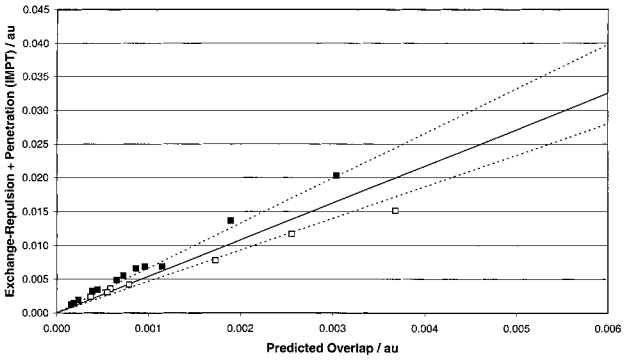

Figure 4. IMPT exchange−repulsion plus penetration energy (E_{erp}) plotted against the total model overlap for 19 configurations of the NMA dimer. The solid line represents the K_{erp} value used in the OM-ERP model for NMA. The upper and lower dashed lines represent K^{X}_{erp} and K^{O}_{erp}, respectively, from the KO-ERP model. Points for which more than half the predicted overlap involves oxygen are shown as open squares, other points as filled squares; this distinction is made binary for visualization purposes only.

empirical device, we chose to generate a set of potentials using one K value (K^{O}) for atom pair overlaps involving oxygen, and another value (K^{X}) for overlaps not involving oxygen. We call these the KO-models. The ratio of the K values was chosen to optimize the correlation coefficient between overlap and E_{er}. For simplicity, the same ratio (K^{O}/K^{X}) was used for the KO-ER and KO-ERP models, as justified by comparison of Figure 2a,b. The ratio was, however, reoptimized for the charge-transfer term. A significantly different ratio might reasonably be expected here on theoretical grounds, as charge-transfer energies are likely to be greatest between electronegative and electropositive atoms, particularly those involved in hydrogen bonding. Thus, the overall (K^{O}/K^{X}) ratio for the KO-ERPC models differs slightly from that in the KO-ER and KO-ERP models.

2.3. Models with Two K values and Adjustment for the Power Law (KP-Models). The calculations in Figure 5 were performed in order to determine the limitations of the use of the overlap in determining model repulsion potentials. Figure 5a shows the interaction energy of the doubly hydrogen bonded cyclic formamide dimer (the geometry used for Figures 5a,b is illustrated in Figure 5c), as a function of separation. This benchmark energy is estimated by the IMPT energy (electrostatic plus exchange−repulsion plus penetration plus charge transfer), with the long-range electrostatic term replaced by the 6-31G** MP2 DMA, and the same dispersion term as used in our models. When this is contrasted with the best possible potential based on proportionality to the overlap (eq 2), (obtained by calculating the overlap explicitly at each point, and multiplying by an optimized K_{erp} to convert this to an ERP energy (GM-ERP)), then it is clear that the overlap repulsion rises too steeply. This results in the N···O hydrogen bond distance corresponding to the minimum energy increasing from 3.09 Å (IMPT) to 3.11 Å (GM-ERP). The model overlaps (OM-ERP) show a similar trend to the accurately calculated overlaps, giving a minimum at 3.14 Å, consistent with our results showing that the overlap models overestimate the hydrogen bond lengths.

Thus, the assumption that the repulsion is proportional to the overlap (OM-ERP and GM-ERP potentials) produces repulsive walls that are too steep, an observation consistent with the overlap model being too repulsive at short range. The effective exponential constants for E_{erp} in these potentials, in the region of the minimum, are 4.039 $Å^{-1}$ (IMPT), 4.449 $Å^{-1}$ (GM-ERP), and 4.484 $Å^{-1}$ (OM-ERP). Thus, the overlap model overestimates the exponential decay constant by about 10%. This is related to the underlying assumption that of the overlap model, namely,

$$E_{rep} = KS_{\rho} \tag{2}$$

As both Kim et al.[25] and Wheatley and Price[26] noted, a more accurate relationship is

$$E_{rep} = KS_{\rho}^{\ y} \tag{4}$$

where the power y is usually in the approximate range $0.8 \leq y \leq 1$.[25,26] The assumption that $y = 1$ leads to an overestimate of the repulsion at large overlaps (short range) and to an underestimate at small overlaps (long range).

In terms of our model overlaps, this relationship is equivalent to

$$E_{rep} = KS_{m}^{\ y} = K[\Sigma_{i\in A,k\in B} A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik})]^{y} \tag{17}$$

Unfortunately, raising the total overlap to a power is only equivalent to adding the atom−atom overlaps raised to the same power when $y = 1$. To retain the additivity of the atom−atom contributions to the repulsion, we choose to make the approximation

$$E_{rep} = K[\Sigma_{i\in A,k\in B} A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik})]^{y} \approx \Sigma_{i\in A,k\in B} K[A_{\iota\kappa} \exp(-\alpha_{\iota\kappa} R_{ik})]^{y} \tag{18}$$

This form is computationally convenient, since it can be expressed as

$$E_{rep} \approx \Sigma_{i\in A,k\in B} K[(A_{\iota\kappa})^{y} \exp(-y\alpha_{\iota\kappa} R_{ik})] \tag{19}$$

with $K(A_{\iota\kappa})^{y}$ and $(y\alpha_{\iota\kappa})$ as the new preexponential and exponential decay constants, respectively. The approximation is small when y is close to 1, and can be partially absorbed into the reoptimization of K. Thus, to give more realistic exponential decay constants, we started from our existing KO-model parameters, optimized y and refitted the proportionality constants K^{O} and K^{X}. For formamide, the RMS percentage error in the predicted exchange−repulsion energy over the 21 dimer configurations was lowest at $y = 0.93$ (optimized to 2 s.f.). The

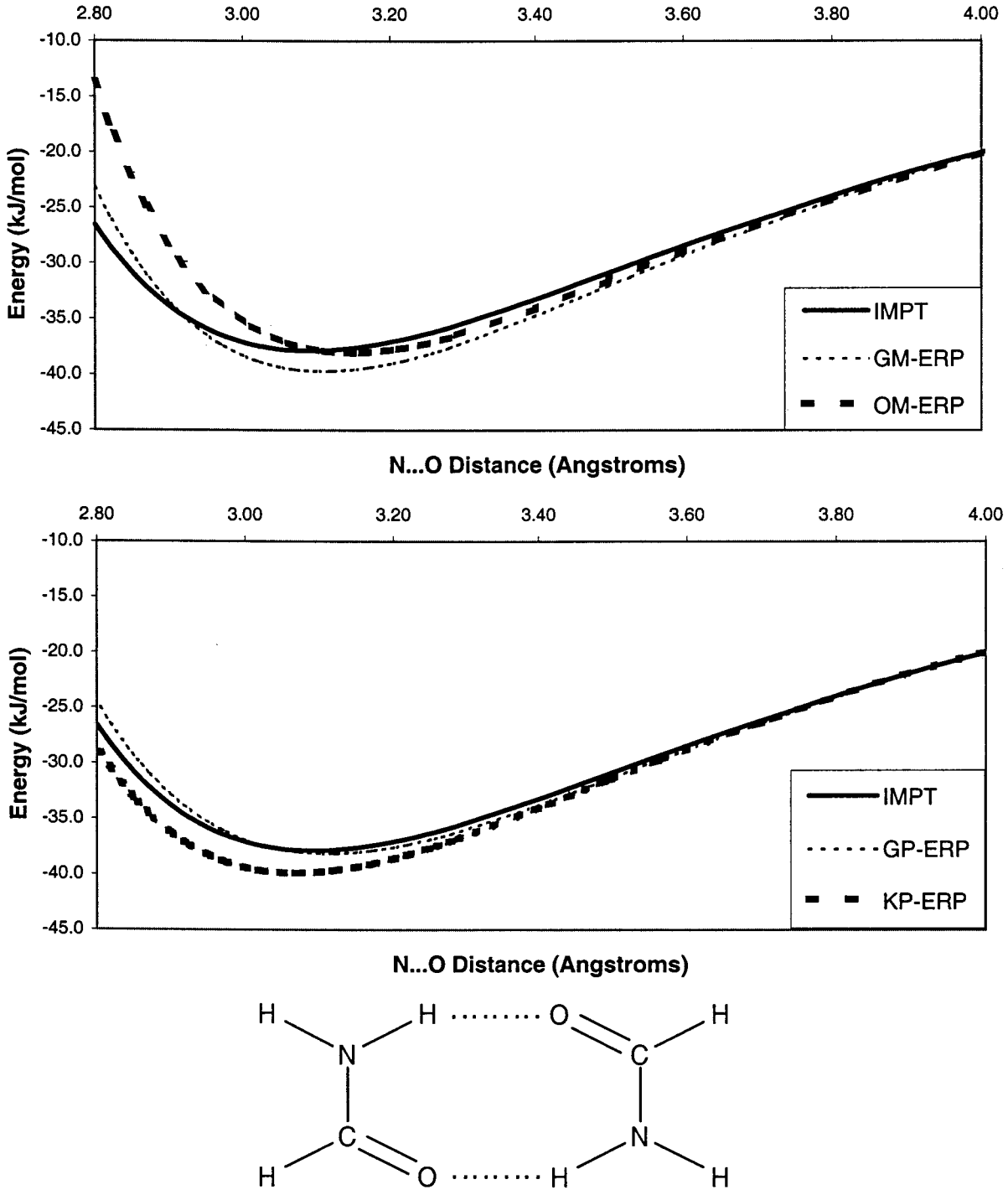

Figure 5. (a) The OM-ERP model illustrated for the case of the doubly hydrogen bonded cyclic formamide dimer. The solid curve ("IMPT") represents the IMPT energy, with the long-range electrostatic term replaced by the 6-31G** MP2 DMA and the model dispersion added; this "IMPT" energy is our benchmark. The light dashed curve ("GM-ERP") is the energy obtained by calculating the overlap explicitly at each point, and multiplying by an optimized K_{erp} to convert this to an ERP energy. The heavy dashed curve ("OM-ERP") is the ERP energy calculated using the overlap model. All three potentials illustrated contain identical long-range electrostatic (6-31G** MP2 DMA) and dispersion (Ioannou and Amos C_6)[50] terms and thus differ in the ERP energy alone. The horizontal axis is the N···O hydrogen bond distance. (b) The KP-ERP model illustrated for the case of the doubly hydrogen bonded cyclic formamide dimer. The solid curve ("IMPT") again represents the IMPT energy. The light dashed curve ("GP-ERP") is the energy obtained by calculating the overlap explicitly at each point, raising it to the power of 0.93, and multiplying by an optimized K_{erp} to convert this to an ERP energy. The heavy dashed curve ("KP-ERP") is the ERP energy calculated using the KP-ERP model. (c) The cyclic formamide dimer geometry on which Figure 5a,b are based. The two identical N...O distances are varied between 2.8 and 4.0 Å.

corresponding values of y were found to be 0.93 for acetamide and 0.95 for NMA. These values were similar to or slightly larger than the values of y obtained by direct fitting of $\ln(E_{rep})$ against $\ln(S_\rho)$, which are shown in Table 1. The approximation of maintaining atom−atom additivity (eq 18) contributed an RMS error of 3.6% for formamide.

This method was used to derive a set of repulsion models (KP-ER, KP-ERP, and KP-ERPC) for each of the three molecules. The efficacy of introducing the power dependence is illustrated for the formamide dimer in Figure 5b. Using the accurately calculated GMUL overlap raised to the power of 0.93, with the proportionality constants derived by fitting to the

exchange−repulsion plus penetration energies ("GP-ERP"), gives a potential curve which closely follows the IMPT benchmark. The corresponding atom−atom model (KP-ERP) has larger errors (as would be expected from having to approximate the overlap), but does give considerably better agreement with the steepness of the repulsive wall and the position of the minima than was obtained for OM-ERP. The minima are at N···O hydrogen bond distances of 3.09 Å (IMPT), 3.11 Å (GP-ERP), and 3.07 Å (KP-ERP). The effective exponential decay constants for E_{erp} in these potentials, in the region of the minimum, are 4.039 $Å^{-1}$ (IMPT), 4.137 $Å^{-1}$ (GP-ERP), and 4.151 $Å^{-1}$ (KP-ERP). Thus, as we expect, the KP-models give a more faithful representation of the IMPT exponential decay constants than do the OM-models.

3. Results

3.1. Analysis of the Potential Derivation. One advantage of the overlap model is that no prior assumptions about transferability need to be made, but as the overlaps are analyzed it becomes apparent which atoms are so similar that they can be treated as a single type. For formamide, all atoms were considered as distinct atom types, though the parameters describing the two polar hydrogens (H3 and H4) turned out to be very similar. This also proved to be the case for all methyl hydrogens in acetamide and NMA, which were grouped together as a single type (H8). Hence the number of atom types could be reduced to those shown in Figure 1. The potentials for each pair of atom types were determined independently for each molecule, allowing a final analysis of the transferability of the parameters by comparing corresponding pair potentials for different molecules.

The method of derivation of the potentials is detailed in Table 1, along with an analysis of how well the accurately calculated overlap S_ρ correlates with the IMPT energies. It is clear that the overlap correlates well with the penetration and exchange−repulsion energies, but a little less well with the charge-transfer energies. The correlation is somewhat poorer for NMA than for formamide and acetamide.

A detailed analysis of the fits and parameters of both the total overlap and various models for the overlap against the IMPT short range interaction energies, which cover a range from about 1 kJ/mol to 100 kJ/mol, is given for each molecule in Table 2. A selection of correlations is shown in Figures 2−4. Table 2 shows that the overlap is a useful predictor of the short-range energy contributions, and that raising the overlap to a power slightly less than unity further improves the correlation for the exchange−repulsion plus penetration energy. The charge transfer is somewhat less accurately represented by the overlap. There is a significant error in representing the repulsive wall inherent in using the overlap, as illustrated for a specific potential curve in Figure 5 for the total overlap (GM and GP) models. This intrinsic error is comparable in size to the additional error introduced by modeling the overlap by a sum of isotropic atom−atom exponential functions (which is done to produce a potential that can be used to simulate the crystal structures). Although the trends in the errors in the model potentials relative to more reliable IMPT estimates of the short-range repulsion are reasonable, and the results of comparing potential curves as in Figure 5 are encouraging, the real test of how important the residual errors are comes from the practical use of the model potentials. Here we use them to attempt to reproduce the experimental crystal structures by static lattice energy minimization.

3.2. Reproduction of the Crystal Structures with the Basic Overlap Model $E_{rep} = KS_m$. The most basic overlap models OM-ERP and OM-ERPC reproduced the crystal structure[36] of formamide with acceptable accuracy (Table 3), with RMS errors of 2.9% and 4.0%, respectively, in the cell edges. The cell angle β was somewhat overestimated, at around 115° compared to the experimental 101°, but the potential energy surface was very flat. The OM-ERP model, which gave the best fit overall, predicted a packing about 10% denser than the crystal structure (the minimization was effectively at 0 K, the crystal structure at 90 K). The N···O hydrogen bond lengths, however, were overestimated. The calculated lattice energies were −65.5 (OM-ERP) and −71.2 (OM-ERPC) kJ/mol, in fairly good agreement with the experimental sublimation energy of 71.7 kJ/mol.[32] The minimization causes some untwisting in the hydrogen bonded sheets, such that the molecular planes of molecules in the same sheet are now close to parallel, but offset in the perpendicular direction. A fish-scale pattern is apparent, only in the minimized structure, when the sheets are viewed side-on along the *b*-direction. The OM-ER model, however, failed to produce a well-defined local minimum for formamide.

An overestimate of the hydrogen bond lengths, combined with an overestimate of the overall packing density, indicates that empirically adjusting the value of K is unlikely to produce an improved potential. A smaller *K* value would scale all atom−atom repulsions and so increase the density as it decreased the hydrogen bond lengths. Thus, we compared a "relative hydrogen bond length", defined as the average crystal hydrogen bond length divided by the cube root of the cell volume, and normalized such that the experimental structure has a value of 1.000. The values were 1.106 and 1.104 for the structures resulting from the OM-ERP and OM-ERPC models, respectively.

For acetamide, all our models successfully reproduced minima in the regions of both polymorphic crystal structures, as reported in Tables 4 and 5. The RMS errors in the reproduction of the cell edges of the *R*3*c* structure (Table 4) were 6.7%, 5.9%, and 6.1% for the OM-ER, OM-ERP, and OM-ERPC models, respectively. The three models gave lattice energies of −56.5, −65.4, and −69.8 kJ/mol, respectively. The latter two are within expected[54,55] margins of error in comparison with of the 77.2 kJ/mol experimental sublimation energy, which de Wit et al.[33] assigned to a structure solved at 23 K[37,56] and thereby uniquely identified with the *R*3*c* polymorph. The OM-ERP model gives a density for this *R*3*c* polymorph 3.3% too low and a relative hydrogen bond length of 1.081; OM-ERPC predicts a density only 0.6% too high, and a relative hydrogen bond length of 1.078. Our minimized structures were systematically compressed along the *c*-direction, by about 8% when corrected for changes in the overall packing density, with commensurate increases of about 4% in the *a*- and *b*-directions. As most of the hydrogen bonds in the structure are roughly aligned with the *ab*-plane, this is closely related to the overestimated relative hydrogen bond lengths.

The *Pccn* polymorph (Table 5) is experimentally less densely packed, its density falling midway between the predictions of the OM-ER and OM-ERP models. This probably results from the structure having been solved at room temperature, rather than the low-temperature regime of the *R*3*c* crystal data. DMAREL minimization is a 0 K method, as thermal effects are excluded, and so is expected best to reproduce low-temperature structures. The RMS errors in the cell edges were 5.7% (OM-ER), 6.1% (OM-ERP), and 6.7% (OM-ERPC). Our calculations tended to compress the *Pccn* structure along the *a*-direction and expand it in the *c*-direction. The relative hydrogen bond lengths were between 1.082 and 1.091 for all

TABLE 2: Calculating ER, ERP, and ERPC Energies from the Overlap Model

calculating	from	RMS%	correl[a]	K[b]	K^O[b]	K^X[b]	y	model
			(a) Formamide					
IMPT ER	GMUL S_ρ	15.1%	0.9926	8.588			1.00	GM-ER
IMPT ERP	GMUL S_ρ	19.2%	0.9909	6.354			1.00	GM-ERP
IMPT ERPC	GMUL S_ρ	21.1%	0.9839	5.476			1.00	GM-ERPC
IMPT ER	GP $(S_\rho)^y$	12.5%	0.9929	5.777			0.93	GP-ER
IMPT ERP	GP $(S_\rho)^y$	16.1%	0.9914	4.275			0.93	GP-ERP
IMPT ERPC	GP $(S_\rho)^y$	23.0%	0.9842	3.684			0.93	GP-ERPC
GMUL S_ρ	OM S_m	18.9%	0.9791					
IMPT ER	OM S_m	24.6%	0.9514	7.857			1.00	OM-ER
IMPT ERP	OM S_m	26.5%	0.9496	5.813			1.00	OM-ERP
IMPT ERPC	OM S_m	31.7%	0.9323	4.985			1.00	OM-ERPC
IMPT ER	KO S_m	12.1%	0.9961		6.627	10.843	1.00	KO-ER
IMPT ERP	KO S_m	14.9%	0.9954		4.905	8.026	1.00	KO-ERP
IMPT ERPC	KO S_m	15.4%	0.9947		4.066	7.226	1.00	KO-ERPC
IMPT ER	KP $(S_m)^y$	8.1%	0.9962		3.971	6.609	0.93	KP-ER
IMPT ERP	KP $(S_m)^y$	12.5%	0.9960		2.940	4.894	0.93	KP-ERP
IMPT ERPC	KP $(S_m)^y$	14.1%	0.9945		2.412	4.426	0.93	KP-ERPC
			(b) Acetamide					
IMPT ER	GMUL S_ρ	14.1%	0.9955	9.349			1.00	GM-ER
IMPT ERP	GMUL S_ρ	17.8%	0.9927	6.984			1.00	GM-ERP
IMPT ERPC	GMUL S_ρ	19.0%	0.9906	6.201			1.00	GM-ERPC
IMPT ER	GP $(S_\rho)^y$	13.5%	0.9952	6.247			0.93	GP-ER
IMPT ERP	GP $(S_\rho)^y$	15.3%	0.9929	4.669			0.93	GP-ERP
IMPT ERPC	GP $(S_\rho)^y$	20.4%	0.9906	4.144			0.93	GP-ERPC
GMUL S_ρ	OM S_m	13.5%	0.9958					
IMPT ER	OM S_m	18.5%	0.9851	8.731			1.00	OM-ER
IMPT ERP	OM S_m	20.7%	0.9822	6.522			1.00	OM-ERP
IMPT ERPC	OM S_m	24.0%	0.9782	5.785			1.00	OM-ERPC
IMPT ER	KO S_m	15.1%	0.9933		6.655	9.106	1.00	KO-ER
IMPT ERP	KO S_m	18.3%	0.9901		4.969	6.799	1.00	KO-ERP
IMPT ERPC	KO S_m	17.1%	0.9894		4.134	6.081	1.00	KO-ERPC
IMPT ER	KP $(S_m)^y$	7.1%	0.9955		4.145	5.514	0.93	KP-ER
IMPT ERP	KP $(S_m)^y$	8.4%	0.9933		3.097	4.120	0.93	KP-ERP
IMPT ERPC	KP $(S_m)^y$	8.8%	0.9926		2.591	3.684	0.93	KP-ERPC
			(c) *N*-methylacetamide					
IMPT ER	GMUL S_ρ	16.5%	0.9703	8.682			1.00	GM-ER
IMPT ERP	GMUL S_ρ	21.9%	0.9396	6.328			1.00	GM-ERP
IMPT ERPC	GMUL S_ρ	24.1%	0.9321	5.501			1.00	GM-ERPC
IMPT ER	GP $(S_\rho)^y$	12.9%	0.9714	6.366			0.95	GP-ER
IMPT ERP	GP $(S_\rho)^y$	18.8%	0.9410	4.642			0.95	GP-ERP
IMPT ERPC	GP $(S_\rho)^y$	21.8%	0.9335	4.048			0.95	GP-ERPC
GMUL S_ρ	OM S_m	13.7%	0.9981					
IMPT ER	OM S_m	20.1%	0.9672	7.431			1.00	OM-ER
IMPT ERP	OM S_m	24.6%	0.9382	5.418			1.00	OM-ERP
IMPT ERPC	OM S_m	25.9%	0.9311	4.711			1.00	OM-ERPC
IMPT ER	KO S_m	12.3%	0.9972		6.335	9.013	1.00	KO-ER
IMPT ERP	KO S_m	14.8%	0.9918		4.665	6.637	1.00	KO-ERP
IMPT ERPC	KO S_m	14.6%	0.9909		4.037	5.814	1.00	KO-ERPC
IMPT ER	KP $(S_m)^y$	9.6%	0.9961		4.284	6.295	0.95	KP-ER
IMPT ERP	KP $(S_m)^y$	8.9%	0.9950		3.160	4.644	0.95	KP-ERP
IMPT ERPC	KP $(S_m)^y$	8.1%	0.9945		2.716	4.086	0.95	KP-ERPC

[a] The column headed "correl" gives the correlation coefficient R between the quantities in the first two columns as calculated by Excel.[70] [b] The values of K, K^O, and K^X quoted are those obtained when both overlap and energy are expressed in atomic units.

three models. The lattice energies calculated for the *Pccn* structure suggest that it is about 2 kJ/mol less stable than the *R*3*c* polymorph; the empirical potential FIT[41] gives an energy difference of 3.7 kJ/mol.

The crystal structure of NMA was successfully reproduced by all three OM-models. The OM-ER model gave a large relative hydrogen bond length (1.116). The OM-ERP model did slightly better on this parameter (1.097) and also gave a lattice energy of −68.9 kJ/mol, compared with the experimental sublimation energy of 70.8 kJ/mol.[34] The OM-ERPC model gave a relative hydrogen bond length of 1.086 and a lattice energy of −74.8 kJ/mol. All three models gave similar RMS errors in the cell edges of 4.0%, 3.6% and 4.1% respectively. The minimized structures are relatively expanded along the *a*-direction, which is close to both hydrogen bonding directions in the structure, and compressed in the *b*-direction, perpendicular to the hydrogen bonds.

Our OM-models were able to find minima that maintained the crystal symmetry and gave cell lengths reasonably close to all four experimental crystal structures, except that the OM-ER model failed to locate a minimum for formamide. There is a general pattern that the hydrogen bond lengths were overestimated by up to about 0.3 Å. The OM-ERP and OM-ERPC models gave lattice energies that were in reasonable agreement with the experimental sublimation energies, but the lattice energies for OM-ER, where there was no allowance for the attractive short range terms, were insufficiently attractive.

3.3. Reproduction of Crystal Structures for Models with Two *K* Values. Allowing the proportionality between the overlap and the repulsion to depend on whether an oxygen atom was involved (KO-models) not surprisingly improved the fit to the IMPT energies quite substantially for all three amides (Table 2 and Figures 2−4). These KO-models gave a better prediction of the cell edges of all four crystal structures than the corre-

TABLE 3: Reproductions of the 90 K Crystal Structure of Formamide (FORMAM02) by Lattice Energy Minimization with Various Model Potentials

	exptl	FIT[c]	OM-ER	OM-ERP	OM-ERPC	KO-ER	KO-ERP	KO-ERPC	KP-ER	KP-ERP	KP-ERPC	NM-ERP
LE_{init}/kJ mol^{-1}		−60.93	−31.83	−52.05	−60.25	−41.43	−59.13	−67.41	−38.92	−57.25	−66.09	−69.35
LE/kJ mol^{-1}	−71.7[b]	−63.22	fails[d]	−65.49	−71.20	−56.97	−66.31	−72.55	−52.35	−61.73	−68.60	−74.65
space group	$P2_1/n$	$P2_1/n$		$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$	$P2_1/n$
a/Å	3.604	3.657		3.452	3.369	3.678	3.519	3.444	3.772	3.595	3.511	3.390
b/Å	9.041	9.135		9.279	9.180	9.414	9.234	9.128	9.500	9.300	9.179	9.132
c/Å	6.994	6.788		7.035	6.893	7.267	7.000	6.840	7.306	6.995	6.787	6.670
β/°	100.5	107.3		115.2	115.3	113.9	113.8	113.3	112.2	111.7	110.6	112.4
vol/Å^3	224.1	216.5		203.9	192.7	230.0	208.1	197.4	242.4	217.3	204.8	190.9
ρ/g cm^{-3}	1.335	1.382		1.467	1.552	1.301	1.437	1.515	1.234	1.377	1.461	1.567
ρ % excess		3.5%		9.9%	16.3%	−2.6%	7.7%	13.4%	−7.6%	3.1%	9.4%	17.4%
N···O (dimer)/Å	2.948	2.935		3.145	3.082	3.205	3.081	3.000	3.214	3.071	2.969	2.940
N···O (other)/Å	2.883	2.906		3.104	3.042	3.163	3.043	2.964	3.175	3.036	2.935	2.931
rel HB length[a]	1.000	1.013		1.106	1.104	1.083	1.076	1.067	1.067	1.058	1.043	1.062
RMS % cell edges		2.0%		2.9%	4.0%	3.5%	1.8%	2.9%	4.7%	1.7%	2.4%	4.4%

[a] Average hydrogen bond length (N···O) divided by the cube root of the cell volume and normalized such that the experimental structures have values of 1.000. [b] −1 × sublimation energy from NIST database,[32] originally from ref 33. [c] The parameters for the empirical model are taken from the FIT parameter set.[41] [d] The OM-ER model failed to locate a minimum for formamide.

TABLE 4: Reproductions of the 23 K *R3c* Crystal Structure of Acetamide (ACEMID05) by Lattice Energy Minimization

	exptl	FIT	OM-ER	OM-ERP	OM-ERPC	KO-ER	KO-ERP	KO-ERPC	KP-ER	KP-ERP	KP-ERPC	NM-ERP
LE_{init}/kJ mol^{-1}		−71.65	−23.08	−49.29	−58.02	−44.36	−65.20	−74.95	−38.52	−60.79	−70.83	−74.89
LE/kJ mol^{-1}	−77.2[a]	−72.43	−56.49	−65.39	−69.76	−61.69	−72.08	−79.20	−55.99	−66.27	−73.32	−76.2
space group	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*	*R3c*
a/Å	11.492	11.599	12.402	12.111	11.985	12.196	11.895	11.706	12.264	11.926	11.714	11.571
b/Å	11.492	11.599	12.402	12.111	11.985	12.196	11.895	11.706	12.264	11.926	11.714	11.571
c/Å	12.892	12.519	12.509	12.001	11.787	12.421	11.907	11.678	12.873	12.293	12.037	12.229
vol/Å^3	1474.5	1458.7	1666.2	1524.4	1466.1	1600.1	1459.0	1385.9	1676.8	1514.1	1430.4	1418.1
ρ/g cm^{-3}	1.197	1.210	1.060	1.158	1.204	1.103	1.210	1.274	1.053	1.166	1.234	1.245
ρ % excess		1.1%	−11.4%	−3.3%	0.6%	−7.9%	1.1%	6.4%	11.0%	6.1%	3.0%	1.2%
N···O (1)/Å	2.898	2.907	3.256	3.138	3.087	3.159	3.034	2.953	3.189	3.045	2.954	2.901
N···O (2)/Å	2.875	2.948	3.288	3.173	3.124	3.196	3.075	2.998	3.218	3.080	2.993	2.939
rel HB length	1.000	1.018	1.089	1.081	1.078	1.071	1.062	1.052	1.063	1.052	1.041	1.025
RMS % cell edges		1.8%	6.7%	5.9%	6.1%	3.1%	5.3%	3.7%	5.5%	4.1%	4.1%	3.0%

[a] −1 × sublimation energy from NIST database,[32] originally from ref 33.

TABLE 5: Reproductions of the *Pccn* Crystal Structure of Acetamide (ACEMID) by Lattice Energy Minimization

	exptl	FIT	OM-ER	OM-ERP	OM-ERPC	KO-ER	KO-ERP	KO-ERPC	KP-ER	KP-ERP	KP-ERPC	NM-ERP
LE_{init}/kJ mol^{-1}		−66.15	−32.30	−51.54	−57.95	−47.63	−63.00	−70.13	−42.32	−59.00	−66.47	−69.18
LE/kJ mol^{-1}		−68.71	−54.46	−63.21	−67.52	−58.90	−68.82	−75.44	−53.51	−63.24	−69.78	−72.34
space group	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*	*Pccn*
a/Å	7.760	7.536	7.400	7.107	6.982	7.420	7.130	7.012	7.652	7.323	7.187	7.248
b/Å	19.000	18.687	19.212	18.679	18.440	19.091	18.569	18.300	19.331	18.759	18.457	18.401
c/Å	9.510	9.566	10.334	10.094	9.993	10.098	9.838	9.668	10.176	9.889	9.710	9.647
vol/Å^3	1402.2	1347.2	1469.2	1340.1	1286.7	1430.4	1302.6	1240.6	1505.2	1358.4	1288.0	1286.5
ρ/g cm^{-3}	1.119	1.165	1.068	1.171	1.220	1.097	1.205	1.265	1.043	1.155	1.218	1.220
ρ % excess		4.1%	−4.6%	4.6%	9.0%	−2.0%	7.7%	13.0%	−6.8%	3.2%	8.9%	9.0%
N···O (1)/Å	2.969	2.927	3.299	3.182	3.132	3.202	3.079	3.000	3.223	3.084	2.993	2.930
N···O (2)/Å	3.013	2.889	3.258	3.142	3.093	3.157	3.035	2.955	3.181	3.040	2.949	2.894
N···O (3)/Å	2.942	3.003	3.321	3.207	3.157	3.225	3.105	3.027	3.259	3.126	3.042	3.008
N···O (4)/Å	2.873	2.866	3.245	3.138	3.093	3.147	3.032	2.956	3.170	3.037	2.951	2.900
rel HB length	1.000	1.008	1.082	1.088	1.091	1.064	1.068	1.065	1.049	1.052	1.048	1.032
RMS % cell edges		1.9%	5.7%	6.1%	6.7%	4.4%	5.3%	5.5%	4.2%	4.0%	4.7%	4.3%

sponding OM-models (Tables 3−6). For formamide, the KO-ER model successfully minimized, while the KO-ERP model predicted the cell edges with an RMS error of only 1.8%. There was a universal decrease in the relative hydrogen bond lengths, on average corresponding to about a quarter of the excess (the average for the ERP and ERPC models decreasing from 1.091 for OM to 1.065 for KO). A decrease was expected as the N···O and HP···O potentials were less repulsive than before, due to the reduced value of K^O compared to the previous *K*.

3.4. Reproduction of Crystal Structures for Models Using $E_{rep} = KS_m^y$. The KP-models, as we would expect from the optimization of the power *y*, all gave potentials with a significantly better overall RMS fit to the IMPT data than those generated by KO-models or OM-models. The overall trend in accuracy KP > KO > OM reflects the numbers of parameters available to be optimized (which are three, two, and one, respectively for the ER and ERP models and four, three, and one, respectively for the CT models). However, the extra parameters are justified by the empirical observations which are the basis of the overlap model, and it is encouraging that the improved fits to the IMPT data led to increasingly good reproductions of the crystal structures. The KP-models did well in reproducing the four crystal structures. The RMS percentage errors in the cell edges were 1.7%, 4.1%, 4.0%, and 1.8%, respectively, for formamide, two acetamide polymorphs and NMA, using the KP-ERP model. This was a substantial improvement on OM-ERP. The assumption that $y = 1$ was at least partially responsible for the excessive short-range repulsion found with the OM-models, and hence for the large relative hydrogen bond lengths. Thus, it was no surprise that the relative hydrogen bond lengths for the KP-models were more realistic, with the average for eight ERP and ERPC models being 1.048.

TABLE 6: Reproductions of the 110 K Crystal Structure of *trans*-*N*-Methylacetamide (METACM02) by Lattice Energy Minimization

	exptl	FIT	OM-ER	OM-ERP	OM-ERPC	KO-ER	KO-ERP	KO-ERPC	KP-ER	KP-ERP	KP-ERPC
LE_{init}/kJ mol^{-1}		−62.93	−44.47	−62.94	−69.42	−49.00	−65.77	−71.93	−46.26	−63.67	−70.22
LE/kJ mol^{-1}	−70.8[a]	−64.79	−58.05	−68.87	−74.76	−58.24	−69.03	−75.12	−54.53	−65.39	−71.83
space group	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$	$Pna2_1$
a/Å	9.650	9.768	10.286	9.948	9.781	10.203	9.866	9.691	10.206	9.834	9.624
b/Å	6.330	6.505	6.199	6.003	5.922	6.245	6.044	5.959	6.386	6.168	6.083
c/Å	7.170	7.152	7.246	7.074	6.990	7.290	7.127	7.048	7.331	7.151	7.062
vol/Å^3	438.0	454.4	462.0	422.5	404.9	464.5	425.0	407.0	477.8	433.8	413.4
ρ/g cm^{-3}	1.109	1.068	1.051	1.149	1.199	1.045	1.142	1.193	1.016	1.119	1.174
ρ % excess		−3.7%	−5.2%	3.6%	8.1%	−5.8%	3.0%	7.6%	−8.3%	1.0%	5.9%
N···O (cryst)/Å	2.820	2.930	3.203	3.056	2.985	3.141	2.993	2.915	3.127	2.960	2.864
rel HB length	1.000	1.026	1.116	1.097	1.086	1.092	1.072	1.060	1.077	1.053	1.035
RMS % cell edges		1.8%	4.0%	3.6%	4.1%	3.5%	2.9%	3.5%	3.6%	1.8%	2.4%

[a] −1 × sublimation energy from NIST database,[32] originally from ref 34.

TABLE 7: Intermolecular Potential Parameters for KP-ERP Models[a]

parameter (units)		$C_{6,ik}$/kJ mol^{-1}Å^6	B_{ik}/Å^{-1}			A_{ik}/kJ mol^{-1}		
type	type	dispersion	formamide	acetamide	NMA	formamide	acetamide	NMA
C1	C1	1449.4	3.764 84	*3.764 84*	*3.845 81*	237 154.1	*199 657.6*	*239 607.4*
C1	C7	1522.6		3.915 77	3.991 47		446 156.4	649 883.6
C1	C9	1494.0			3.968 93			619 185.2
C7	C7	1599.9		3.345 08	3.427 19		249 070.9	312 245.5
C7	C9	1569.6			3.434 81			291 270.0
C9	C9	1539.9			3.464 09			277 762.5
C1	N2	1415.4	3.893 65	4.484 65	*3.977 38*	424 633.5	1 638 239.2	*434 434.5*
C7	N2	1486.9		3.619 05	3.849 26		472 151.5	970 495.0
C9	N2	1419.9			3.873 75			910 376.9
C1	O5	1345.0	3.994 68	4.787 40	4.856 70	294 739.2	1 222 094.3	1 461 002.4
C7	O5	1413.6		3.756 33	3.848 16		429 071.3	497 458.7
C9	O5	1428.1			3.828 71			400 976.5
N2	N2	1383.7	3.900 34	3.972 12	4.811 66	898 422.2	1 014 713.0	11 516 097.1
N2	O5	1315.1	4.087 93	4.110 88	4.588 39	750 754.2	855 234.1	2 480 327.3
O5	O5	1252.2	4.290 27	4.255 76	4.364 79	944 150.8	799 553.0	924 177.5
C1	H3	392.0	3.968 79	4.912 73	*4.054 14*	25 100.5	249 056.6	*24 164.5*
C1	H4	392.0	3.966 23			22 439.2		
C1	H6	424.3	3.781 68			35 103.0		
C1	H8	439.9		4.162 78	4.678 39		45 985.0	173 891.3
C7	H3	411.6		3.532 45	3.499 54		23 132.2	21 922.7
C7	H8	461.8		3.435 82	3.591 47		29 136.9	42 361.0
C9	H3	435.3			3.469 44			18 454.9
C9	H8	435.3			3.622 25			42 223.4
N2	H3	382.2	3.952 11	3.966 97	4.013 31	56 430.1	52 922.8	51 083.4
N2	H4	382.2	3.949 72			55 410.1		
N2	H6	413.9	3.793 24			81 886.7		
N2	H8	428.9		3.740 87	4.317 12		51 691.6	205 425.3
O5	H3	362.6	4.178 36	4.175 05	3.878 04	44 464.4	48 406.6	21 540.8
O5	H4	362.6	4.178 36			46 492.5		
O5	H6	392.0	3.969 86			61 060.5		
O5	H8	406.5		3.843 29	4.163 91		38 811.4	73 673.2
H3	H3	106.7	4.210 38	4.224 78	*4.300 93*	5 754.0	4 895.9	*5 366.7*
H3	H4	106.7	4.207 24			5 800.0		
H4	H4	106.7	4.204 12			5 781.7		
H3	H6	115.3	3.874 66			6 034.1		
H4	H6	115.3	3.872 45			6 192.7		
H3	H8	119.3		3.804 12	3.807 69		3 616.8	3 535.5
H6	H6	125.1	3.628 80			6 960.8		
H8	H8	134.3		3.547 31	3.989 94		3 480.8	9 063.8

[a] Parameters are given for the conventional exp-6 atom−atom potential, $A_{ik}\exp(-B_{ik}R_{ik}) - C_{6,ik}/R_{ik}^6$, where (eq 19) $\boldsymbol{A}_{ik} = K(A_{ik})^y$ and $B_{ik} = y\alpha_{ik}$. A small number of parameters for acetamide (for one atom type pair) and NMA (for 3 atom type pairs) were derived using overlap data from formamide as the atoms are so buried in the larger molecules that insufficient close contacts can be sampled. These are shown in italics. Other than this, the parameters for each molecule were derived independently.

About half of the excess relative hydrogen bond length of the OM-models has therefore been eradicated in the KP-models.

3.5. Transferability. The potential parameters for the KP-ERP models for the three amides are given in Table 7. Given that the parameters were determined independently for each molecule, and that variations in the α_{ik} will produce compensating variations in A_{ik}, there is a reasonable similarity between the chemically similar atoms. Some of the variations will come from the fitting of an approximate functional form to a necessarily limited number of randomly generated atom−atom overlaps in the repulsive region, but there will also be some contribution from actual variations in the charge density associated with each atom type between the three molecules.

To test the significance of these variations in the parameters, we use the KP-ERP parameters derived for NMA to attempt to reproduce the crystal structures of formamide and both acetamide polymorphs. In formamide, the nonpolar hydrogen (H6) is given the parameters from the NMA methyl hydrogens (H8). This model is referred to as NM-ERP and was able to reproduce minima reasonably close to the experimental crystal structures

of formamide and both experimental polymorphs of acetamide (Tables 3−5). The NMA potentials are somewhat less repulsive than the formamide-based and acetamide-based equivalents, and thus predicted denser structures, shorter hydrogen bonds and more attractive lattice energies. This is reflected in the NM-ERP minimum for formamide, where the density is 17.4% too great, but the lattice energy is correctly predicted to within the expected experimental uncertainty. The relative hydrogen bond length is 1.062 and the RMS error in the cell edges 4.4%. The NM-ERP model correctly predicted the lattice energy of *R*3*c* acetamide to within the expected experimental uncertainty and gave an RMS error of only 3.0% in the cell edges. The density was only 1.2% too large and the relative hydrogen bond length 1.025, the best value found for any of our models. For the room-temperature structure of the *Pccn* polymorph of acetamide, the NM-ERP model predicted a density 9.0% too great and a relative hydrogen bond length of 1.032, with a 4.3% RMS error in the cell edges. Thus, overall quality of the crystal structure reproductions using the transferable potential was very encouraging.

3.6. Comparison with Other Amide Potentials. The errors in the reproduction of the crystal structures with these nonempirical potentials will arise from both inaccuracies in the model intermolecular potential and also approximations inherent in using lattice energy minimization to model the crystal. Thus, Tables 3−6 include a comparison with the crystal structure reproduction using the same minimization method and molecular and electrostatic models, but using repulsion−dispersion parameters (FIT)[41] that have been developed by empirical fitting to crystal structures and lattice energies. We find that only the most successful of our ab initio based models (KP-ERP in most cases) give as good a quality of crystal structure reproduction as FIT. Most of the nonempirical OM potentials give a larger overestimate of the density than can be attributed to the neglect of temperature effects. The lattice energies predicted by the nonempirical models which include penetration, or penetration and charge transfer, are generally of similar quality to those calculated using FIT, even though FIT had been empirically fitted to other lattice energies.

The nonempirical potentials can also be compared with the DISCOVER[57] and AMBER[58] potentials, using the results of Pillardy et al.[31] for their reproduction of the formamide crystal structure, albeit starting from a different determination of the crystal structure[59] and using a different minimization protocol. Minimizations using the DISCOVER potential reduced the symmetry from $P2_1/n$ to $P\bar{1}$, and lost the puckering of the sheets as the structure expanded in the *b*-direction and contracted in the *c*-direction. This corresponds to an RMS error in the cell lengths of 6.4% relative to FORMAM02, or 6.7% relative to their starting structure.[59] The overall density, however, is well predicted, with errors of +1.0% (FORMAM02) and −1.5% relative to the two crystal structure values. The hydrogen bond lengths predicted by DISCOVER were also too long, with a relative hydrogen bond length of approximately 1.024, but the calculated lattice energy of −68.5 kJ/mol is very reasonable. AMBER retained the $P2_1/n$ symmetry, but compressed the cell in all directions, with an RMS error of 8.3% relative to either crystal structure, giving a predicted packing which was too dense by 10.6% or 7.2%, and an excessively attractive lattice energy of −92.8 kJ/mol.

Thus, the OM-ERP potential for formamide, while inferior to FIT, is at least comparable in quality with DISCOVER and AMBER. The KP-ERP potential is comparable in quality to FIT and reproduces the formamide crystal structure better than either DISCOVER or AMBER.

4. Discussion

4.1. Value of Simple Overlap Model for Repulsive Potentials. For all three molecules, the overlap model produced repulsion potentials (OM-models) that gave a good description of the IMPT results, showing that the assumption that repulsion is proportional to overlap is a reasonable working approximation. The crystal structure predictions using the OM-models, where penetration or penetration and charge transfer were included via the proportionality constant, give quite reasonable predictions of the crystal structures compared with widely used current models. Since only one proportionality constant is involved, this could have been empirically fitted to the crystal structure data, to give results that should be no worse than those predicted by the nonempirical approach where the K constant was derived from IMPT calculations. The main deficiency of the crystal structure predictions was an overestimate of the hydrogen bond lengths. Thus, the basic overlap model appears to be very suitable for estimating repulsion potentials for molecules containing unusual functional groups, where no reliable empirical parameters are available.

4.2. Improvements on the Overlap Model Approximations. The errors in the OM repulsive potentials arise first from the proportionality assumption and second from the modeling of S_m by isotropic atom−atom exponential functions. These errors are generally comparable in magnitude. It would be possible to fit the overlap data better, for example, by using anisotropic atom−atom model potentials[27] to give a model potential that more accurately represented the calculated total overlaps. However, in this study we have investigated the limitation of the assumption of the overall proportionality between the total overlap and the repulsion energy (E_{er}, E_{erp}, or E_{erpc}, Table 1). It appears that the tendency to overestimate hydrogen bond lengths in the OM-models is due to approximating the variation of repulsion with overlap as $(S_\rho)^y$ with $y = 1$. We found that the optimum value of the power y in our KP-models was between 0.93 and 0.95 for the various amides for reproducing the IMPT data for repulsive energies less than about 100 kJ/mol. The OM assumption that $y = 1$ leads to a model that is too repulsive at the shortest separations and insufficiently repulsive at the larger separations around the minimum. Also, the IMPT data supported the use of a different proportionality constant for interactions involving oxygen atoms, introducing just one additional parameter, in contrast to the many that would be required if a different K were used for each pair of atom types. Combining these two refinements (separate K^O and K^X values and an approximate implementation of the power law) produced KP-models that gave very encouraging reproductions of the crystal structures, considerably reducing the excess relative hydrogen bond length.

4.3. Possible Further Improvements to the Overlap Model Methodology. Although the values of y seem reasonably transferable between these amides, it is not yet clear whether KP-models can produce parameters that are generally transferable to a range of molecules with greater accuracy than the simpler OM-models. Previous tests of the overlap model[25,26] have found $y < 1$, but it is a property of the molecules concerned[26] and possibly of basis set. There is generally close correspondence between the K values for different atom types, but all are subject to errors in the wave functions used to calculate the overlaps and in the IMPT repulsion energies. Indeed, the values of y which best described the dependency of the IMPT energies E_{er}, E_{erp}, and E_{erpc} on the accurately calculated

GMUL overlap (S_ρ) were generally somewhat smaller than those which optimized the KP-models (Table 1). Thus, the fitting process produces variations in y comparable to its deviation from unity; this is at least partly because larger values of y minimize the errors due to nonadditivity of the atom−atom terms. The IMPT calculations used as a benchmark for the short-range energies are limited in quality by the basis set and they give interaction energies at the SCF level. A considerable, but not unimaginable, increase in computational resources, would allow this methodology to be applied more accurately by using large basis set SAPT[60] calculations to calculate the components of the short-range energy, and to use more points on the potential energy surface to improve the fitting.

4.4. Improvements in the Representation of the Different Contributions to the Intermolecular Forces. An advantage of the nonempirical potential approach is that one can improve the theoretical basis and accuracy of the individual contributions to the intermolecular energy separately. In considering the ER, ERP, and ERPC models, we have shown that the penetration energy can be very effectively included in the model potential by reducing the proportionality constant in the overlap model, and that this is also a reasonable method of incorporating the charge-transfer term. Including these effects in the model potential is a theoretical improvement. Practically, the effects on the reproduction of the crystal structures are small compared with the errors in the lattice energy minimization method, but the addition of the penetration term does lead to better predictions of the lattice energies. The further addition of the charge-transfer term has a small effect on the crystal structure modeling, which is not always an improvement. This is consistent with the charge-transfer being less well represented by the overlap, and with the small magnitude of these terms. For example the component energies at the (ERP level) IMPT minimum for the formamide dimer (Figure 5, $d_{N\cdots O} = 3.09$ Å) are E_{es} (multipolar), −43.7; E_{er}, 30.3; E_{pen}, −7.3; E_{disp} (model), −17.2; E_{ct}, −5.6; E_{pol}, −7.1 kJ/mol. The polarization energy is omitted from our model potential, mainly because of its nonadditivity, but its inclusion would further stabilize the lattice.

There is some scope for improving the dispersion potential. We considered the possibility that the absence of dispersion terms higher than C_6 was contributing to the overestimation of relative hydrogen bond lengths, and investigated adding a C_8 term to an OM-ERP model of formamide. We found that, while the extra attractive energy compacted the structure, including the hydrogen bonds, the relative hydrogen bond lengths were virtually unchanged. Development of an effective multiplying function for the C_6R^{-6} terms to incorporate both the higher order terms C_8 and C_{10} terms and the damping effect, has been reported[61] for Ar···Ar, and will probably represent the best way forward for modeling dispersion in the future. However, our results with our best short-range model potentials are already sufficiently good, allowing for the other approximations used in the calculations (temperature and zero point motion effects) and experimental error, that a more accurate potential may not give significantly better results for the crystal structures. Similarly, the approximations in comparing the calculated lattice energies with heats of sublimation obscure evaluation of the cancellation of errors, such as between the absence of polarization and damping of dispersion.

5. Conclusions

We have developed a methodology for deriving model atom−atom intermolecular potentials from the wave functions of the isolated molecules, which can be used for organic molecules containing many different types of atoms. The resulting potentials reproduce the crystal structures and lattice energies of three amides with comparable accuracy to widely used empirically fitted model potentials, showing that the approach is suitable for obtaining model potentials for simulations.

The main novelty of these potentials is that the short-range repulsive terms are derived from considering the overlap of the molecular charge distribution, divided into atom−atom contributions by the GMUL partitioning of the molecular charge densities. Each set of atom−atom overlaps is fitted to give an atom−atom exponential model, without the need to assume transferability or combining rules to reduce the number of parameters. The fitting of the atom−atom overlaps provides some check on the adequacy of the functional form of the model potential. In this case, it seemed unlikely that using an anisotropic atom−atom repulsion model would give a significant increase in accuracy. We were also able to check that the Gaussian basis set used to calculate the overlaps adequately represented the exponential decay of the charge density.[11]

In the simplest version, the short-range potential is assumed to be proportional to the overlap, and thus the one proportionality constant can be fitted to experimental data, thereby effectively modeling all the contributions to the short-range energy. We have determined these constants by fitting to ab initio estimates of the exchange−repulsion, penetration, and charge-transfer energy at a limited number of points. This has confirmed that the exchange−repulsion[26] and penetration[28] energies are both very well correlated with the overlap, and that the overlap can also represent the charge-transfer energy. The resulting potentials are reasonable, except that the hydrogen bond lengths were overestimated.

Analysis of the IMPT potential energy surfaces showed that the inexactness of the proportionality between overlap and repulsion energy provides a limit to the accuracy attainable with the basic overlap model. Better potentials can be obtained by allowing some variation in the proportionality constant for different types of interaction (in this case, two types depending on whether the interaction involved oxygen) and considering a power law relationship. These completely ab initio derived potentials reproduce the amide crystal structures and energies as well as the model repulsion potentials that have been empirically fitted to such experimental data. The parameters also show a chemically reasonable degree of transferability, and hence this approach could be used to build up model potentials for a wide range of functional groups. Such potentials have the advantage over empirically fitted model potentials that all possible types of van der Waals contacts are equally treated, not just the types of contacts sampled in the empirical fitting. Thus, the repulsive potential model in less favorable regions of the potential energy surface should be better described than by the usual assumption of combining rules.

Although the derivation of the more accurate KO and KP model potentials does require about 20 points on the intermolecular potential energy surface to be calculated, this is orders of magnitude fewer than would be required for directly fitting a model potential to a surface involving so many types of interactions. Our methodology involves dividing the repulsion into atom−atom contributions via partitioning of the molecular charge density, and then separately fitting the exponential decay and preexponential constants for each atom−atom type. This avoids many of the problems involved in deriving atomic parameters for model repulsion potentials from calculated intermolecular potential energy surfaces, where it is very hard to decouple the parameters for different atom pairs.[10,29]

Calculating the overlap between two molecular charge densities,[9,11,14,27,30,62] is computationally very inexpensive relative to ab initio supermolecule,[10,16,63,64] density functional theory,[65] or intermolecular perturbation theory[12,13,29,66] (either IMPT[17] or SAPT[60]) energy evaluations. It is also competitive with other approximate methods[67,68] of estimating the interaction energy in the short-range region. This makes the overlap model particularly practicable for organic molecules. This is partly because of the relatively large size of the smallest model organic molecules needed for the assumption of transferability of parameters to be chemically reasonable. The cost of the overlap calculations for organic molecules the size of uracil could be significantly reduced by utilizing the full Gaussian multipole description only for those atom pairs in close contact.[22] Indeed, our results suggest that directly using the accurately calculated total overlap (the GM-ERP model) would estimate repulsion energies quite simply and relatively cheaply, which would have the advantage of eliminating the modeling of the overlap for computational approaches which did not require a simple model potential.

Thus, this study has further developed the overlap model as a systematic, nonempirical method of model potential derivation, by considering the inclusion of penetration and charge transfer, evaluating the effects of the power law governing the variation of repulsion energy with overlap, and comparing the results for a series of amide molecules. The results, along with those of the previous studies on oxalic acid,[11] an oxyboryl derivative[14] and smaller systems,[26,27,30] suggest that the approach will be very useful for model potential development for organic molecules. We are currently investigating its use for developing model potentials for chlorinated organic compounds that reflect the anisotropy in the repulsive wall.[69]

Acknowledgment. J.B.O.M. thanks AstraZeneca and Avecia for financial support under their Strategic Research Fund initiative. We thank Dr. David Buttar (Avecia), Dr. Ron Roberts (AstraZeneca), Dr. Richard Wheatley (Nottingham University), Dr. Irene Nobeli (UCL), Helen Tsui (UCL), and Graeme Day (UCL) for helpful discussions.

References and Notes

(1) Ng, K.-C.; Meath, W. J.; Allnatt, A. R. *Mol. Phys*. **1979**, *37*, 237.
(2) van Mourik, T.; Dunning, T. H. *J. Chem. Phys*. **1999**, *111*, 9248.
(3) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc*. **1974**, *96*, 5319.
(4) Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E. *Prot. Eng*. **1995**, *8*, 677.
(5) Muegge, I.; Martin, Y. C. *J. Med. Chem*. **1999**, *42*, 791.
(6) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. *J. Comput. Chem*. **1999**, *20*, 1165.
(7) Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol*. **2000**, *295*, 337.
(8) Millot, C.; Soetens, J.-C.; Martins Costa, M. T. C.; Hodges, M. P.; Stone, A. J. *J. Phys. Chem. A* **1998**, *102*, 754.
(9) Buck, U.; Siebers, J. G.; Wheatley, R. J. *J. Chem. Phys*. **1998**, *108*, 20.
(10) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.
(11) Nobeli, I.; Price, S. L. *J. Phys. Chem. A* **1999**, *103*, 6448.
(12) Cabaleiro-Lago, E. M.; Rios, M. A. *Mol. Phys*. **1999**, *96*, 309.
(13) Cabaleiro-Lago, E. M.; Rios, M. A. *J. Chem. Phys.* **1999**, *110*, 6782.
(14) Tsui, H. H. Y.; Price, S. L. *Cryst. Eng. Comm*. **1999**, 7.
(15) Stone, A. J.; Tong, C. S. *J. Comput. Chem*. **1994**, *15*, 1377.
(16) Brdarski, S.; Karlstrom, G. *J. Phys. Chem. A* **1998**, *102*, 8182.
(17) Hayes, I. C.; Stone, A. J. *Mol. Phys*. **1984**, *53*, 83.
(18) Stone, A. J. *Chem. Phys. Lett*. **1981**, *83*, 233.
(19) Le Sueur, C. R.; Stone, A. J. *Mol. Phys*. **1994**, *83*, 293.
(20) Stone, A. J. *Chem. Phys. Lett*. **1989**, *155*, 102.
(21) Mitchell, J. B. O. *Theor. Chim. Act*. **1996**, *94*, 287.
(22) Wheatley, R. J.; Mitchell, J. B. O. *J. Comput. Chem*. **1994**, *15*, 1187.
(23) Price, S. L. *Rev. Comput. Chem*. **2000**, *14*, 225.
(24) Fowler, P. W.; Lazzeretti, P.; Zanasi, R. *Mol. Phys*. **1989**, *68*, 853.
(25) Kim, Y. S.; Kim, S. K.; Lee, W. D. *Chem. Phys. Lett*. **1981**, *80*, 574.
(26) Wheatley, R. J.; Price, S. L. *Mol. Phys*. **1990**, *69*, 507.
(27) Nobeli, I.; Price, S. L.; Wheatley, R. J. *Mol. Phys*. **1998**, *95*, 525.
(28) Wheatley, R. J.; Meath, W. J. *Mol. Phys*. **1993**, *79*, 253.
(29) Fraschini, E.; Stone, A. J. *J. Comput. Chem*. **1998**, *19*, 847.
(30) Wheatley, R. J.; Price, S. L. *Mol. Phys*. **1990**, *71*, 1381.
(31) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski C.; Scheraga, H. A. *J. Am. Chem. Soc*. **2000**, *122*, 907.
(32) NIST Database http://webbook.nist.gov/.
(33) de Wit, H. G. M.; van Miltenburg, J. C.; de Kruif, C. G. *J. Chem. Thermodyn*. **1983**, *15*, 651.
(34) Roux, M. V.; Jiminez, P.; Davalos, J. Z.; Castano, O.; Molina, M. T.; Notario, R.; Herreros M.; Abboud, J.-L. M. *J. Am. Chem. Soc*. **1996**, *118*, 12735.
(35) Allen, F. H.; Kennard, O. *Chemical Design Automation News* **1993**, *8*, 1 and 31.
(36) Stevens, E. D. *Acta Crystallogr. B* **1978**, *34*, 544.
(37) Zobel, D.; Luger, P.; Dreissig, W.; Koritsanszky, T. *Acta Crystallogr. B* **1992**, *48*, 837.
(38) Hamilton, W. C. *Acta Crystallogr*. **1965**, *18*, 866.
(39) Hamzaoui, F.; Baert, F. *Acta Crystallogr. C* **1994**, *50*, 757.
(40) Willock, D. J.; Leslie, M.; Price, S. L.; Catlow, C. R. A. *J. Comput. Chem*. **1995**, *16*, 628.
(41) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. *J. Phys. Chem*. **1996**, *100*, 7352.
(42) Williams, D. E.; Cox, S. R. *Acta Crystallogr. B* **1984**, *40*, 404.
(43) Mitchell, J. B. O.; Price, S. L. *J. Comput. Chem.* **1990**, *11*, 1217.
(44) Beyer, T.; Price, S. L. *J. Phys. Chem. B* **2000**, *104*, 2647.
(45) Mooij, W. T. M.; van Eijck, B.; Price, S. L.; Verwer, P.; Kroon, J. *J. Comput. Chem.* **1998**, *19*, 459.
(46) Abramov, Y. A.; Volkov, A.; Wu, G.; Coppens, P. *J. Phys. Chem. B.* **2000**, *104*, 2183.
(47) Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdhry, B. Z.; Price, S. L. *J. Mol. Struct.* **1999**, *485*−*486*, 349.
(48) Aakeroy, C. B.; Nieuwenhuyzen, M.; Price, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 8986.
(49) Amos, R. D. with contributions from Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Koga, N.; Laidig, K. E.; Laming, G.; Lee, A.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M. D.; Tozer, D. J. *CADPAC6: The Cambridge Analytical Derivatives Package*; Cambridge University: Cambridge, 1995.
(50) Ioannou, A. G. Ph.D. Thesis, Cambridge University, 1998.
(51) Sadlej, A. *Theor. Chm. Act*. **1991**, *79*, 123.
(52) Price, S. L.; Stone, A. J.; Alderton, M. *Mol. Phys*. **1984**, *52*, 987.
(53) Stone, A. J. *Chem. Phys. Letts*. **1993**, *211*, 101.
(54) Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method*; Springer: Berlin, 1986.
(55) Gavezzotti, A.; Filippini, G. In *Theoretical Aspects and Computer Modelling of the Molecular Solid State*; Gavezzotti, A., Ed.; John Wiley & Sons: Chichester, 1997; Chapter 3, pp 61−97 (with particular reference to the appendix on p 97).
(56) Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K.; DeFrees, D. J.; Binkley, J. S.; Pople, J. A. *Acta Crystallogr. A* **1981**, *37*, C80.
(57) Ewig, C. S.; Thacher, T. S.; Hagler, A. T. *J. Phys. Chem. B* **1999**, *103*, 6998.
(58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc*. **1995**, *117*, 5179.
(59) Torrie, B. H.; O'Donovan, C.; Powell, B. M. *Mol. Phys*. **1994**, *82*, 643.
(60) Jeziorski, B.; Mosynski, R.; Szalewicz, K. *Chem. Rev*. **1994**, *94*, 1887.
(61) Hodges, M. P.; Stone, A. J. *Mol. Phys.* **2000**, *98*, 275.
(62) Beu, T. A.; Buck, U.; Siebers, J. G.; Wheatley, R. J. *J. Chem. Phys.* **1997**, *106*, 6795.
(63) van Mourik, T.; Price, S. L.; Clary, D. C. *J. Phys. Chem A* **1999**, *103*, 1611.
(64) Hobza, P.; Havlas, Z. *Theor. Chim. Act.* **1998**, *99*, 372.
(65) Mok, D. K. W.; Handy, N. C.; Amos, R. D. *Mol. Phys*. **1997**, *92*, 667.
(66) Cabaleiro-Lago, E. M.; Rios, M. A. *J. Chem. Phys.* **1998**, *109*, 8398.
(67) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Phys. Chem. A* **1996**, *105*, 1968.
(68) Sorescu, D. C.; Rice, B. M.; Thompson, D. L. *J. Phys. Chem. B* **1997**, *101*, 798.
(69) Price, S. L.; Stone, A. J.; Lucas, J.; Rowland, R. S.; Thornley, A. E. *J. Am. Chem. Soc*. **1994**, *116*, 4910.
(70) *Microsoft Excel 97 SR-1*; Microsoft Corporation: Redmond, WA, 1985−1997.