

# The Nature of the $\text{N}-\text{H}\cdots\text{O}=\text{C}$ Hydrogen Bond: An Intermolecular Perturbation Theory Study of the Formamide/Formaldehyde Complex

John B. O. Mitchell and Sarah L. Price\*<sup>†</sup>

University Chemical Laboratory, Lensfield Rd., Cambridge CB2 1EW, United Kingdom

Received 14 March 1990; accepted 18 June, 1990

We have used Hayes-Stone Intermolecular Perturbation Theory (IMPT) to study the variation with distance and orientation of the various components of the interaction energy of the  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonded trans-formamide/formaldehyde complex, a model system for hydrogen bonding in proteins. The directionality of the total interaction energy is similar to that of the electrostatic component alone. We have analysed our data in terms of two model atom-atom intermolecular potentials, using an isotropic functional form and an anisotropic one. The anisotropic form gives an excellent representation of the IMPT potential energy surface, considerably better than the isotropic model, and is comprised entirely of theoretically justified, physically meaningful terms.

## INTRODUCTION

The crucial role played by  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonds in determining the structures and properties of proteins and nucleic acids means that a quantitative understanding of  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonding is of great importance, not only to chemists, but also to molecular biologists<sup>1</sup> and pharmacologists.<sup>2</sup> Such a quantitative understanding, once expressed in terms of the functional form and parameters of a model interaction potential, is the major scientific input into computer simulations. Model potentials have usually been based on Lennard-Jones 12-6 or similar isotropic atom-atom repulsion-dispersion terms, plus an atomic point charge electrostatic model. Such models take no account of the nonsphericity of the electron density around the nuclei in molecules. Their potential parameters are optimized by fitting to experimental<sup>3,4</sup> or *ab initio*<sup>5</sup> data. In order to describe hydrogen bonding, empirical adjustments to the functional form, such as enforcing "lone pair" electrostatic directionality,<sup>4</sup> introducing an empirical hydrogen bonding term,<sup>6</sup> or ignoring the repulsion between the hydrogen and its acceptor<sup>7,8</sup> have often been used. It should, however, be possible to describe hydrogen bonding in a much more satisfactory and theoretically justified way by using accurate models for the behavior of the electrostatic and other components

of the potential. For instance Reiher<sup>5</sup> used accurate *ab initio* calculations to adjust the functional form and parameters of the CHARMM force-field,<sup>6</sup> and improved the description of hydrogen bonded systems while removing the empirical hydrogen bonding term. In recent years, the theory of intermolecular forces has been developed<sup>9,10</sup> beyond the isotropic atom-atom model and applied to small molecules to give anisotropic atom-atom forms for all the major contributions to the potential.<sup>11</sup> The challenge now is to apply this knowledge to produce nonempirical, theoretically justified, accurate yet computationally convenient model potentials to describe the interactions between biological molecules.

It is not, at present, possible to perform accurate *ab initio* calculations of the interactions between large molecules at sufficiently many points to characterize the potential energy surface. Theoretical studies of  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonds are, therefore, restricted to smaller systems which contain the same functional groups in similar bonding environments, thus retaining the essential chemistry. Small protein-like molecules, such as single amino acids or dipeptides, will be good models and have been used in studies of protein hydration.<sup>12-14</sup> However, we cannot study  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonding by using *ab initio* calculations of the interaction of a pair of small peptides, as this would need unobtainable computational resources, particularly since we require many points on the potential energy surface. The formamide molecule, with an almost planar OCNH group, contains the essential features of the peptide linkage and is often

\*To whom all correspondence should be addressed.

<sup>†</sup>Present address: Department of Chemistry, University College London, 20 Gordon St, London WC1H 0AJ, U.K.

used in its modeling. It is not, however, a perfect model since the nature of the bonded environment affects the properties of atoms and groups in molecules,<sup>15-18</sup> as one would expect from the elementary principles of organic chemistry. Calculations on the formamide dimer<sup>5,19</sup> and the formamide/formaldehyde complex<sup>20</sup> have previously given some insight into the properties of the N—H···O=C hydrogen bonds found in proteins, and in this study we use the formamide/formaldehyde complex as our model system.

## PART I: HAYES–STONE IMPT CALCULATIONS

### Hayes–Stone Intermolecular Perturbation Theory (IMPT)

#### *Perturbation Theory and Intermolecular Forces*

We chose to use perturbation theory rather than the supermolecule approach because it partitions the energy into its components (supermolecule methods can only do this by Morokuma analysis,<sup>21,22</sup> which adds considerably to the computational cost), is considerably cheaper (requiring only about a quarter of the time of a supermolecule calculation, even at the uncorrelated level), and can easily and economically take account of electron correlation, and hence the dispersion interaction.

There is no unique formulation of perturbation theory for the study of intermolecular forces. The simplest approach is the Rayleigh–Schrödinger expansion, which is complete only at long range, giving at first order simply the electrostatic energy and, at second order, polarization and dispersion terms. However, to study hydrogen bonding we require a theory that will be valid not just at long range, but also in the region of the potential well. We would like such an intermediate range method to be accurate over the whole of the region where the potential is attractive.<sup>23</sup> Closer in than this, we will reach a region where the interaction is not small and so perturbation theory, based on the reference state of separated monomers, can never be valid. Thus a suitable theory will accurately describe the potential in the whole of the region where the overlap of the two monomer charge distributions is either small or moderate. This gives rise to two problems:<sup>24,25</sup> firstly the nonorthogonality of the orbitals from the different fragments (*A* and *B*) means that they cannot be eigenfunctions of a single unperturbed supersystem Hamiltonian, and secondly the supersystem orbitals have to be made antisymmetric with respect to any permutation of electrons. There are many proposals in the literature for tackling these problems and the relation-

ship of our chosen approach, Hayes–Stone IMPT, to other methods has been discussed by Hayes, Stone, and Hurst.<sup>24-26</sup> Hayes–Stone IMPT uses determinantal wavefunctions to ensure antisymmetry and works directly with the nonorthogonal orbitals of the monomer calculations. Using these nonorthogonal functions entails some elaborate matrix algebra<sup>27</sup> and also means that the Hamiltonian cannot be simply broken down into a zeroth-order reference part and a perturbation. However, the zeroth-order energy is simply defined as the sum of the separated monomer energies. The theory can be written down either in terms of an expansion of the secular determinant or as a more conventional perturbation expansion.<sup>24-26</sup> The IMPT method has previously been successfully applied to a number of chemical systems.<sup>28,29</sup> The mathematical and computational details have been dealt with by Hayes and Stone<sup>24</sup> and will not be repeated here. What follows is intended as a brief explanation of the physical significance of each term.

#### *First-Order Terms*

**Electrostatic Energy (ES).** The electrostatic energy is simply the classical Coulombic interaction energy between the two undistorted monomer charge distributions (including both electrons and nuclei)

$$E_{ES} = \iint \rho^A(\mathbf{r}_1)\rho^B(\mathbf{r}_2)|\mathbf{r}_1 - \mathbf{r}_2|^{-1}d\mathbf{r}_1d\mathbf{r}_2$$

where  $\rho^A(\mathbf{r}_1)$  and  $\rho^B(\mathbf{r}_2)$  are the two fragment charge distributions.

At long range this energy can be well described using distributed multipole analysis (DMA),<sup>30,31</sup> as indeed it could be by any of several multicentre multipole expansions in the literature.<sup>32-41</sup> It is pairwise additive and can be attractive or repulsive, although for formamide/formaldehyde in the configurations considered here it is always attractive. At short range, a penetration term<sup>23,42</sup> must be added to the DMA electrostatic energy to describe the classical effect of the interpenetration of the charge clouds (but not the Pauli exclusion principle type repulsion, which is found in the exchange-repulsion energy). The penetration decays approximately exponentially with distance (i.e., much faster than the  $1/R^n$  multipolar long range term). It can be either attractive or repulsive, although in the work reported here it is almost always attractive.

**Exchange-Repulsion Energy (EXREP).** The exchange-repulsion energy is a short-range term made up of two components. The first of these is the exchange energy, a quantum mechanical energy lowering due to the possibility of the exchange of electrons of parallel spin between the

two molecules. It is given by

$$E_{EX} = - (1/2) \sum_{ab} \langle \phi_a(\mathbf{r}_1) \phi_b(\mathbf{r}_2) | 1/r_{12} | \phi_b(\mathbf{r}_1) \phi_a(\mathbf{r}_2) \rangle,$$

the quantity in Dirac brackets being an exchange integral between orbital  $a$  on fragment  $A$  and orbital  $b$  on fragment  $B$ . The exchange energy is attractive and pairwise additive.<sup>24</sup>

The second component, the repulsion energy, is a consequence of the Pauli exclusion principle preventing electrons with parallel spins from occupying the same region of space. It is non-additive, repulsive and about twice as large as the exchange energy. The two components are usually considered together as their sum. This sum, the exchange-repulsion energy, is nonadditive, anisotropic<sup>43</sup> and repulsive.

### Second-Order Single Excitation Terms

The second-order single excitation terms can be calculated with negligible extra computational effort once the first order energy has been obtained. These terms are of two kinds, polarization (induction) and charge transfer. If  $A$  or  $B$  denote orbitals belonging to fragment  $A$  or  $B$  respectively, then polarization is an  $A \rightarrow A$  or  $B \rightarrow B$  excitation and charge transfer is an  $A \rightarrow B$  or  $B \rightarrow A$  excitation. In the limit of a complete basis set, the charge transfer terms would become spurious<sup>24,44</sup> since the basis functions of one fragment could describe the orbitals of the other, and thus all single excitations from occupied to virtual orbitals could be described as polarization. Nevertheless, the separation of the terms is well defined by IMPT, within a given basis set, and ascribes all the basis set superposition error (BSSE) at this level of theory to the charge transfer term. In addition, the IMPT polarization energy corresponds fairly well to that calculated by other methods (later section). Hence we have retained the separate identities of the polarization and charge transfer terms.

**Polarization (Induction) Energy (PL).** Excitations from an occupied to a virtual orbital on the same fragment, i.e.,  $A \rightarrow A$  or  $B \rightarrow B$  excitations, give rise to the polarization energy. This term describes the response of one monomer to the electric field of the other. Since it describes a distortion of the charge distribution which would not occur unless it were energetically favorable, the polarization energy is always attractive, as one would expect from a second-order perturbation theory term. It is nonadditive (polarization energy is quadratic in the local electric field<sup>45</sup>) and is subject to some short range modification due to overlap. A further technical point is that Hayes–Stone IMPT is an uncoupled perturbation theory and so does not allow for the perturbation changing the zeroth-order Fock operator.<sup>23</sup> This

causes the polarization (and also the dispersion) energy to be slightly inaccurate.

**Charge Transfer Energy (CT).** The charge transfer term comprises excitations from an occupied orbital on one fragment to a virtual orbital on the other i.e., an  $A \rightarrow B$  or  $B \rightarrow A$  excitation. In practice, we find that some charge transfer occurs in each direction, both terms contributing to the interaction energy even if the net transfer of charge between the fragments is small. Charge transfer energies are attractive.

The charge transfer term contains all the BSSE up to the second order single excitation level. An approximate correction for this BSSE can be made by performing a calculation on fragment  $A$  in which the virtual orbitals of fragment  $B$  are available (and also the corresponding calculation on the other fragment). There are two possible ways of evaluating the correction.<sup>24,26</sup> One [denoted by (F) for ‘Fock’] assigns to the virtual orbitals of fragment  $B$  an energy corresponding to the expectation value of the Fock operator for fragment  $A$ . The other [denoted by (E) for ‘Eigenvalue’] assigns them the appropriate eigenvalues from the fragment  $B$  monomer calculation. The corrections are of similar magnitude, though those of (E) are generally somewhat larger and are possibly overestimates since, on a few occasions, the charge transfer energy in one direction would become repulsive if correction (E) were applied. Thus we use correction (F) for the main part of this work.

### Second-Order Double Excitation Terms

The calculation of second-order double excitation terms requires an approximate doubling of the computational expense required to evaluate the single excitation terms (1200 rather than 600 seconds CPU time on the Convex-C210). We will use an extension of the previous notation<sup>24</sup> to describe the terms; for instance  $AB \rightarrow AA$  is an excitation of one electron in an occupied orbital of fragment  $A$  and one electron in an occupied orbital of fragment  $B$  to virtual orbitals of fragment  $A$ . There are five possible kinds of term, two of which are real contributions to the interaction energy and are included in these calculations. These two terms are free of BSSE.<sup>24</sup>

**Dispersion  $AB \rightarrow AB$  (DISP).** The dispersion energy describes the interactions due to correlated fluctuations in the charge distributions of the two monomers. It is the only interaction between two neutral spherical atoms at long range. The interaction between two atoms is a sum of terms of the form  $C_n R_{ab}^{-n}$  with the leading term being the one in  $R_{ab}^{-6}$ . The form is modified by the overlap of the charge distributions at short range.<sup>24</sup> It is attractive and pairwise additive. The

dispersion energy will be underestimated by all but the largest basis sets, though the inclusion of the charge transfer correlation term tends to compensate for this, and it will also be somewhat affected by the uncoupled nature of Hayes-Stone IMPT.

**Charge Transfer Correlation  $AB \rightarrow AA$  or  $AB \rightarrow BB$  (CTC).** This part of the interaction involves the correlation of a transferred electron with an electron in the acceptor fragment. In the limit of a complete basis set, this would merely be counting again parts of the interaction which had already been considered as dispersion, exactly analogously to the situation with charge transfer and polarization. However, for 6-31G\* and smaller basis sets, it is believed that, within the framework of Hayes-Stone IMPT, one should include charge transfer correlation explicitly.<sup>46</sup>

The three remaining kinds of term are all "spurious"<sup>24</sup> in the sense that they do not represent part of the intermolecular interaction, but describe intramolecular correlation and BSSE type effects. However, the improvement in the description of the monomer charge distribution resulting from the inclusion of intramolecular correlation would affect the electrostatic and exchange energies.<sup>47</sup> We have not taken this effect into account.

In the work reported here, we consider neither the effects of triple excitations at second order (not necessarily zero, due to nonorthogonality, but inevitably very small<sup>25</sup>), nor of third or higher orders of perturbation theory (which we can safely neglect since Hayes<sup>25</sup> found them to give very small contributions so long as Møller-Plesset denominators are used, which is the case here).

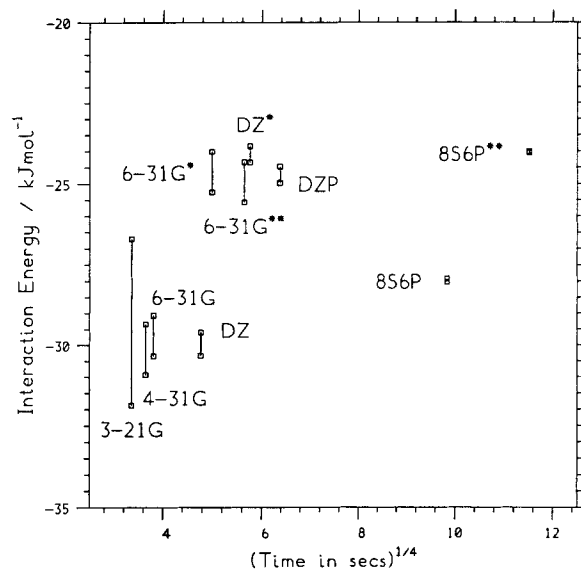
### Choice of System and Basis Set and a Comparison with Supermolecule Results

The hydrogen bond in the formamide/formaldehyde complex was chosen as a model of the  $N-H \cdots O=C$  linkages found in proteins. The trans conformation, in which the hydrogen bond is the only intermolecular van der Waals contact, is a simple model for peptide hydrogen bonds, with a planar OCNH unit hydrogen bonded to a carbonyl group. Since there are no secondary intermolecular contacts, one can approach the unattainable goal of abstracting the properties of the notional "isolated  $N-H \cdots O=C$  hydrogen bond." These properties should be transferable to systems with various secondary contacts, provided that these other interactions are treated separately. The formamide/formaldehyde complex has a significant economic advantage over larger systems; for instance an IMPT calculation including dispersion at the 6-31G\* level is a factor of about three cheaper than for the next smallest possible model, the formamide dimer.

### Choice of Basis Set

The choice of basis set was made by comparing the IMPT interaction energies obtained for the cis hydrogen bonded conformation (at the geometry predicted by the Buckingham-Fowler electrostatic plus hard sphere model<sup>48</sup>) using a variety of basis sets. This structure has two separate intermolecular contacts fairly rigidly determining its geometry. This is thus essentially basis set independent, unlike the geometry of the single-contact trans conformation on which the subsequent calculations were performed. We found that the attractive interaction energy decreases with increasing basis size, even when basis set superposition error (BSSE) corrections are applied, as shown in Figure 1. The addition of *d* polarization functions to heavy atoms makes the interaction energy significantly less attractive in all cases. Subsequent addition of *p* polarization functions to the hydrogens has a much smaller effect (in the opposite direction). We chose to use a 6-31G\* basis set for the investigation of the variation of the interaction energy with intermolecular distance

Basis Set Dependence of Interaction Energy



**Figure 1.** IMPT interaction energy to second-order single excitation level plotted as a function of the fourth root of the total CPU time taken for the calculation. Since the time depends on the number of basis functions in an approximate  $N^4$  manner, this fourth root is a measure of basis size. Its advantage is that it distinguishes between basis sets with an equal number of contracted functions, such as 3-21G, 4-31G, and 6-31G. For each basis set, the upper square on the diagram represents the result obtained using BSSE correction (E) and the lower square that obtained with correction (F). "DZ\*" is a double zeta basis set with polarization functions on heavy atoms only. "8S6P\*\*", the largest basis set used, is 13s8p/8s6p on heavy atoms, 10s/6s on hydrogen with polarization functions on all atoms. All basis sets were taken from the CAD-PAC library.<sup>52</sup>

and orientation. This gave a very similar interaction energy to larger basis sets in the test calculations while being small enough to allow many points on the potential energy surface to be calculated.

### Comparison of IMPT with Supermolecule SCF Results

The IMPT results up to the second-order single excitation (i.e., uncorrelated) level can reasonably be compared with the results of SCF (self-consistent field) supermolecule calculations. We have done this at the lowest energy linearly hydrogen bonded configuration (see later section) of the formamide/formaldehyde complex. The results, shown in Table I, provide an interesting comparison of the IMPT BSSE corrections with the Boys–Bernardi counterpoise method<sup>49</sup> used to correct supermolecule calculations for BSSE.

These data show that the (F) corrected IMPT and counterpoise corrected supermolecule results are very similar, differing by just 0.65 kJ/mol. The BSSE corrections used in the two methods are somewhat different—the counterpoise correction includes both occupied and virtual orbitals as ghosts, whereas corrections (F) and (E) include only virtual orbitals. The inclusion of occupied orbitals among the ghosts in the counterpoise correction has usually been taken to mean that it overestimates the BSSE, though there has been some debate about this in the literature.<sup>50,51</sup> It is encouraging that both corrected IMPT calculations predict a larger interaction energy than does the counterpoise corrected supermolecule method (which probably overestimates BSSE and so underestimates the magnitude of the interaction energy) and a weaker interaction than the uncorrected supermolecule method (which ignores BSSE and so overestimates the magnitude of the interaction energy).

The validity of IMPT at short range was tested by means of a similar comparison with supermolecule calculations, carried out for smaller  $d_{\text{NO}}$  values. This has shown that, at least at the non-correlated level, IMPT is still reliable as far in as 2.55 Å; all three IMPT results (uncorrected and

with each BSSE correction) lay between the uncorrected and the counterpoise corrected supermolecule values. Thus we feel confident that our results are meaningful for  $d_{\text{NO}} \geq 2.55$  Å.

We cannot test the accuracy of the dispersion and charge transfer correlation terms against correlated supermolecule calculations, as this would be prohibitively expensive. Indeed, although our calculations probably give the best available estimate of the dispersion energy contribution in this system, we expect that the description of the double excitation part is the largest source of uncertainty in the total interaction energy. Other inaccuracies in the IMPT results may result from imperfect BSSE corrections, possible double counting in the charge transfer and charge transfer correlation terms, omission of third- and higher-order contributions and the uncoupled nature of the perturbation theory.

## Results and Discussion

The following results were obtained, unless explicitly stated otherwise, from Hayes–Stone IMPT calculations including BSSE correction (F) and the ('genuine') second-order double excitation terms (dispersion and charge transfer correlation) performed with 6-31G\* basis sets using the *ab initio* package CADPAC.<sup>52</sup> The basis sets were taken from the CADPAC library. Throughout this work, the monomers were treated as rigid.

The relative orientation of the two monomers is defined by the angles  $\alpha$  and  $\beta$  (Figs. 2(a) and 2(b)) and  $d_{\text{NO}}$ , the intermolecular nitrogen-oxygen distance in the hydrogen bond, with the N—H···O linkage assumed (except where indicated otherwise) to be linear.

### Locating Minimum Energy Structures

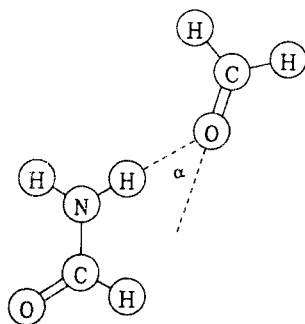
We minimized the energy with respect to variations in  $d_{\text{NO}}$  (to an accuracy of  $\pm 0.01$  Å) and  $\alpha$  (to  $\pm 1^\circ$ ) by a grid searching technique to find the lowest energy structure, subject to the constraints of linearity and planarity ( $\beta = 0^\circ$ ), for the trans configuration of the formamide/formaldehyde complex. This was found to have

**Table I.** IMPT and supermolecule methods compared for the trans-formamide/formaldehyde complex.<sup>a</sup>

Method	BSSE correction	Interaction energy
IMPT 2nd order	none	−13.65 kJ/mol
IMPT 2nd order	(F)	−13.11 kJ/mol
IMPT 2nd order	(E)	−12.68 kJ/mol
Supermolecule	none	−15.61 kJ/mol
Supermolecule	counterpoise	−12.46 kJ/mol

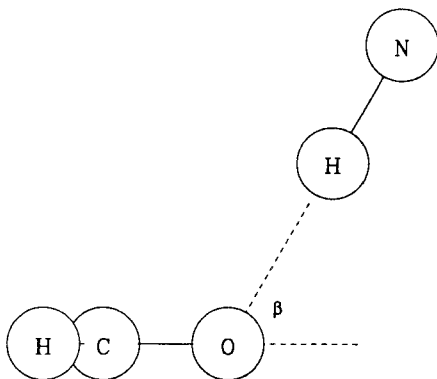
<sup>a</sup>The IMPT calculations were performed at the second-order single excitation level and the supermolecule calculations at the SCF level, both with 6-31G\* basis sets. The supermolecule facility is available along with IMPT as part of CADPAC.<sup>52</sup>

## Definition of Alpha



**Figure 2(a).** The in plane angle  $\alpha$  used to describe the orientation of the formamide/formaldehyde complex. This figure shows a positive value of  $\alpha$ , in fact  $\alpha = 42^\circ$ . We can extend the definition of  $\alpha$  from planar to non-planar cases by projecting onto the x-y plane.

## Definition of Beta



**Figure 2(b).** The angle,  $\beta$ , of elevation of the hydrogen bonding hydrogen above the carbonyl plane, subtended at the oxygen. This oxygen is also in the plane of the formamide molecule (except in the  $\alpha$ -helix model). This angle is that defined by Baker and Hubbard<sup>55</sup> and is related to Taylor et al.'s  $\theta_H$ <sup>75</sup> by  $\beta = 90^\circ - \theta_H$ .

$d_{NO} = 2.89 \text{ \AA}$  and  $\alpha = 42^\circ$ . We then relaxed the constraints and found that the minimum energy was now obtained with a planar but nonlinear structure having an  $\angle HNO$  angle of  $6^\circ$  and an  $\angle NHO$  angle of  $170.8^\circ$ . Though a small nonlinearity is favorable, the linear structure has a total interaction energy ( $-27.07 \text{ kJ/mol}$ ) very similar to that of the nonlinear one ( $-27.34 \text{ kJ/mol}$ ). The values of the various energy terms at the nonlinear minimum are given in Table II. The electrostatic and exchange-repulsion terms are by far

the largest in magnitude, but the other components are not negligible. The exchange-repulsion is almost equal in magnitude to the sum of the penetration energy and the nonelectrostatic attractive contributions, so that the DMA electrostatic term is very similar to the total interaction energy; this is the case at points with  $d_{NO}$  close to the optimum value of  $2.89 \text{ \AA}$  and also at long range ( $d_{NO} \geq 6 \text{ \AA}$ ) where all terms except the electrostatic one are close to zero.

We have calculated the position of the minimum in the formamide/formaldehyde interaction surface, subject to the constraints of linearity and planarity, at several levels of theory. The results in Table III show that the polarization, charge transfer and charge transfer correlation terms have a marked effect on the optimum value of  $\alpha$ , and are as important as the dispersion term in overcoming the repulsion to give a short  $N \cdots O$  distance.

It is interesting to note that an isotropic atom-atom model including point charges, based on a parameterization<sup>5</sup> to *ab initio* supermolecule calculations on hydrogen bonded complexes, gives a similar (incorrect)  $\alpha$ -dependence to that obtained by truncating the IMPT expansion at first order, with a single minimum near  $\alpha = 0^\circ$ . However, this model potential does give good values for the optimum  $d_{NO}$  and the total interaction energy.

## The Effects of Variations in the Geometry

## Directionality Within the Carbonyl Plane.

Figures 3(a) to (d) show the variations of the various energy terms with the angle  $\alpha$  (for  $\alpha = -80^\circ$  to  $+80^\circ$ ) at  $d_{NO} = 2.89 \text{ \AA}$ , for the planar complex with the  $N-H \cdots O$  unit kept linear. The most striking feature is the flatness of the total energy as a function of  $\alpha$ , varying by no more than  $1 \text{ kJ/mol}$  over a range of about  $110^\circ$  roughly corresponding to the region between the two  $sp^2$  oxygen lone pairs, and then rising sharply at larger  $|\alpha|$ . To a reasonable approximation, the angular variation of the total energy mirrors that of the electrostatic term, with a distinctive double minimum structure. This domination of the electrostatic term, combined with the comparative flatness of the overall angular potential, is in good agreement with the IMPT results obtained

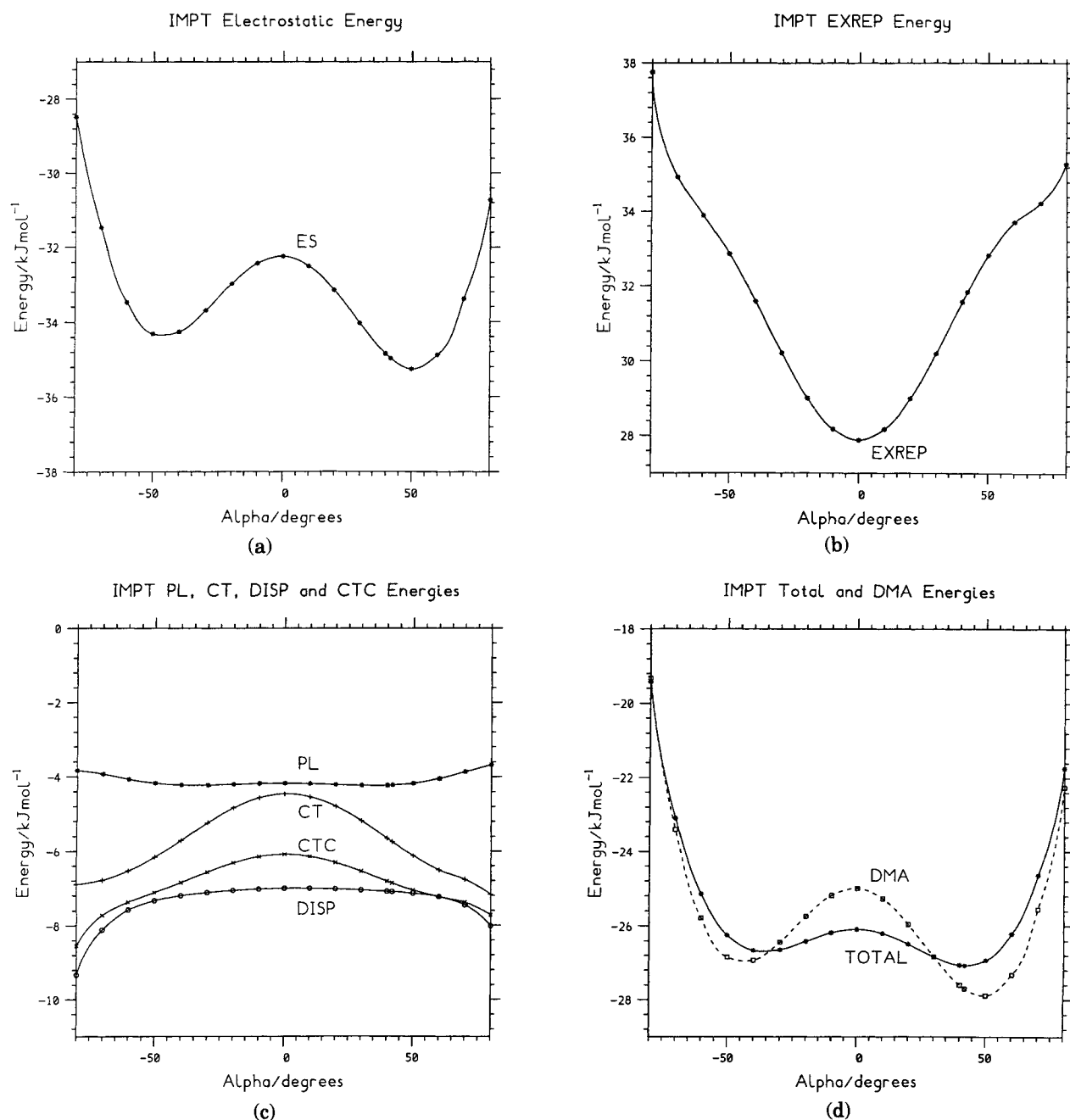
**Table II.** Components of the trans-formamide(A)/formaldehyde(B) interaction energy (kJ/mol) at the hydrogen bonded minimum.

ES	=	-34.72	=	-27.67 (DMA)	-	7.06 (pen)
EXREP	=	30.99	=	-33.93 (exch)	+	64.92 (rep)
PL	=	-4.13	=	-2.15 (A $\rightarrow$ A)	-	1.98 (B $\rightarrow$ B)
CT	=	-5.76	=	-0.89 (A $\rightarrow$ B)	-	4.87 (B $\rightarrow$ A)
DISP	=	-7.01				
CTC	=	-6.70	=	-4.03 (AB $\rightarrow$ AA)	-	2.67 (AB $\rightarrow$ BB)
TOTAL	=	-27.34				

**Table III.** Constrained minima at different levels of theory.<sup>a</sup>

	$d_{\text{NO}}/\text{\AA}$	$\alpha/^\circ$	$E/\text{kJmol}^{-1}$
IMPT 6-31G* 1st order	3.34	4	-13.82
2nd order SE(0)	3.20	35	-18.20
2nd order SE(F)	3.19	32	-17.71
2nd order SE(E)	3.19	29	-17.31
2nd order SE(F) + DISP	3.06	33	-21.88
2nd order SE(E) + DISP	3.06	31	-21.50
2nd order SE(F) + DISP + CTC	2.89	42	-27.07
2nd order SE(E) + DISP + CTC	2.89	42	-26.64
LJ + C (12-6 + point charges)	2.94	-2	-26.99

<sup>a</sup>“(0)” means no BSSE correction, ‘(F)’ means including BSSE correction (F), ‘(E)’ means including BSSE correction (E). SE means single excitations, ‘+DISP’ means including dispersion, ‘+DISP + CTC’ means including dispersion and charge transfer correlation. ‘LJ + C’ is a Lennard-Jones 12-6 plus point charge model based on Reiher’s reparameterization of CHARMM.<sup>5</sup> For each minimum,  $d_{\text{NO}}$  is found to an accuracy of  $\pm 0.01 \text{ \AA}$  and  $\alpha$  to  $\pm 1^\circ$ .

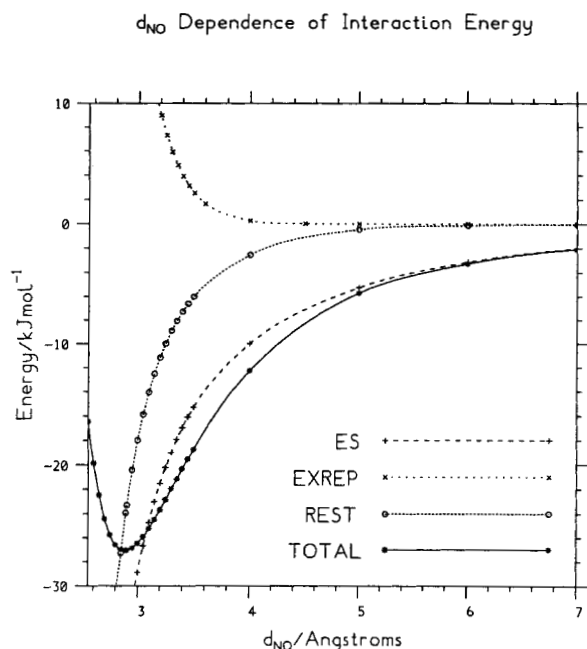


**Figures 3(a) to (d).** The electrostatic (3(a)); exchange-repulsion (3(b)); polarization, charge transfer, dispersion and charge transfer correlation (3(c)); and total (3(d)) IMPT interaction energies as functions of  $\alpha$  for planar, linear hydrogen bonds with  $d_{\text{NO}} = 2.89 \text{ \AA}$ . The broken line in Figure 3(d) represents the DMA contribution to the electrostatic energy.

by Hurst et al. for a variety of hydrogen bonded complexes.<sup>29</sup> A similar result, the angular variation of the total energy being essentially that of the electrostatic component, was found by Umeyama and Morokuma<sup>53</sup> for more strongly hydrogen bonded systems, such as  $(\text{HF})_2$  and  $(\text{H}_2\text{O})_2$ , using the supermolecule based method of Morokuma analysis.<sup>21,22</sup> The exchange-repulsion term strongly favors angles close to  $\alpha = 0^\circ$ , while the charge transfer and charge transfer correlation terms disfavor (though to a lesser extent) such orientations relative to those with extreme values of  $\alpha$ . The dispersion energy is almost flat over the region  $\alpha = -40^\circ$  to  $+40^\circ$ , but is more attractive for large  $|\alpha|$  values. The polarization energy is extremely flat, though in fact it has a double minimum. The electrostatic energy has minima at  $\alpha = -46^\circ$  and  $+51^\circ$ , of which that at  $\alpha = +51^\circ$  is the lower, and a central maximum at  $\alpha = -1^\circ$ , 3.03 kJ/mol above the lower minimum. The electrostatic energy becomes much less attractive at extreme values of  $\alpha$ . To quantify our observation that the total energy is very flat as a function of  $\alpha$ , we note that it has the lower of its two minima ( $-27.07$  kJ/mol) at  $\alpha = +42^\circ$  and the other ( $-26.69$  kJ/mol) at  $\alpha = -36^\circ$ . The central maximum ( $-26.09$  kJ/mol), at  $\alpha = 0^\circ$ , is only 0.98 kJ/mol above the minimum. This shows that the  $\alpha$  directionality of  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonds could be changed to fit the steric requirements of a given system, or to optimize interaction energies between other parts of the molecules, at very little cost to the hydrogen bonding part of the interaction energy.<sup>48</sup>

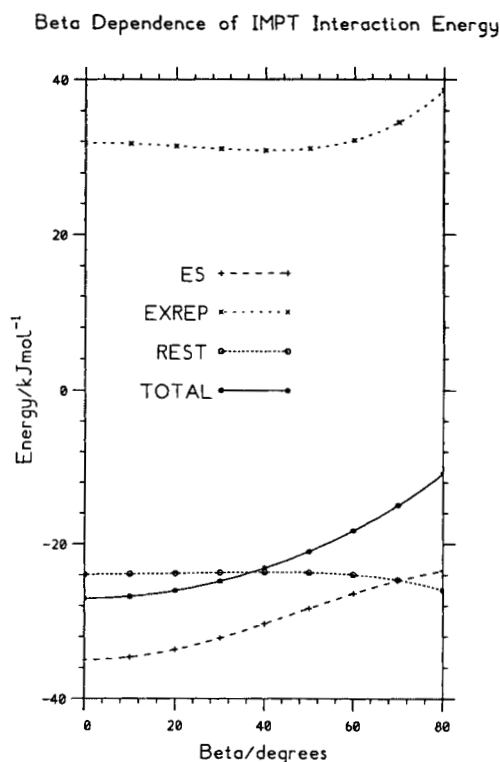
**Variation with Intermolecular Separation.** Figure 4 shows the total IMPT energy as a function of  $d_{\text{NO}}$  at  $\alpha = 42^\circ$  for the planar complex with the  $\text{N}-\text{H}\cdots\text{O}$  atoms collinear. The general form of the variation with distance is the familiar potential well. The total interaction energy tends slowly to zero from below at long range, where the multipolar electrostatic energy dominates, and increases rapidly at short range, where the exchange-repulsion dominates. The sensitivity of the optimum  $d_{\text{NO}}$  value to the level of theory is apparent. Going from first-order to second-order moves the minimum significantly inwards towards shorter  $d_{\text{NO}}$  distances because the additional terms are increasingly attractive as  $d_{\text{NO}}$  is reduced. The  $d_{\text{NO}}$  value predicted when double excitation terms are included, 2.89 Å, is in excellent agreement with the average experimental  $d_{\text{NO}}$  values for intermolecular hydrogen bonds in crystal structures of small molecules, 2.89 Å,<sup>54</sup> and for protein hydrogen bonds in  $\beta$ -sheets, 2.91 Å.<sup>55</sup>

**The Effect of Nonplanarity.** We have investigated the favorability of nonplanarity by distorting the structure so that the complex is no longer planar, changing  $\beta$  while keeping  $\alpha$  and



**Figure 4.** The variation of the IMPT interaction energy with  $d_{\text{NO}}$ . The behavior of the total energy and of the principal components is shown. ( $\alpha = 42^\circ$ ,  $\beta = 0^\circ$ ,  $\omega = 0^\circ$ ).

$d_{\text{NO}}$  fixed at  $42^\circ$  and 2.89 Å respectively and the  $\text{N}-\text{H}\cdots\text{O}$  unit linear. The results are shown in Figure 5. We found that deviation from planarity ( $\beta = 0^\circ$ ) was unfavorable, with a change in total energy of 2.25 kJ/mol for a deviation of  $46^\circ$ , primarily due to a reduction in the large and attractive electrostatic energy. The electrostatic term



**Figure 5.** The total IMPT interaction energy as a function of the angle of nonplanarity,  $\beta$  ( $\alpha = 42^\circ$ ,  $d_{\text{NO}} = 2.89$  Å,  $\omega = 0^\circ$ ).



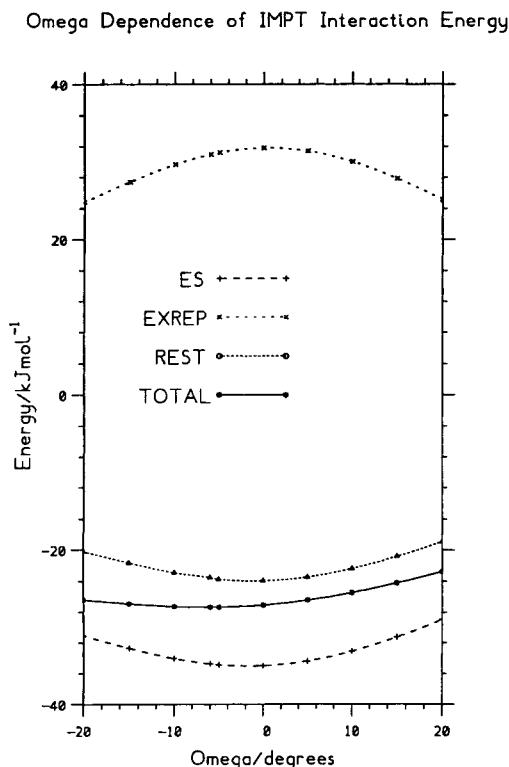
mirrors the variation of the total extremely well, the sum of the remaining terms being almost constant over the range  $\beta = 0^\circ$  to  $50^\circ$ .

**The Effect of Nonlinearity.** To investigate the effect of nonlinearity on the hydrogen bonding energy, we kept  $d_{\text{NO}}$  constant at 2.89 Å,  $\alpha$  at  $42^\circ$ ,  $\beta$  at  $0^\circ$  and varied the  $\angle\text{HNO}$  angle  $\omega$  from  $0^\circ$ ; the results are shown in Figure 6. As the deviation from linearity was increased, the two largest effects were that the electrostatic energy became less attractive and that the exchange-repulsion became smaller. However, the smaller changes in the other terms were also important in determining the optimum nonlinearity,  $|\omega| = 6^\circ$ .

### Models for Protein Hydrogen Bonds

We have also used the formamide/formaldehyde complex as a model for the hydrogen bonds found in the common units of protein secondary structure. We used appropriate average geometries<sup>55</sup> for  $\alpha$  helices (a typical example<sup>56</sup> helped us to calculate the coordinates) and for both antiparallel and parallel  $\beta$ -sheets. The geometries, which include nonplanar nonlinear hydrogen bonds, and results are given in Table IV. The overall trans formamide/formaldehyde minimum is included in the table for comparison.

In a  $\alpha$ -helix structure,  $d_{\text{NO}}$  is considerably longer than in either kind of  $\beta$  sheet. Thus it is



**Figure 6.** The total IMPT interaction energy as a function of the  $\angle\text{HNO}$  angle  $\omega$ , a measure of nonlinearity, near the minimum energy structure ( $\alpha = 42^\circ$ ,  $d_{\text{NO}} = 2.89$  Å,  $\beta = 0^\circ$ ).

not surprising that the interaction energy attributable to the hydrogen bond is greater for the  $\beta$  sheets. However, since the overall energies of the conformations depend on other factors as well as on main chain-main chain hydrogen bonding, one cannot conclude that one structure is more stable than another on the basis of hydrogen bond energies alone.<sup>55</sup> The greater hydrogen bond energy in parallel compared with antiparallel  $\beta$  sheets is interesting, since parallel sheets are less common and thought to be less stable than antiparallel ones.<sup>57</sup> While factors other than hydrogen bonding probably favor the antiparallel conformation,<sup>57</sup> the stability of the hydrogen bond in the parallel case, with an energy less than 1.4 kJ/mol above the overall minimum energy, was somewhat unexpected.

## PART II: IMPLICATIONS FOR MODELING THE N—H···O=C HYDROGEN BOND IN SIMULATIONS

### Criteria for Model Intermolecular Potentials

A model potential allows one to obtain the interaction energy between two molecules as a function of distance and orientation. This must be simple enough to be evaluated rapidly for the large number of configurations needed in a simulation. It is also desirable that the model potential should contain as much good science as possible—that the terms comprising it should each be physically meaningful and describe both the distance and orientation dependence of the site-site interactions with theoretically justified functional forms. For instance, models of the electrostatic term which go beyond the point charge level, such as those which have been developed for peptides<sup>15,58</sup> and for nucleic acid bases,<sup>59</sup> will make possible a quantitative understanding of the electrostatic directionality of biological interactions. Thus, intermolecular potential schemes which use accurate descriptions of the electrostatic term, such as those provided by the DMA,<sup>30,31</sup> SIBFA<sup>42,60,61</sup> or CAMM<sup>37,62</sup> models, should give much better representations of hydrogen bonding than do isotropic atom-atom potentials.

A potential containing distinct and accurately modeled terms describing the electrostatics, repulsion, charge transfer, dispersion, and polarization will allow the different contributions to the interactions of, for example, enzymes and their substrates or drug molecules and their receptors to be analyzed. An appropriate functional form for each term will allow comparisons with other potentials, and any future adjustments, to be made on the basis of a quantitative understanding of well defined, physically meaningful contributions to the interaction energy.

**Table IV.**  $\alpha$ -helix and  $\beta$ -sheet model geometries and results.

	$\alpha$ -helix	$\beta$ -sheet (anti-parallel)	$\beta$ -sheet (parallel)	Overall minimum
Geometry				
$\alpha$	18.0°	20.0°	15.0°	42.0°
$\beta$	27.0°	15.0°	11.0°	0.0°
$ \omega $	18.8°	16.1°	16.2°	6.0°
$\angle \text{COH}$	147.9°	155.2°	161.5°	138.0°
$d_{\text{NO}}/\text{\AA}$	2.99	2.91	2.92	2.89
$d_{\text{HO}}/\text{\AA}$	2.06	1.96	1.97	1.89
Energy/kJmol <sup>-1</sup>				
ES	-21.93	-28.00	-29.29	-34.72
EXREP	15.56	22.73	21.34	30.99
PL	-2.41	-3.22	-3.39	-4.13
CT	-3.07	-3.74	-4.08	-5.76
DISP	-4.79	-5.90	-6.11	-7.01
CTC	-3.83	-5.19	-4.93	-6.70
TOTAL	-20.48	-23.31	-26.46	-27.34
LJ + C	-23.35	-25.35	-26.72	-24.52
ES(DMA)	-18.48	-22.46	-24.05	-27.67

The success of the force-fields which are at present widely in biological modeling has to rely on a cancellation of errors, as they represent the intermolecular forces by models which are known to be inadequate from a theoretical point of view.<sup>10,63</sup> However, it would be most imprudent to rely on an approximate and empirical cancellation which is known not to be reliable in all cases, and we might also find that a clear improvement of the description of one term rendered the total potential less accurate by uncovering a compensating error elsewhere. Thus we seek to model each contribution accurately and to include the anisotropy explicitly where it is required.

### Representation of the IMPT Results By Model Potentials

We consider two representations of the formamide/formaldehyde potential energy surface, one an anisotropic atom-atom model and the other an isotropic atom-atom form. The anisotropic model has five components, but the isotropic model only four, the dispersion and polarization terms having to be considered together. The functional form of the anisotropic model was developed by considering the theory underlying each term, guided by the need to reproduce the IMPT data accurately. The isotropic model was chosen to be broadly representative of those currently in use and to give a fair comparison with the anisotropic model.

#### The Multipolar (DMA) Electrostatic Energy

The IMPT electrostatic energy is exactly equal to  $E_{\text{DMA}}$ , the electrostatic interaction energy calculated using Stone's distributed multipole analysis, at long range. However they differ at short

range due to the effect of penetration. For the isotropic atom-atom potential, a charge only form is used, with the atomic DMA charges being scaled to give the same mean electrostatic interaction energy over 62 points as the full DMA used in the anisotropic model. This gave an rms error of 1.95 kJ/mol (or 8.1%) relative to the full DMA energy. A better fit could have been obtained by empirically adjusting the individual charges, but this would almost certainly have worsened the accuracy of the electrostatic energy for unsampled regions of the surface. The contrast between the orientation dependence of the IMPT electrostatic energy and that given by the "LJ + C" point charge model (Table III) suggests that any atomic point charge model would still give significant errors because of the neglect of the anisotropic features of the charge distribution.

#### The Polarization Energy

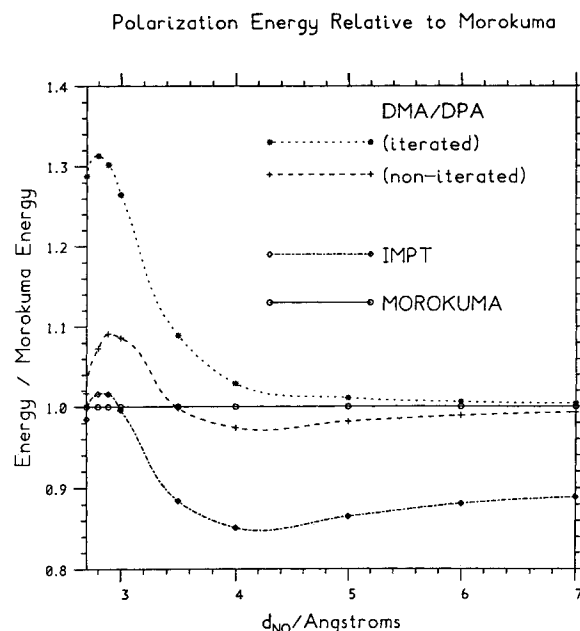
This term represents the energy lowering caused by the electron density of each monomer responding to the field of the other. Although it is the smallest term in the potential, it is the hardest to fit accurately to a simple analytical form.

For the isotropic model, we attempted to fit the polarization energy to an atom-atom  $R_{ab}^{-6}$  form. However, we found that this gave rise to unphysical repulsive terms and hence we fitted the polarization together with the dispersion as one term. In the case of the anisotropic model, we found that the polarization energy could not be modeled in a theoretically justified manner with an accuracy commensurate with that of other terms. The approach which would have produced the lowest rms error would have been to use exactly the same polarization/dispersion term as in the isotropic model. The greater theoretical justification of

the distributed polarizability approach discussed below was, however, considered sufficient to warrant its use in the anisotropic atom-atom model potential.

The polarization energy may be calculated by, among other methods,<sup>42,44</sup> Hayes–Stone IMPT, the supermolecule based Morokuma analysis method, or Stone's DMA and distributed polarizability analysis (DPA) approach. IMPT and Morokuma analysis results differ due to the uncoupled nature of IMPT. The DMA/DPA polarization energy<sup>64,65</sup> is calculated from distributed polarizabilities up to quadrupole rank on every atom. The distributed polarizabilities are obtained using coupled Hartree–Fock perturbation theory<sup>66,67</sup> within CADPAC,<sup>52</sup> the DPA routine having recently been extended to deal with nonlinear molecules by Stone and Le Sueur.<sup>68</sup> The DMA/DPA method is by far the cheapest, since the multipoles and polarizabilities are monomer properties which only have to be calculated once and can then be used to produce electrostatic and polarization energies for as many points as is desired at very little cost. Two versions of the DMA/DPA polarization energy have been evaluated, one is iterated to infinite order and the other noniterated ("first order"). Stone<sup>65</sup> suggests that at this level of description, with polarizabilities up to quadrupole rank, the methods will be roughly equally accurate, but that the noniterated method is more theoretically consistent with the other energy terms (since iterating involves calculating terms which are of higher than second order in the perturbation) and would be more accurate for higher rank polarizabilities. We have used noniterated polarization energies in our anisotropic atom-atom model potential. The DMA/DPA method is expected to be less accurate at shorter range where the absence of higher rank polarizabilities (and thus  $1/R^n$  terms of higher  $n$ ), the neglect of hyperpolarizabilities included in the Morokuma formulation, and the effect of overlap,<sup>23</sup> are more important.

In Figure 7, we compare the polarization energies calculated by the different methods over a range of N···O distances. This brings out the fundamental difficulty that we face in fitting the polarization energy. The IMPT energies which we are fitting to differ from the Morokuma energies by about 10%, which is significant in relation to the very small fitting errors in the other terms in our potential—though not in relation to the total IMPT energy. The DMA/DPA method gives results considerably closer to those of the Morokuma method than to the IMPT results. This reflects two fundamental differences. Firstly, the DMA/DPA and Morokuma analysis methods, unlike IMPT, suffer from the weakness of not including the effect of the Pauli exclusion principle on the



**Figure 7.** The ratio of the polarization energies calculated by the DMA/DPA (polarizabilities to quadrupole on all atoms, both iterated and noniterated energies) and IMPT methods to that calculated by Morokuma analysis for planar, linear configurations of the formamide/formaldehyde complex with  $\alpha = 42^\circ$ . IMPT's underestimation of the size of the polarization energy at most distances can be ascribed to the uncoupled nature of the perturbation theory. The Morokuma polarization energy is itself a decaying function of distance, so the error of around 10% in the IMPT value at long range is numerically fairly insignificant.

polarization energy.<sup>44</sup> The second major difference is, on the other hand, an inaccuracy in the IMPT method; namely that it relies on an uncoupled perturbation theory description of polarization. Thus the comparatively large errors in the fit of the DMA/DPA to IMPT polarization energies are not a fair measure of the worth of the distributed polarizability approach, but rather reflect the theoretical problems in defining and calculating the polarization energy.

#### *The Exchange-Repulsion and Penetration Energy*

This is the short-range, overlap dependent, net repulsive term, corresponding to the difference between the first order IMPT and DMA energies. It is the quantity which, for smaller molecules, Wheatley and Price fitted to an anisotropic atom-atom repulsion model.<sup>43</sup> We consider the repulsion between the N–H of formamide and the O=C of formaldehyde in the hydrogen bond.

For the isotropic model, this repulsion is modeled by

$$E_{\text{EXREP+PEN}} = A_{\text{NO}}\exp(-B_{\text{NO}}d_{\text{NO}}) + A_{\text{HO}}\exp(-B_{\text{HO}}d_{\text{HO}}) + A_{\text{HC}}\exp(-B_{\text{HC}}d_{\text{HC}})$$

This is a standard atom-atom exponential model, the exponents and preexponential coefficients in this case are all freely fitted parameters. We include repulsion between only the atoms comprising the  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bond because, since only the hydrogen bonding region of the potential energy surface is considered, the repulsion between other pairs of atoms is negligible. We found empirically that no  $\text{N}\cdots\text{C}$  repulsion term is necessary, but that  $\text{H}\cdots\text{O}$  and  $\text{H}\cdots\text{C}$  terms definitely are. It thus appears that the technique of neglecting the repulsion from hydrogen bonding protons, used by Spackman<sup>7</sup> and others, in fact demonstrates a deficiency of similar magnitude to that repulsion in their description of the attractive part of the hydrogen bonding interaction. In the  $E_{\text{EXREP}+\text{PEN}}$  term of the anisotropic atom-atom model potential, we also include the effects of the nonsphericity of atoms. The functional form chosen for the anisotropic potential is

$$E_{\text{EXREP}+\text{PEN}} = [A_{\text{NO}}\exp(-B_{\text{NO}}d_{\text{NO}}) + A_{\text{HO}}\exp(-B_{\text{HO}}d_{\text{HO}})] \times [(1 - c_1 - c_2 - c_3 - c_4) + c_1 \cos(\alpha) + c_2 \cos^2(\alpha) + c_3 \cos(\beta) + c_4 \cos^2(\beta)] + A_{\text{HC}} \exp(-B_{\text{HC}}d_{\text{HC}}).$$

This is essentially an atom-atom exponential repulsion model, modified by the inclusion of anisotropic terms, using a functional form based on that of Wheatley and Price.<sup>43</sup> They expressed the overlap, and hence the repulsion, between two sites in terms of an exponential multiplied by a (terminating) series describing the anisotropy. For consistency with that work, we have elected not to include the atom-atom anisotropy inside the exponential, as has been done successfully in other contexts.<sup>10,69</sup> The terms in  $\cos(\alpha)$  describe the lone pair type directionally around the oxygen, the effect of which can be seen in Figure 3(b) and those in  $\cos(\beta)$  describe the out of plane anisotropy, as shown in Figure 5. However, our functional form for the anisotropy is neither rigorous nor complete, since Wheatley and Price's model is only approximately applicable to this system and also we are including only those anisotropic terms which we expect to be most important. Furthermore, we do not allow the  $\text{N}\cdots\text{O}$  and  $\text{H}\cdots\text{O}$  components to have independent anisotropies.

The potential parameters were found by fitting  $E_{\text{EXREP}+\text{PEN}}$  to the IMPT data at 132 different points. Many of these represented planar and linear hydrogen bonds with either  $d_{\text{NO}}$  or  $\alpha$  or both close to their values at the minimum. The accuracy of the fit will thus be biased towards such points as opposed to other parts of the potential energy surface. However, there were sufficient

non-planar points to allow the variation with  $\beta$  to be parameterized and sufficient nonlinear points to allow the  $\text{N}\cdots\text{O}$  and  $\text{H}\cdots\text{O}$  contributions to the exponential repulsion to be separated.

We optimized the A and B parameters for the isotropic atom-atom model using the NAG<sup>70</sup> routine E04HFF to minimize the rms error. For the anisotropic atom-atom model, all ten parameters were optimized simultaneously. This process had to be performed several times, starting from different points, to combat the multiple minima problem. The final values of the parameters for both models are shown in Table V, along with the rms and maximum errors. The inclusion of the anisotropic terms reduces the rms error by a factor of seven and the maximum error by 2.0 kJ/mol, clearly a significant effect.

### The Charge Transfer and CTC Energy

This term is also short ranged and overlap dependent, but is attractive. Its fall off is significantly slower than that of  $E_{\text{EXREP}+\text{PEN}}$  and so it is described by different exponents. Thus the common practice of not having an explicit charge transfer term in a model potential but absorbing it in the repulsion term, so describing both effects with the same exponents, is based on a poor approximation.  $E_{\text{CT}+\text{CTC}}$  is fitted to the IMPT charge transfer and charge transfer correlation energies. The functional form is the atom-atom exponential

$$E_{\text{CT}+\text{CTC}} = X_{\text{NO}} \exp(-Y_{\text{NO}}d_{\text{NO}}) + X_{\text{HO}} \exp(-Y_{\text{HO}}d_{\text{HO}})$$

for the isotropic model and

$$E_{\text{CT}+\text{CTC}} = [X_{\text{NO}} \exp(-Y_{\text{NO}}d_{\text{NO}}) + X_{\text{HO}} \exp(-Y_{\text{HO}}d_{\text{HO}})] \times [(1 - k_1 - k_2 - k_3 - k_4) + k_1 \cos(\alpha) + k_2 \cos^2(\alpha) + k_3 \cos(\beta) + k_4 \cos^2(\beta)]$$

for the anisotropic model. The  $\alpha$ -anisotropies of the charge transfer and charge transfer correlation

**Table V.** Parameters for  $E_{\text{EXREP}+\text{PEN}}$ .

Parameter	Isotropic fit	Anisotropic fit
$A_{\text{NO}}/10^6\text{kJmol}^{-1}$	8.691	0.5101
$A_{\text{HO}}/10^6\text{kJmol}^{-1}$	0.04018	0.02758
$A_{\text{HC}}/10^6\text{kJmol}^{-1}$	0.001264	0.04434
$B_{\text{NO}}/\text{\AA}^{-1}$	4.834	3.8881
$B_{\text{HO}}/\text{\AA}^{-1}$	4.310	4.0678
$B_{\text{HC}}/\text{\AA}^{-1}$	2.004	3.5707
$c_1$	n/a	1.4540
$c_2$	n/a	-1.1935
$c_3$	n/a	-0.2273
$c_4$	n/a	0.3554
Max error/kJmol <sup>-1</sup>	2.74	0.70
<sup>a</sup> RMS error/kJmol <sup>-1</sup>	0.79	0.11

<sup>a</sup>The RMS error is calculated over all 132 points and is unweighted.

tion terms can be seen in Figure 3(c), the similarity between the two is sufficient to warrant considering them together. The theoretical justification for the form of this term is not so strong as that for  $E_{\text{EXREP}+\text{PEN}}$  since, although both terms are dependent on overlap, the dependence of  $E_{\text{CT}+\text{CTC}}$  is more complicated than the simple proportionality or near-proportionality<sup>43</sup> which gives a good model for  $E_{\text{EXREP}+\text{PEN}}$ . However, once scaling and the change of sign are taken account of, the  $\alpha$  variation of the charge transfer and charge transfer correlation is fairly similar to that of the  $E_{\text{EXREP}+\text{PEN}}$  (Fig. 3(b)), which suggests, along with the small rms error obtained, that the functional form chosen for  $E_{\text{CT}+\text{CTC}}$  is a reasonable one.

$E_{\text{CT}+\text{CTC}}$  was fitted to IMPT data at 69 points, fewer than for  $E_{\text{EXREP}+\text{PEN}}$  since the second-order double excitation energy, which includes the charge transfer correlation, was not calculated for all 132 points. The parameters were optimized in a similar way to those for  $E_{\text{EXREP}+\text{PEN}}$ . The results are shown in Table VI for both the isotropic and anisotropic forms. We found empirically that no H···C term is required to fit the IMPT data. In each case, as we would expect,  $X_{\text{NO}}$  and  $X_{\text{HO}}$  are both negative. Again the introduction of anisotropy reduces the rms error by a factor of about seven.

### The Dispersion Energy

The dispersion part of the IMPT interaction energy is modeled by an atom-atom  $R_{ab}^{-6}$  functional form:

$$E_{\text{DISP}} = \sum_{\substack{a \in A; \\ b \in B}} -\gamma_a \gamma_b R_{ab}^{-6}$$

Since dispersion is a long range attraction, the sum is taken over all intermolecular pairs of atoms and the  $\gamma$ 's are assumed to depend on atomic number only. Thus there are just four such coefficients for this system;  $\gamma_{\text{C}}$ ,  $\gamma_{\text{N}}$ ,  $\gamma_{\text{H}}$ , and  $\gamma_{\text{O}}$ . These were obtained by fitting to the IMPT data

**Table VI.** Parameters for  $E_{\text{CT}+\text{CTC}}$ .

Parameter	Isotropic fit	Anisotropic fit
$X_{\text{NO}}/\text{kJmol}^{-1}$	-6.758	-140.5
$X_{\text{HO}}/\text{kJmol}^{-1}$	-6602	-7929
$Y_{\text{NO}}/\text{\AA}^{-1}$	0.4774	1.270
$Y_{\text{HO}}/\text{\AA}^{-1}$	3.409	3.740
$k_1$	n/a	0.1784
$k_2$	n/a	-0.5329
$k_3$	n/a	-0.8621
$k_4$	n/a	0.5321
Max error/kJmol <sup>-1</sup>	3.66	0.39
<sup>a</sup> RMS error/kJmol <sup>-1</sup>	0.90	0.13

<sup>a</sup>The RMS error is calculated over 69 points and is unweighted.

**Table VII.** The dispersion parameters.

Parameter	DISP + PL <sup>a</sup>	DISP <sup>b</sup>
$\gamma_{\text{C}}/(\text{kJmol}^{-1})^{1/2}\text{\AA}^3$	5.51	18.38
$\gamma_{\text{N}}/(\text{kJmol}^{-1})^{1/2}\text{\AA}^3$	42.31	54.22
$\gamma_{\text{H}}/(\text{kJmol}^{-1})^{1/2}\text{\AA}^3$	3.63	4.18
$\gamma_{\text{O}}/(\text{kJmol}^{-1})^{1/2}\text{\AA}^3$	67.00	32.44
Max error/kJmol <sup>-1</sup>	0.71	0.39
RMS error/kJmol <sup>-1c</sup>	0.22	0.12

<sup>a</sup>The fit of  $E_{\text{DISP}+\text{PL}}$  to this form was used in the isotropic model.

<sup>b</sup>The fit of  $E_{\text{DISP}}$  to this form was used in the anisotropic model, with  $E_{\text{PL}}$  being modeled by the DMA/DPA method.

<sup>c</sup>RMS error over the 62 points at which these components were fitted.

at 62 points in such a way as to minimize the rms error. The results are given in Table VII.

This simple model gave an excellent fit to the dispersion energy, as calculated by IMPT. Thus, neither anisotropy, nor short range damping (though we have only fitted to the hydrogen bonding region of the surface with  $d_{\text{NO}} \geq 2.70$  Å), nor  $R_{ab}^{-8}$  terms are necessary to obtain a very good fit to the IMPT dispersion energy. It would be interesting to compare the coefficients obtained here, which are empirically fitted parameters, with theoretical values calculated as integrals of products of polarizabilities at imaginary frequencies;<sup>71</sup> however these are not yet available. In the future, it may be possible to calculate dispersion coefficients in this way using distributed polarizabilities, but for the present we shall use the fitted ones. The small size of  $\gamma_{\text{H}}$  is encouraging since the dispersion interactions involving hydrogen are theoretically expected to be weaker than those involving heavy atoms, as hydrogen atoms should have relatively small distributed polarizabilities at imaginary frequencies. However, one should not read too much into the values of the coefficients, partly because we are assuming a very simple combining rule and partly because our form absorbs the effects of  $R_{ab}^{-8}$  and higher terms. Other workers have obtained good fits to the *ab initio* dispersion energy for systems such as the small dimers<sup>72</sup> ( $\text{H}_2\text{O}$ )<sub>2</sub> and ( $\text{HF}$ )<sub>2</sub> and the larger azabenzene dimers,<sup>73</sup> using essentially the same atom-atom isotropic  $R_{ab}^{-6}$  form as we have done. However, both the above studies considered a much larger number of points than does our work, since we are only interested in a small part of the formamide/formaldehyde potential energy surface. Thus both groups were able to fit a separate parameter for each kind of atom-atom interaction, rather than assuming a geometric mean combining rule as we have done.

For the isotropic potential model, the polarization energy is included with the dispersion energy in this term. Thus both the fit and the theoretical justification are not as good in this case as for the

anisotropic model, where the terms are considered separately.

### Fitting the Total Energy

The isotropic model potential comprising four terms and the anisotropic model potential of five terms have been compared to the total IMPT interaction energy at the 62 points at which the dispersion had been fitted. The rms errors, over the 62 points, of each component and of the total interaction energies, are shown in Table VIII.

The anisotropic model gives an excellent fit, with an rms error of only 1.3%, most of which can be attributed to the polarization energy. The fit could have been made even better by combining the polarization energy with the dispersion, but only at the cost of having a less theoretically sound description. In contrast, although there is some cancellation between the errors in the various components in the isotropic model, the rms error of 2.0 kJ/mol or 8.3% would still give cause for concern in a simulation. The largest single contribution is the error in the electrostatic energy. The plots of the  $\alpha$ -dependences of the relevant components (Figures 3(b) and (c)) suggest that the smaller errors in neglecting anisotropy in the  $E_{\text{EXREP+PEN}}$  and  $E_{\text{CT+CTC}}$  terms may have cancelled to a significant extent.

## CONCLUSIONS

This IMPT study of an  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bond has shown that, as in several other hydrogen

bonded systems,<sup>29,53,74</sup> the orientation dependence of the interaction is dominated by the long-range electrostatic component. In this case, the exchange repulsion is similar in magnitude to the electrostatic contribution at the minimum. The dispersion, polarization, charge transfer, and charge transfer correlation terms are, however, not negligible and play a significant role in determining the hydrogen bond geometry. It appears that the cancellation between the various terms results in the electrostatic energy mirroring the total interaction energy reasonably well at experimentally observable geometries. The results do not suggest that any other simplification, such as omitting some of the smaller terms, would give a better approximation to the behavior of the total interaction energy than the DMA does.

This study shows that there is a marked energetic preference for the  $\text{N}-\text{H}$  donor group to lie in the acceptor oxygen atom's "lone pair plane." A slight nonlinearity of the  $\text{N}-\text{H}\cdots\text{O}$  group is favorable. However, the potential energy surface for planar, near-linear hydrogen bonds is very flat for variations in the  $\text{C}=\text{O}\cdots\text{N}$  angle  $\alpha$  between about  $-50^\circ$  and  $+60^\circ$ , with the central maximum at  $\alpha = 0^\circ$  being less than 1 kJ/mol above the minimum at  $\alpha = +42^\circ$ . This flatness implies that we would not expect to see a marked preference for hydrogen bonding in the lone pair direction ( $\alpha = \pm 60^\circ$ ) arising from the properties of an isolated  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bond. We would therefore not want to impose a strong  $sp^2$  oxygen lone pair directionality in the context of a theoretically based potential, although the fact that

**Table VIII.** Comparison of the IMPT Values of the Components of the Interaction Energy of Formamide/Formaldehyde with our Fitted Isotropic and Anisotropic Atom-Atom Models.<sup>a</sup> Energies are given in kJ/mol.

	DMA	ERP	CT	D + PL	TOTAL	
Isotropic Atom-Atom Model						
Parameters	1	6	4	4	15	
IMPT mean	-23.47	19.64	-10.46	-9.58	-23.87	
IMPT rms	24.05	22.03	11.27	10.14	24.34	
Fit mean	-23.47	19.42	-10.39	-9.55	-23.99	
Fit rms	23.93	21.80	11.17	10.14	24.52	
rms error <sup>b</sup>	1.95	0.92	0.90	0.22	2.02	
error (%)	8.10	4.16	7.96	2.18	8.30	
	DMA	ERP	CT	DISP	PL	TOTAL
Anisotropic Atom-Atom Model						
Parameters	0	10	8	4	0	22
IMPT mean	-23.47	19.64	-10.46	-6.12	-3.45	-23.87
IMPT rms	24.05	22.03	11.27	6.48	3.67	24.34
Fit mean	-23.47	19.63	-10.47	-6.11	-3.66	-24.08
Fit rms	24.05	22.03	11.27	6.48	3.88	24.57
rms error <sup>b</sup>	0.00	0.09	0.11	0.12	0.25	0.32
error (%)	0.00	0.39	0.98	1.81	6.85	1.30

<sup>a</sup>The column headings are as follows: DMA is the multipolar (DMA) electrostatic energy, ERP is exchange-repulsion plus penetration energy, CT is the charge transfer plus charge transfer correlation energy, D + PL is the dispersion plus polarization energy (isotropic model), DISP is the dispersion energy (anisotropic model), PL is the polarization energy (anisotropic model).

<sup>b</sup>The rms error evaluated over 62 points is different from that over 132 and 69 points for  $E_{\text{EXREP+PEN}}$  and  $E_{\text{CT+CTC}}$ .

such an approach has proved empirically useful<sup>4</sup> illustrates the importance of including atom-atom anisotropy.

How then, in the absence of prominent lone pairs at  $\alpha = \pm 60^\circ$ , do we account for the observed preference for  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonding in these directions found in a statistical analysis<sup>75</sup> of such bonds in small molecule structures from the Cambridge Crystallographic Database? The key to this problem is that a  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bond cannot be considered in isolation from the rest of the molecule. Whilst the experimental preference for planar, near-linear hydrogen bonds<sup>75</sup> can be attributed to the energetic advantages of such a geometry (Figures 5 and 6), the flatness of the potential with respect to the in plane angle  $\alpha$  implies that this angle can easily be varied in order to optimise other interactions. A study of the electrostatic interactions in 29  $\text{N}-\text{H}\cdots\text{O}=\text{C}$  hydrogen bonded complexes<sup>48</sup> of small molecules (using a DMA plus hard sphere model, which this work shows should mimic well the orientation dependence of the total interaction energy) demonstrated that structures with two close intermolecular contacts were strongly favored. Thus the statistical preference for hydrogen bonding in the "lone pair direction" ( $\alpha = \pm 60^\circ$ ) shown by these electrostatic minima arises from the steric consequences of optimising the intermolecular electrostatic interactions of the functional groups bonded to the  $\text{N}-\text{H}$  and  $\text{C}=\text{O}$  groups of the hydrogen bond. Thus, when the common occurrence of certain favorable motifs<sup>48,76</sup> is taken into account, the crystal structure results of Taylor et al.<sup>75</sup> for small molecules, and those of Baker and Hubbard<sup>55</sup> for proteins, are seen to be consistent with the hydrogen bonding potential energy surface described in this study.

The detail of the hydrogen bonding potential energy surface needs to be incorporated into biological force-fields if we are to have confidence in the results of simulations. We have shown that a state-of-the-art intermolecular potential, of the anisotropic atom-atom form currently used for small molecules, can give a highly accurate model for the hydrogen bonding region of the trans formamide/formaldehyde potential energy surface. The key feature of this model is the use of atomic multipoles, derived from DMA's of the monomer wavefunctions, to model the electrostatic interaction. Thus the dominant term in the potential is modeled very accurately at all points. The polarization can also be represented without any empirical parameters, using the DMA/DPA approach. The other terms in the potential have been fitted empirically, for lack of any rigorous theoretical model, and thus are unlikely to extrapolate well to the unsampled regions of the

surface. Fortunately, these contributions can be well represented by fairly simple models, provided that the nonsphericity of the charge distribution around the carbonyl oxygen is taken into account.

In contrast, an isotropic atom-atom model constructed on a similar term by term basis gives rise to significant errors, of the order of 2 kJ/mol, in the hydrogen bonding region of the potential energy surface. These errors would probably have been larger had the model been parameterized over the whole surface. Although there are almost endless possible variations on an isotropic atom-atom model, the orientation dependence of the different terms (Figs. 3 to 6) suggests that an acceptable representation of hydrogen bonding is unlikely without the inclusion of anisotropic terms, particularly in the electrostatic component. This makes sense since the electrostatic energy arises precisely because neutral atom spherical charge distributions are distorted on forming molecules. The dispersion energy is fairly well represented by an isotropic atom-atom model, while there is some cancellation between the orientation dependences of the other non-electrostatic terms<sup>29</sup> since the exchange-repulsion and the attractive polarization and charge transfer energies are dependent in similar ways on the overlap of the monomer charge distributions.

The functional forms which we have studied are much more elaborate than those currently used in biological simulations. The success of the anisotropic atom-atom model emphasises that hydrogen bonding will not require "special treatment" (i.e., extra empirical terms) when more accurate models of intermolecular forces are used. However, it is clear that modeling biological systems reliably will require elaborate anisotropic atom-atom model potentials. Such a model would have to be parameterized against accurate *ab initio* quantum chemical calculations on a reasonable number of systems in order to obtain parameters which are transferable, rather than merely describing one particular pair of molecules. However, it is encouraging that so much valuable information, particularly the multipoles determining the electrostatic interaction, can be obtained from monomer wavefunctions. Thus, theoretical chemistry can provide molecular modeling with a sound theoretical basis by replacing outdated empiricism with the products of state-of-the-art research into intermolecular forces.

We would like to thank Dr. A. J. Stone for the use of his IMPT and ORIENT programs and for some very helpful discussions. We are also indebted to R. J. Wheatley for the functional form of the anisotropic repulsion term and to C. R. Le Sueur for the use of her distributed polarizabilities program. JBOM acknowledges the support of SERC in the form of an earmarked studentship and SLP is grateful to the Royal Society for a 1983 University Research Fellowship.

## References

1. T. Alber, S. Dao-pin, K. Wilson, J. A. Wozniak, S. P. Cook and B.W. Matthews, *Nature*, **330**, 41 (1987).
2. D. J. Danziger and P. M. Dean, *Proc. R. Soc. Lond.*, **B236**, 1015 (1989).
3. A. T. Hagler, E. Huler and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
4. A. Vedani and J. D. Dunitz, *J. Am. Chem. Soc.*, **107**, 7653 (1985).
5. W. E. Reiher, Ph.D. Thesis, Harvard University (1985).
6. B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).
7. M. A. Spackman, *J. Chem. Phys.*, **85**, 6587 (1986).
8. J. Hermans, H. T. C. Berendsen, W. F. Van Gunsteren, and J. P. M. Postma, *Biopolymers*, **23**, 1513 (1984).
9. A. J. Stone and S. L. Price, *J. Phys. Chem.*, **92**, 3325 (1988).
10. S. L. Price, *Molec. Sim.*, **1**, 135 (1988).
11. R. J. Wheatley and S. L. Price, *Molec. Phys.* (in press).
12. F. M. L. G. Stamato and J. M. Goodfellow, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **13**, 277 (1986).
13. G. N. J. Port and A. Pullman, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **1**, 21 (1974).
14. C. N. Yoon and M. S. Jhon, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **12**, 33 (1985).
15. C. H. Faerman and S. L. Price, *J. Amer. Chem. Soc.*, **112**, 4915 (1990).
16. S. L. Price and A. J. Stone, *Chem. Phys. Lett.*, **98**, 419 (1983).
17. M. N. Bellido and J. A. C. Rullmann, *J. Comp. Chem.*, **10**, 479 (1989).
18. M. Eisenstein, *Int. J. Quantum Chem.*, **33**, 127 (1988).
19. A. M. Sapse, L. M. Fugler and D. Cowburn, *Int. J. Quantum Chem.*, **29**, 1241 (1986).
20. J. E. Del Bene, *J. Chem. Phys.*, **62**, 1961 (1975).
21. K. Morokuma, *J. Chem. Phys.*, **55**, 1236 (1971).
22. K. Kitaura and K. Morokuma, *Int. J. Quantum Chem.*, **10**, 325 (1976).
23. A. J. Stone, NATO-ASI Summer School 'Hydrogen Bonded Liquids,' Cargese, Corsica (1989) (to be published).
24. I. C. Hayes and A. J. Stone, *Molec. Phys.*, **53**, 83 (1984).
25. I. C. Hayes, Ph.D. Thesis, University of Cambridge (1982).
26. G. J. B. Hurst, Ph.D. Thesis, University of Cambridge (1985).
27. I. C. Hayes and A. J. Stone, *Molec. Phys.*, **53**, 69 (1984).
28. I. C. Hayes, G. J. B. Hurst, and A. J. Stone, *Molec. Phys.*, **53**, 107 (1984).
29. G. J. B. Hurst, P. W. Fowler, A. J. Stone, and A. D. Buckingham, *Int. J. Quantum Chem.*, **29**, 1223 (1986).
30. A. J. Stone, *Chem. Phys. Lett.*, **83**, 233 (1981).
31. A. J. Stone and M. Alderton, *Molec. Phys.*, **56**, 1047 (1985).
32. R. Rein, *Adv. Quant. Chem.*, **7**, 335 (1973).
33. R. Dovesi, C. Pisani, F. Ricca, and C. Roetti, *J. Chem. Soc. Faraday Trans. II*, **70**, 1381 (1974).
34. M. Mezei and E. S. Campbell, *Theoret. Chim. Acta*, **43**, 227 (1977).
35. R. Lavery, C. Etchbest and A. Pullman, *Chem. Phys. Lett.*, **85**, 266 (1982).
36. J. Bentley, in *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, P. Politzer and D. G. Truhlar, eds., Plenum, New York, (1981).
37. W. A. Sokalski and R. A. Poirier, *Chem. Phys. Lett.*, **98**, 86 (1983).
38. J. Fernandez Rico, J. R. Alvarez-Collado and M. Paniagua, *Molec. Phys.*, **56**, 1145 (1985).
39. D. L. Cooper and N. C. J. Stutchbury, *Chem. Phys. Lett.*, **120**, 167 (1985).
40. J. R. Rabinowitz, K. Namboodiri, and H. Weinstein, *Int. J. Quantum Chem.*, **29**, 1697 (1986).
41. F. Vigné-Maeder and P. Claverie, *J. Chem. Phys.*, **88**, 4934 (1988).
42. N. Gresh, P. Claverie and A. Pullman, *Int. J. Quantum Chem. Quantum Chem. Symp.*, **13**, 243 (1979).
43. R. J. Wheatley and S. L. Price, *Molec. Phys.*, **69**, 507 (1990).
44. S. M. Cybulski and S. Scheiner, *Chem. Phys. Lett.*, **166**, 57 (1990).
45. A. D. Buckingham, P. W. Fowler and J. M. Hutson, *Chem. Rev.*, **88**, 963 (1988).
46. A. J. Stone, private communication.
47. S. M. Cybulski, G. Chalasinski and R. Moszynski, *J. Chem. Phys.*, **92**, 4357 (1990).
48. J. B. O. Mitchell and S. L. Price, *Chem. Phys. Lett.*, **154**, 267 (1989).
49. S. F. Boys and F. Bernardi, *Molec. Phys.*, **19**, 553 (1970).
50. M. Gutowski, J. H. Van Lenthe, J. Verbeek, F. B. Van Duijneveldt, and G. Chalasinski, *Chem. Phys. Lett.*, **124**, 370 (1986).
51. D. W. Schwenke and D. G. Truhlar, *J. Chem. Phys.*, **82**, 2418 (1985).
52. R. D. Amos and J. E. Rice, CADPAC: The Cambridge Analytical Derivatives Package, issue 4.0, University of Cambridge (1987).
53. H. Umeyama and K. Morokuma, *J. Am. Chem. Soc.*, **99**, 1316 (1977).
54. R. Taylor, O. Kennard, and W. Versichel, *Acta Cryst.*, **B40**, 280 (1984).
55. E. N. Baker and R. E. Hubbard, *Prog. Biophys. Molec. Biol.*, **44**, 97 (1984).
56. S. Arnott and A. J. Wonacott, *J. Mol. Biol.*, **21**, 371 (1966).
57. J. Richardson, *Nature*, **268**, 495 (1977).
58. W. A. Sokalski, K. Maruszewski, P. C. Hariharan, and J. J. Kaufman, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **16**, 119 (1989).
59. W. A. Sokalski, P. C. Hariharan, and J. J. Kaufman, *Int. J. Quantum Chem. Quantum Biol. Symp.*, **14**, 111 (1987).
60. N. Gresh, P. Claverie, and A. Pullman, *Theoret. Chim. Acta*, **66**, 1 (1984).
61. N. Gresh, A. Pullman, and P. Claverie, *Theoret. Chim. Acta*, **67**, 11 (1985).
62. W. A. Sokalski, and A. Sawaryn, *J. Chem. Phys.*, **87**, 526 (1987).
63. J. A. C. Rullmann and P. T. van Duijnen, Reports in Molecular Theory (to be published).
64. A. J. Stone, *Chem. Phys. Lett.*, **155**, 102 (1989).
65. A. J. Stone, *Chem. Phys. Lett.*, **155**, 111 (1989).
66. R. M. Stevens, R. Pitzer and W. N. Lipscomb, *J. Chem. Phys.*, **38**, 550 (1963).
67. J. Gerratt and I. M. Mills, *J. Chem. Phys.*, **49**, 1719 (1968).
68. A. J. Stone and C. R. Le Sueur, work in preparation.



69. S. L. Price, *Molec. Phys.*, **62**, 45 (1987).
70. Numerical Algorithms Group, Oxford, U.K.
71. A. J. Stone and C. S. Tong, *Chem. Phys.*, **137**, 121 (1989).
72. M. M. Szczesniak, R. J. Brenstein, S. M. Cybulski, and S. Scheiner, *J. Phys. Chem.*, **94**, 1781 (1990).
73. C. Huiszoon and F. Mulder, *Molec. Phys.*, **38**, 1497 (1979).
74. A. D. Buckingham and P. W. Fowler, *Can. J. Chem.*, **63**, 2018 (1985).
75. R. Taylor, O. Kennard, and W. Versichel, *J. Amer. Chem. Soc.*, **105**, 5761 (1983).
76. Z. Berkovitch-Yellin and L. Leiserowitz, *J. Amer. Chem. Soc.*, **102**, 7677 (1980).