Theor Chim Acta (1996) 94: 287–295

Theoretica Chimica Acta

Multipole-based calculation of the polarization energy

John B.O. Mitchell

Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Received April 23, 1996/Final revision received June 5, 1996/Accepted July 9, 1996

Abstract. A novel method for calculating the polarization energy of a biomolecular complex is developed. It is based on direct calculation of the interaction between the unperturbed charge distribution of one molecule and the perturbed charge distribution of the other. This "multipole-based polarization energy" is shown to include charge transfer. The method can alternatively incorporate Gaussian multipoles in place of the usual distributed point multipoles. This definition of the polarization energy is also used as the basis of a partitioning scheme for the total supermolecule interaction energy. Results are presented for the formamide/formaldehyde complex.

Key words: Polarization energy – Charge transfer energy – Distributed Multipole Analysis – Gaussian multipoles

1 Introduction

When a molecule experiences an electric field, its charge distribution will relax in response to that field. The energy associated with this relaxation is known as the polarization (or induction) energy. It is a negative (or zero) quantity, as can be seen from its appearance at second order in perturbation theory, or from the simple argument that, if such a relaxation were to lead to an increase in energy, it would not occur. There are also small contributions to the polarization energy at higher order in perturbation theory; these are not considered here. In the case of a bimolecular complex, the polarization energy has two components, one from the relaxation of each molecule's charge distribution in the electric field of the other.

In Hayes–Stone Intermolecular Perturbation Theory (IMPT) [1], the polarization energy is expressed in terms of excitations from the occupied orbitals of one molecule into the virtual orbitals of the same molecule (A → A transitions). A related energy term, which also appears at second order in perturbation theory, is the charge transfer energy [2]. This is described in IMPT by excitations from the occupied orbitals of one molecule into the virtual orbitals of the other (A → B

Present address: Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX; e-mail: ubcg34a@mail.cryst.bbk.ac.uk; Fax: (+ 44)-171-631-6803; Tel. (+ 44)-171-631-6810

transitions). For a sufficiently large basis set, the charge transfer term would be spurious, since any physical effect of this kind could be mathematically described by occupied and virtual orbitals derived from the basis functions of the same molecule. Hence, in the limit of large basis sets, the charge transfer becomes part of the polarization energy. It is, however, a separate term for practical purposes when using basis sets such as 6–31G* (used here) or smaller.

Applying the general formula of Stone [3] to the case where the number of interacting molecules is two, the polarization energy can be written as follows:

$$E_{\mathrm{pol}} = \tfrac{1}{2}\sum_{tu} [\Delta Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}}) + (Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} \Delta Q_u^{\mathrm{b}})]\,. \tag{1}$$

In this equation, Q represents the undistorted charge distribution, ΔQ the change in the charge distribution on relaxation in the other molecule's electric field, and T is the interaction tensor [3]. The indices a and b denote which of the two molecules the charge distribution is associated with, and the subscripts t and u, which are summed over, represent the different multipolar components of the interaction energy [3].

Polarization energies have been calculated by a number of methods, including IMPT [1], Morokuma analysis [4] and distributed polarizabilities [5]. This work is an attempt to assess the possibility of direct calculation of the polarization energy from Q and ΔQ.

2 Theory

The molecular charge distribution Q is conveniently represented by Distributed Multipole Analysis (DMA) [6]. The multipolar electrostatic energy between the unperturbed monomer charge distributions of molecules A and B, denoted E_{00}^{M} in this paper, is given by [3]

$$E_{00}^{\mathrm{M}} = \sum_{tu} (Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}})\,. \tag{2}$$

It is calculated simply from the DMA multipoles; the "M" superscript denotes the "multipole-based" method of calculation, which excludes the penetration energy [7]. By performing a supermolecule calculation, the relaxed charge distribution of the complex ($Q + \Delta Q$) can be obtained. While the apportioning of this charge distribution between the monomers is somewhat arbitrary, a DMA of the supermolecule wave function provides a well-defined way of achieving this. That part of the supermolecule charge distribution assigned by the DMA procedure to the atoms of molecule A is identified with ($Q^{\mathrm{a}} + \Delta Q^{\mathrm{a}}$) and that part assigned to the atoms of molecule B is identified with ($Q^{\mathrm{b}} + \Delta Q^{\mathrm{b}}$). The relationship between these ΔQ and those of Eq. (1) is discussed below. One can define a multipolar interaction energy between the perturbed supermolecule charge distribution of molecule A and the unperturbed monomer charge distribution of molecule B, and this quantity E_{*0}^{M} is given by

$$\begin{aligned} E_{*0}^{\mathrm{M}} &= \sum_{tu} [(Q_t^{\mathrm{a}} + \Delta Q_t^{\mathrm{a}}) T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}}] \\ &= \sum_{tu} [(Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}}) + (\Delta Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}})] \\ &= E_{00}^{\mathrm{M}} + \sum_{tu} [\Delta Q_t^{\mathrm{a}} T_{tu}^{\mathrm{ab}} Q_u^{\mathrm{b}}]\,. \end{aligned} \tag{3}$$

One may also define the analogous quantity E^{M}_{0*}:

$$E^{\mathrm{M}}_{0*} = E^{\mathrm{M}}_{00} + \sum_{tu} [Q^{\mathrm{a}}_t T^{\mathrm{ab}}_{tu} \Delta Q^{\mathrm{b}}_u] . \tag{4}$$

E^{M}_{*0} and E^{M}_{0*} can both be evaluated easily using Stone's program ORIENT [8] with monomer multipoles for one molecule and supermolecule multipoles for the other, and E^{M}_{00} by using monomer multipoles for each. From the definitions of E^{M}_{00}, E^{M}_{*0} and E^{M}_{0*}, it follows trivially that

$$E^{\mathrm{M}}_{*0} + E^{\mathrm{M}}_{0*} - 2E^{\mathrm{M}}_{00} = \sum_{tu} [(\Delta Q^{\mathrm{a}}_t T^{\mathrm{ab}}_{tu} Q^{\mathrm{b}}_u) + (Q^{\mathrm{a}}_t T^{\mathrm{ab}}_{tu} \Delta Q^{\mathrm{b}}_u)] , \tag{5}$$

and by comparison with Eq. (1) one can write

$$E^{\mathrm{M}}_{\mathrm{pol}} = \tfrac{1}{2}(E^{\mathrm{M}}_{*0} + E^{\mathrm{M}}_{0*} - 2E^{\mathrm{M}}_{00}) . \tag{6}$$

One must, however, be wary of identifying this quantity unambiguously with the previous definition of the polarization energy. The ΔQ^{a} and ΔQ^{b} used here are in fact defined somewhat differently from those in Eq. (1). This method permits charge transfer between the monomers, so that when Q^{a} and Q^{b} represent neutral molecules, $(Q^{\mathrm{a}} + \Delta Q^{\mathrm{a}})$ and $(Q^{\mathrm{b}} + \Delta Q^{\mathrm{b}})$ may have small net charges, which will be equal and opposite. Thus, the polarization energy calculated by the multipole-based method ($E^{\mathrm{M}}_{\mathrm{pol}}$) will include a charge transfer component. The ΔQ used here will also be affected by exchange and the Pauli exclusion principle, effects which do not form part of the usual "classical" induction energy. Furthermore, these ΔQ depend on the apportioning of multipoles originally centred at the overlap centres of two Gaussians in the supermolecule charge distribution; the DMA method re-expresses these multipoles as a new series at the nearest expansion site [6]. Overlap charge density between two Gaussians on different molecules will be assigned to one molecule or the other, on the basis of the nearest atom (the expansion sites used here being the atomic nuclei).

Hayes and Stone originally [1] defined charge transfer in a manner that was susceptible to basis set superposition error (BSSE), but more recently [2] Stone has published a definition that is free of BSSE. The charge transfer incorporated in $E^{\mathrm{M}}_{\mathrm{pol}}$ is somewhat susceptible to BSSE; it is not identical to either of Stone's definitions. The multipole-based method also ignores the effect of charge overlap; E^{M}_{00}, E^{M}_{*0} and E^{M}_{0*} all exclude the penetration energy [7].

One may include the effect of charge overlap on the polarization energy by using Gaussian multipoles [7], instead of the usual distributed point multipoles defined by Stone [6]. The energy components calculated in this way are denoted E^{G}_{00}, E^{G}_{*0}, E^{G}_{0*} and $E^{\mathrm{G}}_{\mathrm{pol}}$, the latter being the "Gaussian polarization energy". These quantities all include a penetration energy component. They are calculated in an exactly analogous manner to the corresponding multipole-based components, using Wheatley's program GMUL [7]. In principle, E^{G}_{00} is the exact electrostatic interaction energy between the monomer wave functions, as calculated by IMPT [1],

$$E^{\mathrm{G}}_{00} = \int \rho^{\mathrm{A}}(\mathbf{r}_1)\, \rho^{\mathrm{B}}(\mathbf{r}_2) r_{12}^{-1}\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{r}_2 , \tag{7}$$

where $\rho^{A}(\mathbf{r}_1)$ is the total (electronic plus nuclear) charge density of molecule A at $\mathbf{r}_1$, and so on. In practice, some small approximations are made to expedite its calculation (see [7] and also below).

In order to assess the results calculated with the multipole-based method, a benchmark is required. The benchmark chosen is the sum of the polarization and charge transfer terms from perturbation theory. The polarization part of the benchmark energy is taken from Morokuma analysis [4]. The difference between the Morokuma and IMPT polarization energies, which has its origin in the uncoupled nature of IMPT [1], is small. However, the Morokuma polarization is identical to the value calculated by distributed polarizabilities [5] at longer range [9], and expected also to be identical to the multipole-based polarization in this region (where the charge transfer would have decayed effectively to zero). Thus, it makes a better benchmark than the IMPT polarization energy. The charge transfer contribution to the benchmark energy is calculated using Stone's recent BSSE-free definition [2].

It is also possible to partition the supermolecule energy into three components. The total supermolecule interaction energy E_{sup} is known from the *ab initio* calculation performed to obtain the supermolecule multipoles (it is the difference between the total SCF energies in the supermolecule and the two monomer calculations). The multipolar electrostatic energy (i.e. excluding penetration) is simply E^{M}_{00}, and is considered to constitute one term of the partitioning. The second term of the partitioning is the multipole-based polarization energy, namely $E^{\text{M}}_{\text{pol}}$, which represents the combined polarization and charge transfer energy. The remaining term, $E^{\text{M}}_{\text{erp}}$, represents the exchange repulsion and penetration; it is calculated simply by subtraction:

$$E^{\text{M}}_{\text{erp}} = E^{\text{C}}_{\text{sup}} - E^{\text{M}}_{00} - E^{\text{M}}_{\text{pol}}. \tag{8}$$

$E^{\text{C}}_{\text{sup}}$ is the BSSE-corrected supermolecule interaction energy. The BSSE is estimated by calculating the difference between Stone's original BSSE-prone [1] and new BSSE-free [2] definitions of the charge transfer energy, both these quantities being readily available. Strictly speaking, it might be more accurate to correct the supermolecule interaction energy via the counterpoise method [10], rather than with a quantity derived from IMPT. The extra computational expense, however, was not considered to be justified. If this correction is not made, it is found that $E^{\text{M}}_{\text{erp}}$ may become significantly negative in some circumstances. Any BSSE component of $E^{\text{M}}_{\text{pol}}$, which is expected to be significantly smaller than the BSSE in the supermolecule interaction energy, is not corrected for.

Where Gaussian multipoles are used, the repulsion term E^{G}_{er} includes exchange repulsion and possibly some BSSE, but the penetration is incorporated in E^{G}_{00}. The repulsion is given by

$$E^{\text{G}}_{\text{er}} = E^{\text{C}}_{\text{sup}} - E^{\text{G}}_{00} - E^{\text{G}}_{\text{pol}}. \tag{9}$$

Compared with the partitionings obtained by IMPT [1] or Morokuma analysis [4], this method has the advantage of representing the interaction concisely in only three terms, but the disadvantage is that the main repulsive term is not obtained directly, but by subtraction. The method, like Morokuma analysis, describes only the SCF energy and a dispersion term would have to be added if this partitioning were to be used as the basis for a model potential.

3 Computational methods

The system chosen for the test calculations is the singly hydrogen bonded formamide/formaldehyde complex (Fig. 1). The geometries chosen have the two molecules coplanar, with the N–H ⋯ O atoms collinear. The distance N ⋯ O is varied from 2.7 to 7.0 Å, the IMPT energy minimum (when dispersion is taken into account) is at 2.89 Å.

All the *ab initio* calculations in this work are performed using the CADPAC [11] suite of quantum chemistry programs, with CADPAC library 6–31G* basis sets. The monomer multipoles, supermolecule multipoles and E^{C}_{sup} are calculated in this way. Only one calculation has to be performed for each monomer, but a separate supermolecule calculation is required for each different geometry. All sets of multipoles are outputted in a format very similar to that required by ORIENT [8]. The supermolecule multipoles are divided into two files, one for the atoms from each monomer. This procedure allows one to use ORIENT to calculate E^{M}_{00}, E^{M}_{*0} and E^{M}_{0*} in a straightforward manner for each geometry, with all multipole/multipole terms up to and including r^{-5} (i.e. up to a total rank of 4) being included.

The calculation of the corresponding Gaussian components E^{G}_{00}, E^{G}_{*0} and E^{G}_{0*} is carried out in an analogous manner using the GMUL program [7]. The

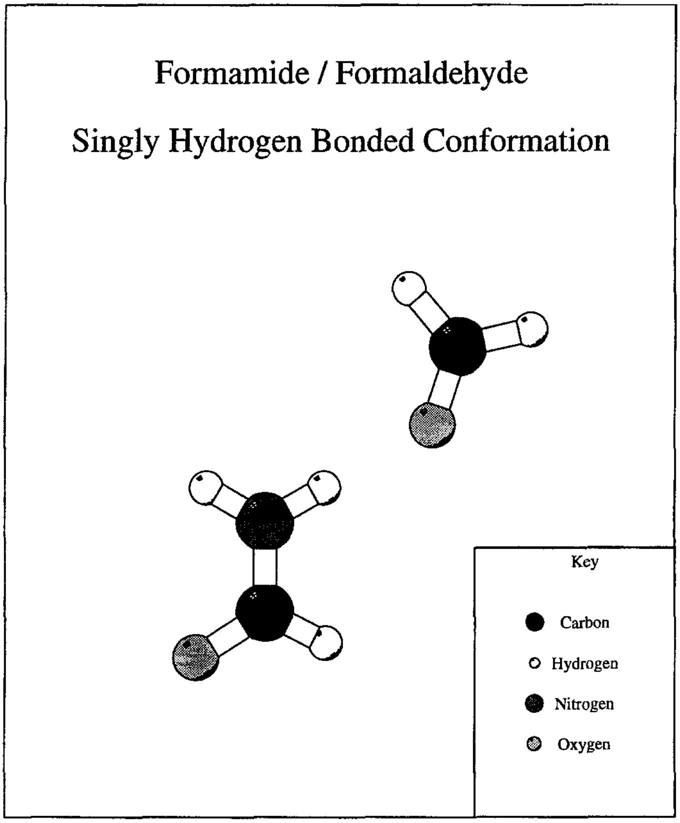

Fig. 1. The formamide/formaldehyde complex used in this work. The two molecules are coplanar, with the N–H ⋯ O atoms collinear. The angle H ⋯ Ô=C is fixed at 138°, corresponding to the minimum energy configuration within the constraints of coplanarity and N–H ⋯ O collinearity [9]. The distance N ⋯ O is varied from 2.7 to 7.0 Å, the IMPT energy minimum (when dispersion is taken into account) is at 2.89 Å. The N–H distance is 1.01 Å. In terms of the parameters defined in Ref. [9], these geometries have $\alpha = 42°$, $\beta = 0°$, $\omega = 0°$. This figure was produced using the program MOLSCRIPT [12]

division of the supermolecule multipoles into two sets has to be carried out by careful manual editing. In order to prevent large amounts of CPU time being used up calculating the very small penetration energies between distant sites, the use of the full Gaussian multipole protocol is restricted to those centres less than 6 bohr (3.175 Å) apart, the other site–site interactions being treated by a point multipole procedure [7]. This is expected to have an insignificant effect on the calculated $E_{\text{pol}}^{\text{G}}$ values. It will have a small but noticeable effect on E_{OO}^{G} at the smaller intermolecular separations.

4 Results

The calculated values of $E_{\text{pol}}^{\text{M}}$ and $E_{\text{pol}}^{\text{G}}$ are shown in Table 1, along with the benchmark energies. The ratios of firstly the multipole-based polarization energy, and secondly the Gaussian polarization energy, to the benchmark are shown as functions of N ··· O distance in Fig. 2. The multipole-based polarization energy exceeds the benchmark energy by up to 45% for geometries in the region of the IMPT minimum, but tends towards it at longer distances. The multipole-based and benchmark energies are very similar for N ··· O distances beyond 5.0 Å. This behaviour is qualitatively similar to that of the ratio of the polarization energies calculated with distributed polarizabilities and by Morokuma analysis, which has also been analysed for this system [9]. The Gaussian polarization energy, which incorporates the effect of charge overlap on the multipoles, is significantly smaller than the benchmark at the shortest distances, but very close to it at the IMPT minimum (probably fortuitously so). As the N ··· O distance is further increased, the Gaussian polarization energy becomes very similar to the multipole-based value (the charge overlap decreasing rapidly with distance), and thus follows the benchmark energy beyond 5.0 Å.

The partitionings of the BSSE-corrected supermolecule interaction energy are shown in Table 2. The variations of the different energy components with N ··· O distance are comparable to those of their IMPT counterparts [9]. In general, the

Table 1. Multipole-based and Gaussian polarization energies compared with the benchmark

	Benchmark energy [kJ mol^{-1}]			Multipole-based polarization energy [kJ mol^{-1}] and ratio to benchmark		Gaussian polarization energy [kJ mol^{-1}] and ratio to benchmark	
d_{NO}[Å]	E_{pol}	E_{CT}	Total	$E_{\text{pol}}^{\text{M}}$	Ratio	$E_{\text{pol}}^{\text{G}}$	Ratio
2.70	− 7.33	− 7.61	− 14.94	− 19.17	1.283	− 9.30	0.623
2.80	− 5.31	− 5.41	− 10.71	− 14.81	1.383	− 9.23	0.862
2.89	− 4.16	− 4.07	− 8.22	− 11.73	1.426	− 8.51	1.034
3.00	− 3.21	− 2.91	− 6.13	− 8.88	1.449	− 7.30	1.192
3.25	− 1.97	− 1.40	− 3.37	− 4.92	1.460	− 4.62	1.372
3.50	− 1.27	− 0.703	− 1.98	− 2.83	1.432	− 2.80	1.416
4.00	− 0.577	− 0.191	− 0.768	− 0.937	1.220	− 0.939	1.223
5.00	− 0.158	− 0.0077	− 0.166	− 0.162	0.979	− 0.162	0.979
6.00	− 0.0569	− 0.00006	− 0.0570	− 0.0569	0.998	− 0.0569	0.998
7.00	− 0.0244	-3×10^{-6}	− 0.0244	− 0.0244	1.000	− 0.0244	1.000

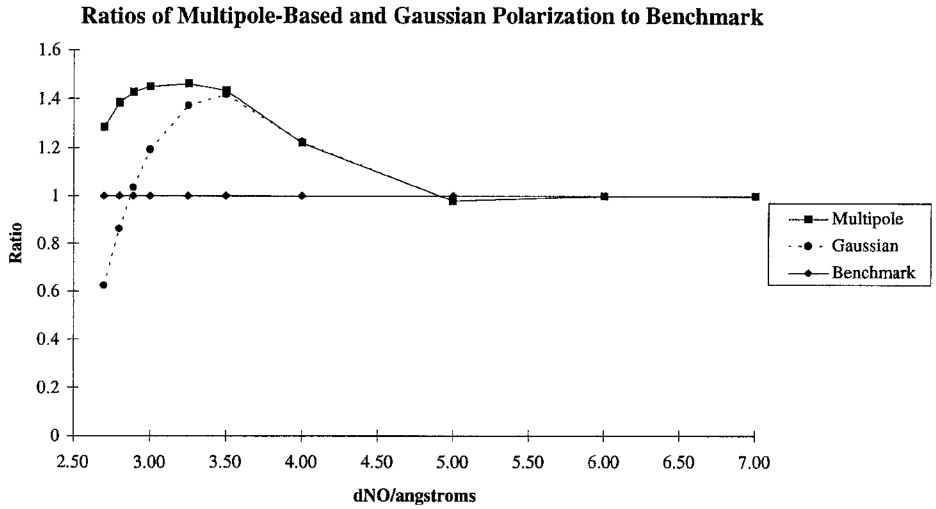

Fig. 2. The ratios of the multipole-based polarization energy (solid curve) and the Gaussian polarization energy (dashed curve) to the benchmark energy, as functions of N ··· O distance

Table 2. Multipole-based and Gaussian partitionings of the BSSE-corrected supermolecule energy

	Supermolecule energy [kJ mol^{-1}]	Components of multipole-based energy partitioning [kJ mol^{-1}]			Components of Gaussian energy partitioning [kJ mol^{-1}]		
d_{NO}[Å]	$E^{\mathrm{C}}_{\mathrm{sup}}$	E^{M}_{00}	$E^{\mathrm{M}}_{\mathrm{pol}}$	$E^{\mathrm{M}}_{\mathrm{erp}}$	E^{G}_{00}	$E^{\mathrm{G}}_{\mathrm{pol}}$	$E^{\mathrm{G}}_{\mathrm{er}}$
2.70	0.05	− 35.10	− 19.17	54.32	− 49.89	− 9.30	59.25
2.80	− 8.64	− 30.89	− 14.81	37.06	− 40.72	− 9.23	41.31
2.89	− 13.39	− 27.71	− 11.73	26.05	− 34.42	− 8.51	29.54
3.00	− 16.49	− 24.44	− 8.88	16.83	− 28.57	− 7.30	19.39
3.25	− 17.63	− 18.84	− 4.92	6.13	− 19.78	− 4.62	6.77
3.50	− 15.69	− 14.94	− 2.83	2.07	− 15.22	− 2.80	2.33
4.00	− 10.95	− 9.99	− 0.94	− 0.02	− 10.01	− 0.94	0.00
5.00	− 5.43	− 5.24	− 0.16	− 0.02	− 5.24	− 0.16	− 0.02
6.00	− 3.19	− 3.14	− 0.06	0.00	− 3.14	− 0.06	0.00
7.00	− 2.06	− 2.04	− 0.02	0.00	− 2.04	− 0.02	0.00

partitioning does appear to work reasonably well. $E^{\mathrm{M}}_{\mathrm{erp}}$, however, becomes slightly negative at 4.0 and 5.0 Å; this would be even more pronounced if the BSSE correction were not made ($E^{\mathrm{M}}_{\mathrm{erp}}$ would then be − 1.54 kJ mol^{-1}, rather than − 0.02 kJ mol^{-1} at 4.0 Å). The Gaussian repulsion energy, $E^{\mathrm{G}}_{\mathrm{er}}$, is also negative at 5.0 Å, but its exclusion of the penetration ensures that it is significantly more repulsive than $E^{\mathrm{M}}_{\mathrm{erp}}$ at shorter distances. The supermolecule interaction energy has its minimum at a greater N ··· O distance than the IMPT one, since the IMPT calculations [9] included the effect of dispersion.

5 Discussion

The multipole-based method gives rise to a polarization energy including charge transfer. The multipole-based polarization energy is seen to depend somewhat on the assignment of overlap charge density between the two molecules in the calculation of the supermolecule multipoles, and it is also affected by exchange and Pauli-exclusion effects. Its evaluation also requires an *ab initio* dimer calculation, as with the IMPT [1] and Morokuma [4] polarization energies; the distributed polarizabilities approach [5] has the considerable advantage of requiring only monomer calculations. The approach of calculating the polarization energy directly from Q and ΔQ has, nonetheless, been shown to be a feasible one.

The multipole-based method provides quite a different approach to the polarization energy from the IMPT [1], Morokuma [4] and distributed polarizabilities [5] methods. It does not reproduce accurately their results, for reasons outlined above. It is clearly not an appropriate method for use in routine situations. It may, however, be of interest in future to investigate more deeply the reasons why the multipole-based method behaves as it does. For instance, the multipole-based approach might be compared with damped methods for calculating the polarization energy, which would avoid the problem of short-range singularities. The differences between the methods also illustrate the difficulty in defining the polarization energy unambiguously.

The use of Gaussian multipoles gives, for the test system, polarization energies significantly smaller in magnitude than those obtained with point multipoles (up to an N ··· O distance of about 3.5 Å). In terms of the partitioning, however, this is more than counterbalanced by the larger electrostatic energy (including penetration). Thus, the sum of the electrostatic and polarization terms is more negative, and the repulsive term (excluding penetration) therefore more positive than with point multipoles.

6 Conclusions

The multipole-based approach to the calculation of the polarization energy directly from perturbed and unperturbed charge distributions has been shown to be feasible, though not appropriate for routine calculation of polarization energies. The polarization energy calculated in this way includes charge transfer, and thus the benchmark with which it is compared is the sum of the (Morokuma) polarization energy and the (IMPT) charge transfer energy. This comparison, however, does not take account of exchange and Pauli-exclusion effects on the multipole-based polarization energy. For the test system of formamide/formaldehyde, the multipole-based polarization energy is an overestimate, exceeding the benchmark by up to 45% in the region of the IMPT minimum. It does, however, agree with the benchmark energy at longer range. A refinement of the method, utilising Gaussian multipoles, gives somewhat better agreement with the benchmark energy for the test system.

Acknowledgements. Dr R.J. Wheatley is thanked for helpful discussions on various aspects of this work. The author was financially supported on a DTI/BBSRC/Glaxo Wellcome Protein Engineering LINK project. A referee provided some insightful comments which have been incorporated into the text. The calculations were carried out on the Convex and Silicon Graphics machines at the Royal Institution.

References

1. Hayes IC, Stone AJ (1984) Molec Phys 53:83
2. Stone AJ (1993) Chem Phys Lett 211:101
3. Stone AJ (1989) Chem Phys Lett 155:102
4. Morokuma K (1971) J Chem Phys 55:1236
5. Le Sueur CR, Stone AJ (1994) Molec Phys 83:293
6. Stone AJ (1981) Chem Phys Lett 83:233
7. Wheatley RJ, Mitchell JBO (1994) J Comp Chem 15:1187
8. Stone AJ (1994) ORIENT: a program for calculating electrostatic interactions, version 3, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW UK
9. Mitchell JBO, Price SL (1990) J Comp Chem 11:1217
10. Boys SF, Bernardi F (1970) Molec Phys 19:553
11. Amos RD (1995) CADPAC: the Cambridge Analytical Derivatives Package, version 6.0, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW UK
12. Kraulis PJ (1991) J Appl Cryst 24:946