A Transferable Distributed Multipole Model for the Electrostatic Interactions of Peptides and Amides

Carlos H. Faerman[†] and Sarah L. Price*[,‡]

Contribution from The University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England. *Received September 12, 1989*

Abstract: The accurate representation of the electrostatic potential around a molecule requires not only point charges at each atom, as generally assumed in molecular modeling studies, but also point dipoles, quadrupoles, etc., to represent the nonsphericity of the atomic charge distribution within the molecule. Such distributed multipole models have been obtained for several amides and several dipeptides with hydrocarbon side chains by a distributed multipole analysis (DMA) of their ab initio SCF wave functions, calculated with a 3-21G basis set. The atomic multipole moments appear to be reasonably transferable to other molecules provided that at least the directly bonded functional groups are the same. This makes it possible to build distributed multipole models to be used for calculating electrostatic interaction energies for polypeptide molecules without requiring an ab initio calculation on the entire molecule, either by using average atomic multipole moments or by transferring the DMAs of peptide fragments. Both schemes are tested in various applications, including the electrostatic potential around an alanine dipeptide in various conformations, and around an undecapeptide cyclosporin derivative. The results show that transferable distributed multipole models enable the electrostatic interactions of polypeptides to be modeled at a new level of accuracy. The main limitations are the errors inherent in any transferable electrostatic model that does not explicitly represent polarization effects.

1. Introduction

It is widely believed that electrostatic forces determine the interactions of biological molecules, including drug-receptor interactions,[1] once the steric constraints are satisfied. If we want to test and utilize this hypothesis, we must use accurate models for the electrostatic potential around a molecule.

Most molecular modeling work uses atomic point charges to represent the electrostatic interactions, which assumes that the molecule can be adequately represented as a superposition of spherical charge distributions. However, the electrostatic potential outside a molecule can be considered as arising from the rearrangement of the valence electrons (from spherical atoms) on bonding, since the electrostatic potential outside a superposition of neutral spherical atoms would be zero. Hence, any atomic point charge model assumes that charge moves from one atom to another on bonding so that the charge distribution of both atoms remains spherical. Such a model cannot represent nonspherical features in the valence-electron distribution such as π electrons and lone-pair density. The errors implicit in assuming an atomic point charge model will be dependent upon the actual molecular charge distribution and the distance from the molecule. However, the errors that remain when atomic point charge models have been optimized by fitting directly to the ab initio electrostatic potential[2-4] may be significant for many applications. The addition of extra charge sites is a useful expedient for improving point charge models for small molecules, but this approach lacks flexibility and is unattractive for large molecules because of the rapid increase in the number of site-site distances that must be summed over to evaluate intermolecular energies.

A recent development in the modeling of the electrostatic potential around small molecules is the use of distributed multipole models,[5-10] where the charge density is represented by a set of point charges, dipoles, quadrupoles, and higher moments on many sites in the molecule, usually with a site on every atom. Distributed multipoles can, in principle, give an exact representation of the electrostatic potential arising from an ab initio charge density throughout the region that is sampled in realistic simulations, assuming that the effects of any overlap of the charge distributions are modeled separately or are absorbed into the repulsion model.[11] The higher multipole moments (dipole, quadrupole, etc.) represent the nonspherical features in the valence-electron distribution around each atom, such as lone pairs and π electrons. Such nonspherical features are often invoked in rationalizing organic reaction mechanisms and hydrogen-bonding geometries and so anisotropic model potentials should be used in computer simulations of such processes. Multisite multipolar electrostatic models have already been applied to the nucleic acid bases,[10,12,13] and it is clear that the modeling of biological processes would benefit from the use of these more accurate models.

We, therefore, want to use the theoretically well-founded distributed multipole models for the electrostatic interactions of peptides. Polypeptides are large, flexible molecules, composed of relatively few different organic functional groups. Hence, the basic concept of organic chemistry suggests that the important features in the charge distribution of a molecular fragment should be the same in different molecules, so it should be possible to build distributed multipole models for polypeptides from ab initio calculations on smaller molecules. The aim of this paper is to investigate the feasibility of such an approach.

In the first part of this paper, we report sets of distributed multipoles for 21 molecules made of amide, aldehyde, and saturated hydrocarbon (CH_n,n = 1-3) groups. These results are analyzed to establish the degree of transferability of the atomic multipole moments and thus to derive empirically some indication of the factors needed to determine transferable atomic types and to obtain a set of average atomic multipole moments. This study also shows how distributed multipole models for larger fragments, such as entire peptide residues, can be defined so as to be reasonably transferable between different polypeptides.

In the second part of the paper, we use the averaged sets of atomic multipoles for the various atomic types, and also the

[†] Current address: Biotechnology Research Institute, 6100 Avenue Royalmount, Montreal, Quebec, H4P 2R2 Canada.

[‡] Current address: Department of Chemistry, University College London, 20 Gordon St., London WC1H 0AJ, England.

(1) Dean, P. M. *Molecular foundations of drug-receptor interaction*; Cambridge University Press: Cambridge, 1987; pp 96-103.
(2) Momany, F. A. *J. Phys. Chem.* **1978**, *82*, 592-601.
(3) Cox, S. R.; Williams, D. E. *J. Comput. Chem.* **1981**, *2*, 304-323.
(4) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129-145.
(5) Stone, A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047-1064.
(6) Sokalski, W. A.; Sawaryn, A. *J. Chem. Phys.* **1987**, *87*, 526-534.
(7) Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934-4948.
(8) Rico, J. F.; Alvárez-Collado, J. R.; Paniagua, M. *Mol. Phys.* **1985**, *56*, 1145-1155.
(9) Cooper, D. L.; Stutchbury, N. C. J. *Chem. Phys. Lett.* **1985**, *120*, 167-172.
(10) Pullman, A.; Perahia, D. *Theor. Chim. Acta* **1978**, *48*, 29-36.
(11) Stone, A. J.; Price, S. L. *J. Phys. Chem.* **1988**, *92*, 3325-3335.
(12) Eisenstein, M. *Int. J. Quantum Chem.* **1988**, *33*, 127-158.
(13) Sokalski, W. A.; Hariharan, P. C.; Kaufman, J. J. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1987**, *14*, 111-126.

0002-7863/90/1512-4915$02.50/0

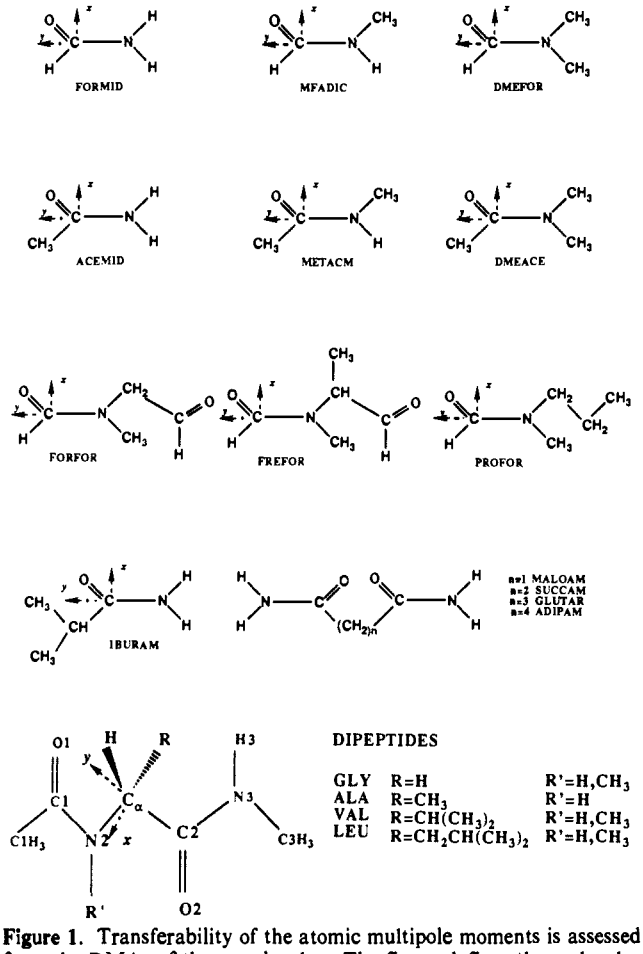

Figure 1. Transferability of the atomic multipole moments is assessed from the DMAs of these molecules. The figure defines the molecular axes for the DMAs deposited as supplementary material and the notation for the dipeptide atoms used throughout this study.

distributed multipoles of various molecular fragments, to build electrostatic models for sample test molecules, from small amides to an undecapeptide. The electrostatic energies calculated from the transferable models are compared with those evaluated from the distributed multipoles obtained directly from the wave functions of the molecules concerned, in order to assess the accuracy of the transferable electrostatic models. We also examine the effect of conformational changes on the distributed multipoles and corresponding electrostatic potential around the alanine-dipeptide to estimate the limits of any model that does not allow for the changes in the polarization of the charge density with conformation.

2. Part I. Transferability of Atomic Multipole Moments

Methodology. We have calculated the SCF wave functions of the molecules shown in Figure 1. This set includes several amides and the acetyl-*N*-methylamide derivatives (dipeptides) of various peptide residues with hydrocarbon side chains. The molecules were chosen to give a large number of examples of a very limited range of functional groups. The geometries were defined with the standard bond lengths and angles of the AMBER force field,[14] with the torsion angles being taken from the crystal structures and a standard CH bond length of 1.063 Å.[15] The use of standard geometries is common in molecular modeling; in this study it was necessary in order to maximize the degree of transferability of the atomic multipole moments. (The use of these atomic multipole moments for molecules with nonstandard geometries is examined in part II.)

The ab initio calculations were performed with the CADPAC[16] suite of programs, on the Cray 1S at London and the Cray XMP at Rutherford. The calculations used a 3-21G basis set,[17] a choice that was mainly dictated by practical limitations, as we wanted to include molecules with up to 34 atoms in the survey and make comparisons with a 3-21G calculation on a undecapeptide. Although 3-21G charge densities are far from definitive, they are on a better theoretical basis than the charge models in most peptide force fields. The problem of the basis set dependence of ab initio based electrostatic models is well-known, and various empirical scalings of the models to give better agreement with larger basis set calculations, or experimental data, have been suggested.[2,3] Since the inclusion of electron correlation effects also has a significant effect on electrostatic properties, the problem of defining which level of ab initio study should give reliable results for molecules such as those in Figure 1 is unresolved. In this study, we circumvent this important problem by making comparisons with ab initio data that were calculated using the same basis set throughout and so we are, strictly speaking, studying the transferability of "3-21G moments". However, the primary aim of this work is to establish how best to transfer atomic multipole moments from calculations on small molecules to give reasonable models for the charge densities of polypeptides. Only when the criteria for transferable multipole moments, and their limitations, are established will it be worthwhile to calculate more accurate atomic multipole moments for suitable small model molecules.

There are several methods of partitioning the charge density between the sites[5-10] to give a multisite multipolar model for the ab initio charge distribution. In the present work we use the distributed multipole analysis (DMA) of Stone,[5] as it allows considerable flexibility in the choice of sites and was designed to optimize the convergence of the multipole series at each site. However, the distinctions between the various methods of obtaining a distributed multipole representation of a charge distribution are less important than the general differences between atomic point charge and atomic multipolar models. The DMA is defined in terms of the Gaussian primitives ϕ_i that comprise the basis set. The charge density of the molecule can be expressed in terms of these primitives and a density matrix by the expression

$$\rho(\mathbf{r}) = \sum_{ij} \rho_{ij}\, \phi_i(\mathbf{r})\, \phi_j(\mathbf{r})$$

Now, if we assume that any penetration effects are modeled separately,[11] the charge distribution associated with the overlap of each pair of orbitals can be represented exactly by a set of point multipoles of order up to $l_i + l_j$ at the center of the Gaussian given by the product of ϕ_i and ϕ_j, where l_i and l_j are the angular momentum quantum numbers of orbitals ϕ_i and ϕ_j. In the DMA, a set of sites for the multipole analysis is chosen, usually with a site on every atom. The overlap multipole moments that are not already centered on one of the chosen sites, which include those arising from the terms involving orbitals centered on different atoms, are then represented at the nearest site, with the formula[5] that represents a multipole at one point by an infinite multipole series at another. Thus, the DMA method is very similar to an extension of Mulliken analysis[18] to give higher multipole moments that complete the description of the charge density, except that, in Mulliken analysis, the contributions from the overlap of orbitals on different atoms are shared equally between the atoms.

If the molecular charge density is calculated with a basis set that contains only s and p orbitals, the highest multipole moment that has to be represented at another site is a quadrupole moment (Q_{2k} in the spherical tensor notation[19] used throughout this work). Thus, provided that the distances between the overlap centers and the expansion sites are small, the multipole series at each site will converge rapidly after the quadrupole moment. The addition of

(14) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
(15) Nyburg, S. C.; Faerman, C. H.; Prasad, L. *Acta Crystallogr.* **1987**, *B43*, 106-111.
(16) Amos, R. D.; Rice, J. E. *CADPAC: The Cambridge Analytical Derivatives Package*, Issue 4.0; University of Cambridge: Cambridge, 1987.
(17) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939-947.
(18) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833-1840.
(19) Price, S. L.; Stone, A. J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987-1001.

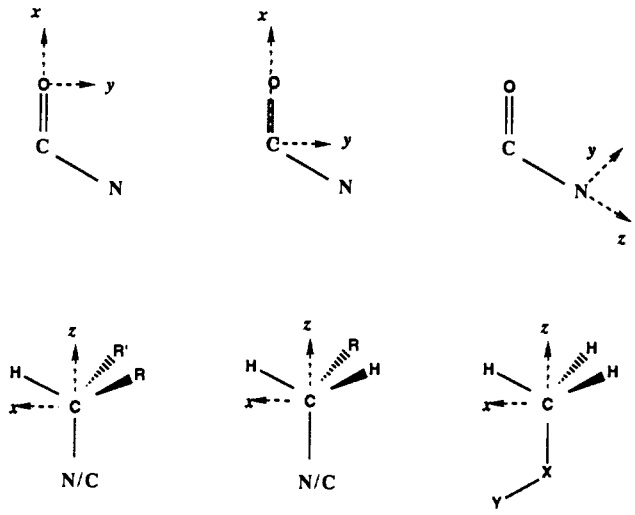

Figure 2. Right-handed orthogonal local axis systems used to define the average multipole moments in Table I. The carbonyl C and O atoms have the x axis along the C→O bond and the amide N (or C for aldehydes) in the xy plane at $y > 0$. The amide N atoms have the z axis in the C→N direction, with the O in the yz plane at $y > 0$. All the united-atom CH_n groups (n = 1–3) had the z axis along the bond from the N/C atom closest to the center of the molecule. A hydrogen at $x > 0$ defines the xz plane. For CH_2 this hydrogen was chosen so that the other hydrogen was at $y > 0$. For CH_3, the hydrogen in the xz plane was the one of highest local symmetry, e.g., the hydrogen coplanar with the peptide group.

sites at the bond centers will clearly improve the convergence of the series, and thus improve the accuracy of the electrostatic energies calculated with a truncated series of multipoles at each site. Conversely, if there is little charge density near a given site, such as a hydrogen atom, then the site can be removed and its contribution represented at the nearest atom with little loss in accuracy and a saving in the number of sites to be summed over in evaluating electrostatic energies. In this study, we have simplified the electrostatic model by using multipole sites on all atoms except those hydrogen atoms that are bonded to carbon atoms. In addition, the hydrogen atoms that are bonded to nitrogen are restricted by allowing them to carry only a charge (Q_{00}). The small higher multipoles that would have been at these hydrogen sites are shifted to the next nearest site.

In order to compare the anisotropic multipole moments on the atoms in different molecules, they must be transformed to the same set of local atomic axes. The local axes are defined by the positions of the bonded atoms, since the nonspherical features in the charge distribution, such as lone pairs, are positioned relative to the intramolecular bonds. Such an atomic local axis system would be a natural choice for simulations of flexible molecules with anisotropic atom–atom model potentials. The axis system for each atom in a given hybridization is defined in Figure 2. The formulas for transforming the multipole moments from the molecular to the local axis system can be derived with standard angular momentum theory,[20] since we have used spherical tensor definitions for the multipole moments. These equations were expanded, simplified and translated into Fortran code by the algebraic manipulation program REDUCE.[21] The Fortran code for transforming all the multipole moments up to octupole Q_{3k} is available from the authors. Further discussion of the rotation of multipole moments can be found in a recent paper by Magnasco.[22] The rotation matrices D^j_{km} for $j \leq 3$ have recently been tabulated in terms of quaternions and direction cosines, as well as Euler angles.[23]

(20) Brink, D. M.; Satchler, G. R. *Angular Momentum*; Clarendon Press: Oxford, 1968.
(21) Hearn, A. C. *REDUCE user's manual*; University of Utah Report UCP-19; University of Utah: Salt Lake City, 1973.
(22) Magnasco, V.; Figari, G.; Costa, C. *J. Mol. Struct. (THEOCHEM)* **1988**, *164*, 49–66.
(23) Lynden-Bell, R. M.; Stone, A. J. *Mol. Simuln.* **1989**, *3*, 271–281.

3. Analysis of the Transferability of Atomic Multipole Moments

Results. The DMAs for the molecules in Figure 1, referred to the molecule fixed axes, are deposited as supplementary material. The charges on a given atom within the set of molecules show a significant variation; for example, the charge on an amide nitrogen ranges from −0.268 to −0.822e. Thus, the definition of transferable atomic types in terms of the atomic number and the hybridization state is not realistic and cannot be used as the basis for transferable model potentials. However, corresponding atoms in different molecules, such as the nitrogen atom in $CONH_2$ groups, do have very similar charges, so a transferable scheme can be defined by a more subtle definition of atomic types.

This is achieved by sorting the sets of transformed atomic multipole moments, derived from the supplementary material, into groups with very similar multipoles, to give the classification shown in Table I. The data naturally fall into distinct groups, with significant differences in the charges between members of different groups, in such a way that it is not necessary to define quantitative criteria for the sorting. The atoms that have similar charges also have similar higher multipole moments in the local axis system. The average multipole moments and their standard deviations (σ_{n-1}) for each group are given in Table I, and in effect, the classification into atoms that seem reasonably transferable is done in order to achieve the smallest possible standard deviations (without having every atom in a group of its own). The subdivision into atomic types suggested by the multipole moments corresponds to the atoms being bonded to different functional groups. For example, the difference in the average charge between an amide nitrogen bonded to two hydrogens and one bonded to a hydrogen and a methyl group is 0.258e, which is over 8 times the sum of the standard deviations for the two atomic types. Thus, the ab initio results suggest that the definition of transferable atomic types must include the specification of the bonded functional groups. In this context the atomic multipoles show that we should consider a $C_\alpha H$ in a peptide residue as a different neighboring functional group from a CH_3 unit, although the bonded atom is an sp^3 carbon atom in both cases.

This subdivision is very clear for the carbonyl and amide groups, with differences of the order of 0.1e in the charges between the different groups. The carbonyl atoms bear much larger charges in an amide group than in an aldehyde group, and the charges for a carbonyl group differ according to whether it is in a HCON or CCON unit. This subdivision is not an artifact of the absence of a hydrogen site, as the changes in the multipoles on the oxygen with and without a site on the proton are insignificant, showing that the need for a separate class comes from the difference in the charge distribution of the oxygen. The amide nitrogen CONHR has a very different set of multipoles according to whether R is H, CH_3, or a peptide C_α group. The differences are smaller for $CON(CH_3)R$. The one example where R is a hydrocarbon chain (from *N*-methyl-*N*-propylformamide) has a DMA that is closer to that of R = C_α than of R = CH_3 and might have been grouped in with the *N*-methyl peptide nitrogen atoms.

The multipoles for the saturated hydrocarbon fragments can also be grouped in a similar way, with CH, CH_2, or CH_3 units that are bonded to different functional groups having markedly different multipoles. However, this subdivision, although the best suggested by the data, is somewhat less satisfactory for the hydrocarbon fragments than for the heteroatoms; often a hydrocarbon group from an amide, with the same neighboring functional groups as one from a dipeptide, has a somewhat different, though small, charge and even larger differences in the higher multipole moments. There was also some evidence that farther neighbors could be important, for example, being bonded to −(CO)H instead of −(CO)C appeared to affect the CH_n multipole moments significantly, but where this is based on only one example, it is just noted as an exception in Table I. Thus, the effects of more distant parts of the molecule appear to be quite important for the hydrocarbon fragments.

Implications for Transferable Multipolar Electrostatic Models. The atomic multipole moments on an atom in a given hybridization

Table I. Groupings of Atomic Multipole Moments Referred to a Local Axis System,[a] Average Values and Standard Deviations within Each Group in Atomic Units

	Q_{00}	Q_{10}	Q_{11c}	Q_{11s}	Q_{20}	Q_{21c}	Q_{21s}	Q_{22c}	Q_{22s}	Q_{30}	Q_{31c}	Q_{31s}	Q_{32c}	Q_{32s}	Q_{33c}	Q_{33s}
Amide Oxygen Oa (22 Occurrences) CCON																
$\bar{Q}_{lk}$	−0.947	0.002	0.242	0.009	0.270	−0.001	−0.003	0.251	0.027	0.003	0.141	0.022	−0.003	0.012	−0.321	−0.137
σ_{n-1}	0.012	0.008	0.018	0.033	0.017	0.014	0.016	0.030	0.043	0.024	0.022	0.036	0.014	0.027	0.056	0.038
Amide Oxygen Ob (6 Occurrences) HCON																
$\bar{Q}_{lk}$	−0.780	0.000	0.033	−0.178	0.064	0.001	−0.001	0.333	0.392	0.003	0.501	0.263	0.002	0.003	−0.192	−0.548
σ_{n-1}	0.008	0.001	0.010	0.013	0.016	0.002	0.003	0.013	0.024	0.006	0.011	0.007	0.004	0.008	0.058	0.014
Carbonyl Oxygen Oc (2 Occurrences) $(CH_n)COH$																
$\bar{Q}_{lk}$	−0.623	−0.004	0.011	−0.154	0.124	0.005	−0.007	0.254	0.351	0.010	0.682	0.237	0.004	0.010	−0.007	−0.549
σ_{n-1}	0.010	0.006	0.008	0.010	0.004	0.007	0.009	0.001	0.017	0.014	0.017	0.003	0.005	0.015	0.008	0.018
Amide Carbon Ca (22 Occurrences) CCON																
$\bar{Q}_{lk}$	1.080	0.003	0.360	0.065	−0.392	−0.003	−0.007	0.003	−0.144	−0.004	−0.110	0.625	−0.049	0.070	−0.674	−0.039
σ_{n-1}	0.054	0.024	0.051	0.023	0.053	0.021	0.046	0.070	0.049	0.104	0.086	0.150	0.057	0.099	0.103	0.184
Amide Carbon Cb (6 Occurrences) HCON																
$\bar{Q}_{lk}$	0.852	0.001	0.310	0.289	−0.302	0.002	0.002	−0.107	0.255	0.002	0.179	0.555	−0.002	0.005	0.135	0.528
σ_{n-1}	0.018	0.002	0.031	0.012	0.029	0.005	0.004	0.035	0.031	0.005	0.017	0.033	0.006	0.012	0.056	0.046
Carbonyl Carbon Cc (2 Occurrences) $(CH_n)COH$																
$\bar{Q}_{lk}$	0.733	−0.006	0.132	0.262	−0.658	0.000	−0.018	0.265	0.377	−0.015	0.835	0.246	0.028	0.003	0.276	−0.400
σ_{n-1}	0.019	0.008	0.010	0.021	0.029	0.001	0.025	0.003	0.008	0.022	0.018	0.088	0.039	0.005	0.021	0.012
Amide Nitrogen Na (7 Occurrences) $CONH_2$																
$\bar{Q}_{lk}$	−0.805	−0.022	0.002	0.023	0.177	−0.002	−0.032	−0.786	−0.002	0.385	0.003	−0.022	0.068	0.007	−0.001	−0.037
σ_{n-1}	0.015	0.027	0.005	0.015	0.045	0.006	0.043	0.031	0.006	0.098	0.007	0.062	0.086	0.022	0.005	0.034
Amide Nitrogen Nb (9 Occurrences) $CONH(CH_3)$																
$\bar{Q}_{lk}$	−0.547	0.213	0.004	0.414	0.178	−0.008	0.374	−1.067	−0.002	0.276	0.006	0.029	0.236	0.009	−0.007	−0.034
σ_{n-1}	0.016	0.029	0.006	0.012	0.049	0.016	0.038	0.020	0.013	0.080	0.047	0.055	0.066	0.034	0.037	0.034
Amide Nitrogen Nb′ (4 Occurrences) CONHCHRCO																
$\bar{Q}_{lk}$	−0.622	0.023	−0.041	0.255	0.316	0.016	0.208	−1.115	−0.156	0.814	0.076	−0.146	0.162	−0.073	0.314	−0.249
σ_{n-1}	0.017	0.009	0.044	0.026	0.035	0.060	0.022	0.027	0.127	0.047	0.031	0.069	0.121	0.140	0.335	0.052
Amide Nitrogen Nc (2 Occurrences) $CON(CH_3)_2$																
$\bar{Q}_{lk}$	−0.289	0.406	0.000	0.019	0.148	0.000	0.042	−1.462	0.000	0.339	0.000	−0.047	0.246	0.000	0.000	−0.093
σ_{n-1}	0.018	0.018	0.000	0.020	0.060	0.000	0.034	0.057	0.000	0.267	0.000	0.219	0.009	0.000	0.000	0.101
Amide Nitrogen Nc′ (5 Occurrences) $CON(CH_3)CHRCO$																
$\bar{Q}_{lk}$	−0.323	0.256	−0.019	−0.158	0.143	−0.004	0.125	−1.555	−0.048	0.933	−0.052	−0.019	0.109	0.040	0.117	−0.116
σ_{n-1}	0.033	0.024	0.064	0.091	0.192	0.085	0.202	0.018	0.147	0.748	0.242	0.164	0.163	0.156	0.377	0.226
Amide Nitrogen Nc″ $CON(CH_3)(CH_2)$ (PROFOR)																
Q_{lk}	−0.333	0.237	0.000	−0.106	−0.065	0.000	0.196	−1.537	0.000	0.469	0.000	−0.245	0.034	0.000	0.000	0.135
Carbon CHa (5 Occurrences) N(CH)CO in Dipeptides																
$\bar{Q}_{lk}$	0.112	−0.341	0.063	0.148	0.447	0.321	0.257	0.237	−0.075	−0.346	−0.585	−0.144	1.831	−0.643	0.524	−0.240
σ_{n-1}	0.026	0.005	0.041	0.070	0.043	0.036	0.094	0.017	0.072	0.234	0.109	0.145	0.077	0.499	0.107	0.084
Does Not Include NCHCHO from FREFOR																
Q_{lk}	−0.041	−0.199	0.139	0.296	0.459	0.307	−0.067	0.249	−0.236	−0.759	−0.476	0.288	1.277	0.047	0.889	0.119
Carbon CHb **(CH)**$(CH_n)_2CO$ C2 (IBURAM)																
Q_{lk}	−0.120	0.168	−0.192	−0.033	−0.147	0.153	−0.041	−0.050	−0.072	0.336	−0.895	−0.033	0.678	−0.028	0.696	−0.012
Carbon CHc (4 Occurrences) **(CH)**$(CH_n)_3$																
$\bar{Q}_{ik}$	0.073	−0.065	−0.108	0.013	0.129	−0.002	−0.086	−0.063	0.000	−0.755	−0.497	0.201	0.848	−0.088	0.993	0.013
σ_{n-1}	0.019	0.009	0.024	0.065	0.026	0.029	0.086	0.096	0.108	0.128	0.107	0.079	0.163	0.135	0.084	0.114
Carbon CH_2a (2 Occurrences) N**(CH$_2$)**CO (GLY)																
$\bar{Q}_{lk}$	0.145	−0.274	−0.001	0.050	0.290	0.328	0.586	0.146	−0.165	−0.448	−0.263	−0.706	0.806	−1.883	1.375	−0.208
σ_{n-1}	0.000	0.003	0.008	0.012	0.023	0.010	0.012	0.009	0.018	0.011	0.014	0.003	0.023	0.008	0.041	0.011
Does Not Include N**CH$_2$**CHO from FORFOR																
Q_{lk}	0.007	−0.070	0.147	0.249	0.257	0.152	0.258	0.184	−0.333	−1.244	−0.152	−0.258	0.611	−1.109	1.881	−0.056
Carbon CH_2b (2 Occurrences) CO**(CH$_2$)**(CH_n) in Diamides																
$\bar{Q}_{lk}$	−0.166	−0.107	0.079	0.136	−0.056	0.195	0.338	0.197	−0.344	−0.572	−0.188	−0.324	0.717	−1.249	1.324	−0.005
σ_{n-1}	0.023	0.004	0.007	0.012	0.014	0.011	0.019	0.012	0.021	0.017	0.047	0.082	0.039	0.069	0.089	0.000
Does Not Include CO**CH$_2$** from SUCCAM																
Q_{lk}	−0.054	−0.171	0.130	0.225	−0.009	0.148	0.256	0.179	−0.312	−0.775	−0.119	−0.205	0.785	−1.367	1.713	−0.006
Carbon CH_2 CO**(CH$_2$)**CO (MALOAM)																
Q_{lk}	−0.267	0.127	0.074	0.182	−0.493	0.404	0.341	0.117	−0.449	0.076	−0.707	−0.272	0.612	−1.456	1.578	0.319
Carbon CH_2c (4 Occurrences) (CH_n)**(CH$_2$)**(CH_n)																
$\bar{Q}_{lk}$	0.034	−0.063	−0.054	−0.022	0.067	0.007	−0.031	−0.049	−0.015	−1.046	−0.080	−0.009	0.282	−0.023	1.558	0.082
σ_{n-1}	0.046	0.059	0.031	0.115	0.101	0.057	0.112	0.016	0.122	0.149	0.383	0.741	0.154	0.559	0.319	0.170
Does Not Include $COCH_2$**CH$_2$**CH_2CO from GLUTAR																
Q_{lk}	0.212	−0.062	−0.044	0.076	0.051	0.013	−0.023	−0.054	−0.094	−1.454	−0.194	0.335	0.430	0.749	2.199	0.008
Carbon CH_2 N**(CH$_2$)**(CH_2) from PROFOR																
Q_{lk}	0.177	−0.173	−0.034	−0.057	0.353	0.055	0.093	0.003	−0.005	−1.174	−0.202	−0.342	0.270	−0.489	1.470	−0.044

Table I (Continued)

	Q_{00}	Q_{10}	Q_{11c}	Q_{11s}	Q_{20}	Q_{21c}	Q_{21s}	Q_{22c}	Q_{22s}	Q_{30}	Q_{31c}	Q_{31s}	Q_{32c}	Q_{32s}	Q_{33c}	Q_{33s}
Carbon CH_3a (19 Occurrences) N(CH_3)																
$\bar{Q}_{lk}$	0.124	−0.090	0.108	−0.001	0.215	0.073	0.000	0.165	0.001	−1.544	−0.180	−0.001	0.232	0.004	2.542	−0.007
σ_{n-1}	0.017	0.013	0.049	0.019	0.033	0.106	0.013	0.043	0.026	0.161	0.166	0.024	0.224	0.048	0.295	0.025
Carbon CH_3b (10 Occurrences) CO(CH_3)																
$\bar{Q}_{lk}$	−0.041	0.186	0.056	0.008	−0.588	0.005	0.011	0.107	−0.010	−0.675	0.396	−0.021	0.013	−0.026	2.804	−0.004
σ_{n-1}	0.010	0.006	0.025	0.013	0.055	0.023	0.010	0.060	0.015	0.093	0.195	0.018	0.059	0.037	0.148	0.012
Carbon CH_3c (8 Occurrences) (CH_n)(CH)(CH_3)$_2$ (VAL, LEU)																
$\bar{Q}_{lk}$	−0.011	−0.080	0.017	0.002	−0.001	0.001	−0.004	−0.007	−0.006	−1.465	−0.001	−0.040	0.156	0.040	2.474	−0.017
σ_{n-1}	0.006	0.009	0.024	0.044	0.010	0.085	0.057	0.034	0.066	0.067	0.261	0.122	0.080	0.134	0.077	0.056
Carbon CH_3d (5 Occurrences) other X(CH)(CH_3)																
$\bar{Q}_{lk}$	0.051	−0.091	−0.017	0.003	0.023	−0.021	−0.011	−0.021	−0.010	−1.546	0.097	−0.016	0.079	0.038	2.600	−0.009
σ_{n-1}	0.022	0.008	0.023	0.022	0.032	0.085	0.079	0.012	0.064	0.131	0.198	0.215	0.197	0.149	0.160	0.031
Hydrogen H (27 Occurrences) NH																
$\bar{Q}_{lk}$	0.334															
σ_{n-1}	0.011															

[a] All the DMAs in the supplementary material are included, with only half of the diamides (MALOAM SUCCAM, GLUTAR, and ADIPAM) contributing to the averages. The local axis system for each atom l defined in Figure 2.

state show significant differences, of the order of 0.1e in the charge, for relatively small changes in the nature of the bonded atoms. Thus, we cannot construct a transferable atom-based electrostatic model without using the nature of the bonded functional groups to define the atomic types. This definition takes into account the short-range inductive effects in a chemical bond and, thus, is in accord with chemical intuition. The short-range inductive effect is clearly seen in the DMAs of the azabenzenes,[24] where there is a marked pattern of each carbon atom donating about one-fourth of an electron to any neighboring nitrogen atom. A study of the atomic multipole moments of purines and pyrimidines[12] (derived by stockholder partitioning of the density) also found reasonable transferability of the functional groups, provided the neighboring atoms were the same, the worst case being the CH groups in the ring. Similarly, the net charges and dipole moments of methyl and methylene groups in $CH_3(CH_2)_mCH_3$ (m = 0–4), derived by Bader partitioning of the charge density, also show a high degree of transferability[25] when the neighboring groups were the same. It is encouraging that this sort of pattern seems also to hold for the molecules in this study and that in many cases quite a reasonable degree of transferability is obtained (i.e., fairly small standard deviations on all multipole moments) by subdividing only according to the bonded functional groups.

The SCF charge distribution, and hence the DMA, also includes the effects of polarization through space from adjacent nonbonded groups. Such effects would often change significantly with the conformation of the molecule. The r^l term in the definition of the multipole moments Q_{lk} means that the higher multipole moments will be more sensitive to the outer edges of the charge distribution and thus to adjacent nonbonded atoms. This probably accounts for there being larger standard deviations associated with the higher moments. However, the relatively small size of the standard deviations in the amide and carbonyl group multipole moments suggests that the effects of neighbors beyond those used to define the atomic types is fairly small. This is also seen in the approximate symmetry of the atomic charge distributions. For example, the N atom in $CONH_2$ (N type a) has approximate C_{2v} symmetry, which if exact, would ensure that the multipoles Q_{lk} with k odd are zero. These multipoles are indeed all very small, in marked contrast with those on N in $CONH(CH_3)$ (N type b), where the differences in the dipoles on the nitrogen associated with the NH and NC bonds produce a significant dipole component in the y direction (Q_{11s}).

The effects of polarization are more significant for the saturated hydrocarbon fragments. The use of the united-atom approach results in significant octupole moments (Q_{3k}) on the carbon atoms, which represent the bonded hydrogen atoms. (The first nonzero multipole moment of CH_4 is an octupole.) If the CH_3 groups had exact C_{3v} symmetry, all the multipoles with $k \neq 0$ or ±3 would be zero. This is only approximately true for CH_3 groups bonded to other saturated hydrocarbon fragments. Examination of the DMAs calculated with sites on all the atoms shows that the three hydrogen atoms rarely carry an equal charge. This is also apparent in dipole-conserving charges[26] and potential derived charges[27] for methyl groups in similar molecules, and it reflects differences in the polarization of the hydrogens by neighboring nonbonded atoms. In this work, the united-atom approach has the advantage of enabling us to include the polarization effects of the bonded functional groups by defining the local axis system to take account of the CH_3 torsion angle, for example, to orient the CH_3 group so that the same hydrogen is closest to the carbonyl oxygen. An all-atom approach would require the definition of conformation-dependent hydrogen atomic types. Either way, the polarization of the hydrogen electron density will inevitably produce problems in using transferable models for hydrocarbon fragments. In simulations where the CH_3 was allowed to rotate, it would not be possible to include such polarization effects without an explicit polarization model, and the hydrogen atoms would have to be made equivalent by imposing C_{3v} local symmetry.

Thus, the charge distribution around each atom, as quantified by a DMA, is primarily determined by the nature of the bonded functional groups through short-range inductive effects. The changes in the atomic charge distribution caused by more distant neighbors, probably mainly by through-space polarization, are at least 1 order of magnitude smaller. However, although the standard deviations in the multipoles in Table I are quite small, even a small charge of 0.01e produces an electrostatic potential of 5 kJ mol^{-1} at a point 3 Å away. Hence, the degree of transferability shown by the atomic multipole moments is not sufficient to ensure that a transferable atomic model will give the electrostatic potential to an accuracy of better than 10 kJ mol^{-1} in the region of importance for molecular recognition.

The use of larger fragments, for example entire peptide residues, is clearly a more attractive method of building an electrostatic model for a polypeptide, as more of the local environment is conserved, and so some of the effects of the more distant parts of the fragment are taken into account. The analysis of the atomic multipole moments shows that the DMAs of molecular fragments should only be transferred between molecules when the functional groups bonded to the fragment are the same. Thus, our observations have important implications for any scheme in which electrostatic models for large systems are built from calculations on smaller model molecules. A model molecule that consisted of the required fragment with the valences satisfied by hydrogen

(24) Price, S. L.; Stone, A. J. *Chem. Phys. Lett.* **1983**, *98*, 419–423.
(25) Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; MacDougall, P. J.; Wiberg, K. B. *J. Chem. Phys.* **1987**, *87*, 1142–1152.
(26) Bellido, M. N.; Rullman, J. A. C. *J. Comput. Chem.* **1989**, *10*, 479–487.
(27) Williams, D. E. *J. Comput. Chem.* **1988**, *9*, 745–763.

or another standard atom would give a very poor model charge distribution. It would be much more accurate to use a smaller fragment of the model molecule, which had been calculated in the same bonding environment, in the model for the large system. The dipeptide molecules include peptide residues (i.e., the $-NHC_\alpha HRCO-$ fragments) with the correct neighboring functional groups and so the DMAs of the peptide fragments can be used in modeling polypeptides. Thus, we can compare the use of transferable peptide fragments (using the DMAs in the supplementary material) with the use of transferable atomic multipole moments (using the average values in Table I) for modeling the electrostatic potential around polypeptides.

4. Part II. Assessment of Transferable Distributed Multipole Models

We now use the average sets of atomic multipoles of Table I and, for larger molecules, also the peptide fragment multipoles from the original dipeptide DMAs, to build transferable ATOM and PEPTIDE models for calculating the electrostatic potential around various molecules. The accuracy of these models can be assessed by comparing the resulting electrostatic energies with the "exact" energies calculated from the DMA of the molecule's wave function and also from the corresponding Mulliken charges. Thus, the errors involved in transferring the atomic multipoles between molecules can be compared with those resulting from truncating the analysis of the charge density at the first term to give an atomic point charge model.

The first problem in building a transferable electrostatic model from fragments is that it is unlikely to result in the molecule having exactly the correct net charge (zero for neutral molecules). The sum of the fragment charges should be small, but nevertheless, the excess charge needs to be removed, or it will make a significant unphysical contribution to the electrostatic potential. Therefore, we use the following recipe to neutralize the transferable distributed multipole models. If the sum of the average charges Q_{00}^i on the N atomic sites is q, and the sum of the corresponding standard deviations of these charges σ^i is s, we would probably find that $|q| \lesssim s$. (A more rigorous statistical treatment is not appropriate as the variations in the charges on different atoms are not independent.) The charge on each atom i is adjusted according to its likely variation, so each atom has a charge of $Q_{00}^i - (q/s)\sigma^i$, which makes the model molecular charge distribution neutral. In principle, similar adjustments can be made to give agreement with the total dipole of the molecule, but dipole moments are not available for most molecules of interest.

The aim of this paper is to explore the possibilities of transferable multipole models, rather than to recommend a definitive set. Hence, we did not continually extend our database of DMAs until we had a statistically significant number of occurrences for every atomic type that we have decided is necessary, but only for those that are most important in modeling peptides. Also, the ATOM models use the averaged atomic multipole moments straight from Table I, without any further adjustments, such as giving CH_3 groups C_{3v} symmetry or ignoring certain small multipole components. (The systematic simplification of Table I to eliminate negligible multipole components is not straightforward, since the choice of axis system determines which multipole components are small.)

The analysis of the transferability of atomic multipole moments suggests that the DMAs of the dipeptides could be used in forming a transferable peptide-distributed multipole model. A prerequisite for any model that builds a polypeptide charge distribution from a set of residue charge distributions is that each residue should be neutral (or carry the correct integer charge for the ionized residues). This not only is a convenient method of ensuring that the resulting polypeptide has the right net charge but also enables the long-range electrostatic energy to be calculated more accurately, since it is possible to truncate the summation range so that it includes only a neutral unit. The DMAs of the dipeptide molecules show that this is a very reasonable approximation, as the net charges on the peptide residues are +0.008, −0.002, +0.036, +0.009, and −0.015e on the glycine, alanine, valine, leucine, and isoleucine residues, respectively. The N-methylated peptide residues carry larger net charges of +0.046, +0.075, and +0.042e for *N*-MeGly, -Val, and -Leu, respectively, but even these charges are sufficiently small that it is reasonable to consider modeling the peptide as a neutral unit. It is also reasonable to neutralize each peptide residue by distributing the net charge on just the backbone atoms, as these are more likely to be affected by the variations in the neighboring residues. This is also convenient, as it is then unnecessary to have estimates of the standard deviations for the side-chain atoms. These neutralized peptide fragments provide the transferable PEPTIDE model that is also investigated in the next section.

Table II. Minima in the Electrostatic Interaction Energy of *N*-Methylacetamide/Formamide Complex

	ab initio DMA[a]	average atomic DM		Mulliken charges	
minimum	$U_{Estatic}$ (kJ mol^{-1})	$U_{Estatic}$ (kJ mol^{-1})	corresp[b] min?	$U_{Estatic}$ (kJ mol^{-1})	corresp[b] min?
M1	−37.9	−37.1	yes	−33.3	yes
M2	−36.2	−35.6	yes	−31.6	yes
M3A	−34.3	−34.6	yes	−30.3	*c*
M3B	−33.6	−34.2	no	−29.8	no
M4A	−30.0	−30.9	no	−16.6	no
M4B	−30.2	−31.1	yes	−16.4	no

[a] The minima are calculated with a DMA of the 3-21G wave function with multipoles up to rank 4 on every atom. The orientations in the minima are shown in Figure 3. [b] This shows whether the electrostatic model also has a minimum in essentially the same position. [c] The Mulliken charge model has a global minimum of −33.5 kJ mol^{-1} with a single hydrogen bond to the *N*-methylacetamide oxygen (cf., M3), but with the formaldehyde molecule almost perpendicular to the plane of the *N*-methylacetamide and on the nitrogen side of the acceptor carbonyl.

The following studies were chosen to assess first the errors involved in using the averaged atomic multipoles and then the additional errors involved in so modeling molecules in different conformations and with nonstandard bond lengths. The calculations were performed with the program ORIENT,[28] which uses molecule-fixed axes, and so the atomic multipoles had to be transformed back from the local axis system to the appropriate orientation for the molecular fragment in the molecule. This would not be necessary for a program that worked in terms of the local axis system.

5. Hydrogen Bonding in the *N*-Methylacetamide/Formamide Complex

Hydrogen bonding plays a major role in determining the structure and properties of peptides, and the hydrogen bond is mainly electrostatic in origin. Hence, it is very important that any electrostatic model for peptides gives a good prediction of the electrostatic contribution to hydrogen bond energies. To test this, the minima in the electrostatic energy of the *N*-methylacetamide/formamide (METACM/FORMID) complex (in the orientations allowed by a hard-sphere model of the molecular shapes) are calculated with both a full DMA of the molecules obtained from the original 3-21G wave functions and the ATOM model which uses the same geometries and the average DMAs of Table I. The sums of the average atomic charges, prior to neutralization by the above method, are −0.065e for formamide and 0.003e for *N*-methylacetamide, which is reasonable in comparison with the sums of the standard deviations in the charges of s = 0.063 and s = 0.120, respectively. The details of the hard-sphere model and the minimization procedure are given in a study of the electrostatic directionality of N—H···O═C hydrogen bonding.[29]

The minima in the electrostatic energy of the *N*-methylacetamide/formamide complex are shown in Figure 3. The two lowest energy minima have two contacts between the hard spheres representing the molecular shape, and the other four also have a N—H···O═C hydrogen bond. Table II shows that the transferable atom electrostatic model gives a good prediction of the electrostatic energy at these positions with errors of less than 1

(28) Price, S. L.; Stone, A. J. *J. Chem. Phys.* **1987**, *86*, 2859–2868.
(29) Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* **1989**, *154*, 267–272.

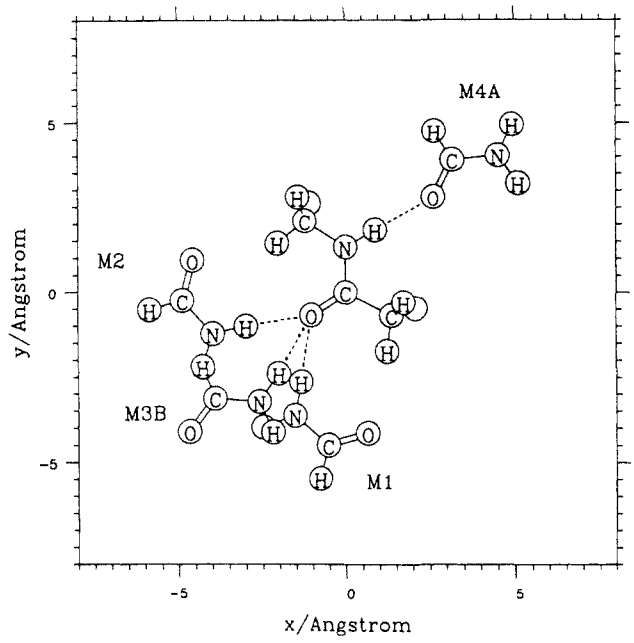

Figure 3. Minimum electrostatic energy structures, in accessible orientations for the *N*-methylacetamide formamide complex. The *N*-methylacetamide is shown at the origin, and the various relative positions of the formamide are shown in an orthogonal projection. The minima M3A and M4B are omitted for clarity, as they have the same hydrogen-bond geometry as M3B and M4A, respectively, but are rotated around the hydrogen bond so that the plane of the formamide molecule is approximately perpendicular to that of the *N*-methylacetamide.

kJ mol^{-1}. It also gives a satisfactory prediction of the positions of the minima in the electrostatic interaction energy, with identical positions for the lowest energy minima, though the positions of some of the local minima differ in the orientation of the parts of the formamide molecule not involved in the hydrogen bond. Hence, the errors involved in using the average DMAs, and the neutralization procedure, are fairly minor.

In contrast, the Mulliken charges derived from the molecular wave functions give a qualitatively different picture of the electrostatic forces between these molecules. Mulliken charges predict that the electrostatic energy for minima where the hydrogen bond involves the formamide carbonyl group is almost half that found for hydrogen bonds to the *N*-methylacetamide carbonyl oxygen, whereas the full DMA gives a difference of less than 7 kJ mol^{-1}. The Mulliken charges predict a global minimum for a single-contact nonplanar, nonlinear hydrogen bond, in complete disagreement with the accurate DMA results and the observed directionality of N—H···O=C hydrogen bonds.[30]

Thus, a transferable ATOM multipole model is much more accurate than the use of Mulliken charges, showing that the approximation of assuming transferability is far more realistic than approximating the molecular charge densities by this point charge model.

6. Prediction of the Potential Around the Isoleucine Dipeptide

This study investigates the predictive power of a transferable distributed multipole model for a molecule that is not included in the work in part I but, nevertheless, is closely related to many of the molecules in Figure 1.

The wave function and DMA of the isoleucine dipeptide (defined by Figure 1 with R = $CH(CH_3)CH_2CH_3$) were calculated with the same basis set and standard geometry as in part I, and the corresponding DMA is reported in the supplementary material. The charges on most of the atoms are within the average value $\bar{Q}_{00} \pm \sigma$ for the corresponding atomic type (Table I), though three atoms (O2, N3, N2) have charges that are slightly outside this limit. The significant differences from the average values are for $C_\beta H$, $C_{\gamma 1}H_3$, and $C_\delta H_3$, which is consistent with the divisions into transferable types being less clear-cut for the hydrocarbon fragments and the small samples of some of these types. $C_\beta H$ in the isoleucine dipeptide has a charge of 0.116*e*, compared to the values of 0.095, 0.083, 0.058, and 0.057*e* for the CH units in the side chains of the Val, *N*-MeVal, Leu, and *N*-MeLeu dipeptides, respectively. The isoleucine side-chain methyl groups, $C_{\gamma 1}H_3$ and $C_\delta H_3$, have charges of −0.028 and 0.002*e*, which are rather different from the charges on the terminal CH_3 in the Val and Leu side chains (−0.011 ± 0.006*e*) or those in the other group, X-(CH)CH_3 (0.051 ± 0.022*e*). These results are satisfactory considering the aims and limitations of Table I, but they confirm the conclusion from part I that the variations in the charge distributions in hydrocarbon fragments, due to subtle balances of electronegativity and polarization effects, limit the accuracy of any transferable model based on small fragments. Given the small size of the charges on the hydrocarbon side chain, which might be relatively unimportant in determining the electrostatic interactions of the dipeptide, it is still worth investigating the accuracy of the ATOM model, although this scheme would be difficult to apply more accurately to hydrocarbons. The net charge of the ATOM model, prior to neutralization, is q = 0.168*e*, which is less than the sum of the standard deviations s = 0.349*e*.

A second transferable model (PEPTIDE) is also constructed from the DMA of the isoleucine residue, by neutralizing the net charge of −0.014*e* over the backbone atoms C2, O2, C_α, N2, and H2 and then by using the average atomic multipoles (Table I) to represent the end groups, with the charge on C1 being decreased by 0.003*e* to make the net charge on the end groups (C1, O1, $C1H_3$, N3, H3, $C3H_3$) exactly zero. Thus, the PEPTIDE model is almost identical with the original DMA for the isoleucine residue and very similar to the ATOM model for the end groups, the differences being only in the charges on the backbone atoms from different neutralization procedures. The ATOM model differs most from the original DMA or the PEPTIDE model in the hydrocarbon side chain, with charge differences of up to 0.05*e*.

Whether these relatively small differences between the electrostatic models are practically significant depends on the accuracy with which they predict the electrostatic interactions of the molecule. This is tested by comparison with the electrostatic potential evaluated from a DMA of the isoleucine dipeptide wave function that had all multipoles up to hexadecapole (Q_{4k}) on every atom and can be taken as the accurate electrostatic potential corresponding to this wave function, in the absence of penetration effects. The potential is compared on a 1-Å cubic grid of points around the molecule, excluding any point that is closer than 1 Å from the van der Waals surface (i.e., closer than 3 Å from any C atom, 2.4 Å from O, or 2.5 Å from N). Table III gives the errors in the potential at the closest grid point to each atom and also the rms errors for all the grid points in the ranges 1–2 and 2–5 Å from the van der Waals surface.

The transferable ATOM model gives rise to significant errors in the electrostatic potential in the regions that would be sampled by other molecules in "contact" with the dipeptide, with an rms error of 10 kJ mol^{-1} for points 1–2 Å from the van der Waals surface. There are significant errors near the hydrocarbon side chain, as well as close to the more polar backbone atoms. The largest error of −27.5 kJ mol^{-1} is at a grid point that is close to several atoms. The PEPTIDE model gives an error of only −5.3 kJ mol^{-1} at this point so the problems with the ATOM model must be predominantly due to its poor representation of the hydrocarbon side chain. The PEPTIDE model is much more successful in all regions, with an rms error of only 3 kJ mol^{-1} close to the van der Waals surface. Thus, the use of larger fragments, which include more polarization effects, can give a significant improvement in an electrostatic model, even when the molecules are so closely related that the differences are relatively small and mainly in hydrocarbon fragments. This is despite the assumption that it is valid to neutralize the peptide residue and the rest of the molecule independently.

Table III also shows the errors arising from approximate representations of the wave function. Comparing the potential

(30) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1983**, *105*, 5761–5766.

Table III. Comparison of Predicted Electrostatic Potentials at a Grid around the Isoleucine Dipeptide[a]

atom	dist/(Å)	closest grid point (Å)			potential	errors model − potential (kJ mol^{-1})			
						DMA	ATOM	PEPTIDE	Mulliken
C2, C1[b]	3.038	0.0	−2.0	−3.0	−71.3	−2.5	−27.6[c]	−5.3	−48.0[c]
O2	2.464	−3.0	−1.0	1.0	−42.1	−0.4	−3.0	3.1	−43.6
N2	2.609	−3.0	2.0	0.0	93.0	0.6	14.7	10.5[c]	−38.4
$C_\alpha H$	3.162	3.0	1.0	0.0	45.2	−0.2	9.7	1.7	−18.2
$C_\beta H$	3.202	−2.0	0.0	4.0	19.7	0.1	8.6	1.3	−17.8
$C_{\gamma 1}H_2$	3.054	−1.0	5.0	1.0	16.2	−0.9	15.9	2.3	−7.5
$C_\delta H_3$	3.044	−3.0	4.0	3.0	28.7	−1.6	14.7	0.9	−0.3
$C_{\gamma 2}H_3$	3.032	0.0	−1.0	5.0	16.5	0.9	7.7	0.7	2.3
O1	2.440	2.0	−1.0	−3.0	−95.1	0.9	−16.2	−0.6	−31.8
$C1H_3$	3.005	−3.0	4.0	−2.0	39.5	−2.6[c]	1.4	2.8	16.2
N3	2.698	3.0	−2.0	−2.0	5.0	1.9	−13.8	−2.0	−42.3
$C3H_3$	3.035	4.0	−5.0	0.0	23.3	1.8	−7.9	−2.8	11.1
rms value for 416 pts 1–2 Å from vdW surface					32.9	0.6	10.1	3.0	14.1
2384 pts 2–5 Å					13.8	0.1	5.4	1.5	5.1

[a] The electrostatic models being compared with the potential evaluated from a DMA with multipoles up to hexadecapole on all atoms are as follows: DMA, use of DMA with restrictions on hydrogen sites as used throughout this study; ATOM, use of average atomic DMAs from Table I; PEPTIDE, peptide residue taken from the DMA of the dipeptide and neutralized, end groups from average atomic DMAs; Mulliken, use of Mulliken charges from the corresponding wave function. [b] The closest atom to this grid point is O1. [c] Largest error for model on entire grid.

calculated from the DMA used in this work, with that calculated from a more accurate DMA with multipoles upto hexadecapole on all atoms, shows that it is an excellent approximation to omit the DMA sites on the hydrogen atoms in CH_n groups and only have charges on the hydrogens bonded to nitrogen, even for points very close to the atoms. In contrast, the Mulliken charges derived from the molecule's wave function predict the potential with an rms error of 14 kJ mol^{-1} close to the molecule, with errors greater than 40 kJ mol^{-1} near some atoms. Since the higher multipoles have most effect at short range, it is not surprising that the Mulliken charges do particularly badly near the van der Waals surface. However, it is worth noting that the errors involved in representing a wave function by its Mulliken charges are much larger close to the molecule than those involved in assuming the transferable ATOM model.

Thus, the transferable ATOM model gives a moderately good prediction of the electrostatic potential, which will be adequate for many purposes, despite its poor representation of the intramolecular polarization effects within the hydrocarbon side chain. The transferable PEPTIDE model is very successful. However, this approach does require the identification of virtually neutral fragments within model molecules that are sufficiently large to have the correct bonding environment for the molecule. Otherwise, some scheme fragment coupling procedure is required, such as those devised by Bellido and Rullman to couple charge models for the backbone and side chains of dipeptides[26] or the superposition procedure of Vigné-Maeder,[31] which will further reduce the accuracy of the electrostatic model.

7. Effect of Conformation on Distributed Multipoles

One of the major assumptions implicit in any transferable force field is that the charge distribution associated with each atom, and thus its potential parameters, does not change with the conformation of the molecule. This is clearly an approximation that ignores changes in the intramolecular polarization of the molecular charge density with conformation. For example, two low-energy conformations, with different torsion angles, may have different functional groups in the strong electric field around a carbonyl oxygen atom, and the charge density on these groups will be polarized by this electric field. The changes in the induced multipole moments will be particularly large when hydrogen bonds are formed. However, it may be a reasonable approximation to neglect these polarization effects for the type of conformation that would be sampled within a molecular modeling study, though clearly the approximation would be very poor if, for example, there was significant overlap of the charge density of nonbonded atoms or if the intramolecular bonds were very elongated. To test the transferability of the multipoles with changes in the torsion angles in a dipeptide, we have calculated the SCF wave function for the alanine dipeptide in three reasonably natural conformations. The first conformation is that of the dipeptide structure used in part I, where the torsion angles $\phi = -84°$ and $\psi = 159°$ are taken from the crystal structure and the conformation is denoted CRYSTAL. The second conformation, TORSION, has angles $\phi = -175°$ and $\psi = 175°$ coresponding to a minimum in a molecular mechanics study of this molecule.[32] Another of these minima, corresponding to an intramolecular hydrogen bond, was optimized with the AMBER force field, and the resulting conformation with $\phi = 75°$ and $\psi = -65°$ is denoted H-BOND. The TORSION and H-BOND geometries are determined only by (model) intramolecular forces, whereas the experimental CRYSTAL geometry is a balance of intra- and intermolecular forces.

Table IV. Comparison of Charges in the DMA for the Ala Dipeptide in Different Conformations

Charges on Atoms in Ala Residue (e)

conformer	C2	O2	N2	H2	$C_\alpha H$	$C_\beta H_3$
CRYSTAL	1.060	−0.958	−0.635	0.325	0.148	0.058
TORSION	1.102	−0.958	−0.585	0.346	0.094	0.027
H-BOND	1.151	−0.971	−0.602	0.328	0.032	0.095
av $\bar{Q}_{00}$[a]	1.080	−0.947	−0.622	0.334	0.112	0.051
σ_{n-1}[a]	0.054	0.012	0.017	0.011	0.026	0.022

Charges on End Group Atoms (e)

conformer	C1	O1	$C1H_3$	N3	H3	$C3H_3$
CRYSTAL	1.047	−0.935	−0.036	−0.531	0.314	0.142
TORSION	1.038	−0.946	−0.024	−0.561	0.322	0.145
H-BOND	1.038	−0.969	−0.014	−0.610	0.394	0.128
av $\bar{Q}_{00}$[a]	1.080	−0.947	−0.041	−0.547	0.334	0.124
σ_{n-1}[a]	0.054	0.012	0.010	0.016	0.011	0.017

[a] Values from Table I for corresponding atomic types, averaged over charges in different molecules.

The charges from the DMAs for all three conformations of the alanine dipeptide are given in Table IV. The variation of most of the multipoles with conformation is on the order of the standard deviations in Table I (and similar to the variations in dipole-conserving charges with dipeptide conformation[26]), with the notable exception of the N3, H3, and O1 atoms involved in the internal hydrogen bond. The atomic charge on H3 increases by about 0.08e on hydrogen bonding, though the net change in the charge in the N–H bond is much smaller. The hydrogen bond also changes the charge, and higher multipoles, on the proton acceptor O1. Although this is the most dramatic change, the charges on the other atoms in the TORSION and H-BOND geometries often lie outside the range of the average value $\bar{Q}_{00} \pm \sigma$, though they are generally within $\bar{Q}_{00} \pm 2\sigma$. This result implies that the changes in the Ala dipeptide charges with conformation

(31) Vigné-Maeder, F. *Chem. Phys. Lett.* **1987**, *133*, 337–342.

(32) Pettitt, B. M.; Karplus, M. *J. Am. Chem. Soc.* **1985**, *107*, 1166–1173.

Table V. Comparison of Electrostatic Potential (V) around the Ala Dipeptide in Different Conformations for Various Electrostatic Models

potential around conformer	CRYSTAL		TORSION		H-BOND	
Grid Definition						
grid limits[a] (Å)	1–2	2–5	1–2	2–5	1–2	2–5
grid spacing (Å)	0.5	1.0	0.5	1.0	0.5	1.0
no. points	2851	2099	2842	2084	2828	2079
rms value V (kJ mol^{-1})	43.4	18.7	33.8	13.7	37.4	16.2
max V (kJ mol^{-1})	122.1	53.6	131.8	53.5	139.4	56.4
min V (kJ mol^{-1})	−125.5	−70.4	−114.7	−61.0	−119.2	−70.2
Electrostatic Model						
CRYSTAL DMA						
rms error (kJ mol^{-1})	0	0	1.8	0.6	6.1	2.8
max error $\|\Delta V\|$ (kJ mol^{-1})	0	0	6.6	2.5	21.4	11.7
TORSION DMA						
rms error (kJ mol^{-1})	3.7	1.6	0	0	8.4	3.8
max error $\|\Delta V\|$ (kJ mol^{-1})	14.7	7.7	0	0	26.2	14.7
H-BOND DMA						
rms error (kJ mol^{-1})	10.0	4.1	11.5	4.8	0	0
max error $\|\Delta V\|$ (kJ mol^{-1})	44.9	18.5	49.5	21.3	0	0
Transferable ATOM Model with Average Multipoles[b]						
rms error (kJ mol^{-1})	3.4	1.6	3.2	1.4	6.8	3.1
max error $\|\Delta V\|$ (kJ mol^{-1})	12.4	5.6	9.3	4.9	19.5	11.1
Transferable PEPTIDE Model with CRYSTAL DMA for Side Chain[c]						
rms error (kJ mol^{-1})	3.0	1.5	2.9	1.2	4.7	1.9
max error $\|\Delta V\|$ (kJ mol^{-1})	9.5	4.8	9.0	4.2	14.9	7.7
MULLIKEN Charges from Wave Function of Corresponding Conformer						
rms error (kJ mol^{-1})	14.5	5.1	15.3	5.9	15.0	4.8
max error $\|\Delta V\|$ (kJ mol^{-1})	49.7	22.9	50.1	22.8	65.7	28.5

[a] All points from the cubic grid within this range of the van der Waals surface of the molecule. [b] The net charge q = 0.011e was neutralized as described in the text. [c] The net charge of −0.002e on the peptide residue was neutralized on C2, and the end groups were constructed from the average atomic multipoles as described for the isoleucine dipeptide ATOM model.

are very similar to the variations in charges on an atom between different molecules with changes in the farther functional groups. If this was generally true of peptide charge distributions, it would be futile to seek a significantly more accurate definition of transferable atomic types than those given in Table I. However, within a DMA, the changes in the charges with conformation could be partially counterbalanced by changes in the higher moments. For example, the charge shift on hydrogen bonding produces a change in the net dipole moment of the NH bond, which is slightly counteracted by a change in the dipole moment on the N atom. The higher multipole moments often show larger variations than the charges, reflecting their greater sensitivity to the distortion of the outermost regions of the charge distribution. The variation of the octupole moments on the CH_n groups will also reflect changes in the polarization of the bonded hydrogen atoms and so, for example, $|Q_3|$ for $C_\beta H_3$ is larger for TORSION, where this methyl group is close to O1 than it is in the CRYSTAL or H-BOND conformations.

Thus, the distributed multipoles certainly change with conformation, though the changes are fairly small except where the atoms are involved in a very strong electrostatic interaction, such as the hydrogen bond, where we would expect significant polarization. The significance of these distortions depends on the effect they have on the electrostatic potential around the model. Table V shows the error in rotating the multipoles calculated from a wave function in one conformation, so that they can be used to predict the potential around the molecule in another conformation, as rms errors on various grids of points outside the molecule. These results are contrasted with those obtained from the transferable ATOM model, the transferable PEPTIDE model (from the residue from CRYSTAL), and the Mulliken charges from the different wave functions.

The clear conclusion from Table V is that significant errors can be introduced by assuming that the charge distribution of a molecule (as represented by atomic multipoles) is independent of its conformation. The use of the DMA calculated in the TORSION geometry for the CRYSTAL geometry, and vice-versa, gives quite a reasonable prediction of the potential, with rms errors of less than 5 kJ mol^{-1} even when only 1–2 Å from the van der Waals surface. In contrast, if the H-BOND DMA is used to model these conformers, there are very significant rms errors of greater than 10 kJ mol^{-1} because the potential is sampled close to the atoms that were polarized in the hydrogen bond. The errors involved in using the DMAs from the non-hydrogen-bonded conformers to model the potential around H-BOND are smaller since the atoms involved in the hydrogen bond are buried within the molecule. Hence, transferring an electrostatic model between similar low-energy conformers will be a fair approximation unless there are changes in any strong electrostatic interactions, such as hydrogen bonds.

The transferable ATOMIC and PEPTIDE models give very similar predictions of the potential around the different conformers to the CRYSTAL and TORSION DMAs, the transferable PEPTIDE model being marginally better as would be expected from the more accurate balancing of the short-range inductive effects within the peptide residue. Thus, the errors in assuming transferability between conformations appear to be comparable to the errors in assuming transferability between molecules when the neighboring functional groups are the same.

By far the worst electrostatic model examined in Table V is the use of the Mulliken atomic charges of the corresponding wave function, with rms errors of around 15 kJ mol^{-1} close to the molecule, where the errors in any point charge model are likely to be largest. Thus, the errors in a transferable electrostatic model, assuming transferability either between conformations or between molecules, are much smaller than the errors involved in using the Mulliken charges to represent the ab initio charge distribution.

8. Electrostatic Potential around an Undecapeptide

The purpose of transferable electrostatic models is to enable us to predict the electrostatic interactions of polypeptides that are too large for an ab initio calculation on the molecule itself to be feasible. Thus, the possibilities for testing transferable models in the applications that they are aimed at are intrinsically limited. However, an ab initio SCF calculation has been performed on a cyclic undecapeptide,[33] a derivative of the immunosuppressive cyclosporin, $C_{63}H_{113}N_{11}O_{12}$, with a 3-21G basis set, as a benchmarking exercise for a direct SCF[34] program. The calculation took 400 h on an FPS-164, and so such calculations will not be routine in the foreseeable future.

Cyclosporin has the peptide sequence (1) MeBmt, (2) Abu, (3) Sar, (4) MeLeu, (5) Val, (6) MeLeu, (7) Ala, (8) D-Ala, (9) MeLeu, (10) MeLeu, and (11) MeVal. A derivative form of the MeBmt peptide was used in the crystal structure and the ab initio calculation. All the NH groups are involved in an internal hydrogen bond, and the 9–10 peptide linkage is cis. A comparison of the DMA of the cyclosporin wave function with the atomic multipole moments found in this study is facilitated by the use of the same basis set. The most marked differences in the charges are in the peptide linkages O4C4–N5, O9C9–N10, O2C2–N3, and O7C7–N8, which have C–N bond lengths of 1.13, 1.20, 1.44, and 1.43 Å, respectively, in the crystal structure[35] used in the ab initio calculation. These are unphysically distorted from the normal peptide bond length of 1.335 Å used in our calculations, so that it is not surprising that the charges are significantly different; for example, the charge on O4 is −0.746e compared with the average value in Table I of −0.947e. Thus, the poor quality of the crystal structure prevents a rigorous comparison of either

(33) Price, S. L.; Harrison, R. J.; Guest, M. F. *J. Comput. Chem.* **1989**, *10*, 552–567.
(34) Almlöf, J.; Faegri, K.; Korsell, K. *J. Comput. Chem.* **1982**, *3*, 385–399.
(35) Petcher, T. J.; Weber, H.-P.; Rüegger, A. *Helv. Chem. Acta* **1976**, *59*, 1480–1488.

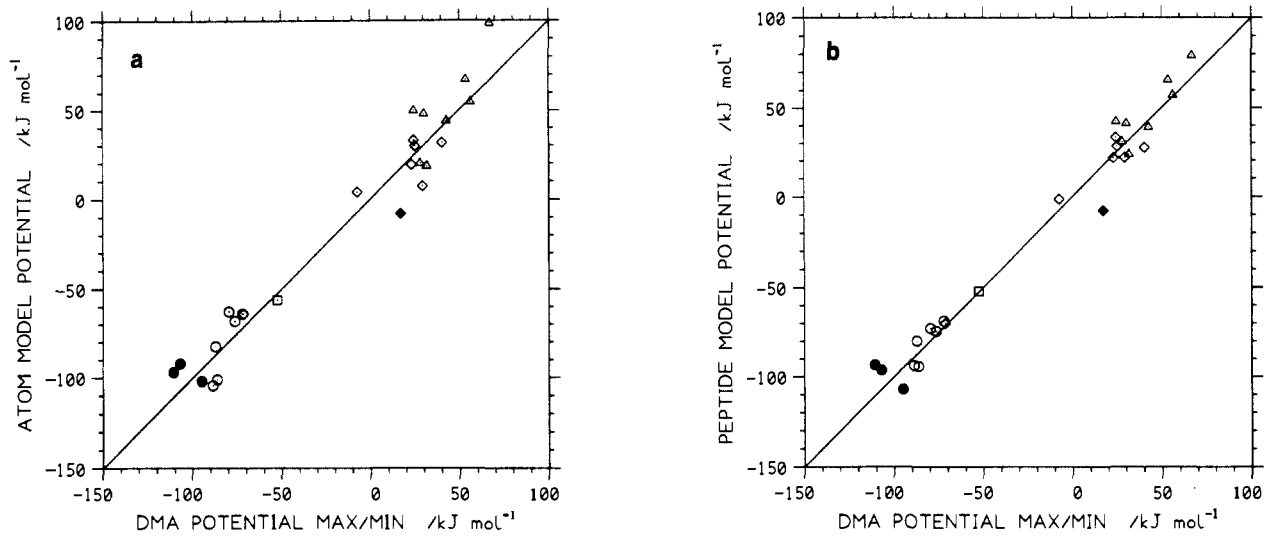

Figure 4. Plot of the electrostatic potential predicted by (a) the transferable ATOM model and (b) the transferable PEPTIDE model, against the potential evaluated from the DMA of the cyclosporin derivative wave function, at the maxima and minima of the DMA potential at 1 Å above the van der Waals surface. The type of the potential maxima and minima are denoted by the main contact: O, carbonyl oxygen; □, ether oxygen; △, *N*-methyl; ◇, other hydrocarbon groups. The shaded symbols are the minima near the grossly distorted functional groups mentioned in the text.

the multipole moments or the potential predicted by the various transferable models. However, it is possible to test whether the use of transferable multipole models, with the error involved in transferring them to a molecule with some very different bond lengths, can provide a realistic model of the electrostatic potential around this polypeptide. This is worthwhile as many molecular modeling studies have to be based on similar quality structural data.

The transferable ATOM model for this molecule is made by rotating the average atomic multipole moments for the appropriate atomic type from Table I into the correct orientation and position for the atoms in residues 2–11. Thus, the structure used in the ab initio calculation is also used in the electrostatic models, except that all CH_n units are represented as having a tetrahedral geometry from the use of the united-atom approximation. However, the model does not describe the changes in the charge distribution for each atom resulting from geometric distortions. Each residue is individually neutralized, from the standard deviations of all the atoms, although the distortions in some of the peptide linkages had resulted in significant charge flow between the residues. Since the unusual ether-linked side chain of the first residue contains atoms not included in our analysis, the multipole model for this residue is borrowed from the cyclosporin wave function DMA, with the net charge of $-0.061e$ being neutralized on the backbone atoms.

The transferable PEPTIDE model is constructed in a similar fashion, except that the appropriate peptide fragments from the dipeptide DMAs are used to provide the multipoles on each atomic site for residues 3–11 and the original cyclosporin DMA for residues 1 and 2. All residues are individually neutralized by adjusting the charges on the backbone atoms only. The DMA of the D-Ala residue for use in the PEPTIDE model is obtained by inverting the structure and the multipoles of the L-Ala dipeptide; the D-C_α multipoles for the ATOM model are derived by the symmetry transformation of the averge C_α multipoles in Table I. The cis 9–10 peptide linkage is approximated in both models by rotating the multipoles on N10 about the local z axis so that the bonded C_α is in the yz plane at $y > 0$. The average multipole moments of the nitrogen atoms in $CONH_2$ and $CON(CH_3)_2$ show approximate C_{2v} symmetry so the relative position of the C_α and CH_3 group is much more important than that of the carbonyl group in defining the asymmetry of the nitrogen charge distribution in the plane perpendicular to the amide group.

The values of the electrostatic potential predicted by the transferable models are compared in Figure 4 with the potential calculated from the DMA of the entire molecule, at the positions of its maxima and minima at a distance of 1 Å above the van der Waals surface. More precise details of these minima can be found in the original study.[33] These results are quantified in Table VI, which also includes the rms errors in the potential on various grids. The transferable PEPTIDE model gives a very reasonable prediction of the electrostatic potential around the undecapeptide, with rms errors of 7 and 3 kJ mol^{-1} for points 1–2 and 2–5 Å, respectively, from the van der Waals surface. It seems likely that the structural distortions make a significant contribution to these errors, as by far the largest error in Figure 4b is for the shallow maximum in contact with a methyl group that has a HCH bond angle of 87° in the crystal structure, whereas the methyl group is tetrahedral in the electrostatic models. The transferable ATOM model is less satisfactory, presumably because it does not represent the internal polarization effects within the hydrocarbon side chains as consistently as the PEPTIDE model, and these side chains dominate the van der Waals surface of the molecule.

It is noteworthy that both transferable multipolar models give a better representation of the potential around this molecule than the Mulliken charges, particularly close to the molecule. This is despite the distortions in the charge density from unphysical bond lengths and internal hydrogen bonds being implicitly included in the Mulliken charges and omitted from the transferable models,

Table VI. Comparison of the Electrostatic Potential around a Cyclosporin Derivative As Calculated from Various Electrostatic Models

set of points	max/min	grid[a]	grid[a]
dist from vdW surface (Å)	1	1–2	2–5
no. points	26	1290	5466
rms electrostatic potential V (kJ mol^{-1})	62	24	10
max V (kJ mol^{-1})	67	64	39
min V (kJ mol^{-1})	−110	−97	−55
Transferable ATOM Model			
rms error (kJ mol^{-1})	14	10	5
max error $\|\Delta V\|$ (kJ mol^{-1})	32	46	23
Transferable PEPTIDE Model			
rms error (kJ mol^{-1})	10	7	3
max error $\|\Delta V\|$ (kJ mol^{-1})	25	31	14
Mulliken Charges			
rms error (kJ mol^{-1})	30	13	7
max error $\|\Delta V\|$ (kJ mol^{-1})	93	67	38

[a] Points were derived from a 1-Å cubic grid of points around the molecule, omitting any points outside the given range from the van der Waals surface of the molecule. The van der Waals surface was defined by a radius of 2 Å for all C atoms, 1.4 Å for O, and 1.5 Å for N atoms.

although this is known to introduce serious errors. In particular, the Mulliken charges give a qualitatively misleading prediction of the relative energies of the different types of maxima and minima,[33] whereas both transferable multipolar electrostatic models give a useful model for even this grossly distorted structure.

This preliminary application is encouraging, as it shows that it is worthwhile to construct a transferable distributed multipole model to study the potential around a polypeptide, even when the input structure is poor and internal interpeptide polarization effects are large. A study of the atomic multipole moments of small molecules[36] (with a different method of splitting the charge distribution) found that the changes with physically reasonable structural variations were small. This, and the results for the alanine and isoleucine dipeptides, suggests that the transferable models should be even more accurate when applied to more realistic polypeptide geometries.

9. Conclusions and Discussion

Transferability. The concept of transferable functional group properties is central to chemistry and needs to be quantitative if molecular modeling calculations are to be used to understand biochemical processes. We have shown that it is possible to define a transferable atom, or transferable fragment, electrostatic model that can give reasonably good predictions of the electrostatic interactions in peptides. The condition for reasonably transferable atomic multipole moments is that the directly bonded functional groups must be the same. This is chemically reasonable, as it recognizes that the short-range inductive effects along the bonds to the fragment must be approximately the same in the two molecules for the fragment charge distributions to be virtually identical. However, it implies that larger molecules must be used as model systems for the atoms in polypeptides than has often been assumed in theoretical studies. In particular, a dipeptide is the smallest realistic model system for the peptide residue, as the backbone atoms in dipeptides, and, by inference, polypeptides, have a rather different charge distribution than that found in the monopeptides. Unfortunately, many of the commonly used force fields, such as AMBER[14,37] and that of Hagler, Huler, and Lifson,[38] assume that atomic charges can be transferred from monopeptides, such as formamide and *N*-methylacetamide. (This was also noted in a study which used a dipole-conserving charge analysis to examine the charge distribution in formamide, 10 dipeptides, and *N*-formylpentaglycine.[26])

The transferability of the atomic multipole moments is less accurate when the functional groups involved have weak inductive powers and are readily polarizable, for example, hydrocarbon fragments. Thus, it is better to build electrostatic models for polypeptides by transferring the multipoles of entire peptide residues from dipeptide calculations. This ensures that more of the polarization effects from nearby nonbonded functional groups are implicitly included in the model. However, in transferring large fragments between molecules, the prime requirement must be that the bonding environment of the transferred fragments is the same. It would be a very poor approximation to use a large fragment where the ends were tied off with a standard atom, instead of a smaller fragment in the appropriate bonding environment.

The accuracy of these transferable electrostatic models in some practical applications is assessed in part II. The study of different conformers of the Ala dipeptide is particularly revealing. It shows that the variations in the atomic multipoles with conformation are very similar to the variations in the corresponding atomic multipoles between different molecules with the same bonded functional groups. This suggests that the variations in the atomic multipoles are mainly due to through space polarization from farther neighbors. Assuming that atomic multipoles are transferable between conformations or between molecules can give comparable errors in the predicted electrostatic potential. This implies that we are close to the limits of accuracy of any electrostatic model that does not change with conformation.

The errors in the electrostatic potential around a peptide that arise from neglecting changes in intramolecular through-space polarization (with conformation or with molecule) can be less than an rms error of 5 kJ mol^{-1} in the region 1–2 Å from the van der Waals surface, provided that conformations involving severe polarization effects, such as hydrogen bonding, are avoided. (This estimate is derived from the alanine dipeptide study.) Thus, the electrostatic energy of interaction of a polypeptide with a neutral molecule, in a geometry that does not involve hydrogen bonding, can probably be predicted within reasonable "chemical accuracy" by these transferable models. However, the errors are considerably larger when hydrogen bonding is involved, and so some model for polarization effects will be required in a peptide intermolecular force field for it to be able to model molecular recognition processes with quantitative accuracy.

Absolute Accuracy. The same type of ab initio wave function was used throughout this work. Thus, in addition to the transferability errors being investigated, there are also errors present due to the use of a modest basis set and SCF wave functions. SCF calculations generally overestimate the polarity of the molecular charge distribution and, thus, overestimate the electrostatic energies. The modest basis set (3-21G) lacks polarization functions and so has limited flexibility to describe the anisotropy of the atomic charge distributions. This basis set is likely to underestimate the polarizability of the molecule, and thus, the effects of changes in the intramolecular polarization with conformation may be underestimated. However, the qualitative conclusions of this study are unlikely to be affected by improvements in the ab initio method, and although the numbers in the DMAs will change, the corresponding electrostatic potential is unlikely to change by orders of magnitude and thus alter the semiquantitative conclusions. Indeed, our results suggest that the choice of representation (distributed multipoles or point charges), and the assumptions about the transferability of atomic charge distributions between molecules and conformations, could be more important in determining the absolute accuracy of an ab initio based model for the charge distribution of a polypeptide, than the choice of basis set.

Distributed Multipole versus Atomic Point Charge Models. An important new feature of this study is the use of a distributed multipole model, rather than a point charge model, to represent the charge distribution of a peptide. Thus, we are using a theoretically well-founded model of the charge density that includes the electrostatic effects of nonspherical features, such as lone pairs and π electrons. The extra complexity involved in using an anisotropic site–site electrostatic model presents no real problems. The DMA representation is capable of predicting electrostatic properties to essentially the accuracy of the ab initio wave function, assuming that penetration effects are negligible or modeled separately. Although we have used a truncated expansion, and limited the number of sites, this introduces errors (Table III) that are negligible relative to the transferability errors that we are investigating. Thus, the errors we have found in the electrostatic potential arising from assuming transferability, either between conformations or molecules, should be close to the intrinsic errors arising from changes in the wave function and are, therefore, unlikely to be reduced significantly by the use of another type of electrostatic model.

All the results in part II show that the use of Mulliken net atomic charges to represent the molecular charge distribution produces much larger errors than any of our transferable models. The Mulliken charges were used as a bench mark because they are a standard model derived by direct analysis of the wave function, as is a DMA, and Mulliken charges are frequently used in molecular modeling,[39] particularly within commercial packages. However, the deficiencies of Mulliken charges are well-known,

(36) Liang, J.-Y.; Lipscomb, W. N. *J. Phys. Chem.* **1986**, *90*, 4246–4253.
(37) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230–252.
(38) Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319–5335.
(39) Sordo, J. A.; Probst, M.; Corongiu, G.; Chin, S.; Clementi, E. *J. Am. Chem. Soc.* **1987**, *109*, 1702–1708.

and many alternative methods of partitioning the charge density to give atomic point charges, such as a dipole-conserving analysis,[40] are better for calculating intermolecular electrostatic energies. An alternative approach to deriving point charge models is to fit the charges directly to the electrostatic potential outside the molecule. This requires the evaluation of the potential, by integration over the ab initio charge density, at a large grid of points, and thus is computationally expensive for large molecules. The published errors from such potential fitting studies[2-4,27] suggest that the optimum atomic point charge model will predict the electrostatic potential with errors that are not negligible in comparison with those resulting from the use of transferable models. In addition, there is no reason to suppose that the errors involved in assuming transferability between molecules, or conformations, will be smaller for potential-derived charges than for a more complete distributed multipole representation. (Indeed, the potential-derived charges for a few amides[27] suggest that they may be less transferable.) Thus, the way forward for the accurate modeling of electrostatic interactions is first to go to a distributed multipole model and, when it is necessary to assume that an electrostatic model is transferable between molecules, use the recipes derived in this study.

Future Developments. We have established a successful transferable electrostatic model for amides and peptides with hydrocarbon side chains. In order use this approach to model a wide range of polypeptides, we need atomic multipole moments for all the naturally occurring amino acids. Since these include charged side chains and aromatic rings, which would probably be even more susceptible to intramolecular polarization than hydrocarbon side chains, the transferable PEPTIDE model will be more appropriate than a transferable ATOM model. Fortunately, the required calculations on the dipeptides of all the protein residues can now be performed with the direct SCF method,[34] and this work is in progress. This will provide a model for examining the intermolecular electrostatic interactions of polypeptides at a new level of accuracy and reliability. However, analogous critical studies of the other terms in inter- and intramolecular force fields, and improvements in simulation methods, are needed for molecular modeling to achieve its full potential for providing insight into molecular processes.

Acknowledgment. We thank J. B. O. Mitchell and R. J. Wheatley for helpful discussions, and Dr. A. J. Stone and Dr. R. D. Amos for help with the use of their program suites ORIENT and CADPAC, respectively. We also thank the SERC for support for C.H.F. on a postdoctoral research assistantship and the Royal Society for the award of a 1983 University Research Fellowship to S.L.P.

Supplementary Material Available: Table of multipole moments of amides and dipeptides (8 pages). Ordering information is given on any current masthead page.

(40) Thole, B. T.; van Duijnen, P. Th. *Theor. Chim. Acta* **1983**, *63*, 209-221.

Local Structure in Ribonuclease A. Effect of Amino Acid Substitutions on the Preferential Formation of the Native Disulfide Loop in Synthetic Peptides Corresponding to Residues Cys^{58}-Cys^{72} of Bovine Pancreatic Ribonuclease A

K.-H. Altmann[†] and H. A. Scheraga*

Contribution from the Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301. *Received November 14, 1989*

Abstract: The possible existence of local structure in a 15-residue peptide fragment of bovine pancreatic ribonuclease A has been examined. Three peptides, corresponding to the amino acid sequence 58-72 of the native protein and two homologues thereof with amino acid substitutions at residues 66 ($Lys^{66} \rightarrow Gln$) and 67 ($Asn^{67} \rightarrow Ala$), respectively, have been synthesized by classical solution procedures. These peptides, all containing three Cys residues at positions 58, 65, and 72, were studied in disulfide-exchange equilibration experiments under strongly oxidizing conditions at 23 °C and pH 8.0. By use of an RP-HPLC method for the separation of the various species, it was demonstrated that the intramolecular equilibrium constnt K_{exp} for the two possible isomers of the native sequence, containing intramolecular eight-residue cyclic disulfides, was 3.58 ± 0.10 in favor of the disulfide bond between Cys^{65} and Cys^{72}, which is the disulfide bond present in the native protein. Amino acid substitutions at positions 66 ($Lys^{66} \rightarrow Gln$) or 67 ($Asn^{67} \rightarrow Ala$) did not result in marked changes in the standard free energy difference between the native and nonnative eight-residue disulfide loops, K_{exp} being 3.02 ± 0.17 for the Gln^{66} and 4.49 ± 0.23 for the Ala^{67} homologue. Possible reasons for this behavior are discussed.

It has been known for three decades that all the information required for a protein to attain its native three-dimensional structure (under appropriate solution conditions) is contained solely in its amino acid sequence.[1] However, the question as to *how* proteins fold, i.e., as to what are the detailed molecular mechanisms that lead from the unfolded polypeptide chain to a biologically active conformation, still awaits a final resolution. Regarding the early stages of the folding process, it has become increasingly clear that these are dominated by short-range interactions[2] resulting in the formation of locally ordered structures along the overall unfolded polypeptide chain.[3-12] These early-forming "chain-

[†] Present address: Institut de Chimie Organique de l'Université, Rue de la Barre 2, CH-1005 Lausanne, Switzerland.

(1) White, F. H., Jr.; Anfinsen, C. B. *Ann. N.Y. Acad. Sci.* **1959**, *81*, 515-523.
(2) Scheraga, H. A. *Pure Appl. Chem.* **1973**, *36*, 1-8.
(3) Wetlaufer, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 697-701.
(4) Anfinsen, C. B.; Scheraga, H. A. *Adv. Protein Chem.* **1975**, *29*, 205-300.
(5) Rose, G. D.; Winters, R. H.; Wetlaufer, D. B. *FEBS Lett.* **1976**, *63*, 10-16.

0002-7863/90/1512-4926$02.50/0