Theoret. Chim. Acta (Berl.) 62, 17–28 (1982)

THEORETICA CHIMICA ACTA

The Calculation of Molecular Electrostatic Potential from a Multipole Expansion Based on Localized Orbitals and Developed at Their Centroids: Accuracy and Applicability for Macromolecular Computations

Catherine Etchebest, Richard Lavery, and Alberte Pullman

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S. 13, rue Pierre et Marie Curie, F-75005 Paris, France

An improvement of a multipole expansion based on localized orbitals and termed LMTP is presented and its ability to generate accurate electrostatic potentials is demonstrated. The possibilities of using this expansion in studying the potential of different conformational states of a molecule without the necessity of recalculating its molecular wavefunction is described and the construction of macromolecular potentials by the superposition of the potentials of subunits is reconsidered.

Key words: Molecular electrostatic potential – Multipole expansion – Localized orbitals – Molecular conformation.

1. Introduction

The molecular electrostatic potential (MEP) has, in recent years, been used widely as a guide to the reactive properties of molecules [1–4]. It is particularly informative in the case of large biomolecules which are normally difficult to treat theoretically. Such molecules have been extensively studied with the aid of this property in our laboratory and very interesting results concerning the origin of their reactive behavior have been obtained, notably in the case of the nucleic acids [3–6]. These studies have naturally involved the development of new ways of treating such large systems. The technique, due to Pullman and coworkers [4, 7–8], that we have employed consists of the division of the macromolecule

0040-5744/82/0062/0017/$02.40

into smaller units whose *ab initio* SCF wavefunctions can easily be calculated. The electrostatic potential of these subunits is then obtained and the MEP of the macromolecule is calculated by the superposition of the potentials of the subunits, appropriately positionned in space.

A great deal of time can be saved in the calculation of the macromolecular electrostatic potentials by using a procedure, developed in our laboratory [9, 10], in which, whenever possible, the electron density distribution of a subunit is replaced by a multicentered multipole expansion. Our early computations used an expansion termed the OMTP or Overlap MulTiPole expansion consisting of a monopole, a dipole and a quadrupole centered at every atom of the molecule and at the mid-point of every atom pair. This expansion yields potentials in very good agreement with the exact potentials down to a distance of roughly 2 Å from the constituent atoms of the molecule [11]. Below this distance, penetration effects become important and the exact expression of the potential must be employed.

Despite the time saved by the use of the OMTP expansion, the calculation of the potential of very large macromolecules or over large areas is still rather expensive. For this reason we have attempted to develop a multipole expansion which would be cheaper to use, while still yielding accurate potentials. A preliminary presentation of such an expansion termed LMTP, or Localized MulTiPole expansion was given recently [12]. Its essential advantage lies in the fact that it is developed from localized molecular orbitals (LMO's), using the centroid of each such orbital as the point for the multipolar development for this MO. This yields an expansion with many fewer centers than the corresponding OMTP development and, in addition, the choice of the position of each center of multipoles as the centroid of the corresponding localized orbital eliminates the dipole terms. In our exploratory work [12] we have truncated the expansion at each center beyond the quadrupole term, as in the previous OMTP representation, and found that this procedure yielded sufficiently accurate potentials at long and intermediate distances. Nearer the molecule it appeared necessary, however, to introduce a hybrid technique which calculates the potential due to the closest localized orbitals exactly, while maintaining the multipole representation for the rest of the orbitals.

In the hope of avoiding this costly procedure we have looked into the possibility of improving the LMTP expansion by the addition of an octupole at each center of multipoles. We will show that while this only moderately increases the time necessary for the calculation of potentials, it allows accurate values to be obtained much closer to the molecule than before.

Faster computation of electrostatic potentials, is, moreover, not the only advantage associated with the LMTP expansion. Because of the representation of each localized orbital of the molecule separately, certain manipulations of the multipole expansion become feasible. The first such manipulation which will be described involves the representation of different conformations of a molecule and consequently of its potential, without the need for recalculating its molecular

wavefunction. Such a procedure is difficult to envisage with the OMTP expansion because of the lattice-work of multipole centers linking all the non-bonded atom pairs in the molecule. This implies that, if it is desired to, for example, rotate a functional group of the molecule, not only the bond center multipoles of the group must be moved, but also all the non-bonded centers linking this group to the other parts of the molecule.

The second manipulation involves the transfer of partial LMTP expansions from one molecule to another. This offers a possibility of simplifying the construction of larger molecule multipole expansions from smaller constituent molecules. This procedure is particularly easy within the LMTP formalism because of the integer monopoles which are associated with each multipole center and the improved transferability of localized MO's compared to the atomic overlap distributions of the OMTP expansion.

2. Method

Let us recall the principal methodological steps employed in developing the LMTP expansion of a molecule. Firstly, its ab initio SCF wavefunction is calculated using, in general, a standard atomic orbital basis of gaussian orbitals (3s for hydrogens, (7s, 3p) for second row atoms and (10s, 6p, 1d) for third row atoms contracted to minimal [13]). The resulting molecular orbitals are then localized using the technique of Boys [14], which maximizes the distances between the centroids of the orbitals. Next, a single center multipole expansion is developed for each occupied LMO, the position for the expansion being chosen as the centroid of the corresponding orbital. This choice, by the definition of the centroid, yields a null dipole term in the expansion. The monopole of the resulting expansion also is very simple, being just -2 for each doubly occupied MO.

In our preliminary treatment [12] the single center expansion was truncated after the calculation of a quadrupolar term which, for convenience, was reduced to two axial quadrupoles. We have now included the calculation of an octupole for each center.

The potential due to the octupoles was calculated using the expression given in reference [15]:

$$V_S^{\Omega}(P) = -\Omega : \nabla\nabla\nabla \frac{1}{|r_{SP}|} \tag{1}$$

where Ω is a tensor of the third order defined by:

$$\Omega = \int \rho \boldsymbol{R} \otimes \boldsymbol{R} \otimes \boldsymbol{R}\, d\tau \tag{2}$$

with ρ being the electronic density for the LMO considered and $\boldsymbol{R}$ the vector joining the LMO centroïd where the octupole is calculated and the elementary volume $d\tau$. r_{SP} is a vector joining the multipole center at point S to the point of potential calculation, P.

In a cartesian coordinate system, with unit vector e_1, e_2, e_3 along the x, y and z axes, the above expression becomes:

$$V_S^{\Omega}(P) = \sum_{i=1,3} \left[\frac{9\boldsymbol{r}\cdot\boldsymbol{e}_i}{r^5} - \frac{15(\boldsymbol{r}\cdot\boldsymbol{e}_i)(\boldsymbol{r}\cdot\boldsymbol{e}_i)(\boldsymbol{r}\cdot\boldsymbol{e}_i)}{r^7} \right] \Omega_{iii}$$
$$+ \sum_{i=1,3} \sum_{\substack{j=1,3 \\ j\neq i}} \left[\frac{9\boldsymbol{r}\cdot\boldsymbol{e}_i}{r^5} - \frac{45(\boldsymbol{r}\cdot\boldsymbol{e}_i)(\boldsymbol{r}\cdot\boldsymbol{e}_i)(\boldsymbol{r}\cdot\boldsymbol{e}_j)}{r^7} \right] \Omega_{iij} \tag{3}$$
$$- \frac{90(\boldsymbol{r}\cdot\boldsymbol{e}_i)(\boldsymbol{r}\cdot\boldsymbol{e}_j)(\boldsymbol{r}\cdot\boldsymbol{e}_k)}{r^7} \Omega_{ijk}$$

where Ω_{iii} is a tensor element of the type Ω_{xxx}, Ω_{iij} is an element of the type Ω_{xxy} and Ω_{ijk} is the element Ω_{xyz}.

A simplification can be introduced for the core molecular orbitals of all atoms other than hydrogen, since the quadrupole and higher terms of the expansions for these, almost spherical, orbitals are extremely small. Thus we reduce them to solely the monopole of -2. Moreover, because the centroids of these orbitals are almost coincident with the atomic nuclei, these monopoles may simply be subtracted from the corresponding nuclear monopoles.

In our previous study [12] with the expansion limited to the quadrupoles we have observed that for planar molecules better results were obtained when the localization of the molecular orbitals of sigma and pi symmetry was performed separately. After the introduction of the octupole terms it was found that this procedure did not improve the quality of the resulting potentials, as was previously the case, and it has consequently been dropped. This has the advantage that all molecules may be localized in the same way.

Tests of the new LMTP expansion are presented for subunits of B-DNA, and for a simple dipeptide model.

3. Results and Discussion

(a) *The Addition of Octupole Terms*

In order to test the quality of the improved LMTP expansion we have chosen to study the electrostatic potentials of the B-DNA subunits, that is, a phosphate, a sugar and the bases guanine, adenine, cytosine and thymine. These molecules, being relatively large and having, in the case of the bases, conjugated pi systems, represent a much more demanding test of a multipolar expansion than the small, saturated molecules on which tests are often performed. The geometries are those of Arnott and Hukins [16].

We illustrate the results for two bases of the nucleic acids, cytosine and adenine and also for the phosphate group, in Figs. 1, 2 and 3 respectively. In each case, the left hand side (denoted "a") is the LMTP isopotential map for the subunit, the potentials being calculated in the plane of the bases and in the plane containing

the phosphorus atom and the anionic oxygens for the phosphate. The right hand side (denoted "b") is a difference map showing the differences calculated between the LMTP potential and the exact potential for each subunit. It may be seen that the errors in the LMTP potentials are small and are moreover very localized in small zones close to the molecule. Comparison with the results of reference [12] shows that the addition of an octupole to each center in the LMTP expansion greatly improves the quality of the resulting electrostatic potentials. It was found that for all the molecules studied, the multipolar potentials could be calculated

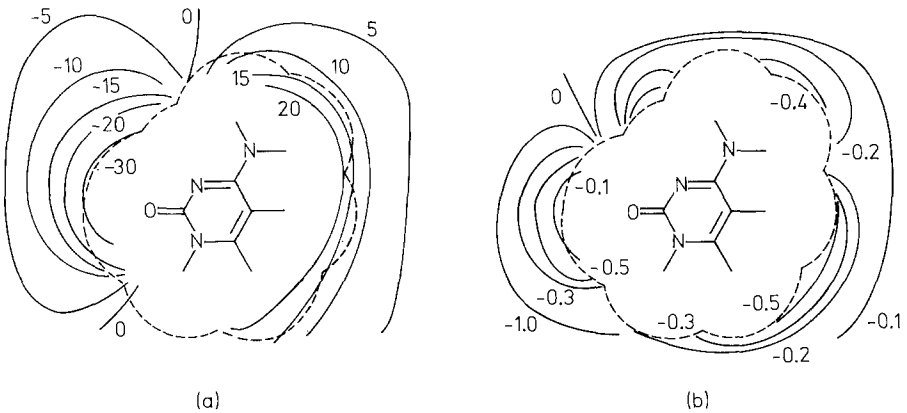

Fig. 1a, b. Potentials for cytosine in the plane of the base (kcal/mole) (the dotted line in this figure and in Figs. 2 and 3 indicates the 2.5 Å approach limit from the multipole centers). **a.** LMTP potentials. **b.** Potential difference between the LMTP and the exact potentials. (LMTP-exact)

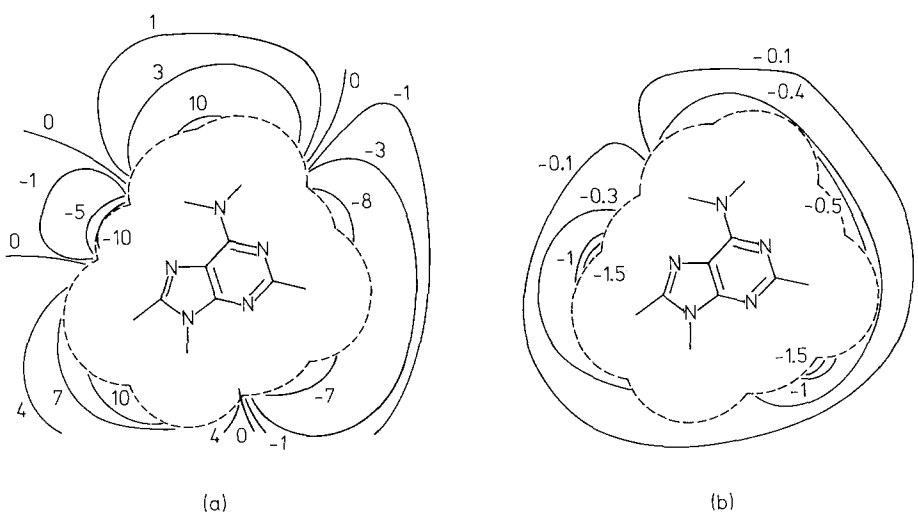

Fig. 2a, b. Potentials for adenine in the plane of the base (kcal/mole). **a.** LMTP potentials. **b.** Potential difference between the LMTP and the exact potential

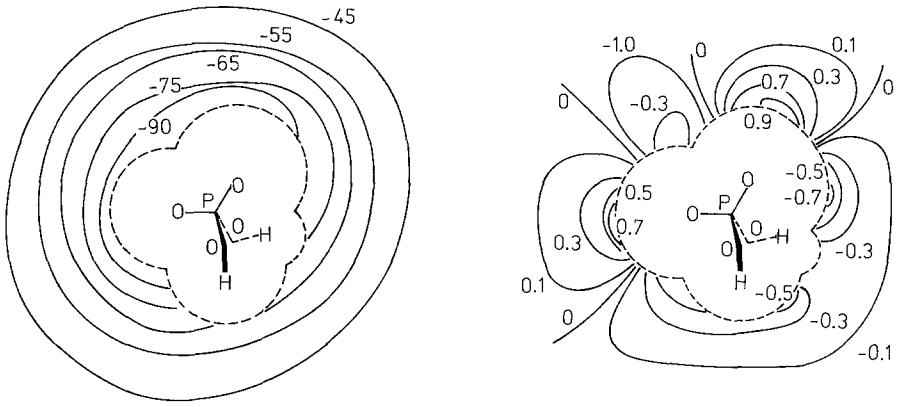

Fig. 3a, b. Potentials for phosphate in a plane containing the phosphorus atom and the anionic oxygens (kcal/mole). **a.** LMTP potentials. **b.** Potential difference between the LMTP and the exact potentials

to within 2.5 Å of any center of the associated LMTP expansion without incurring errors of greater than 2 kcal/mole and that, generally, the errors were less than 1 kcal/mole. This approach limit, which is illustrated in Figs. 1–3 by the dotted lines, is very much closer to the molecule than the limit of 4 Å below which it was previously necessary to use the hybrid LMTPE technique and is only slightly greater than the 2 Å limit (measured from the atoms of the molecule) employed with the OMTP expansion [14].

Table 1. Comparison of the characteristics of the different multipole expansions. (results refer to potential calculations in a plane of 256 points for the base cytosine)

	Expansion			
Property	OMTP	LMTP without octupoles	LMTP with octupoles	Point charges (using Mulliken atomic charges)
Number of multipole centers	91	34	34	13
Computation time[a]	1.0	0.19	0.28	0.12
Multipoles: Monopoles	91	34	34	13
Multipoles: Dipoles	91	—	—	—
Multipoles: Quadrupoles	91	21	21	—
Multipoles: Octupoles	—	—	21	—
ΔV_{max}[b] (kcal/mole)	0	3.9	1.5	6.0
Points with errors[b] between 1 kcal/mole and ΔV_{max}	0	37	9	242

[a] expressed as a fraction of the time for the calculation with the OMTP expansion.
[b] error in calculated electrostatic potentials down to the 2.5 Å approach limit.

In Table 1 we make a comparison of the number of centers and multipole terms in the LMTP expansion (before and after the addition of the occtupoles) and OMTP expansions and of the relative times of calculation for the plane of potential of cytosine given in Fig. 1a. (The times are given as fractions of the time necessary for the potential calculation with the OMTP expansion). We also give the maximal error ΔV^{max} in potential for each type of expansion and the number of points of potential calculated for which the error was between ΔV^{max} and 1 kcal/mole. This latter value enables the spatial extent of significant errors in potential to be compared. It appears from this table that the LMTP expansion, even in its extended form, represents a considerable time saving. This naturally becomes relatively more and more important as the size of the molecule studied increases.

We have also included, in the last column of Table 1, a comparison with a very simple expansion used rather frequently in studying large molecules, consisting of just the Mulliken partial charges situated on each atom of the molecule. Although, as the table shows, this expansion is, in computational terms, very economical, it is seen that approaching the molecule to the same limit as we have adopted for the LMTP expansion, namely 2.5 Å from the multipole centers, results in errors in the calculated potentials of up to 6 kcal/mole. It is therefore clear that this expansion must be used with caution if other than qualitative studies of potential are to be made. Table 1 shows moreover that the errors with this representation remain relatively more important at long distance.

(b) *Changes in Molecular Conformation*

We now turn to the use of the LMTP expansion in studying the electrostatic potential of different conformations of a molecule. This possibility was tested on a simple dipeptide model which is illustrated in Fig. 4. The geometry utilized is that of ref. [17]. Three possible pairs of values for the torsional angles ϕ and ψ about the bonds N—C_α and C_α—C′ were considered. These values, given in Table 2, correspond to the well-known polypeptide conformations, the α-helix,

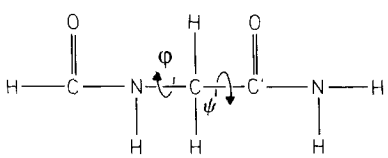

Fig. 4 Dipeptide model employed in testing the LMTP expansion

Table 2. Torsional angles defining the conformations of the dipeptide model studied. (angles defined using the convention of ref. [17])

Conformation	ϕ (degrees)	ψ (degrees)
α-helix	132	123
2_7-ribbon	105	250
β-pleated sheet	40	315

the 2_7-ribbon and the β-pleated sheet [17]. Wavefunctions were calculated for each of the conformations of the model and subsequently the LMTP expansions were developed. As the values of the torsion angles for the 2_7-ribbon conformation are intermediate between those of the other two conformations, it was chosen as the reference conformation from which we attempt to construct the two other forms.

In order to perform this step, a computer program was written which could modify the LMTP expansion of a molecule so as to follow changes in molecular conformation. These changes, which are restricted to changing torsion angles around single bonds, result in the displacement of certain multipole centers and of appropriate rotations of the associated quadrupoles and octupoles. This program was used to construct the α-helix and the β-pleated sheet conformations from the intermediate 2_7-ribbon conformation and the potentials of the resulting "constructed" conformations were compared with the potentials calculated from their own LMTP expansions.

The success of this procedure may be judged from the results in Figs. 5a and 5b which contain respectively the LMTP isopotential map for the $N—C_\alpha—C'$ plane of the true α-helical conformation and a map of the difference in potential between these results and those for the "constructed" dipeptide. The errors due to using the modified LMTP expansion can been seen to be very small, the largest difference with the true potentials down to the 2.5 Å approach limit being 2.7 kcal/mole. This is particularly encouraging as the change in the torsion angles between the 2_7-ribbon and the α-helix is quite large and, moreover, the two carbonyl oxygens of the dipeptide model approach one another much more closely in the latter conformation (3.1 Å) than in the former (4.4 Å). This might be expected to cause a considerable change in the polarization of the dipeptide electron distribution from one conformation to the other and thus invalidate the

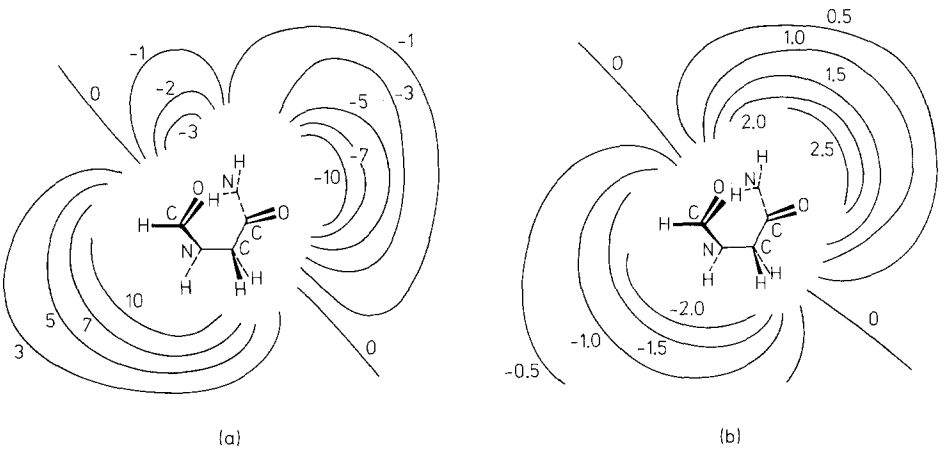

Fig. 5a, b. Potentials for the dipeptide model in the α-helix conformation. Calculated in the plane of the $N—C_\alpha—C'$ atoms (kcal/mole). **a.** LMTP potentials. **b.** Potential difference between the above results and those of the molecule obtained by rotation from the 2_7-ribbon conformation

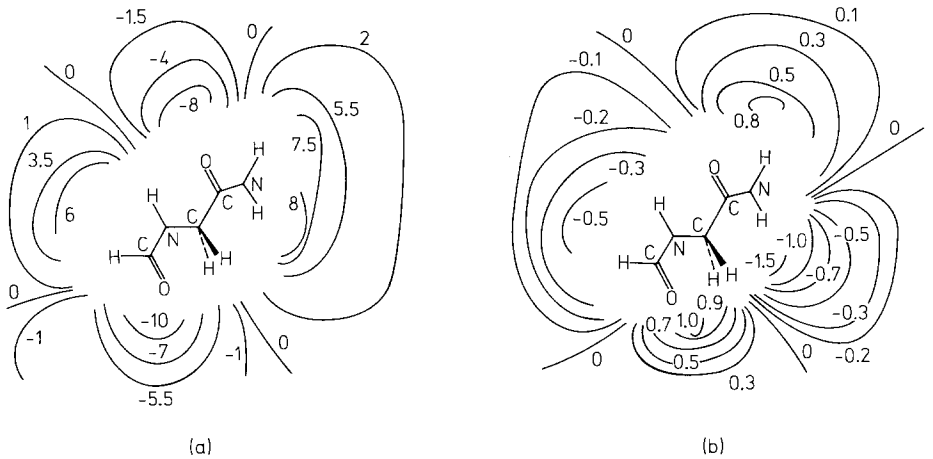

Fig. 6a, b. Potentials for the dipeptide model in the β-pleated sheet conformation. Calculated in the plane of the N—C_α—C' atoms (kcal/mole). **a.** LMTP potentials. **b.** Potential difference between the above results and those of the molecule obtained by rotation from the 2_7-ribbon conformation

simple "reconstruction" we have attempted, which of course neglects such polarization. The results for the construction of the less modified β-pleated sheet conformation, presented in Figs. 6a and 6b, are better, the maximal error in the potentials resulting from the rotated LMTP expansion being only 1.7 kcal/mole down to the prescribed approach limit of 2.5 Å.

Let us further point out that the errors observed after the creation of a new conformation were found to be essentially due to small discrepancies in the positions of the centroids of the constructed LMTP expansion compared to those in the true expansion. In this connection it is interesting to note that at, for example, the approach limit of 2.5 Å from a monopole of −2, a displacement of only 0.01 Å results in a change of potential of more than 1 kcal/mole.

It thus seems that this technique can be very useful in studying the effects of limited conformational changes on electrostatic potential or on electrostatic interaction energies calculated from multipole expansions, particularly when the molecules involved are large. It enables such changes to be investigated without the need for recomputing a new and expensive ab initio SCF wavefunction for every conformation studied. Since conformational change is very important for the understanding of many macromolecular properties, we hope that this procedure will find frequent applications. It needs to be stressed however that one must not misuse it by constructing conformations too far removed from the reference conformation or, alternatively, conformations where atomic contacts risk causing important changes in the associated molecular electronic distribution.

(c) *Superposition of Multipole Expansions*

The final section of this publication concerns an aspect of the calculation of the MEP of macromolecules by the superposition of the potentials of smaller

subunits. This technique [4, 7–8], which our laboratory has widely employed in studying biological macromolecules, involves cutting a number of bonds in the macromolecule to form the subunits. The free valencies caused by these cuts are absorbed by the addition of hydrogen atoms, which we will term "fictive hydrogens". After calculating the wavefunctions of the resulting small molecules and developing multipole expansions of their associated electronic distributions, the macromolecular potentials are obtained by superposing the potentials of the subunits appropriately oriented in space. This superposition clearly involves the overlap of the bonds carrying the fictive hydrogen atoms of the subunits and might be expected to lead to perturbed macromolecular potential in these regions. Tests performed on the nucleic acids and on other large molecules constructed from subunits [7, 18] have shown that there are indeed perturbations, but that these are generally small and confined to the vicinity of the junctions.

With the development of the LMTP expansion a new way of considering this problem presents itself. Because the centers of these expansions represent doubly occupied and localized molecular orbitals and thus both have an integer monopole (equal to -2) and correspond to an electronic distribution covering a restricted volume of space, they can easily be separated and reassembled to form new expansions. Thus, it is possible, once the LMTP expansions of the subunits have been developed, to remove the centers representing the "fictive" hydrogen nuclei and the bonds to these atoms and to replace them with centers representing the bond existing in reality between the subunits in the macromolecule. The multipole center representing this bond may be obtained from an LMTP expansion developed for an appropriate model compound containing a similar bond.

This possibility was tested for the dipeptide molecule (in its α-helix conformation) of the preceeding section by cutting it into two parts at the $C_\alpha - C'$ bond. Although this subdivision is unnecessary for such a small molecule, it will serve as an illustration of the technique. Thus, after subdivision, saturation of the subunit free valencies by fictive hydrogens and development of the LMTP expansions from the electron distributions of the subunit SCF wavefunctions, the centers representing the fictive hydrogen nuclei and their bonds were replaced by a center representing the $C_\alpha - C'$ bond drawn, in this case, from the LMTP expansion for the complete dipeptide. The dipeptide potentials were then calculated by the superposition of potentials of the two subunits and of the linking bond, each appropriately oriented to reproduce the α-helix conformation.

The potential difference map which compares the potentials of the true dipeptide with those for the constructed molecule, in the plane passing through the $N - C_\alpha - C'$ atoms, is given in Fig. 7a. It can be noted that the differences are always less than 1 kcal/mole and that, in particular, the potentials closest to the replaced bond are very accurate.

In fact, in this case the replacement of the "fictive" hydrogens by the linking bond is really unnecessary. As Fig. 7b shows, almost the same error margin is

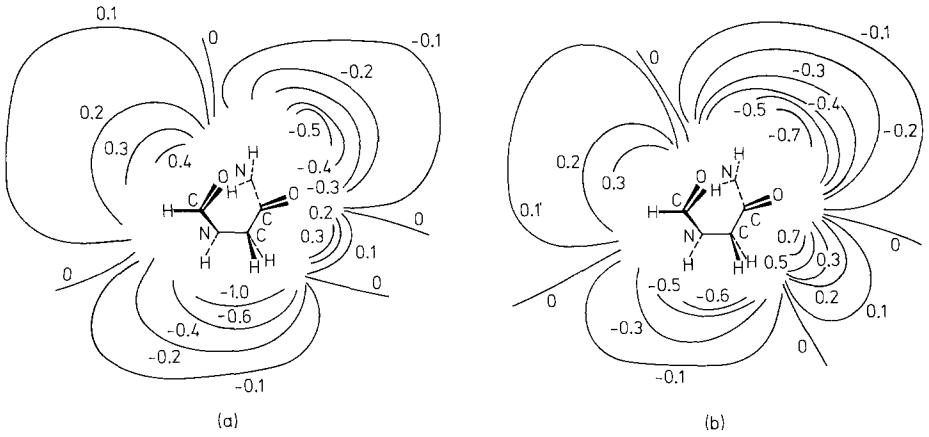

Fig. 7a, b. Potential difference maps for the dipeptide model in the α-helix conformation. Calculated in the plane of the N—C_α—C' atoms (kcal/mole). **a.** Difference between the true α-helix LMTP potentials (Fig. 5a) and those obtained by constructing the molecule from two subunits. **b.** Difference between the true α-helix LMTP potentials and those obtained by constructing the molecule from two subunits, but without replacing the overlapping fictive hydrogens

obtained by superposing the potentials of the two unmodified subunits, although the error close to the superposed bond is slightly more important.

Consequently, it seems, at least in some cases, that the "fictive" hydrogens may often have only a very minor perturbing influence on the macromolecular potentials obtained by superposition and that the proposed substitution scheme with linking bond centers will be unnecessary. It should be understood however that the position of subdivision must be chosen with care to avoid disturbing too greatly the electronic distributions of the subunits and that this should be verified by appropriate tests before the macromolecular investigations are undertaken. In the event of significant error, the use of the LMTP expansion provides an easy and elegant way to correct them.

4. Conclusions

We have presented an extension of a localized multipolar expansion, termed LMTP, which enables, by the introduction of octupole terms, accurate electrostatic potentials to be calculated down to 2.5 Å from the constituent centers of the expansion. The small number of these centers enables the potential calculations to be made very economically.

It has also been shown that because each center of the LMTP expansion represents an individual localized molecular orbital, these centers may be manipulated in two ways:

(1) they may be translated and rotated to follow certain conformational changes in the molecule studied and thus enable the effects of these changes on the

electrostatic potential to be investigated without the need for expensive repetition of the SCF wavefunction calculation.

(2) they may be separated and interchanged between different expansions, and in this way a more elegant superposition of small subunits molecules to form macromolecules may be achieved.

Work is now underway to exploit fully the possibilities of studying conformational changes offered by the LMTP expansion, notably through investigations of the mode of action of ion transporters. The economy of this formalism is also being used to facilite calculations of the energy of interaction between small reactive species and the nucleic acids which would otherwise have been computationally prohibitive. The results of these studies will be published shortly.

References

1. Scrocco, E., Tomasi, J.: Top. Curr. Chem. **42**, 95 (1977)
2. Scrocco, E., Tomasi, J.: Int. J. Quant. Chem. **11**, 115 (1978)
3. Pullman, A., Pullman, B.: in Chemical applications of atomic and molecular electrostatic potentials. ed. Politzer, P., Truhlar, D. G.: Plenum (1981)
4. Pullman, A., Pullman, B.: Q. Rev. Biophys. **14**, 289 (1981)
5. Pullman, B., Perahia, D., Cauchy, D.: Nucl. Acid. Res. **6**, 3821 (1979)
6. Pullman, B., Lavery, R., Pullman, A.: Eur. J. Biochem. **124**, 229 (1982)
7. Pullman, A., Berthod, H.: Theoret. Chim. Acta (Berl.) **48**, 269 (1978)
8. Pullman, A., Zakrzewska, K., Perahia, D.: Int. J. Quant. Chem. **16**, 395 (1979)
9. Port. G. N. J., Pullman, A.: FEBS Lett. **31**, 70 (1973)
10. Pullman, A., Perahia, D.: Theoret. Chim. Acta (Berl.) **48**, 29 (1978)
11. Goldblum, A., Perahia, D., Pullman, A.: Int. J. Quant. Chem. **15**, 121 (1979)
12. Lavery, R., Etchebest, C., Pullman, A.: Chem. Phys. Letters **85**, 266 (1982)
13. Pullman, B., Gresh, N., Berthod, H., Pullman, A.: Theoret. Chim. Acta (Berl.) **44**, 151 (1977)
14. Boys, S. F.: Rev. Mod. Phys. **32**, 306 (1960)
15. Bottcher, C. J. F.: Theory of electric polarization vol. 1, Elsevier (1973)
16. Arnott, S., Hukins, D. W. L.: Biochem. Biophys. Res. Comm. **47**, 1504 (1972)
17. Dickerson, R. E.,, Geis, I.: The structure and action of proteins, Benjamin (1969)
18. Pullman, A. in New horizons in biological chemistry, Symposium in honour of Prof. K. Yagi, Japan Scientific Societies Press (1981)

Received May 3, 1982