Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins

PIOTR CIEPLAK,[†] WENDY D. CORNELL,[‡] CHRISTOPHER BAYLY,[#] and PETER A. KOLLMAN*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received 11 August 1994; accepted 21 December 1994

ABSTRACT

We present the derivation of charges of ribo- and deoxynucleosides, nucleotides, and peptide fragments using electrostatic potentials obtained from *ab initio* calculations with the 6-31G* basis set. For the nucleic acid fragments, we used electrostatic potentials of the four deoxyribonucleosides (A, G, C, T) and four ribonucleosides (A, G, C, U) and dimethylphosphate. The charges for the deoxyribose nucleosides and nucleotides are derived using multiple-molecule fitting and restrained electrostatic potential (RESP) fits,[1,2] with Lagrangian multipliers ensuring a net charge of 0 or ± 1. We suggest that the preferred approach for deriving charges for nucleosides and nucleotides involves allowing only C1′ and H1′ of the sugar to vary as the nucleic acid base, with the remainder of sugar and backbone atoms forced to be equivalent. For peptide fragments, we have combined multiple conformation fitting, previously employed by Williams[3] and Reynolds et al.,[4] with the RESP approach[1,2] to derive charges for blocked dipeptides appropriate for each of the 20 naturally occuring amino acids. Based on our results for propyl amine,[1,2] we suggest that two conformations for each peptide suffice to give charges that represent well the conformationally dependent electrostatic properties of molecules, provided that these two conformations contain different values of the dihedral angles that terminate in heteroatoms or hydrogens attached to heteroatoms. In these blocked dipeptide models, it is useful to require equivalent N—H and C=O charges for all amino acids with a given net charge (except proline), and this is

[†]Current address: Dept. Chem., U. of Warsaw, Pasteura 1, 02-093 Warsaw, Poland

[‡]Current address: EMBO Labs, Heidelberg, Germany

[#]Current address: Merck Frost Inc., C. P. 1005 Point Claire, Dorval Quebec, H9R 4P8 Canada

*Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 16, No. 11, 1357–1377 (1995)

CCC 0192-8651 / 95 / 111357-21

accomplished in a straightforward fashion with multiple-molecule fitting. Finally, the application of multiple Lagrangian constraints allows for the derivation of monomeric residues with the appropriate net charge from a chemically blocked version of the residue. The multiple Lagrange constraints also enable charges from two or more molecules to be spliced together in a well-defined fashion. Thus, the combined use of multiple molecules, multiple conformations, multiple Lagrangian constraints, and RESP fitting is shown to be a powerful approach to deriving electrostatic charges for biopolymers.

Introduction

There are two desirable properties of atomic charge models to be used in molecular mechanical studies of complex molecules. These are transferability and accuracy. There are a variety of ways to achieve these properties, but two different approaches are highlighted here.

The first approach is to derive charges empirically; the most elaborated application of this to peptides and proteins is the OPLS model.[5] By carrying out Monte Carlo calculations on representative liquids, partial charges on atom types can be derived which optimize the agreement between calculation and experiment. Transferability is assumed, which, based on Monte Carlo calculations on a number of related liquids, is often a reasonable assumption. The main disadvantages of this approach are the requirements of Monte Carlo simulations on requisite liquids, the fact that such methods cannot be easily extended to excited states,[6] the difficulty of assessing when the charge "transferability" breaks down, and the subjective judgments that must be made in that regard in charge derivation.

The other main approach to deriving partial charges is based on the use of quantum mechanical calculations. The actual use of intermolecular interactions in this derivation[7] is impractical in general for deriving charges, but the molecular charge distribution has been useful in this regard. It is also clear that the use of the quantum mechanical electrostatic potential or field is an essential element in the derivation of charges that accurately represent the molecular multipole moments.[8-11] Thus, the partial charge models most often involve a least-squares fit between the model and the quantum mechanical potential. This method has the advantage that, with current computer power, charges can be derived for many molecules of significant size in a reasonable amount of time. The charges derived can be dependent on the *ab initio* basis set or semiempirical methodology, but a reasonable model of choice is the use of *ab initio*–derived charges using a 6-31G* basis set,[4] which uniformly overestimates molecular polarity. This overestimate makes such models relatively well balanced with empirical solvent models such as TIP3P[11] or SPC[12] water, which include polarization effects implicitly because they have been empirically calibrated to reproduce the density and enthalpy of vaporization of the liquid.

These electrostatic potential derived (ESP) charges have suffered from two main disadvantages. First, they have not been transferable, with chemically similar atoms often having variable charges. Second, derivation of charges for large polymers becomes impractical, even with powerful computers. Although there are real dependencies of partial charges on molecular conformation, these cannot be easily handled within the current framework of two-body additive molecular mechanical potentials of biopolymers. What have been more problematic are spurious fluctuations of the charges obtained in the least-squares fitting procedure for the statistically poorly determined (buried) centers. A way out of the first problem has been offered by the recent development and implementation of multiple-conformation fitting[1,2,4] and RESP charges.[1,2] The use of these two techniques offers a significant improvement in the quality and applicability of electrostatically determined charges.

How then should one derive charges for biopolymers? That is the focus of this article. Weiner et al. derived electrostatic potential–based charges for monomers of proteins and nucleic acids and then pieced these together, adjusting charges on junction atoms to ensure unit charges.[9,10] This is a reasonable approach given that one chooses appropriate atoms (i.e., nonpolar and nonconjugated) for these adjustments. However, it suffers

from being unaesthetic, nonalgorithmic, and not easily generalizable.

Recently, Bayly et al. have developed new software to allow simultaneously multiple conformations, multiple molecules, restraints, and Lagrangian constraints in the derivation of the charge model.[1] This allows one simultaneously to fit the charges of different nucleosides and dimethylphosphate to derive, in a clean, algorithmic way, the charges of all the fragments needed for simulations of DNA and RNA. This also allows the derivation of a set of charges for all of the naturally occurring amino acid dipeptides and, thus, all the charges necessary for the simulations of proteins. Although there are still some subjective decisions that need to be made in applying these algorithms, they can be clearly defined at the beginning and consistently followed throughout. Thus, we feel that this work offers a new, more powerful, and general approach for the derivation of charges for organic molecules and biopolymers.

Methodology

GENERAL REMARKS

We begin the derivation of the charges by calculating electrostatic potentials (ESP) at a grid of points[8] around appropriate components of the nucleic acids and proteins. The Hartree-Fock method with the 6-31G* (ref. 13) basis set and the Gaussian 90 program[14] were used. In the present case, we performed calculations for dimethylphosphate (DMP) in its gauche–gauche (g^+, g^+) conformation geometry optimized at the 6-31G* level, as well as A and B standard forms[15] of deoxyribonucleosides, standard A forms[15] of ribonucleosides, and all amino acids with appropriate $CH_3—CO—$ and $—NH—CH_3$ blocking groups (dipeptides) optimized using molecular mechanics[16] with the Weiner et al.[9,10] force field. For each amino acid, we calculated ESPs for two sidechain conformers (or four, in the case of proline).

Faster computers and the direct Hartree-Fock approach enabled calculations on larger systems than previously considered (e.g., nucleosides and dipeptides of each of the naturally occuring amino acids). This reduced the number of components for which *ab initio* calculations needed to be done; also, using larger fragments, we decrease the possibility of force field inaccuracies arising from building larger residues from smaller ones. The electrostatic potentials were subsequently used in our RESP fitting procedure.[1,2]

Table I contains important geometrical parameters characterizing the nucleic acid fragments: sugar puckering parameters q and W,[17] and angles γ and χ. In all cases, the HO5′—O5′ and HO3′—O3′ bonds were kept in the trans position (dihedral angle 180°) with respect to the heavy atoms to which they are linked. The molecules were geometry optimized using the molecular simulation program AMBER[16] and its force field.[10]

The RESP charges for the amino acids were fit using two different conformations for each amino acid (or four, in the case of proline). With the exception of the proline residue, each amino acid was represented in both its alpha-helical and its extended (beta-sheet) forms. Sidechain χ values were chosen based on the Protein Data Bank[18] analysis of McGregor et al.,[19] which correlates backbone and sidechain conformations for each of the amino acids. The acetyl and *N*-methyl blocked amino acids were minimized with AMBER[16] using the Weiner et al. all-atom force field.[9,10] Each molecule was first minimized with backbone and sidechain constraints. The constraints were then removed (except for the alpha-helix backbone, where constraints of $\phi = -60°$ and $\psi = -40°$ were used throughout) and the molecule was allowed to minimize freely. Molecules which did not remain in the desired local minimum were either reminimized with an intermediate step employing a smaller subset of the original constraints followed by a final free minimization; or, in the case that that strategy also failed, the final minimization employed constraints on the necessary dihedrals. All peptide bonds were in the trans conformations, with the exception of proline, as described later.

We assigned χ values for each amino acid as follows. First, using the data from Table 1 in McGregor et al.[19] for residues in the center of alpha-helices and beta-sheets, we assigned a different χ_1 (t, g^+, g^-) for the alpha-helix and the beta-sheet conformations for a given residue. We follow the convention used in McGregor et al.[19] that g^+ corresponds to 300°. The χ_1's were chosen to maximize the total number of occurences of these backbone–χ_1 combinations, where the χ_1 for the alpha-helix differs from the χ_1 for the beta-sheet. Specifically, one calculates the percentage given for that χ_1 within either the alpha-helix (center) or beta-sheet (center) category, multiplied by the total number of occurences of residues in that secondary structure category.

The χ_2 value was then assigned according to the data presented in Tables 3 and 4 of McGregor et al.[19] Once again, we chose the most common χ_2

for each backbone–χ_1 pair as long as that yielded a different χ_2 for each conformation. When the same χ_2 was preferred by each of the two backbone–χ_1 combinations, different χ_2's were assigned to maximize the total number of occurences of the two backbone–χ_1–χ_2 combinations.

Sidechain hydrogens attached to oxygen or sulfur (Thr, Ser, Cys, Tyr) were placed according to their minimum energy-minimized conformation of $t/g^+/g^-$ or syn/anti. When this preference was the same for the two conformations of an amino acid, then the hydrogen was placed uniquely on each conformation to yield the lowest overall energy for the two conformations added together.

There were five exceptions to the preceding rules. First, minimized conformations which had hydrogen bonds between the sidechain and backbone atoms were eliminated. This was accomplished either by minimizing with constraints on some of the dihedrals or, when that was not sufficient to eliminate the hydrogen bond, by choosing an alternative conformation.

Second, in the case of cysteine, the second most common pair of backbone–χ_1 conformations was used since the extended conformation did not stay in its local minimum when a χ_1 of $-60°$ was used. The second pair occured 42% of the time as compared to 46% of the time for the most common pair, so they were considered to occur with nearly equal frequency.

Third, McGregor et al.[19] tabulated data for the most common overall conformations of methionine, arginine, and lysine. For these three molecules, we chose sidechain conformations for the alpha-helix and beta-sheet backbones that (1) were among the most commonly observed, (2) had different χ_1 values, and (3) had different χ_n values, where χ_n had a heteroatom in the first or fourth positions. The rationale for this choice came from our work on propylamine,[2] where it was shown to be bene-

TABLE I.
The Most Important Geometrical Parameters Characterizing the Molecules Considered in *Ab Initio* Calculations.

A Deoxynucleosides	q (Å) (Ref. 17)	W(°) (Ref. 17)	γ(°)	χ(°)	
ADE	0.39	17.8	60.0	203.0	
CYT	0.38	25.0	60.1	206.1	
GUA	0.39	18.8	59.6	203.9	
THY	0.38	25.9	60.0	209.0	
B Deoxynucleosides					
ADE	0.38	151.9	58.5	210.0	
CYT	0.38	149.2	58.9	209.7	
GUA	0.38	151.4	58.5	209.9	
THY	0.38	149.1	58.4	215.7	
A Ribonucleosides					Angle C3′—C2′—O2′—H2′
ADE	0.40	12.4	59.0	191.2	−42.0
CYT	0.40	18.1	58.9	195.6	−42.0
GUA	0.40	14.3	58.7	197.6	−40.0
URA	0.40	19.0	60.3	193.7	−40.0
DMP	Angle C1—O1—P—O2	Angle O1—P—O2—C2			
	73.1	73.1			

q, W are the amplitude and phase of sugar ring puckering according to the Cramer and Pople definition.[17] The nucleosides were optimized using the molecular mechanical AMBER[16] program and its force field,[9,10] whereas dimethylphosphate was fully optimized using the Gaussian 90 program.[14]

ficial to allow alternative conformations around the central N—C—C—C bond in order to derive the most robust set of charges for this molecule.

Fourth, in proline the peptide bond is found to be in the cis rather than the trans conformation approximately 20% of the time.[20] For this reason, four different conformations were used for the proline residue, representing both the cis and trans peptide bonds as well as two different backbone conformations. The backbone conformations were assigned based on data in the PDB survey by MacArthur and Thornton.[21] They found that phi–psi plots of both trans and cis proline exhibited two distinct minima, corresponding to conformations labeled "alpha" and "beta." For trans proline, the mean ϕ and ψ values for the alpha conformation were $\phi = -61°$ and $\psi = -35°$, and for the beta conformation they were $\phi = -65°$ and $\psi = 150°$. For cis proline, the mean ϕ and ψ values for the alpha conformation were $\phi = -86°$ and $\psi = -1°$, and for the beta conformation they were $\phi = -76°$ and $\psi = 159°$. These were the conformations used for the backbones. The minimum energy ring pucker was chosen for each backbone conformation (within either the cis or trans set) since it was different in each case.

Finally, the cystine residue was treated in its disulfide bridged form, so it was only represented by one molecule comprised of two residues, each having a different backbone conformation. We assigned the dicystine backbone and sidechain conformations based on data in the PDB survey carried out by Thornton.[22] She found that right-handed ($\chi_3 = +90°$) and left-handed ($\chi_3 = -90°$) disulfides occured in relatively equal numbers. The right-handed disulfides displayed a greater variety of conformations than did the left-handed ones, however, with 70% of the left-handed disulfides occuring with $\chi_2 = \chi_2' = -80°$ and $\chi_1 = \chi_1' = -60°$. We therefore chose to use the predominant left-handed conformation for assigning the sidechain dihedrals. Thornton[22] found that cystine residues occured primarily with random coil backbones (59%) but also occured frequently in alpha-helical (25%) and beta-sheet (18%) conformations. Based on these data, we gave one of the cystine residues an extended backbone and the other an alpha-helical one. The final minimization was carried out without constraints on the ϕ, ψ dihedrals of the alpha-helical backbone in order to allow it to relax somewhat. The minimized values of ϕ and ψ were $-60°$ and $-27°$. The conformations chosen for each amino acid are described in Table II.

RESP FITTING METHODOLOGY

The procedure of RESP fitting to obtain atomic charges has been described in ref. 1, 2. The term *RESP* refers to the restrained ESP charge fitting using the following equation:

$$f(q_1, \ldots, q_{natoms}) = \chi^2_{esp} + \chi^2_{hyp.restr} + \lambda_1 g_1 + \cdots + \lambda_w g_w \quad (1)$$

where

$$\chi^2_{esp} = \sum_{i=1}^{ESPpoints} \left(V_i - \sum_{j=1}^{natoms} \frac{q_j}{r_{ij}} \right)^2 \quad (2)$$

and

$$\chi^2_{hyp.restr} = a \sum_{j=1}^{natoms} \left(\left(q_j^2 + b^2 \right)^{1/2} - b \right) \quad (3)$$

In the preceding equations, V_i is the quantum mechanically calculated electrostatic potential (ESP) at point i, q_j are the resultant charges, a is a scale factor defining the asymptotic limits of the strength of the hyperbolic restraint according to eq. (3), and b defines the tightness of the hyperbola around the minimum. A value of 0.1 was found to be appropriate for b.[1] The g_i are additional constraints imposed on resultant charges, and λ_i are Lagrange multipliers. The minimum of the $f(q_1, \ldots, q_{natoms})$ function is sought by requiring that

$$\frac{\partial f}{\partial q_k} = 0, \frac{\partial f}{\partial \lambda_l} = 0, \text{ for each } k, l \quad (4)$$

which leads to the matrix equation of the type

$$\mathbf{Aq} = \mathbf{B} \quad (5)$$

which must be solved for q. This is done iteratively when using nonzero hyperbolic restraints, since the left-hand side of eq. (5) (matrix **A**) depends on charges **q**.

The RESP fitting scheme, which we have suggested to be useful, involves a two-stage procedure with hyperbolic restraints, which we denoted as (wk.fr./st.eq.) in earlier works.[1,2] In the first stage, a weak hyperbolic restraint ($a = 0.0005$) to a target value of 0.0 is applied to all heavy atoms. Hydrogen atoms are not restrained because they are never buried within a molecule and are always well defined by the ESP points. In the second stage, charges on all atoms were kept frozen to their values obtained in the first stage, except for those in methyl and methylene groups. CH_3 and

TABLE II.
AMBER Minimized Conformations Used for Each Amino Acid.

Amino Acid	Backbone	ϕ	ψ	χ_1	χ_2	χ_3	χ_4
Val	C_5	208	154	61 (g^-)			
	α_R	−60*	−40*	181 (t)			
Ser	C_5	197	174	47 (g^-)	299 (g^+)		
	α_R	−60*	−40*	187 (t)	43 (g^-)		
Asn	C_5	217	160	297 (g^+)	263		
	α_R	−60*	−40*	180 (t)	269		
Cys	C_5	206	141	172 (t)	179 (t)		
	α_R	−60*	−40*	284 (g^+)	303 (g^+)		
Cyx	C_5	218	157	290 (g^+)	303 (g^+)	276 (g^+)	
	α_R	−60*	−27	293 (g^+)	302 (g^+)	276 (g^+)	
Asp	C_5	218	160	299 (g^+)	85		
	α_R	−60*	−40*	160 (t)	92		
Thr	C_5	220	165	46 (g^-)	300 (g^+)		
	α_R	−60*	−40*	333 (g^+)	185 (t)		
Ile	C_5	207	154	184 (t)	62 (g^-)		
	α_R	−60*	−40*	300 (g^+)	176 (t)		
Leu	C_5	209	150	181 (t)	63 (g^-)		
	α_R	−60*	−40*	302 (g^+)	179 (t)		
Hid	C_5	217	162	314 (g^+)	270*	182	
	α_R	−60*	−40*	182* (t)	82*	178	
Hie	C_5	217	161	279 (g^+)	271*	177	
	α_R	−60*	−40*	186 (t)	95	182	
Hip	C_5	208	145	312 (g^+)	271*	176	
	α_R	−60*	−40*	184 (t)	82	188	
Phe	C_5	217	160	308 (g^+)	119		
	α_R	−60*	−40*	185 (t)	264 (g^+)	82	
Tyr	C_5	217	160	299 (g^+)	96	177	
	α_R	−60*	−40*	180 (t)	256	357	
Trp	C_5	216	161	299 (g^+)	100		
	α_R	−60*	−40*	191 (t)	282		
Met	C_5	205*	153*	187 (t)	176 (t)	180 (t)	
	α_R	−60*	−40*	298 (g^+)	306 (g^+)	302 (g^+)	
Pro(t)	β	295*	150*	28	326		
	α	299*	325*	334	35		
Pro(c)	β	305	143	334	35		
	α	274*	1*	33	324		
Glu	C_5	206*	153*	297 (g^+)	189 (t)	65	
	α_R	−60*	−40*	192* (t)	88	90*	
Gln	C_5	205	153	192 (t)	177 (t)	120	
	α_R	−60*	−40*	308 (g^+)	310 (g^+)	123	
Arg	C_5	214	145	298 (g^+)	186 (t)	181 (t)	177 (t)
	α_R	−60*	−40*	179 (t)	160 (t)	291 (g^+)	283 (g^+)
Lys	C_5	205	153	188 (t)	180 (t)	198 (t)	73 (g^-)
	α_R	−60*	−40*	293 (g^+)	189 (t)	177 (t)	183 (t)

CH_2 groups are refitted with the hydrogens within a given group constrained to have equivalent charges. The hyperbolic restraint applied during the second stage is twice as strong as the one in stage one ($a = 0.001$). The two-stage restrained ESP charges exhibit less conformational dependence compared to the standard ESP charges, result in excellent conformational energies, and give good results for hydrogen bonding energies and free energies of solvation.[1,2]

The necessity of a two-stage fit arises from the need to constrain atoms which are not symmetrically equivalent within the static conformation of the molecule used for the calculation but which become equivalent under dynamical conditions when rotation can occur. One example of this

would be the three methyl hydrogens in methanol. If these inequivalent atoms are forced to have the same charge during a one-stage fit, the charge on the oxygen is reduced to a value which does not yield good free energies of solvation or interaction energies. The two-stage fit then allows for the "best" charges to be fit on the heteroatoms during the first stage, with the maximum number of degrees of freedom available to the molecule. Then methyl and sometimes methylene hydrogens are constrained to be equivalent in the second stage of the fit.

THE ROLE OF LAGRANGE CONSTRAINTS

The role of Lagrange constraints (conditions) in eq. (1) is manifold. In the standard RESP procedure described in the previous section, they were used for two purposes. In the first and simplest case, they were used to keep the sum of charges equal to the total molecular charge. In the second case, they were used to force identical charges on equivalent atoms during the fit. The most common example applies to methyl hydrogens in the second stage, as mentioned earlier; however, chemically equivalent atoms which are not refit in stage 2 can be constrained to have equivalent charges in stage 1. The two oxygens in the sidechain of aspartic acid are an example of the latter situation. When two different methyl groups in a molecule were defined to be symmetric, the two carbons were constrained to have the same charge during the first stage of the fit, but each hydrogen was allowed to optimize freely.

In this article, Lagrange multipliers are shown to have some additional uses. They will be used for equivalencing atomic charges on the same atoms of different conformers of the same molecule. This was extensively tested in our earlier study for propylamine[2] and will be applied here to equivalence some atoms in the A and B conformers of nucleosides or different conformers of sidechains in amino acids. Multiple-conformation fitting has been shown to be useful in dealing with nonphysical conformational variation of charges. Unlike Reynolds et al.,[4] however, we do not use Boltzmann weighting for different conformers, since we do not know the relative energies in solution or the dielectric environment of a protein.

The Lagrangian multiplier method will also be applied to equivalence some charges on atoms during multiple-molecule fitting. This will be used to force similar groups of atoms in different residues to have the same atomic charges. This strategy was employed to derive charges for sugar atoms in different nucleosides and —CO—NH— backbone atoms in amino acids. Using this approach in creating our database is especially important with respect to its further application for any free-energy perturbation calculations.[23]

Finally, Lagrange constraints can be used in splicing together two fragments from different molecules. This use arises, for example, when one wants to build the nucleotide residues containing base, sugar, and phosphate by combining dimethylphosphate (DMP) and the nucleosides. In this way, the charges on the CH_3 group of DMP and the O3′—HO3′ atoms on nucleosides can be required to have a net charge equal to zero and then those groups can be removed from the system when the whole nucleotide residue is created. In the case of amino acids, one can force the blocking groups of the dipeptide to have zero charges, thus leading to the appropriate net unit charge for the central (NH—CHR—CO—) fragment.

The aforementioned features [i.e., restrained electrostatic potential (RESP) and multimolecular and multiconformational fitting] will all be employed in our charge development for DNA, RNA, and proteins. The software for carrying out such manipulations is free and available upon request and will also be distributed with the AMBER software package.[16]

Derivation of the Nucleic Acid Charges

DATABASE STRUCTURE

In addition to deriving the charges for the nucleic acids, we decided to modify the structure of the AMBER database for nucleic acids.[10] Previously, the nucleic acids were built within the AMBER program separately, treating terminal atoms (H, OH), linker groups (PO_2—), and base–sugar (from O5′ to O3′) as separate residues. The charges on the sugar part of nucleosides were kept the same, independent of the nucleic acid base to which they were attached. For the first and last residues, the charges on the O5′—HO5′ and O3′—HO3′ end atoms were modified to take into account the difference in charge on O3′ and O5′ when it is attached to a negative phosphate or a hydrogen. In the present approach, the nucleic acid strands are built from the residues containing four fragments for each nucleic acid base. In our new database, we have the following residues, with the names pointing to their position in the

nucleic acid strand: "X5E," where X = A, T, G, C, or U and is located at the 5′-end; "X," which is any central residue in a nucleotide chain; "X3E," which stands for O3′-end residues; and "XN," the nucleoside itself. The type of residues (X5E) and "XN" have O5′—HO5′ atoms instead of the phosphate group. Using this approach, we closely follow the way the crystallographic databases are prepared and constructed.[18] This has also influenced the procedure undertaken in the charge derivation.

SMALLER VERSUS LARGER RESIDUES IN DERIVING THE ESP

For the sake of comparison, we have also performed *ab initio* calculations for smaller constituents of the nucleic acids (i.e., *N*-methyl-bases, ribose, and deoxyribose, with C1′ substituted with a formamide —NH—CHO group). The formamide group resembles the immediate chemical environment for sugars linked to the bases. This allows us to compare the charge distributions obtained from the calculations for the whole nucleosides with those for smaller components. According to the approach used previously[10] (labeled "old scheme"), the larger residues were built from smaller fragments, removing some end groups and redistributing their charges into the atoms creating the new bond. In our case, creating charges for the larger residue—nucleoside from *N*-methyl-base and formamido-sugar—was done by keeping all ESP (electrostatic potential fitted) charges for all atoms as they are in separate molecules and by redistributing the sum of the charges on the $-CH_3$ from the *N*-methyl-base and the —NH—CHO from the formamido-sugar into the C1′—H1′ atoms. To obtain the charges on the sugar–phosphate group joining atoms, one can "patch" charges on HO5′—O5′, HO3′—O3′ in the sugar together with the two O—Me groups in DMP to modify appropriately the charges on O5′ and O3′ atoms.

For further comparison of the influence of ESP charges on the size of residue taken into consideration, we have also performed test calculations for the O(P)—Me nucleotides in their B forms (see Fig. 1). Calculations for this size of molecules involved from 349 (for cytosine) up to 394 (for guanine) basis functions when using the 6-31G* basis set. Table III shows the results of those calculations only for deoxyadenine nucleoside.

Looking at the comparison compiled in Table III, one can see that the charges on some atoms vary substantially depending on the scheme used

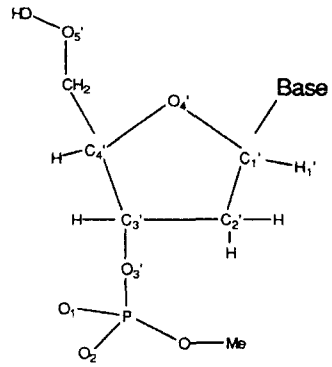

FIGURE 1. The scheme of O—Me nucleotides considered in calculations of ESP charges.

in the calculations. This is true especially for the N1 or N9 base atoms, C2′, C3′, as well as on C1′ and H1′ atoms. The large magnitude of C2′ and C3′ charges in the A-form of deoxynucleosides is probably caused by the proximity of the base in this conformation, which overpolarizes those atoms. Such an effect is eliminated in separate (i.e., "old" scheme) sugar and base calculations. In the ribonucleoside cases (results not reported here in detail but available upon request from the authors), the C2′ and C3′ atoms do not exhibit such a large fluctuation such as those reported in Table III for deoxynucleosides. The charges on C1′ and H1′ atoms also fluctuate to a significant extent depending on the scheme used to obtain them. Additionally, in light of the extensive tests done previously in our laboratory on the electrostatic potential charge fitting,[1,2] the charges on C1′ and H1′ should not have such large values since those atoms are buried inside the residues and their charges are poorly determined in the least-squares fitting procedure. This suggests that the previously employed approach of patching charges of smaller molecules onto larger residues with some method (usually subjective) of redistributing charges is not necessarily a fully justified way of constructing electrostatic representation of molecules.

Because of the aforementioned problems, we prepared the new database of nucleic acid components charges in the way which will be described later.

MULTIMOLECULAR FITTING AND EQUIVALENCING

The first question which we wanted to answer is: What should be equivalenced during the nucleic acid charge fit? We did multimolecular RESP charge fitting for the following cases: (1) for the

TABLE III.
Comparison of the Atomic Charges on Some Important Atoms Obtained Using the Standard ESP and RESP Fit (ref. 8).

Residue	Atom	"Old" Scheme	New Scheme		Whole Nucleotide Calculation
			B-Form	A-Form	B-Form
Deoxy-ADE	N9	−0.132	−0.300	−0.294	−0.278
	C1′	0.236	0.478	0.513	0.424
	H1′	0.294	0.095	0.069	0.126
	C2′	−0.174	−0.275	−0.521	−0.243
	C3′	0.115	0.169	0.556	0.101
	C4′	0.382	0.316	0.335	0.480
	O4′	−0.582	−0.530	−0.556	−0.598
	O3′ (end)	−0.708	−0.708	−0.783	—
	O3′ (Phos)	−0.552	—	—	−0.543
	O5′ (end)	−0.653	−0.663	−0.675	−0.672
	O5′ (Phos)	−0.552	—	—	−0.524
	P	1.247	—	—	1.315
	O1P, O2P	−0.802	—	—	−0.828

Comparison is made between (a) old scheme, involving separate deoxyribose and base calculations, (b) new approach based on the whole nucleoside fit, and (c) whole nucleotide fit.

four ribonucleosides (A, C, G, U), (2) for four deoxyribonucleosides with their sugars in the C3′–endo conformations (A, C, G, T, in A-DNA form), (3) for four deoxyribonucleosides with the sugars in C2′–endo conformations (B-DNA form), and (4) for all eight deoxy (i.e., A and B) forms of the nucleosides together. In all of those four cases, we tried three different schemes of intermolecular equivalencing. In the first approach, all sugar atoms were equivalenced between all fitted molecules; in the second, all sugar atoms except C1′, H1′ were equivalenced; and in the final scheme, the atoms O4′, C1′, H1′, C2′, H2A′, H2B′ were allowed to vary in each nucleoside. The results of this calculations for the case (3), involving four deoxyribonucleosides with the sugars in C2′–endo conformations, are summarized in Table IV. One can see that using RESP approach and allowing C1′ and H1′ atoms to vary during the multimolecular fit produces the lowest rrms (relative root mean square deviation error) of the fit. Even though all the rrms values are small, such a model seems to be most appropriate from a physical point of view.

Another question arose concerning the issue of how to fit deoxyribonucleotides: whether one should use A- together with B-DNA forms in multimolecular fitting. After performing some tests, we decided to use only B-forms of nucleotides because of relatively large resultant charges on C2′, C3′ atoms in the A-forms and because it is more important that we accurately represent the biologically more important B-form of DNA. This charge behavior on C2′ and C3′ atoms is likely due to the intramolecular interactions between bases and sugar atoms in C3′–endo sugar puckering, as noted earlier. Table IV also includes the results of separate nucleoside fitting using the B-form. One can see that using intermolecular equivalencing in the fit causes smoothing of the variation of the charge on the C1′ sugar atom and has the important effect of averaging fluctuations in the C2′ and C3′ charges as well as those at other atoms. Also, the rrms error of the fit is the lowest in this case.

At this point, we chose to examine the effect of using the RESP fit charges versus the "old" standard ESP scheme,[8] where no restraints were used. The data in Table V (in which, as an example, results for deoxyadenosine are presented) show how RESP fitting causes buried atoms to have much smaller and more reasonable charges from the electrostatic point of view. This is particularly true for the C1′ sugar atoms. With the old approach, one finds very high positive charges on these atoms, whereas with RESP with or without intermolecular equivalencing they become small. Also, more positive charge ends up on the H1′ than C1′, which is appropriate from the point of view of electronegativity.

We also studied the issue of amino–NH_2 group fitting in the nucleic acid bases. The question is whether hydrogens in this important group should be equivalenced in the first or second stage of RESP fitting[1,2] (i.e., should they be treated in the same way as CH_2 and CH_3 groups) and whether

TABLE IV.
Comparison of Different Sugar Atom Equivalencing in Multimolecular RESP Fit for Deoxynucleosides.

Atom	No Sugar Atoms Equivalencing				All Sugar Atoms Equivalencing	All Sugar Atoms Equivalencing Except C1′ Group	All Sugar Atoms Equivalencing Except C1′, C2′, O4′ Groups
	A	C	G	T			
HO5′	0.435	0.432	0.442	0.441	0.444	0.444	0.438
O5′	−0.618	−0.618	−0.622	−0.631	−0.638	−0.638	−0.624
C5′	−0.025	−0.027	−0.033	−0.048	−0.001	0.000	−0.035
H5A′	0.077	0.079	0.075	0.094	0.074	0.073	0.083
H5B′	0.077	0.079	0.075	0.094	0.074	0.073	0.083
C4′	0.195	0.168	0.233	0.135	0.163	0.164	0.172
H4′	0.109	0.115	0.096	0.124	0.114	0.114	0.114
O4′	−0.401	−0.389	−0.429	−0.369	−0.372	−0.373	−0.376 ± 0.010
C1′	0.056	−0.002	0.131	0.027	0.039	0.039 ± 0.028	0.030 ± 0.035
H1′	0.187	0.189	0.153	0.195	0.183	0.181 ± 0.009	0.184 ± 0.012
C3′	0.102	0.122	0.124	0.122	0.107	0.107	0.099
H3′	0.097	0.085	0.087	0.096	0.087	0.086	0.090
C2′	−0.092	−0.060	−0.087	−0.080	−0.085	−0.085	−0.056 ± 0.016
H2A′	0.067	0.057	0.057	0.064	0.069	0.069	0.060 ± 0.005
H2B′	0.067	0.057	0.057	0.064	0.069	0.069	0.060 ± 0.005
O3′	−0.665	−0.663	−0.679	−0.664	−0.668	−0.668	−0.665
HO3′	0.434	0.434	0.433	0.436	0.441	0.441	0.437
rrms		0.076			0.075	0.075	0.076

Results are based on four deoxynucleosides fitting with their sugars in the C2′-endo conformations. In each case, rrms is calculated for the four nucleosides together.

TABLE V.
Comparison of Using Standard (ref. 8) ESP and RESP Charges for the ADE – Deoxynucleoside.

	ADE			Sugar	
	ESP	RESP		ESP	RESP
N9	−0.258	−0.018	HO5′	0.456	0.444
C8	0.293	0.141	O5′	−0.674	−0.638
H8	0.156	0.192	C5′	0.047	0.000
N7	−0.647	−0.606	H5A′	0.051	0.073
C5	0.000	0.082	H5B′	0.053	0.073
C6	0.838	0.683	C4′	0.279	0.164
N6	−1.034	−0.900	H4′	0.067	0.114
HN6A	0.460	0.412	O4′	−0.464	−0.373
HN6B	0.446	0.412	C1′	0.347	0.048
N1	−0.849	−0.775	H1′	0.118	0.182
C2	0.631	0.590	C3′	0.164	0.107
H2	0.060	0.056	H3′	0.062	0.086
N3	−0.804	−0.748	C2′	−0.206	−0.085
C4	0.489	0.372	H2A′	0.089	0.069
			H2B′	0.085	0.069
			O3′	−0.706	−0.668
			HO3′	0.452	0.441

Data are obtained in multimolecular fitting of the four nucleosides together. Sugar atom charges are eqivalenced with exception of C1′, H1′ atoms. Sugars in the nucleosides are in its C2′-endo conformations. The rrms for the ESP and RESP fit is 0.0647 and 0.0746, respectively.

they should have the same charges. Should they be intramolecularly equivalent or not? The comparison of different approaches (not presented in detail here; i.e., equivalencing amino hydrogens in the first stage or in the second stage of the fit or not equivalencing them at all) reveals that the rrms of the fit is almost independent of the approach chosen. Thus we decided to equivalence hydrogens intramolecularly in all amino groups in the first stage of the RESP fit (i.e., not to treat them as hydrophobic CH_3 and CH_2 groups, for which it was shown that the second stage refitting was necessary[1,2]).

"JOINTS"

As mentioned earlier, our new database is built from whole nucleotide fragments, whereas ESPs (electrostatic potentials around molecules) were derived for nucleosides and dimethylphosphate (DMP) only. To make the "joints" between those residues, we use additional Lagrange constraints, which force the sum of the charge on the group of atoms to become zero (see Fig. 2). For example, for the residue type "X5E," which is located at the beginning of the nucleic acid strand, it is required that the sum of charges of the group I and II is zero: $q_{\mathrm{I}} + q_{\mathrm{II}} = 0.0$. For the residue type "X3E," this condition could be simply set to $q_{\mathrm{III}} + q_{\mathrm{IV}} = 0.0$. For any "middle"-type residue X, those two conditions should be fulfilled at the same time. As the result of imposing those additional Lagrange conditions, the charges on "joint" O3′ or O5′ and phosphate group atoms changed their charges insignificantly. The rrms error of the fit does not change upon imposing those additional constraints in comparison to the unconstrained fitting.

SUMMARY

The following decisions were made in order to get the final set of nucleic acid charges. RESP two-stage charge fitting is used where four nucleosides in the B-form, in the case of deoxyribo, and four nucleosides in the A-form, in the case of ribo, were fitted together. In the first stage, (1) charge constraints were applied to the sum of charges between CH_3 groups of DMP and O5′H and/or O3′H groups in order to create appropriate "joints"; (2) sugar atoms were intermolecularly equivalenced with the exception of C1′ and H1′ atoms; (3) O1P, O2P in phosphate groups were equivalenced, (4) H in NH_2 groups were equivalenced, and (5) hyperbolic restraints with force constant $a = 0.0005$ were applied to all heavy

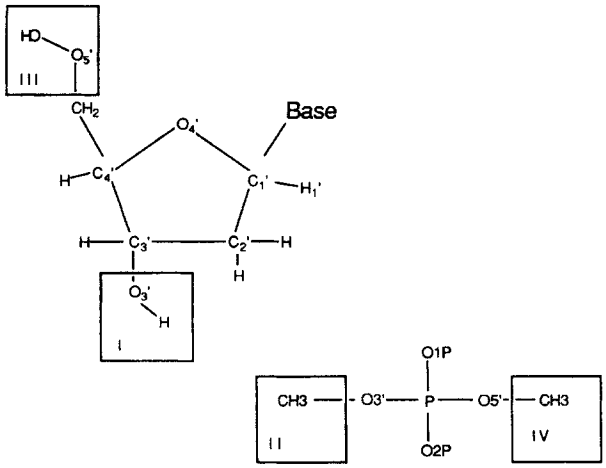

FIGURE 2. The scheme used for charge fitting on nucleoside–phosphate bonds.

atoms.[1,2] Hydrogens were not restrained. In the second stage of fitting, CH_2, CH_3 (e.g., thymine) groups were refit with a force constant $a = 0.001$ for hyperbolic restraints applied to their carbon atoms with no restraints put on hydrogens, with intra- and intermolecular equivalencing applied to the C2′(H_2) groups in deoxynucleotides and C5′(H_2) for both ribo- and deoxyribonucleotides.

The final charges obtained using the aforementioned protocol are collected in Table VI. This table contains all necessary data used to build our charge database for further AMBER calculations.[16] In view of the results of the tests discussed earlier, showing a small effect of imposing additional "joint" conditions on sugar atom charges, we assumed the following approach to create database charges for all types of residues used in AMBER. In any case, we used the results of the aforementioned fit with two "joint" conditions for linkages between DMP and sugars (i.e., $q_{\mathrm{I}} + q_{\mathrm{II}} = 0.0$, and $q_{\mathrm{III}} + q_{\mathrm{IV}} = 0.0$). For the X type of residues, the values of the charges on the O3′ and O5′ end oxygens were taken to be equal to the O3′ and O5′ (see Fig. 2) charges of dimethylphosphate, respectively, in the "joint" fitting. For the X5E and X3E types of residues, only their O3′ and O5′ charges were taken to be equal to O3′ and O5′ charges of DMP, respectively, whereas their other end-group charges [i.e., on O5′—H (X5E) and O3′—H (X3E)] were assumed to have their actual values obtained in the aforementioned fit. The latter statement also applies to the derivation of the charges of XN type of residues (i.e., to the whole nucleosides themselves). Based on the results of our tests showing minuscule dependence of sugar charges on adding "joint" constraints and the fact that there is no "charge

TABLE VI.
The Final Set of Charges for Nucleic Acid Components Obtained in RESP Multimolecular Fitting with Additional "Joints" Conditions for Sugar – Phosphate Linkages.

Deoxy-ADE		Deoxy-CYT		Deoxy-GUA		Deoxy-THY	
HO5′ (end)	0.4422	HO5′ (end)	0.4422	HO5′ (end)	0.4422	HO5′ (end)	0.4422
O5′ (end)	−0.6318	O5′ (end)	−0.6318	O5′ (end)	−0.6318	O5′ (end)	−0.6318
O5′ (Phos)	−0.4954	O5′ (Phos)	−0.4954	O5′ (Phos)	−0.4954	O5′ (Phos)	−0.4954
C5′	−0.0069	C5′	−0.0069	C5′	−0.0069	C5′	−0.0069
H5A′	0.0754	H5A′	0.0754	H5A′	0.0754	H5A′	0.0754
H5B′	0.0754	H5B′	0.0754	H5B′	0.0754	H5B′	0.0754
C4′	0.1629	C4′	0.1629	C4′	0.1629	C4′	0.1629
H4′	0.1176	H4′	0.1176	H4′	0.1176	H4′	0.1176
O4′	−0.3691	O4′	−0.3691	O4′	−0.3691	O4′	−0.3691
C1′	0.0431	C1′	−0.0116	C1′	0.0358	C1′	0.0680
H1′	0.1838	H1′	0.1963	H1′	0.1746	H1′	0.1804
N9	−0.0268	N1	−0.0339	N9	0.0577	N1	−0.0239
C8	0.1607	C6	−0.0183	C8	0.0736	C6	−0.2209
H8	0.1877	H6	0.2293	H8	0.1997	H6	0.2607
N7	−0.6175	C5	−0.5222	N7	−0.5725	C5	0.0025
C5	0.0725	H5	0.1863	C5	0.1991	C5M	−0.2269
C6	0.6897	C4	0.8439	C6	0.4918	H5MA	0.0770
N6	−0.9123	N4	−0.9773	O6	−0.5699	H5MB	0.0770
HN6A	0.4167	HN4A	0.4314	N1	−0.5053	H5MC	0.0770
HN6B	0.4167	HN4B	0.4314	H1	0.3520	C4	0.5194
N1	−0.7624	N3	−0.7748	C2	0.7432	O4	−0.5563
C2	0.5716	C2	0.7959	N2	−0.9230	N3	−0.4340
H2	0.0598	O2	−0.6548	HN2A	0.4235	H3	0.3420
N3	−0.7417	C3′	0.0713	HN2B	0.4235	C2	0.5677
C4	0.3800	H3′	0.0985	N3	−0.6636	O2	−0.5881
C3′	0.0713	C2′	−0.0854	C4	0.1814	C3′	0.0713
H3′	0.0985	H2A′	0.0718	C3′	0.0713	H3′	0.0985
C2′	−0.0854	H2B′	0.0718	H3′	0.098	C2′	−0.0854
H2A′	0.0718	O3′ (end)	−0.6549	C2′	−0.0854	H2A′	0.0718
H2B′	0.0718	HO3′ (end)	0.4396	H2A′	0.0718	H2B′	0.0718
O3′ (end)	−0.6549	O3′ (Phos)	−0.5232	H2B′	0.0718	O3′ (end)	−0.6549
HO3′ (end)	0.4396	P	1.1659	O3′ (end)	−0.6549	HO3′ (end)	0.4396
O3′ (Phos)	−0.5232	O1P	−0.7761	HO3′ (end)	0.4396	O3′ (Phos)	−0.5232
P	1.1659	O2P	−0.7761	O3′ (Phos)	−0.5232	P	1.1659
O1P	−0.7761			P	1.1659	O1P	−0.7761
O2P	−0.7761			O1P	−0.7761	O2P	−0.7761
				O2P	−0.7761		

Ribo-ADE		Ribo-CYT		Ribo-GUA		Ribo-URA	
HO5′ (end)	0.4295	HO5′ (end)	0.4295	HO5′ (end)	0.4295	HO5′ (end)	0.4295
O5′ (end)	−0.6223	O5′ (end)	−0.6223	O5′ (end)	−0.6223	O5′ (end)	−0.6223
O5′ (Phos)	−0.4989	O5′ (Phos)	−0.4989	O5′ (Phos)	−0.4989	O5′ (Phos)	−0.4989
C5′	0.0558	C5′	0.0558	C5′	0.0558	C5′	0.0558
H5A′	0.0679	H5A′	0.0679	H5A′	0.0679	H5A′	0.0679
H5B′	0.0679	H5B′	0.0679	H5B′	0.0679	H5B′	0.0679
C4′	0.1065	C4′	0.1065	C4′	0.1065	C4′	0.1065
H4′	0.1174	H4′	0.1174	H4′	0.1174	H4′	0.1174
O4′	−0.3548	O4′	−0.3548	O4′	−0.3548	O4′	−0.3548
C1′	0.0394	C1′	0.0066	C1′	0.0191	C1′	0.0674
H1′	0.2007	H1′	0.2029	H1′	0.2006	H1′	0.1824
N9	−0.0251	N1	−0.0484	N9	0.0492	N1	0.0418
C8	0.2006	C6	0.0053	C8	0.1374	C6	−0.1126

TABLE VI.
(Continued)

Ribo-ADE		Ribo-CYT		Ribo-GUA		Ribo-URA	
H8	0.1553	H6	0.1958	H8	0.1640	H6	0.2188
N7	−0.6073	C5	−0.5215	N7	−0.5709	C5	−0.3635
C5	0.0515	H5	0.1928	C5	0.1744	H5	0.1811
C6	0.7009	C4	0.8185	C6	0.4770	C4	0.5952
N6	−0.9019	N4	−0.9530	O6	−0.5597	O4	−0.5761
HN6A	0.4115	HN4A	0.4234	N1	−0.4787	N3	−0.3549
HN6B	0.4115	HN4B	0.4234	H1	0.3424	H3	0.3154
N1	−0.7615	N3	−0.7584	C2	0.7657	C2	0.4687
C2	0.5875	C2	0.7538	N2	−0.9672	O2	−0.5477
H2	0.0473	O2	−0.6252	HN2A	0.4364	C3′	0.2022
N3	−0.6997	C3′	0.2022	HN2B	0.4364	H3′	0.0615
C4	0.3053	H3′	0.0615	N3	−0.6323	C2′	0.0670
C3′	0.2022	C2′	0.0670	C4	0.1222	H2A′	0.0972
H3′	0.0615	H2A′	0.0972	C3′	0.2022	O2′	−0.6139
C2′	0.0670	O2′	−0.6139	H3′	0.0615	HO2′	0.4186
H2A′	0.0972	HO2′	0.4186	C2′	0.0670	O3′ (end)	−0.6541
O2′	−0.6139	O3′ (end)	−0.6541	H2A′	0.0972	HO3′ (end)	0.4376
HO2′	0.4186	HO3′ (end)	0.4376	O2′	−0.6139	O3′ (Phos)	−0.5246
O3′ (end)	−0.6541	O3′ (Phos)	−0.5246	HO2′	0.4186	P	1.1662
HO3′ (end)	0.4376	P	1.1662	O3′ (end)	−0.6541	O1P	−0.7760
O3′ (Phos)	−0.5246	O1P	−0.7760	HO3′ (end)	0.4376	O2P	−0.7760
P	1.1662	O2P	−0.7760	O3′ (Phos)	−0.5246		
O1P	−0.7760			P	1.1662		
O2P	−0.7760			O1P	−0.7760		
				O2P	−0.7760		

rrms for the deoxyribo- and ribonucleotides fit is 0.0340 and 0.0331, respectively.

transfer" between CH_3 groups of DMP and nucleosides in the "joint" fit, we decided to use the appropriate charges for XN residues presented in Table VI.

Derivation of the Amino Acid Charges

BASIS FOR EVALUATING DIFFERENT CHARGE SETS

A known weakness of the Weiner et al.[9,10] and some other protein force fields is that the energy calculated for the C_{7eq} and C_{7ax} conformations of *N*-acetyl and *N*-methylamide blocked alanine and glycine dipeptides is significantly too stable compared to the results from high-level quantum mechanical calculations. A reason for this discrepancy is that at the time that the Weiner et al. force field was developed (mid-1980s), computer limitations were such that no high-level quantum mechanical data were available for these molecules for use in calibration of the force field. A number of such calculations have recently been carried out[24-30] at varying levels of theory and employing either the methyl-blocked residues described earlier or ones in which those methyl groups are replaced by hydrogen atoms. Thus, a key motivation in the evaluation of the charge models described later is to choose those that, first, are representative of important conformations, and, second, come closest to reproducing the quantum mechanical conformational energies.

CONFORMATIONAL ENERGIES FROM SINGLE- AND MULTIPLE-CONFORMATION CHARGE SETS

RESP charges were first calculated using the potentials generated for some of the quantum mechanically optimized[31] conformers of the alanine analog. Previous studies[1,2,4] have shown that multiple conformation fits produce charge sets which perform better at reproducing the electrostatic potentials of more of the low-energy conformations than does a single conformation fit. We thus wished to examine the conformational energies which re-

TABLE VII.
Alanyl Dipeptide Conformational Energies Calculated Using RESP Charges Derived from Single- and Multiple-Conformation Fits.[a,b]

Conf.	E(QM)	E(MM) Charge Model				
		C_{7eq}	C5	α_R	C5 / α_R	C_{7eq} / C5 / α_R
C_{7eq}	0.0	0.0	0.0	0.0	0.0	0.0
C_{7ax}	2.1	0.7	0.6	1.0	0.7	1.0
C5	1.5	4.2	4.1	2.5	3.5	3.4
α_R	3.9	7.0	6.7	4.3	5.6	5.8

[a]Energies in kcal / mol.
[b]A single Lagrange constraint was imposed resulting in overall neutrality of the molecule, and the two amide groups were allowed to optimize independently and have different charges. In multiple-conformation fits, all corresponding atoms are constrained to have equivalent charges between the two or three conformations.

sulted from charge sets derived from different conformations. We investigated the molecular mechanical conformational energies calculated from charges derived from single conformation fits of the C_{7eq}, C_5, and α_R conformations as well as from

multiple conformation fits for C_5 and α_R and C_{7eq}, C_5, and α_R. The C_{7ax} conformation was not considered in this analysis because it is not found frequently in proteins. The conformational energies calculated are presented in Table VII. All five charge sets result in conformational energies with similar trends. The C_{7ax} conformation is from 1.1 to 1.5 kcal/mol too low in energy, and the C_5 conformation is from 1.0 to 2.7 kcal/mol too high in energy relative to the quantum mechanical values. The α_R conformation is the most sensitive to the charge model used, ranging from 0.4 to 3.1 kcal/mol too high in energy. It is interesting that the highest energy conformation, the α_R, yields charges which produce the best set of energies. This is consistent with results seen earlier for dopamine[32] and propylamine.[2]

In subsequent calculations, we employed a C_5/α_R multiple-conformation model. The single-conformation α_R fit provided better alanine dipeptide energies, but we wanted to use a multiple-conformation fit since such charges perform better in modeling the electrostatic potential of many conformations of the amino acid. The three-conformation model was rejected because its energies were nearly identical to those resulting from the two-conformation model. Furthermore, PDB data was available which provided information on sidechain conformations preferred for the two different types of secondary structure—beta-sheet (C_5) and alpha-helix (α_R). Using these two conformations, it was then a straightforward task to assign sidechain conformations for all of the amino acids. (See the Methodology section.)

It is apparent that even the simplest model for calculating the dipeptide charges (i.e., one with no additional constraints) did not result in molecular mechanical energies in good agreement with the quantum mechanical energies. We therefore proceeded in our development of a charge model for the amino acids by testing the effect of each aspect of the model on the conformational energies calculated.

EFFECTS OF MULTIPLE LAGRANGE CONSTRAINTS AND EQUIVALENCING OF BACKBONE AMIDE ATOMS

Table VIII presents the conformational energies calculated for glycine and alanine dipeptides with and without additional Lagrange constraints to produce three neutral residues and with and without forcing equivalent charges on the amide groups on either side of the alpha carbon (Fig. 3). The constraint of three neutral residues is necessary when deriving charges for the amino acid database since each amino acid residue must have a net charge of +1, 0, or −1. When carrying out the quantum mechanical electrostatic potential calculation, however, one must use a blocked form of the residue to have a chemically reasonable structure. The desirability of employing the constraint to equivalence amide groups is related to the desire to have a consensus set of backbone charges that would be used for all of the amino acids. The necessity for this simplification arises from results obtained by Sun et al.,[23] as discussed later. We thus decided to make the simplifying assumption

that different residues having the same net charge should have common backbone amide atom charges. The question of equivalencing the amide group charges from the blocking groups in the charge fit then arises. These C=O or N—H groups in the blocking groups are mimicking the adjacent residues in the protein or peptide, and, in principle, those adjacent residues should have C=O and N—H charges which are identical to the ones found in the central residue.

The results presented in Table VIII employed the C_5/α_R multiple-conformation fit charges derived from the quantum mechanically optimized structures for both glycine and alanine dipeptides. The first set of molecular mechanical energies corresponds to charges derived with each dipeptide treated as a single residue and with no constrained equivalence of the two backbone amides. These alanine energies differ from the ones in the C_5/α_R column of Table VII because in Table VIII the charges were derived without constraining the two conformations to have common charges on their N-terminal methyls and common charges on their C-terminal methyls. It was necessary to remove these two constraints; otherwise there were not enough degrees of freedom in the fit when the molecule was treated as three neutral residues for the charge fit algorithm to converge.

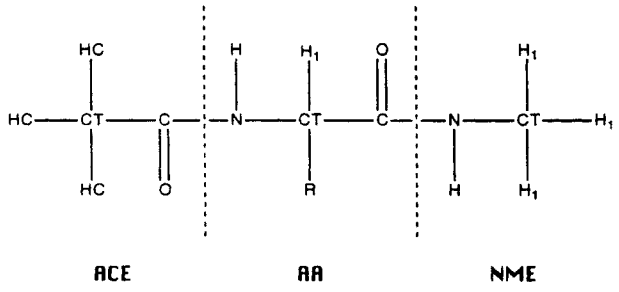

FIGURE 3. The scheme used for fitting the central amino acids. Lagrange constraints are used to define three residues of net integral charge. Blocking groups are neutral.

The two different types of constraints on the fit have similar effects. Both cause the alanine dipeptide C_{7ax} conformation to be lowered in energy by about 0.4 kcal/mol. The C_5 and α_R conformations of alanine both increase in energy with the application of either constraint, the C_5 conformation by 0.5–0.9 kcal/mol and the α_R conformation by 1.7–2.1 kcal/mol. The energy of glycine dipeptide's α_R conformation also goes up in energy in both cases, whereas the C_5 conformation goes up by 0.5 kcal/mol in one case and down in the other. The effect of the simultaneous application of both types of constraints is to lower the energy of the C_5 conformation of glycine dipeptide by about 0.5 kcal/mol and to raise the energy of the α_R conformation by 1.4 kcal/mol. The effect on the alanine dipeptide is to lower the energy of the C_{7ax} conformation by 0.4 kcal/mol, to lower the energy of the C_5 conformation by 0.3 kcal/mol, and to increase

TABLE VIII.
Glycyl and Alanyl Dipeptide Conformational Energies Calculated Using RESP Charges Derived With and Without Multiple Lagrange Constraints and Constrained Equivalence of the Two Amide Groups.[a,b]

			E(MM) Charge Model[c,d]							
	E(QM)		1 res no am eq.		3 res no am eq.		1 res am eq.		3 res am eq.	
Conf.	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala
$C7/_{7eq}$	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C_{7ax}		2.1		0.8		0.8		0.9		0.8
C5	2.0	1.5	4.6	3.4	3.8	3.6	4.8	3.2	3.8	2.4
α_R	4.0	3.9	6.6	5.6	7.3	6.0	6.6	5.6	7.3	5.6

[a] Energies in kcal/mol.
[b] Charges derived from multiple conformation fits of C5 and α_R conformations of alanyl dipeptide. Corresponding atoms constrained to have equivalent charges between the two conformations, with the exception of the terminal methyl groups. Charges for those methyl groups taken from the C5 conformation. Three Lagrange constraints applied to achieve neutrality of the acetyl blocking group, the *N*-methyl blocking group, and the central amino acid residue.
[c] "1 res" and "3 res" signify that the charges were fit to produce either one or three neutral residues from the molecule.
[d] "no am eq" and "am eq" signify that the two amide groups were not constrained to have equivalent charges, or that they were constrained.

the energy of the α_R conformation by 1.7 kcal/mol. The conformation which is most in error is the α_R conformation of glycine dipeptide, which is 3.3 kcal/mol higher in energy than the quantum mechanical reference energy.

EFFECT OF MOLECULAR MECHANICAL VERSUS QUANTUM MECHANICAL GEOMETRY OPTIMIZATION

The foregoing charge calculations were carried out on structures which had been optimized with nearly the same basis set that was used for calculating the electrostatic potential for the charge calculation (6-31G** versus 6-31G*). One might expect such optimized geometries to provide the most reasonable charges. However, such high-level *ab initio* geometry optimizations are fairly costly, so for the remaining amino acids we settled for structures optimized using molecular mechanics—in this case, the Weiner et al. force field.[10] For the sake of better justification of this approach, we compared energies calculated using charges derived from AMBER optimized glycine and alanine dipeptides. The charges were derived from a multiple-conformation fit employing the C_5 and α_R conformations and did not employ the three neutral residue constraint or the equivalent amide constraint. These energies show the same general trends as the ones calculated from the QM optimized geometry charges. For C_{7eq}, C_{7ax}, and C_5, MM and QM geometries give the same relative energies to within 0.3 kcal/mol, but for the α_R conformation the use of the molecular mechanical geometry leads to an energy that is 1 kcal/mol further from quantum mechanically target value. It would therefore seem that the quantum mechanically optimized structures yielded the best charges. We lack sufficient computer resources to carry out the optimizations on the remaining amino acids, however, and therefore were limited to dealing with the molecular mechanically optimized structures.

MULTIPLE-MOLECULE AND MULTIPLE-CONFORMATION FITTING

With the aforementioned tests, we established that the constraints of three neutral residues and equivalent amide groups and the use of molecular mechanically optimized structures produced charge sets which yielded conformational energies in reasonable agreement to the "ideal" charge sets, derived from the quantum mechanically optimized geometries and with no additional Lagrange constraints imposed. We next set out to carry out a multiple-molecule fit for the purpose of deriving a set of consensus charges for the backbone amide atoms. As noted earlier, the use of consensus charges for the amide atoms was motivated by the results obtained by Sun et al.[23] in a free-energy perturbation study involving the perturbation of alanine to valine. They found that the majority of the change in free energy was derived from interactions between water molecules and the atoms in the backbone of each residue. Corresponding backbone atoms had fairly different charges for each residue. While the nature of a given sidechain would be expected to have some effect on the electrostatic character of the backbone, this effect would probably be fairly subtle. The large variation in charge seen by Sun et al. was likely more of an artifact of the ESP fitting procedure. While this variation is reduced with the RESP procedure, we chose to use consensus amide charges in order to avoid the problem. The use of a simplified charge model that has consensus charges on the amide atoms then restricts the charge variation to the sidechains and the alpha carbon and hydrogen.

The first multiple molecule/conformation fit included the amino acids glycine, alanine, serine, valine, asparagine, aspartic acid, and protonated histidine. We chose this group to include the two simplest amino acids as well as a beta-branched and hydrophobic chain, a short and a longer polar chain, and negative and positively charged sidechains. This fit resulted in alpha carbon charges ranging from -0.084 to 0.038 for the neutral amino acids. The alpha carbon charges for aspartic acid and protonated histidine were -0.252 and 0.210, respectively. The larger charges on the alpha carbons of the charged residues suggested that their backbone amide groups were sufficiently different from the neutral residues to merit separate fitting.

We therefore settled on three separate fits to determine consensus charges for the backbone amide atoms. The main fit consisted of the five neutral amino acids from the set of seven—glycine, alanine, serine, valine, and asparagine. The second fit consisted of the two negatively charged amino acids—aspartic acid and glutamic acid. The third fit consisted of the three positively charged amino acids—lysine, arginine, and protonated histidine.

For the fits of the two groups of charged amino acids, the charges on the C=O and N—H groups in the blocking groups were constrained to have the consensus charges derived for the neutral amino acid backbone amide atoms. This modeled the presence of neutral amino acids on either side of the central charged residue. This was an approximation, since charged residues may be found adjacent to each other, but certain simplifying assumptions are necessary when deriving charges in this fashion for residues of a heteropolymer. The consensus amide charges from the neutral amino acid fit were applied to the remaining neutral amino acids through a constrained fit. The charges on the methyl hydrogens in the blocking groups were left free (not constrained to have the same charge within a group) since the blocking groups were discarded after the fit. This allowed the best set of charges to be calculated for the amino acid residue.

Charges for the acetyl and *N*-methyl terminal blocking groups were taken from the C_5 conformation of alanyl dipeptide. Hydrogens in a given methyl group were constrained to have a common charge during a fit where the charges on the remaining atoms were constrained to have the values determined in the five-residue fit. The use of acetyl and NME charges derived from different molecules or conformations was shown to have a minimal effect on the conformational energies calculated.

N- AND C-TERMINAL CHARGED AMINO ACIDS

The final set of charges is given in Table IX. We have also calculated charges for the charged N- and C-terminal versions of the amino acids. These charges were derived by splicing the ammonia group from methylammonium or the carboxylate group from acetic acid onto the blocked versions of the amino acids. Figure 4 illustrates how this procedure was carried out. A Lagrange constraint was applied which forced the charges on the atoms within the two boxed regions together to sum to 0.0. In this way the proper charge was attained on the resulting residue. In addition, the charge on the methyl carbon in methylammonium or acetic acid was constrained to have the same value as the charge on the alpha carbon. Also, in the N-terminal residue fits the N-terminal N and H atoms were constrained to have the same charges for both conformations of a given amino acid even though these atoms were "discarded" after the fit. In the C-terminal fits, the C-terminal C and O atoms were similarly constrained.

Discussion

We have presented an application of multimolecule, multiconformation RESP charge fitting to ribo- and deoxyribonucleic acids. This approach leads to more reasonable charges for buried atoms as compared to the standard ESP approach. Also, the use of a 6-31G* basis set rather than the STO-3G basis set used by Weiner et al.[9,10] has been shown to lead to hydrogen bond energies closer to those found with the highest-level *ab initio* calculations.[24] The use of the splicing approach described earlier allows an algorithmic merging of the charges of separate molecules and inspired a different format for this new nucleic acid force field than for the previous one. Naturally, both will continue to be supported within AMBER, since the previous force field is appropriate for implicit solvent calculations. More extensive molecular mechanical simulations employing these charges will be presented elsewhere, but in calculations on nucleosides and dimethylphosphate, the charges result in a good representation of the energy as a function of W (sugar pucker), γ, χ, ζ, and α.

We have also presented the derivation of charges for the amino acids using RESP with multiple-molecule and -conformational fitting. The amino acid charges also differ from those in the previous force field in that 6-31G*-level calculations were carried out on blocked versions of entire amino acids rather than fitting the backbone and the sidechains separately. Furthermore, Lagrange constraints were employed to obtain residues of the appropriate integral charge and to splice ammonium and carboxylate groups onto the charged N- and C-terminal residues.

A variety of charge models were evaluated to determine their ability to reproduce the quantum mechanical conformational energies of the glycine and alanine dipeptides. It is curious that the RESP charges calculated for alanine and glycine dipeptides using both the C_5 and α_R quantum mechanically optimized conformations for each (no additional Lagrange constraints for multiple residues of integral charge and no equivalencing of amide charge within and between the molecules) did not

TABLE IX.
Fitted RESP Amino Acid Charges from the Multiple Molecule / Multiple Conformation Fits.

Central Amino Acids		N-Terminal Amino Acids		C-Terminal Amino Acids	
		GLY			
N	−0.4157	N	0.2944	N	−0.3821
HN	0.2719	HN1	0.1642	HN	0.2681
		HN2	0.1642		
		HN3	0.1642		
CA	−0.0252	CA	0.0100	CA	−0.2494
HA1	0.0698	HA1	0.0895	HA2	0.1056
HA2	0.0698	HA2	0.0895	HA3	0.1056
C	0.5973	C	0.6163	C	0.7231
O	−0.5679	O	−0.5722	O	−0.7855
				OXT	−0.7855
		ALA			
N	−0.4157	N	0.1416	N	−0.3821
HN	0.2719	HN1	0.1997	HN	0.2681
		HN2	0.1997		
		HN3	0.1997		
CA	0.0337	CA	0.0962	CA	−0.1747
HA	0.0823	HA	0.0889	HA	0.1067
CB	−0.1825	CB	−0.0597	CB	−0.2092
HB1	0.0603	HB1	0.0300	HB1	0.0764
HB2	0.0603	HB2	0.0300	HB2	0.0764
HB3	0.0603	HB3	0.0300	HB3	0.0764
C	0.5973	C	0.6163	C	0.7731
O	−0.5679	O	−0.5722	O	−0.8055
				OXT	−0.8055
		SER			
N	−0.4157	N	0.1849	N	−0.3821
HN	0.2719	HN1	0.1898	HN	0.2681
		HN2	0.1898		
		HN3	0.1898		
CA	−0.0249	CA	0.0567	CA	−0.2721
HA	0.0843	HA	0.0782	HA	0.1304
CB	0.2117	CB	0.2596	CB	0.1123
HB1	0.0352	HB1	0.0273	HB2	0.0813
HB2	0.0352	HB2	0.0273	HB3	0.0813
OG	−0.6545	OG	−0.6714	OG	−0.6513
HOG	0.4275	HOG	0.4239	HOG	0.4474
C	0.5973	C	0.6163	C	0.8113
O	−0.5679	O	−0.5722	O	−0.8132
				OXT	−0.8132
		VAL			
N	−0.4157	N	0.0577	N	−0.3821
HN	0.2719	HN1	0.2272	HN	0.2681
		HN2	0.2272		
		HN3	0.2272		
CA	−0.0875	CA	−0.0054	CA	−0.3439
HA	0.0969	HA	0.1093	HA	0.1438
CB	0.2985	CB	0.3195	CB	0.1940
HB	−0.0297	HB	−0.0221	HB	0.0308
CG1	−0.3192	CG1	−0.3129	CG1	−0.3064

TABLE IX.
(Continued)

Central Amino Acids		N-Terminal Amino Acids		C-Terminal Amino Acids	
		VAL			
HG1	0.0791	HG1	0.0735	HG11	0.0836
HG2	0.0791	HG2	0.0735	HG12	0.0836
HG3	0.0791	HG3	0.0735	HG13	0.0836
CG2	−0.3192	CG2	−0.3129	CG2	−0.3064
HG4	0.0791	HG4	0.0735	HG21	0.0836
HG5	0.0791	HG5	0.0735	HG22	0.0836
HG6	0.0791	HG6	0.0735	HG23	0.0836
C	0.5973	C	0.6163	C	0.8350
O	−0.5679	O	−0.5722	O	−0.8173
				OXT	−0.8173
		ASN			
N	−0.4157	N	0.1801	N	−0.3821
HN	0.2719	HN1	0.1921	HN	0.2681
		HN2	0.1921		
		HN3	0.1921		
CA	0.0143	CA	0.0368	CA	−0.2080
HA	0.1048	HA	0.1231	HA	0.1358
CB	−0.2041	CB	−0.0283	CB	−0.2299
HB1	0.0797	HB1	0.0515	HB2	0.1023
HB2	0.0797	HB2	0.0515	HB3	0.1023
CG	0.7130	CG	0.5833	CG	0.7152
OD1	−0.5931	OD1	−0.5744	OD1	−0.6010
ND2	−0.9190	ND2	−0.8634	ND2	−0.9085
HND1	0.4196	HND1	0.4097	HND1	0.4150
HND2	0.4196	HND2	0.4097	HND2	0.4150
C	0.5973	C	0.6163	C	0.8050
O	−0.5679	O	−0.5722	O	−0.8147
				OXT	−0.8147
		ACE			
		H1	0.1123		
		CH3	−0.3662		
		H2	0.1123		
		H3	0.1123		
		C	0.5973		
		O	−0.5679		
		NME			
		N	−0.4157		
		HN	0.2719		
		CT	−0.1490		
		HT1	0.0976		
		HT2	0.0976		
		HT3	0.0976		

reproduce better the quantum mechanically calculated energies. This results runs counter to ones obtained for butane, simple alcohols, simple amines, ethane diol, and a series of substituted 1,3-dioxanes,[33,34] where RESP charges combined with simple torsional potentials performed quite well at reproducing relative conformational energies. The dipeptide molecules are more complicated, however; with their two amide groups and with 6-31G* charges, the intramolecular hydrogen bonding in the C_{7eq} and C_{7ax} conformations may be exaggerated, leading to an overestimate of the stability of these two conformations relative to C_5 and α_R. Further support for this interpretation comes from simply scaling the charges by 0.88 as part of a nonadditive model for peptide conformational analysis, which leads to significant improvement in the dipeptide energies.[35] It is not clear

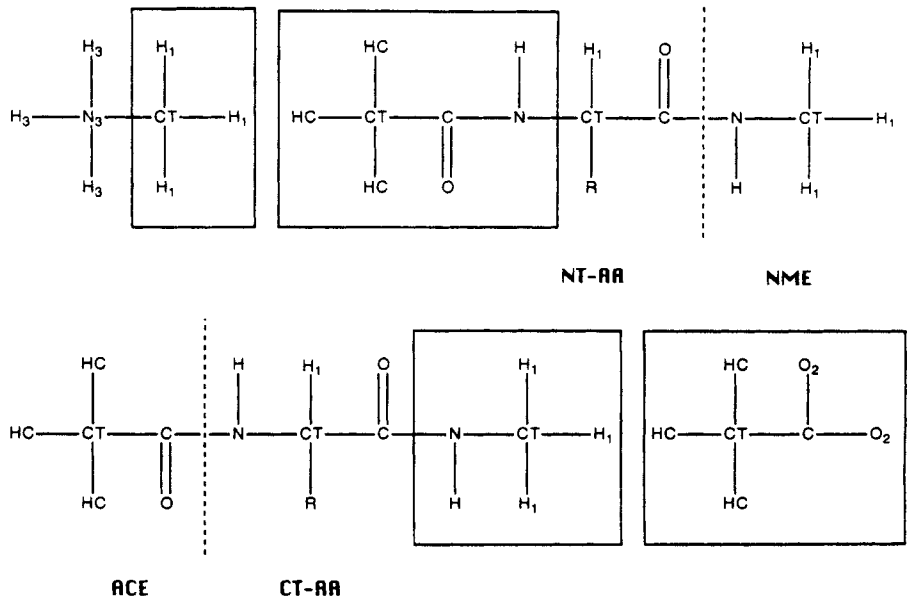

FIGURE 4. The scheme used for fitting the N- and C-terminal amino acids. Lagrange constraints define the appropriate blocking group as neutral. The appropriate charged group is spliced onto the other end of the amino acid by defining a Lagrange constraint which forces the sum of the charges on all atoms within the two boxed groups to sum to zero. The sums of the charges within each boxed group are then equal and opposite, so that the net charge on the charge end group has the same charge as the group that it replaces, ensuring the appropriate net integral charge on the resulting N- or C-terminal amino acid.

what the physical basis is for the fact that C_{7ax} is ~ 2 kcal/mol less stable than C_{7eq} at the quantum mechanical level, but only ~ 1 kcal/mol in the molecular mechanical models. We should emphasize that the dipeptide conformational energies can and have been improved through the addition of torsion parameters. The development of those torsion parameters is described in the next article in this series.[33]

We have presented charges for three forms of the amino acids: central residues and charged N- and C-terminal residues. One can analyze the charges on similar groups of atoms for a given residue type. For the N-terminal residues, the total charge on the ammonium group can be seen to vary from about 0.70 to 0.80. For the C-terminal residues, the total charge on the carboxylate group varies from about -0.79 to -0.85. The charges on the C_α's and H_α's exhibit the effects of induction caused by the adjacent charged group. Some inductive effect is also seen at the β position.

One can also compare charges for a particular amino acid in its three different forms. The charges on the atoms in the serine and valine sidechains are similar for all three versions of each residue. The charges on the asparagine sidechain, however, are less consistent. The charge on the gamma C_γ has a value of about 0.71 for the central and C-terminal residue but a value of 0.58 for the *N*-terminal residue. The charges on the remaining atoms in that sidechain are fairly consistent among the three residue types. We have only gone into detail about the five amino acids used for the neutral amino acid consensus fit; however, we have carried out charge calculations for all of the amino acids and will present the remaining data elsewhere.[33]

Conclusions

We have presented general methods for the derivation of charges for amino acids and nucleic acids for use in molecular mechanics calculations. Our strategy employs multiple-conformation fitting to reduce the conformational dependence of the charges and multiple-molecule fitting to derive consensus charges for certain common atoms, where appropriate. Furthermore, the charges were fit to the electrostatic potential of each molecule with restraints applied to the charges in order to attenuate the charges of statistically ill-determined atoms. The charges have been placed into a database for use in carrying out molecular mechanical simulations on nucleic acids and proteins with a new force field.[33]

Acknowledgments

These calculations were carried out using the facilities of the UCSF Computer Graphics Lab (T. Ferrin, P. I., supported by NIH-RR-1081) and the San Diego Supercomputer Center. Research support was generously provided by our industrial force field consortium partners Searle, Burroughs-Welcome, and Glaxo. We are also grateful for research support from the NIH: GM-29072 and CA-25644 (P. A. K.) and GM-08284 (W. D. C.). P. C. is supported by DARPA (MDA-91-Y-1013) and partially by the Polish Committee for Scientific Research, KBN grant no. CHEM/BST/502/23/95.

References

1. C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.*, **97**, 10269 (1993).
2. W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *J. Am. Chem. Soc.*, **115**, 9620 (1993).
3. S. R. Cox and D. E. Williams, *J. Comp. Chem.*, **2**, 304 (1981); T. R. Stouch and D. E. Williams, *J. Comp. Chem.*, **13**, 622 (1992).
4. C. A. Reynolds, J. W. Essex, and W. G. Richards, *J. Am. Chem. Soc.*, **114**, 9075 (1992).
5. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
6. S. Debolt and P. A. Kollman, *J. Am. Chem. Soc.*, **112**, 7515 (1990).
7. M. Aida, G. Corongiu, and E. Clementi, *Int. J. Quant. Chem.*, **42**, 1353 (1992).
8. U. C. Singh and P. A. Kollman, *J. Comp. Chem.*, **5**, 129 (1984).
9. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7**, 230 (1986).
10. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).
11. W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.*, **79**, 926 (1983).
12. H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunsteren, and J. Hermans, In *Intermolecular Forces*, B. Pullman, Ed., Reidel, Dordrecht, Holland, 1981, p. 331.
13. W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).
14. M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, J. J. P. Stewart, S. Topiol, and J. A. Pople, Gaussian 90, Revision J, Gaussian, Inc., Pittsburgh, PA, 1990.
15. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
16. D. A. Pearlman, D. A. Case, J. W. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman, *AMBER 4.0* (UCSF), University of California, San Francisco 1992.
17. D. Cremer and J. Pople, *J. Am. Chem. Soc.*, **97**, 1354 (1975). [The puckering parameter W reported here is 90° greater than the value calculated according to Cremer and Pople. The convention we use is widely accepted and was defined by C. Altona and M. Sundaralingham, *J. Am. Chem. Soc.*, **94**, 8205 (1972).]
18. (a) F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977); (b) E. E. Abola, F. C. Bernstein, S. H. Bryant, T. F. Koetzle, and J. Weng, In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*, F. H. Allen, G. Bergerhoff, and R. Sievers, Eds., Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, p. 107.
19. M. J. McGregor, S. A. Islam, and M. J. E. Sternberg, *J. Mol. Biol.*, **198**, 295 (1987).
20. T. E. Creighton, *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company, New York, 1984.
21. M. W. MacArthur and J. M. Thornton, *J. Mol. Biol.*, **218**, 397 (1991).
22. J. M. Thornton, *J. Mol. Biol.*, **151**, 261 (1981).
23. Y. Sun, D. Spellmeyer, D. Pearlman, and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 6798 (1992).
24. I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, **116**, 2493 (1994).
25. I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 9255 (1992).
26. I. R. Gould, W. D. Cornell, and I. H. Hillier, *J. Am. Chem. Soc.*, **116**, 9250 (1994).
27. T. Head-Gordon, M. Head-Gordon, M. J. Frisch, C. L. Brooks, and J. A. Pople, *J. Am. Chem. Soc.*, **113**, 5989 (1991).
28. H. Bohm and S. J. Brode, *J. Am. Chem. Soc.*, **113**, 7129 (1991).
29. L. Shafer, S. Q. Newton, M. Cao, A. Peeters, C. Vanalsenoy, K. Wolinski, and F. A. Momany, *J. Am. Chem. Soc.*, **115**, 272 (1993).
30. R. F. Frey, J. Coffin, S. Q. Newton, M. Ramek, V. K. W. Cheng, F. A. Momany, and L. Schafer, *J. Am. Chem. Soc.*, **114**, 5369 (1992).
31. T. H. J. Dunning, *Chem. Phys.*, **55**, 716 (1971).
32. J. J. Urban and G. R. J. Famini, *J. Comp. Chem.*, **14**, 353 (1993).
33. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, Th. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179 (1995).
34. A. E. Howard, P. Cieplak, and P. A. Kollman, in press. *J. Comp. Chem.* **16**, 243 (1995).
35. W. D. Cornell, J. W. Caldwell, and P. A. Kollman, manuscript in preparation.