

An Approach to Computing Electrostatic Charges for Molecules

U. Chandra Singh and Peter A. Kollman*

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

Received 10 February 1983; accepted 17 May 1983

We present an approach for deriving net atomic charges from *ab initio* quantum mechanical calculations using a least squares fit of the quantum mechanically calculated electrostatic potential to that of the partial charge model. Our computational approach is similar to those presented by Momany [*J. Phys. Chem.*, **82**, 592 (1978)], Smit, Derissen, and van Duijneveldt [*Mol. Phys.*, **37**, 521 (1979)], and Cox and Williams [*J. Comput. Chem.*, **2**, 304 (1981)], but differs in the approach to choosing the positions for evaluating the potential. In this article, we present applications to the molecules H₂O, CH₃OH, (CH₃)₂O, H₂CO, NH₃, (CH₃O)₂PO₂⁻, deoxyribose, ribose, adenine, 9-CH₃ adenine, thymine, 1-CH₃ thymine, guanine, 9-CH₃ guanine, cytosine, 1-CH₃ cytosine, uracil, and 1-CH₃ uracil. We also address the question of inclusion of "lone pairs," their location and charge.

INTRODUCTION

One of the fundamental problems in simulations on complex systems using simple analytical potential functions is how to accurately represent the molecular-charge distribution. The ability of molecular mechanics without the use of atomic charges to reproduce the structure and energies of complex hydrocarbons is impressive.¹ However, the pathway for similarly accurate simulations of polar and charged molecules is not so clear. Some approaches use an atom-centered multipole expansion,² and given a sufficiently large expansion, these can describe the molecular-charge distribution nearly as accurately as the wavefunction describing the electronic structure. However, for carrying out molecular mechanics, Monte Carlo, or molecular-dynamics simulations of large and complex molecules, a simpler description of the charge distribution is required. Most studies³ use only the monopoles of the atom-centered multipole expansion (net atomic charges), and this seems a reasonable compromise for complex systems. The question remains: How can one arrive at an appropriate set of site charges for such an expansion? The simplest procedure is to use Mulliken populations from quantum mechanical calculations, but these are based on a simplified model of describing the electron distribution and often yield rather different multipole moments for the mole-

cule than those calculated from the actual wavefunction.⁴ Such Mulliken populations are also often very basis-set dependent. A second approach is to use the molecular electrostatic potential evaluated at points in space around the molecule as a guide and to fit this to point-charge models. Important work in this area has been done by Scrocco and Tomasi,⁵ Smit, Derrison, and van Duijneveldt,⁶ and Cox and Williams.⁷ Our work follows closely the approach by Cox and Williams, with some improvements in methodology. We also present a number of applications that help answer the questions of "how, when, and whether" one should include lone-pair electron sites in the electrostatic model. Finally, we present partial charge models for fragments of nucleic acids and study Watson-Crick hydrogen bonding of the bases using these partial charges. When combined with appropriate Lennard-Jones dispersion and repulsion parameters, the partial charges calculated here are shown to yield H bond energies in very good agreement with gas phase enthalpies of association.

METHODS

One of us (U.C.S.) has combined a number of the best features of several *ab initio* quantum mechanical programs to create the program GAUSSIAN 80 UCSF.⁸ This program has as its

*To whom correspondence should be addressed.

"basis" Gaussian 80.⁹ To this is added: (1) the efficient self-consistent-field (SCF) clamping and extrapolation features from the program GAMESS (from the National Resource of Computational Chemistry¹⁰, (2) the molecular-properties package from GAUSSIAN 79,¹¹ (3) the electrostatic potential program from the Pisa (Italy) group of Tomasi et al.,¹² (4) the program by Connolly to calculate molecular surfaces,¹³ (5) a Levenberg-Marquardt nonlinear least squares program to fit the potential to an analytical model,¹⁴ and (6) the Morokuma energy-decomposition analysis.¹⁵ Thus this program contains many of the essential features for studying molecular-charge distributions and interactions.

Our goal is to find *simple*, analytical, partial-charge models that best reproduce the charge distribution of complex molecules. To find such charges, we first carry out quantum mechanical calculations on the molecules of interest. We then use the Connolly surface algorithm¹³ to calculate a shell or a number of shells of points at a specified distance from the molecule and

then calculate the electrostatic potential at these points. A Z-matrix procedure is used to input an initial partial-charge model in which charges or their location can be varied in any combination, subject to the absolute constraint of units of charge and any selected restraints on the calculated partial-charge dipole moment or quadrupole moment components of the molecule. A Levenberg-Marquardt nonlinear optimization procedure¹⁴ then fits the potential calculated from the partial charges to that calculated quantum mechanically. The algorithm is robust and seems to have a rather wide radius of convergence to the optimal-charge model.

We sought to establish an appropriate criterion for the radii employed in determining the surface points to be used for the electrostatic potential calculations. For H₂O, we examined the use of successive shells of various multiples of the van der Waals radii,¹³ ranging from 1.2 to 2.0 times the radius (smaller radii lead to positive potentials in the lone-pair region). The calculated point-charge models are rather insen-

Table I. Atom-centered point-charge models^a for H₂O.

Property	STO-3G ^b	Clem ^b	Basis Set 4-31G ^c	6-31G [*]	6-31G ^{**}	DZPP ^f
q _O ^g	-0.616(-0.614) ^h	-0.792	-0.938(-0.940) ^h	-0.812	-0.794(-0.786) ^h	-0.752
q _H ⁱ	0.308	0.396	0.469	0.404	0.397	0.376
μ^j	1.732	2.229	2.639	2.283	2.242	2.118
Q _a ^k	1.300	1.674	1.981	1.714	1.678	1.590
Q _b	-0.059	-0.076	-0.090	-0.078	-0.076	-0.073
Q _c	-1.241	-1.598	-1.891	-1.636	-1.602	-1.517
σ^l	0.51	0.57	1.08	1.81	1.54	2.15
s	4.1	3.9	5.6	10.8	10.5	15.3

^aReference 16.

^bReference 17.

^cReference 18.

^dReference 19.

^eReference 19.

^fReference 20.

^gCharge on oxygen.

^hCharge on oxygen found by Cox and Williams (ref. 7).

ⁱCharge on hydrogen.

^jDipole moment from charge model in debyes (the experimental value is shown in Table IV).

^kQuadrupole-moment components in buckinghams relative to center of mass and in principal axis system (experimental values shown in Table IV).

^lRoot mean square deviation between quantum mechanically calculated and point-charge electrostatic potential in kcal/mol.

^mPercent deviation between quantum mechanically calculated and point-charge electrostatic potentials.

sitive to which shell(s) is used within the range of 1.2–2.0 times the van der Waals radius, but the percent deviation between the potential calculated with the analytical model and that calculated quantum mechanically increases as the factor decreases. To be sure that no artifacts from being too close to the atoms entered into the calculations, we used shells of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radius in all of the remaining calculations, using symmetry where possible to reduce the number of surface points. The density of points in these surfaces depended on the size of the molecule, and varied 1–5 points/Å to yield typically 200–300 points/molecule.

Table II. Effect of shells on the point charges.^a

Property ^b	Shells range (van der Waals scale factor) ^c			
	1.2–1.8	1.4–2.0	1.6–2.2	1.8–2.4
q_0	-0.794	-0.794	-0.788	-0.786
q_H	0.397	0.397	0.394	0.393
ν	2.235	2.235	2.217	2.212
Q_a	1.678	1.678	1.664	1.660
Q_b	-0.077	-0.076	-0.076	-0.076
Q_c	-1.601	-1.602	-1.588	-1.584
σ	2.22	1.54	1.16	0.86
ϵ	11.9	10.5	9.8	8.9

^aUsing 6-31G** basis (ref. 6) on H₂O.

^bSee Table I for notation.

^cCreate shells using van der Waals radius of each atom times scale factors which range in increments of 0.2 Å between the tabulated limits.

Table III. 5-point charge model for H₂O.

Property ^a	Basis Set ^b				
	STO-3G	4-31G	6-31G	6-31G**	DZPP
q_0	-0.130	-0.700	-0.538	-0.544	-0.372
q_H	0.250	0.419	0.491	0.489	0.517
q_{LP}^c	-0.185	-0.068	-0.223	-0.217	-0.331
ν	1.732	2.629	2.231	2.189	2.048
Q_a	1.290	2.059	2.275	2.256	2.440
Q_b	0.016	-0.012	0.008	-0.021	-0.024
Q_c	-1.274	-2.047	-2.283	-2.235	-2.416
d^d	0.468	0.908	0.466	0.465	0.443
θ^e	103.73	106.12	70.85	69.28	68.04
σ	0.24	0.72	0.63	0.58	0.46
ϵ	1.8	3.2	3.7	3.5	3.3

^aSee Table I for notation.

^bSee Table I for notation: scale factor 1.4–2.0 times van der Waals radius.

^cCharge on lone pair.

^dDistance of lone pair from oxygen (in Å).

^eAngle in degrees that lone pairs make with HOH bisector (an angle of 109.5° corresponds to approximately “tetrahedral” lone pair with respect to the hydrogens as well as the other lone pair).

Table IV. Constraint values and weights.

Property ^a Molecule	H ₂ O	CH ₃ OH	(CH ₃) ₂ O	CH ₂ O
ux^b	-	1.659(250) ^d	-	-
uy^b	-	-	-	-
uz^b	1.85(440) ^c	-1.154 ^d	1.300(250) ^e	2.300(250) ^f
Qa^g	2.639(20) ^h	-	3.30(20) ⁱ	0.040(10) ^j
Qb^g	-0.050 (20)	-	-2.20 (20)	-0.051 (10)
Qc^g	-2.589 (20)	-	-1.20 (20)	0.011 (10)

^aThe principal axis of the molecules is the Z axis: H₂O, (CH₃)₂O and, CH₂O are in the XZ plane with each molecule at its experimental geometry. Methanol has its C—O bond along the X axis. The geometries used were taken from the Chem. Soc. Special Publications, Nos. 11 and 18, 1958 and 1965, except that ideal tetrahedral CH₃ groups were assumed with $R(C-H) = 1.10 \text{ \AA}$.

^bDipole-moment component in debyes, with constraint weight in parentheses. Experimental values of the dipole moment was used for constraint.

^cExperimental value from ref. 21.

^dUsed calculated components found with 6-31G** basis set.

^eExperimental value from ref. 22.

^fExperimental value from ref. 22.

^gQuadrupole moment in buckinghams with constraint weight in parentheses. The experimental value of the quadrupole moments were used for constraint.

^hExperimental value from ref. 23.

ⁱExperimental value from ref. 22.

^jExperimental value from ref. 22.

RESULTS

H₂O

For H₂O we addressed a number of questions: What is the basis-set dependence of the derived partial charges? How does this depend on the inclusion of lone pairs and/or constraints? How do a 4- and 5-point model compare for H₂O?

In Table I we report the values found for the point-charge fit (atom-centered point-charge model) for H₂O with various basis sets. The calculated charges are compared with those determined by Cox and Williams⁷ and the agreement is very good. Our root mean square fits to the potential are not as low as those of Cox and Williams because we have chosen points closer to the molecule to use in our fitting procedure (they use the van der Waals radius + 1.2 Å). Thus we reinforce their conclusion that, whereas the point-charge models are rather insensitive to the location of the points at which to evaluate the electrostatic potential, the quality of the fit worsens if the points are chosen closer to the molecule and improves if the points are chosen farther away. Our range of points of 1.4–2.0 times the van der Waals radius gives electro-

Table V. 5-point charge H₂O model with constraints.^a

Property ^a	STO-3G	4-31G	Basis Set		DZPP
			6-31G*	6-31G**	
q ₀	-0.433	-0.628	-0.453	-0.453	-0.341
q _H	0.289	0.506	0.508	0.506	0.522
q _{LP}	-0.072	-0.192	-0.282	-0.279	-0.352
μ	1.777	1.986	1.911	1.903	1.901
Q _a	+1.403	2.430	2.450	2.433	2.516
Q _b	0.048	0.235	-0.105	-0.113	-0.066
Q _c	-1.451	-2.195	-2.344	-2.320	-2.460
d	0.710	0.615	0.526	0.517	0.462
ε	100.97	62.41	65.66	65.38	66.55
σ	0.56	4.45	2.35	2.12	1.04
s	4.5	23.6	14.1	12.9	7.4

^aSee Table IV for values chosen for constraint and Tables I and II for other notation.

Table VI. 4-point charge H₂O model.

Property ^a	STO-3G	4-31G	Basis Set ^b		DZPP
			6-31G*	6-31G**	
q _H	0.401	0.548	0.539	0.537	0.543
q _{LP}	-0.401	-0.548	-0.539	-0.537	-0.543
μ	1.777	2.308	2.073	2.048	1.901
Q _a	1.782	2.358	2.420	2.406	2.508
Q _b	0.015	-0.082	-0.037	-0.051	-0.049
Q _c	-1.797	-2.276	-2.383	-2.355	-2.459
d	0.248	0.188	0.283	0.279	0.332
ε	72.14	61.35	66.42	65.49	65.92
σ	1.44	2.45	1.35	1.24	1.05
s	11.4	12.7	8.1	7.5	7.4

^aSee Tables I and III for notation.

^bSee Table I for notation.

Table VII. Partial-charge model for dimethyl ether.

Property ^a	STO-3G	Basis Set ^b		6-31G*	
		4-31G	6-31G		
q ₀ ^c	(-0.2243)	-0.2908	(-0.5399)	-0.4187	(-0.4038) -0.3453
q _C ^d	(-0.1025)	0.1102	(0.1423)	0.1192	(0.0095) 0.0253
q _H ^e	(0.6679)	0.0528	(0.0567)	0.0392	(0.0799) 0.0600
q _H ^f	(0.0733)	0.0408	(0.0355)	0.0254	(0.0562) 0.0437
μ ^g	(1.3735)	1.3276	(2.2230)	1.7281	(1.7968) 1.5337
Q _a ^h	(2.5524)	3.0545	(5.0853)	3.8433	(4.4203) 3.6163
Q _b ^h	(-1.4362)	-1.7583	(-2.9991)	-2.2787	(-2.5267) -2.0843
Q _c ^h	(-1.1162)	-1.2962	(-2.0862)	-1.5646	(-1.8936) -1.5319
σ ⁱ	(1.21)	1.32	(1.45)	2.89	(1.17) 1.83
s	(18.2)	19.8	(13.1)	26.1	(12.8) 20.1

^aMolecular property; values in parentheses without dipole and quadrupole constants described in Table IV.

^bBasis set; see Table I for notation.

^cCharge on oxygen.

^dCharge of carbon.

^eCharge on two hydrogens in XY plane.

^fCharge on the four out-of-plane hydrogens.

^gDipole moment of molecule in debyes.

^hQuadrupole-moment components of molecule in buckinghams in principal axis system.

ⁱrms deviation of quadrupole moment and point-charge model.

^jRelative percentage of rms deviation.

Table VIII. Charge models for dimethyl ether with lone pairs.

Property ^a	STO-3G	Basis Set ^b	
		4-31G	6-31G*
q ₀	0.8464 (1.101) ^c	-0.0546 (2.999) ^c	0.1259 (2.239) ^c
q _C	-0.3529 (-0.307)	-0.1573 (-0.151)	-0.1874 (-0.203)
q _H	0.1200 (0.125)	0.1178 (0.088)	0.1167 (0.103)
q _H ^d	0.1051 (0.091)	0.0890 (0.033)	0.0828 (0.055)
q _{LP}	-0.4005 (-0.550)	-0.1112 (-1.5030)	-0.1580 (-1.129)
μ	1.3593 (1.3179)	2.2126 (1.7111)	1.7856 (1.5205)
Q _a	2.4238 (2.998)	4.9243 (3.750)	4.2915 (3.544)
Q _b	-1.7776 (-2.093)	-3.3991 (-2.5407)	-2.7776 (-2.313)
Q _c	-0.6463 (-0.904)	-1.5252 (-1.203)	-1.5139 (-1.231)
Q _g	0.428 (0.326)	0.906 (0.329) ^f	0.739 (0.329)
σ	109.85 (108.75)	116.0 (102.60)	111.82 (102.97)
s	0.304 (0.59)	0.569 (2.49)	5.65 (1.46)
b	4.5 (9.0)	5.1 (22.5)	6.1 (15.9)

^aSee Table VII for notation.

^bSee Table I for notation.

^cModels derived with constraints (Table IV) in parentheses.

^dCharge on lone pairs.

^eDistance from lone pairs from oxygen.

^fDistance fixed as in 6-31G* model; otherwise lone pair will collapse very near oxygen, with very large and opposite charges on lone pair and oxygen.

^gAngle that lone pairs make with COC bisector.

^hrms deviation of quantum mechanical and model potential in kcal/mol.

ⁱRelative percent-rms deviation.

Table IX. Charge models for methanol.

Property	Model ^a	
	No Lone Pairs	With Lone Pairs
q _C ^b	0.149	-0.176
q _{O,d}	-0.656	-
q _{H,e}	-0.001	0.054
q _{H,f}	0.042	0.101
q _{LP,g}	0.424	0.386
μ _n	-	-0.233
μ _z	1.657	1.639
μ _x	-1.144	-1.149
Q _a	0.744	1.200
Q _b	0.486	-0.019
Q _c	-1.228	-1.219
Q _k	1.40	0.638
d ^l	13.4	6.1
ε _m	-	0.619
ε ₂	-	119.4
ε ₁	-	87.3

^a6-31G** basis set used (ref. 19).

^bCharge on carbon.

^cCharge on oxygen.

^dCharge on hydrogen in XZ plane.

^eCharge on hydrogens in and out of XZ plane.

^fCharge on hydroxyl hydrogen.

^gCharge on lone pair.

^hDipole-moment components in debyes.

ⁱQuadrupole-moment component in buckinghams.

^jrms deviation between quantum mechanical and model electrostatic potentials.

^kRelative percent-rms deviation.

^lLone-pair-oxygen distance in Å.

^mAngle lone pairs makes with C—O bond; one lone pair above and one below COH plane.

ⁿAngle lone pairs make with O—H bond; one lone pair above and one below COH plane.

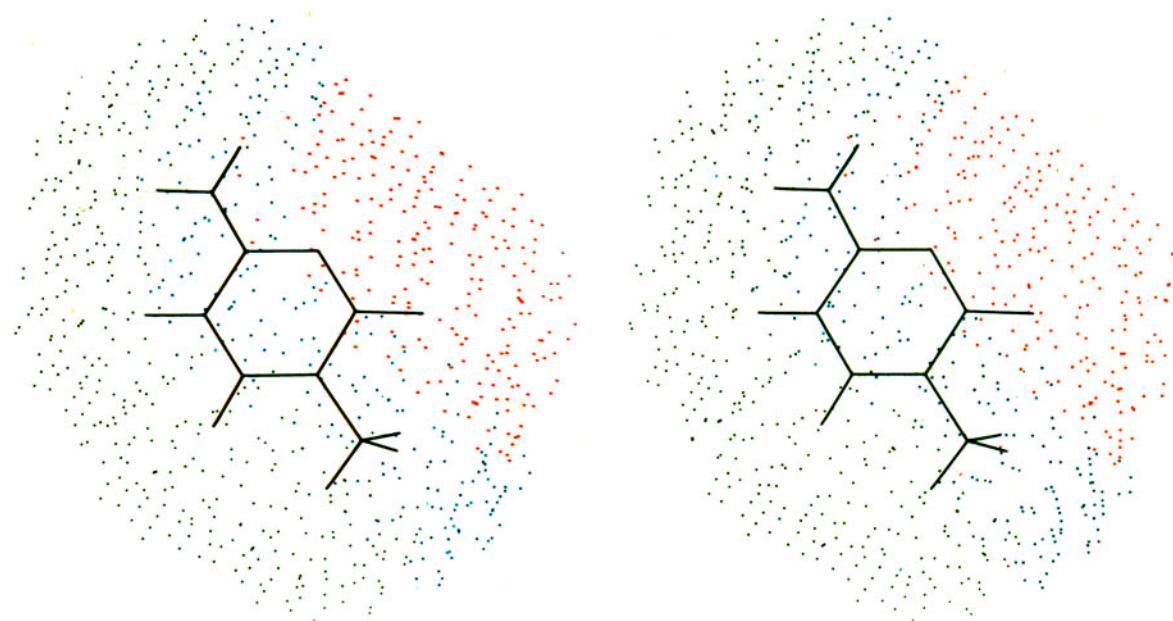


Figure 1. Quantum mechanical potential for 1-CH₃ cytosine. Potential points > 5 kcal/mol are green; potential points between -5 and +5 kcal/mol are blue, and potential values < -5 kcal/mole are red.

Table X. Formaldehyde point-charge models.^a

property	atom centered		Model		5 point ^c	
	no constraint	constraint ^d	6 point ^b	no constraint	constraint	no constraint
q ₀ ^e	-0.463	-0.435	0.152	-0.143	-	-
q _C ^f	0.421	0.421	0.250	0.283	0.262	0.262
q _H ^g	0.021	0.007	0.057	0.035	0.055	0.042
q _{LP} ^h			-0.258	-0.105	-0.186	-0.173
μ^i	2.79	2.56	2.80	2.50	2.80	2.56
Q _a ^j	-0.190	-0.068	-0.023	0.101	-0.025	0.067
Q _b ^j	0.064	0.020	-0.030	-0.099	-0.032	-0.075
Q _c ^j	0.126	0.048	0.053	-0.002	0.057	0.008
σ^k	1.00	1.62	0.50	1.67	0.50	1.38
s ^l	6.5	10.6	3.3	10.9	3.3	9.0
d ^m	-	-	0.459	0.696	0.541	0.547
θ^n	-	-	109.2	113.2	111.2	110.8

^a6-31G** basis set.

^bCharge on all atoms and two lone pairs.

^cNo charge on oxygen and two lone pairs.

^dSee Table IV for constraint values and weights.

^eCharge on oxygen.

^fCharge on carbon.

^gCharge on hydrogen.

^hCharge on lone pair.

ⁱDipole moment in debyes.

^jQuadrupole moment in buckinghams.

^krms deviation of quantum mechanically calculated and model electrostatic potentials.

^lPercent-rms deviation between potentials.

^mDistance of lone pair from oxygen.

ⁿC—O-lone-pair angle.

Table XI. Energies of water dimer and H₂CO ··· HOH (kcal/mol).

Water Dimer			H ₂ CO...HOH			
Geometry ^a	Atom Centered ^b	5 point ^c	Geometry ^d	ab initio ^e	Atom Centered ^f	Extended ^g
Linear, θ=0	-4.7	-4.0	linear, θ=0	-3.5	-3.3	-2.8
θ=30	-4.9 (-5.2)	-4.4	θ=45	-	-3.6	-4.1
θ=40	-4.5	-4.8 (-5.2)	θ=60	-4.8	-3.7 (-4.1)	-4.4 (-4.7)
θ=75	-4.1	-4.8	θ=90	-3.7	-2.2	-2.1
θ=30	-3.9	-3.7	θ=90, φ=90	-1.4	-2.6	-0.4
θ=30, φ=90	-4.3	-3.8				
bifurcated	[-3.8]	[-3.7]				
cyclic	[-3.6]	[-3.5]				

^aSee ref. 25 for description of geometry and results of *ab initio* calculations.^bUsing atom-centered charges from DZPP basis set. Value in parenthesis is completed optimized geometry.

Oxygen van der Waals parameter with well depth of 0.12 kcal/mol and radius of 1.8 Å was used.

^c5-point charge model; DZPP basis set. Value in parenthesis is after complete optimization of geometry.

Oxygen van der Waals well depth of 0.12 kcal/mol and radius of 2.0 Å was used.

^dO—H ··· O linear, θ = angle between C—O and H—O vectors, with external water hydrogen trans to C—O ··· O. φ = 90 means H—O bond approaching in π direction toward oxygen.^eResults of *ab initio* calculations using 6-31G** basis set.^fUsing atom-centered model 6-31G** for H₂CO; DZPP for H₂O; oxygen van der Waals parameters as in atom-centered water dimer; values in parentheses are from complete geometry optimization.^gUsing 6-point 6-31G** charge model for H₂CO; 5-point DZPP for H₂O; van der Waals parameters as used in extended model for (H₂O)₂.Table XII. Charges for 1-CH₃ cytosine.

Property	STO-3G-AA ^a	Basis Set/Model		
		STO-3G-UA ^b	CLEM-AA ^c	CLEM-UA ^d
charges ^e				
N1	-0.187	-0.572	-0.053	-0.723
C2	0.859	0.939	0.910	1.061
N3	-0.860	-0.791	-0.948	-0.848
C4	0.935	0.629	1.126	0.640
C5	-0.576	-0.230	-0.727	-0.220
C6	0.185	0.377	0.043	0.460
O2	-0.508	-0.518	-0.529	-0.543
N4	-0.834	-0.743	-1.079	-0.923
C1M	-0.289	0.235	-0.514	0.286
HN4A	0.351	0.338	0.428	0.405
HN4B	0.329	0.335	0.396	0.406
H5	0.153	-	0.237	-
H6	0.098	-	0.195	-
HC1A	0.116	-	0.167	-
HC1B	0.115	-	0.174	-
HC1C	0.115	-	0.174	-
ux ^f	3.517	3.509	3.293	3.272
uy ^f	-4.181	-4.056	-4.872	-4.667
Q ^g	13.336	13.003	13.504	13.663
Q ^g	-8.548	-8.559	-7.161	-7.942
Q _{RR} ^g	-4.788	-4.444	-6.343	-5.721
σ ^h	0.56	1.64 (4.0) ^j	0.90	2.74
s ⁱ	3.7	11.0 (27.0) ^j	5.7	17.2

^aSTO-3G basis set; all-atom model, i.e. all atoms included in the fit of the model potential to the quantum mechanically calculated one.^bSTO-3G basis set, united-atom model; hydrogen on carbon not used in the fit of the model potential to the quantum mechanical one.^cClementi basis set (ref. 17), all atom model.^dClementi basis set (ref. 17), united-atom model.^eAtom charges found from fit to potential; see ref. 25 for numbering convention; geometries taken from ref. 30.^fDipole-moment components in debyes; molecule in XY plane N1—C2 bond along X axis; geometry from ref. 30.^gQuadrupole-moment components in buckinghams in principal axis system.^hRoot mean square deviation between model and quantum mechanical in kcal/mol.ⁱPercent-rms deviation.^jσ and s values found by taking AA charges and summing C—H hydrogen and carbon charges and placing on carbon to calculate potential.

static potentials of sufficient range (-20 to +20 kcal/mol, where 1 kcal/mol = 4.18 KJ/mol) to describe rather well typical interaction distances and energies of strong and weak interactions, with the exception of proton affinities.

For this reason we believe that our use of Connolly molecular surfaces to determine the electrostatic-potential-point choices is a useful and general one. It leads to partial charges in essential agreement with that from cubic grid algorithm, but more easily ensures that the collection of points chosen for potential evaluation contains approximately equal contributions from all regions of the molecule.

We further explore the dependence of charge model on choice of points for potential evaluation in Table II. As one can see, the point-charge model derived is not very dependent on the "range" of the shells of points chosen.

We next examine the ability of 5-point charge models for H₂O to fit the electrostatic potential. These models contain a partial charge on each hydrogen, one on the oxygen and one in each lone-pair direction with variable distance *d* from the oxygen and angle (relative to the HOH bisector line). Table III contains the results of fitting the quantum mechanical electrostatic potentials to such models. Not surprisingly, the fit of the partial-charge model to the quantum mechanical potential is substantially improved (note σ and s values) over the 3-point model. Most intriguing are the differences in the lone-pair angle θ found in the various models. The cruder basis set STO-

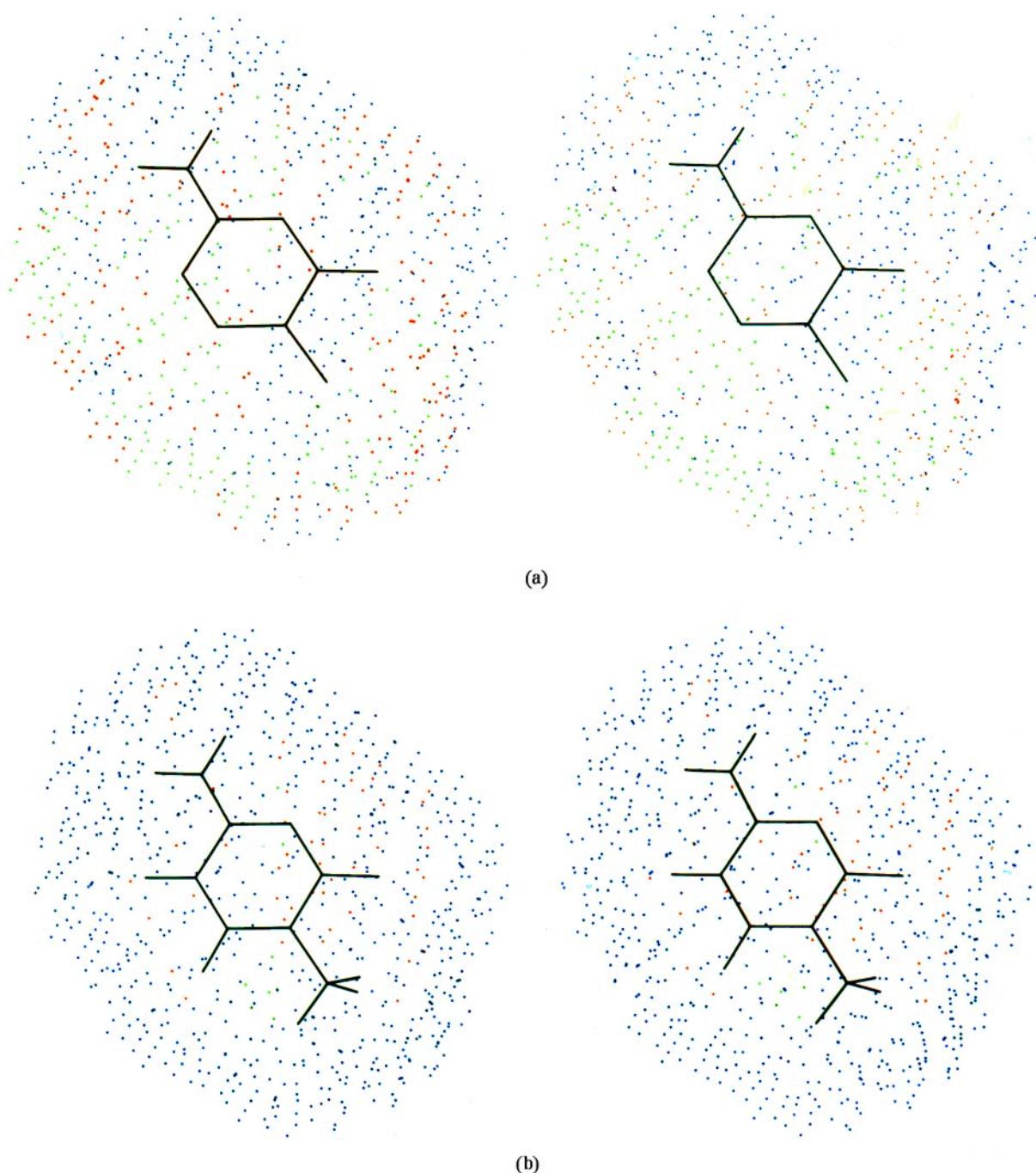


Figure 2. (a) Difference between united-atom model and quantum mechanical potential for 1-CH₃ cytosine. Absolute differences < 1 kcal/mol are blue; absolute differences between 1 and 2 kcal/mol are red, and differences > 2 kcal/mol are green. (b) Difference between all-atom model and quantum mechanical potential. Color code same as for (a).

3G and 4-31G calculations find θ near the tetrahedral value of 109.5°, whereas the better basis sets 6-31G*, 6-31G**, and double zeta plus polarization (DZPP) all find θ between 50° and 70°. The reason for this difference can be traced to the molecular electric moments calculated with the various basis sets. The sto-3G basis calculated a quadrupole moment far less than experiment, whereas the 4-31G basis gives the poorest agreement with the experimental dipole moment.

The other three basis sets yield quadrupole moments in reasonable agreement with experiment and dipole moments that are somewhat high but closer to the experimental value than 4-31G. To derive a 5-point charge model that simultaneously describes well the water molecules' dipole and quadrupole moment requires lone pairs that are *inverted* and on the same side of oxygen as are the hydrogens.

Another option in the fitting program we use

Table XIII. Charges for cytosine.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM-UA
Charges				
N1	-0.642	-0.778	-0.769	-1.060
C2	0.982	0.969	1.121	1.126
N3	-0.892	-0.782	-1.002	-0.839
C4	1.014	0.672	1.252	0.720
C5	-0.623	-0.241	-0.796	-0.237
C6	0.351	0.412	0.327	0.525
O2	-0.538	-0.536	-0.575	-0.573
N4	-0.880	-0.775	-1.155	-0.983
H1	0.310	0.370	0.367	0.481
HN4A	0.361	0.343	0.466	0.416
HN4B	0.342	0.345	0.419	0.424
H5	0.152	-	0.234	-
H6	0.063	-	0.131	-
μ_x	3.873	3.847	3.869	3.827
μ_y	-4.138	-4.048	-4.770	-4.615
Q_a	4.203	5.043	4.512	5.555
Q_b	-0.033	-1.138	1.492	-0.158
Q_c	-4.170	-3.905	-6.005	-5.397
σ	0.56	1.42	0.80	2.33
s	3.4	8.8	4.6	13.3

^aSee Table XII for rotation.

is the ability to constrain the partial-charge model to yield specified values of the dipole or quadrupole moment components; the result of putting on such constraints in the 5-point H₂O model is described in Table IV, with constraint values and weights used in Table V.

Interestingly enough, all models except STO-3G now have inverted lone pairs, and because the constant has significantly altered the lone-pair position in the 4-31G model, the percent-rms deviation for this model is the largest of all. However, this 4-31G model is in much better agreement with the models based on the more extended basis sets than it was without the constraints (Table III), so that this constrained optimization method may be a useful approach in cases where there is some experimental data on the molecular multipole moments, but only a modest (4-31G or double-zeta basis) level of theory can be applied to them.

A final choice of models considered here is a 4-point model, with only hydrogen and lone-pair partial charges. Such models are described in Table VI. Comparing the tables, one sees a significant worsening of the quality of fit of the 4-compared with the 5-point model.

(CH₃)₂O AND CH₃OH

We next extended our studies to the molecules (CH₃)₂O and CH₃OH to see how the point-charge models for these molecules differed from those for H₂O. In the case of (CH₃)₂O, we developed point-charge models using STO-3G, 4-31G, and 6-31G* basis sets. The point-charge models for (CH₃)₂O without lone pairs are presented in Table VII, with and without dipole constraints. As one can see, with dipole constraints the STO-3G

basis gives results that do not differ greatly from those with the most extensive 6-31G* basis.

The addition of "classical" lone pairs (Table VIII) leads to a significant improvement in the fit of the charge model to the electrostatic potential. The lone pair remains in the classical direction, independent of basis set. However, the charge on the oxygen varies widely depending on basis set and whether a dipolar constraint has been applied.

Methanol has lower symmetry than H₂O and (CH₃)₂O, so it is of interest to see what charge models can be derived for this molecule. Since Cox and Williams have derived atom-centered charge models for this molecule, Table IX contains only the results of the 6-31G** models with and without lone pairs.

When lone pairs are added to methanol, not surprisingly, the charge models have a much better fit to the quantum mechanical potential. The lone pairs are in the classical direction and are in a similar location for all basis sets. However, such models only converge to reasonable values of lone-pair charge and distance when there is no charge on oxygen. With a charge on oxygen, the charge on oxygen gets large and positive and the lone pair very negative, with lone-pair "collapse" onto the atom center.

H₂CO

We next sought to extend our studies to formaldehyde, H₂CO, in order to determine lone-pair-containing charge models for this molecule, as well as to see if the lone-pair inversion found for H₂O with the extended basis sets occurred in

Table XIV. Charges for 1-CH₃ uracil.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM-UA
Charges				
N1	-0.159	-0.570	-0.012	-0.702
C2	0.775	0.896	0.845	1.065
N3	-0.768	-0.758	-0.957	-0.970
C4	0.834	0.572	1.018	0.628
C5	-0.529	-0.205	-0.705	-0.237
C6	0.160	0.367	0.065	0.471
O2	-0.472	-0.494	-0.504	-0.540
O4	-0.474	-0.394	-0.533	-0.408
C1M	-0.329	0.239	-0.599	0.264
H3	0.334	0.347	0.403	0.429
H5	0.146	-	0.224	-
H6	0.098	-	0.180	-
HC1A	0.133	-	0.195	-
HC1B	0.126	-	0.190	-
HC1C	0.126	-	0.190	-
μ_x	-1.628	-1.667	-1.671	-1.738
μ_y	-3.264	-3.130	-3.724	-3.510
Q_a	-8.807	-8.460	-9.968	-9.2364
Q_b	5.636	5.146	7.593	6.4667
Q_c	3.171	3.314	2.375	2.7697
σ	0.56	1.70	0.90	2.70
s	5.4	16.3	7.8	23.2

^aSee Table XII for notations.

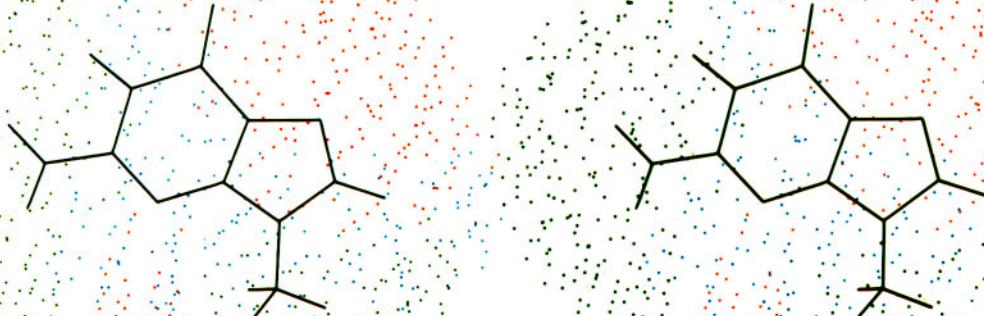


Figure 3. Quantum mechanical potential for 9-CH₃ guanine. See Figure 1 for color code.

H₂CO as well. The partial atom-centered charges we derive for this molecule differ only slightly from those found in the calculations of Cox and Williams.⁷ For example, with a 6-31G ** basis set, we found partial charges of -0.463, 0.421, and 0.021 on oxygen, carbon, and the hydrogens, compared with values of -0.459, 0.423, and 0.018 found previously. When we carry out these calculations adding a constraint to force the charge model to have the experimental dipole and quadrupole moments, there are relatively small (~ 10%) changes change in the charges on the atoms. With the STO-3G basis, the oxygen charges change from -0.28 to -0.33, there is relatively little change in the carbon charges, and the hydrogen charges change from -0.02 to 0.00. In addition, as was found for H₂O, putting on the dipole and quadrupole constraints enables one to force the more limited basis-set calculations to have charges more closely approximating that found with more extended basis sets.

We first considered a model for the charge distribution in H₂CO in which lone pairs were placed in an *sp*² direction around the oxygen (Table X). Interestingly, in the case of STO-3G and 4-31G, the lone pairs nearly collapsed onto the oxygens, with large and compensating charges in the oxygen and lone-pair charges. However, the model in which the lone pairs were held fixed at the distance found in the 6-31G ** calculation had a fit to the quantum mechanical electro-

Table XV. Charges for uracil.^a

Property	Basis Set/Model			
	STO-3G-AA	STO-3G-UA	CLEM-AA	
Charges				
N1	-0.573	-0.762	-0.702	-1.040
C2	0.836	0.883	0.973	1.075
N3	-0.713	-0.675	0.882	-0.851
C4	0.831	0.540	1.019	0.602
C5	-0.518	-0.181	-0.680	-0.205
C6	0.276	0.373	0.288	0.498
O2	-0.489	-0.495	-0.531	-0.547
O4	-0.474	-0.385	-0.534	-0.402
H1	0.304	0.378	0.354	0.478
H5	0.137	-	0.202	-
H6	0.069	-	0.123	-
<i>u</i> _x	-1.288	-1.347	-1.214	-1.303
<i>u</i> _y	-3.112	-2.987	-3.520	-3.323
<i>Q</i> _a	-12.142	-11.846	-12.939	-12.507
<i>Q</i> _b	9.654	9.095	11.583	10.675
<i>Q</i> _c	2.488	2.752	1.356	1.832
<i>σ</i>	0.52	1.36	0.85	2.14
<i>s</i>	4.7	12.3	7.1	17.8

^aSee Table XII for notation.

Table XVI. Charges for 1-CH₃ thymine.^a

Property	Basis Set/Model			
	STO-3G-AA	STO-3G-UA	CLEM-AA	
Charges				
N1	-0.233	-0.739	-0.066	-0.942
C1	0.849	1.113	0.941	1.416
N3	-0.851	-1.012	-1.092	-1.399
C4	0.809	0.980	0.918	1.248
C5	-0.176	-0.595	-0.072	-0.826
C6	0.034	0.551	-0.181	0.772
O2	-0.488	0.529	-0.529	-0.600
O4	-0.464	-0.472	-0.503	-0.523
C5H	-0.382	0.097	-0.627	0.124
HCSA	0.119	-	0.184	-
HCSB	0.111	-	0.170	-
HCSC	0.111	-	0.170	-
C1M	-0.251	0.236	-0.467	0.250
HClA	0.118	-	0.172	-
HClB	0.103	-	0.145	-
HClC	0.103	-	0.145	-
H3	0.355	0.370	0.448	0.479
H6	0.134	-	0.245	-
<i>u</i> _x	-0.981	-0.975	-0.947	-0.944
<i>u</i> _y	-3.238	-3.175	-3.680	-3.581
<i>Q</i> _a	-10.251	-10.368	-11.728	-11.681
<i>Q</i> _b	7.344	7.652	9.253	9.382
<i>Q</i> _c	2.907	2.716	2.475	2.299
<i>σ</i>	0.58	1.23	1.04	2.08
<i>s</i>	6.1	12.9	9.7	19.4

^aSee Table XIII for notation.

Table XVII. Charges for thymine.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM-UA
Charges				
N1	-0.624	-0.858	-0.797	-1.208
C2	0.901	1.034	1.070	1.309
N3	-0.788	-0.862	-0.989	-1.131
C4	0.787	0.858	0.898	1.036
C5	-0.179	-0.474	-0.095	-0.595
C6	0.154	0.485	0.100	0.677
O2	-0.504	-0.523	-0.555	-0.589
O4	-0.458	-0.451	-0.497	-0.491
CSM	-0.370	0.084	-0.615	0.096
HCSA	0.117	-	0.181	-
HCSB	0.107	-	0.165	-
RCSC	0.107	-	0.165	-
H1	0.317	0.367	0.383	0.472
H3	0.335	0.341	0.410	0.432
H6	0.100	-	0.173	-
μ_x	-0.720	-0.726	-0.554	-0.561
ν_y	-3.122	-3.071	-3.495	-3.411
σ_a	-11.337	-11.489	-12.339	-12.399
σ_b	8.739	8.994	10.606	10.753
σ_c	2.598	2.494	1.733	1.646
σ	0.54	0.89	0.83	1.43
s	5.3	8.9	7.6	13.2

^aSee Table XII for notation.

static potential less than 1% worse than that of the collapsed lone pairs. In the case of the 6-31G and 6-31G** basis sets, we found no such collapse, but there still was significant sensitivity of the charges to lone-pair placement. In contrast to the charges derived without lone pairs, in which the 6-31G* and 6-31G** basis sets results were nearly identical, there are significant differences between the charges derived from these two basis sets when lone pairs were included. These differences are reduced when dipole and quadrupole constraints are applied to the calculation. When one turns to 5-point charge models (no charge on oxygen) (Table X), we find physically reasonable charges, which, for the 6-31G* and 6-31G** basis sets, give essentially identical quality fits to the 6-point charge model. However, with STO-3G and 4-31G basis sets, the extra flexibility in the charge model does improve the fit. Only in the case of STO-3G do the charges on the carbon and hydrogens differ greatly in the two models. In none of these models did the lone pairs "invert"; in all models these lone pair charges remained in intuitively reasonable positions. In Table X, we present the results of the above model calculations using the 6-31G** basis set.

Given that one has these charge models, can they be used to study intermolecular interactions?

To carry out such studies, one includes repulsion and dispersion parameters in addition to the electrostatic models. One can evaluate the energy at specific geometries, as well as carrying out energy refinement. In these studies, we used the AMBER molecule mechanics program.²⁴ We

examined the hydrogen bond directionality in $(H_2O)_2$ and $H_2CO \cdots HOH$ with both atom-centered and lone-pair models and the results are presented in Table XI. For $(H_2O)_2$ both models correctly reproduce the fact that the trans linear dimer is the lowest energy structure²⁶ and the bifurcated structure is not a local minimum and is ~ 1.5 kcal/mol higher in energy than the linear structure.²⁷ The lone-pair model has more "lone-pair directionality", i.e. the lowest energy occurs at a larger θ , but both models suggest that the correct orientation of the external hydrogen on the proton-donor molecule is at least as important as a lone-pair approach in that the $\theta = -30$ (indeed all $\theta < 0$) and $\theta = 30$, $\phi = 90$ geometries are higher in energy than any with $\theta < 0$. The importance of the external hydrogen on H bond directionality has also been found in *ab initio* calculations as well.²⁸

In the case of $H_2CO \cdots HOH$, we examined the angular dependence of the interaction energy for HOH approach in the H_2CO plane, as well as the " π " approach of the OH bond toward the carbonyl oxygen. As one can see, both models reproduce the lone-pair directionality and preference of H—O approach in the plane, although the model including lone pairs has a slightly larger minimum energy θ . The biggest difference between the models comes in the calculated relative stabilities of π and σ H bonds, where the atom-centered model finds a much smaller energy difference between H bonds than the extended model. *Ab initio* calculations using the 6-31G** basis set²⁹ on a number of complexes

Table XVIII. Charges for 9-CH₃ guanine.^a

Property	STO-3C-AA	Basis Set/Model		
		STO-3C-UA	CLEM-AA	CLEM-UA
Charges				
N1	-0.729	-0.746	-0.881	-0.952
C2	0.871	0.842	0.988	0.972
N3	-0.709	-0.702	-0.726	-0.712
C4	0.391	0.490	0.333	0.500
C5	-0.060	-0.088	0.003	-0.088
C6	0.690	0.714	0.771	0.865
N7	-0.543	-0.575	-0.615	-0.703
C8	0.266	0.428	0.157	0.570
N9	-0.022	-0.360	0.128	-0.466
N2	-0.778	-0.758	-0.998	-0.979
O6	-0.458	-0.459	-0.492	-0.500
H1	0.336	0.340	0.420	0.432
H8	0.046	-	0.132	-
C9N	-0.459	0.216	-0.747	0.263
HN2A	0.339	0.333	0.414	0.405
HN2B	0.325	0.324	0.392	0.394
HC9A	0.163	-	0.239	-
HC9B	0.164	-	0.241	-
HC9C	0.164	-	0.241	-
μ_x	-3.416	-3.471	-3.187	-3.293
ν_y	5.357	5.294	5.784	5.699
σ_a	25.931	26.062	29.991	29.566
σ_b	-19.728	-19.984	-21.831	-21.784
σ_c	-6.203	-6.078	-8.160	-7.782
σ	0.89	1.22	1.22	1.90
s	5.6	7.7	7.1	11.2

^aSee Table XII for notation.

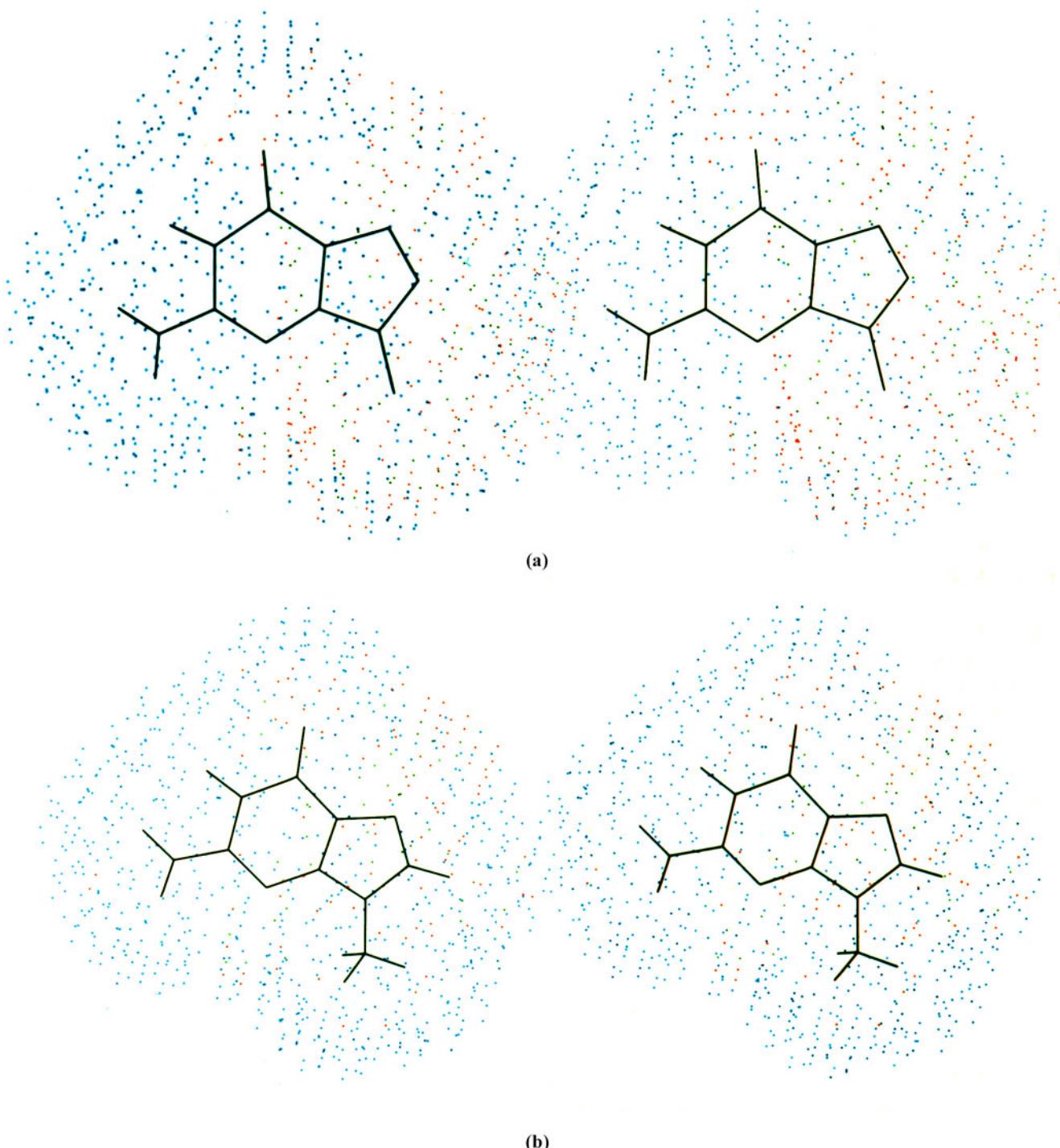


Figure 4. (a) Difference between united-atom model and quantum mechanical potential for 9-CH₃ guanine. See Figure 2(a) for color code. (b) Difference between all-atom model and quantum mechanical potential for 9-CH₃ guanine. See Figure 2(a) for color code.

including formaledhyde (Table XI) suggest that the extended model more accurately represents the relative energies of σ and π H bonds.

NUCLEIC ACID BASES, SUGARS, AND PHOSPHATES

We next determined charge models for the nucleic acid bases, deoxyribose and ribose, and

dimethyl phosphate (Tables XII–XXIV). We examined both the N—CH₃ bases of most direct relevance to DNA and RNA, as well as the bases themselves. Because of the size of these molecules, we used only the more limited STO-3G and the minimal atomic basis of Clementi in these studies. The STO-3G and Clementi basis sets give rather different results for H₂O, with the latter yielding multipole moments more characteristic

Table XIX. Charges for guanine.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM-UA
Charges				
N1	-0.691	-0.716	-0.819	-0.893
C2	0.783	0.797	0.898	0.940
N3	-0.723	-0.721	-0.768	-0.761
C4	0.540	0.557	0.567	0.616
C5	-0.010	-0.039	0.110	0.026
C6	0.655	0.685	0.693	0.774
N7	-0.562	-0.586	-0.653	-0.725
C8	0.357	0.463	0.325	0.640
N9	-0.517	-0.585	-0.646	-0.848
N2	-0.720	-0.725	-0.953	-0.969
O6	-0.456	-0.458	-0.477	-0.487
H1	0.335	0.339	0.416	0.426
H8	0.037	-	0.112	-
H9	0.328	0.345	0.408	0.459
HN2A	0.331	0.330	0.415	0.412
HN2B	0.312	0.314	0.382	0.390
μ_x	-3.569	-3.583	-3.419	-3.459
μ_y	-4.995	4.992	5.184	5.173
μ_z	20.636	19.787	22.791	21.468
Q_a	-14.430	-13.812	-14.505	-13.699
Q_b	-6.206	-5.975	-8.286	-7.769
Q_c	0.92	0.95	1.26	1.46
s	5.4	5.7	7.0	8.1

^aSee Table XII for notation.

of more extended basis sets (Table I), so it is encouraging that the multipole moments of the bases are calculated to be similar for the STO-3G and Clementi basis sets.

The partial charges for the purines with N9-H compared with N9-methyl are quite similar except for the charges at these groups and the neighboring C8 and C4 atoms. Similarly, for the pyrimidines, the N1-H and N1-CH₃ and their neighbors C2 and C6 differ most between the bases and their N1-CH₃ analogs.

Because our current molecular mechanical simulations on proteins and nucleic acids use a united-atom (UA) approach for aliphatic and aromatic CH_n groups, we have also determined a united-atom model for the charges of the bases and their N-CH₃ analogs as well as the all-atom (AA) models. This is an alternative to using the all-atom model, summing up the charges on the C—H hydrogens and placing the net charges on the carbon. We compared the fit to the potential using these two approaches for 1-CH₃ cytosine (Table XII), and it is clear that the united-atom model is superior to the all-atom model in which the C—H charges are summed up (in parentheses in Table XII). Figure 1 shows the quantum mechanically calculated potential for 1-CH₃ cytosine, and Figures 2(a) and 2(b) show the difference between the quantum mechanically calculated potential and the united atom [Fig. 2(a)] and all-atom [Fig. 2(b)] models. Figures 3 and 4 show corresponding potentials for 9-CH₃ guanine, and Figures 5 and 6 show corresponding potentials for 1-NH₂ deoxyribose.

The results of Watson-Crick H-bond calculations using these charges are presented in Table

XXV. With some of the charge models, we added 10–12 H-bond potentials with well depths of 0.5 kcal/mol and varying R* (minimum-energy distance), in order to ensure reasonable H-bond lengths (H ··· O and H ··· N) of 1.8–1.9 Å. As one can see, the agreement of our models with the best available experimental ΔH values is very good and better than all previous point-charge potentials. Also interesting is the ratio of Watson-Crick H-bonding energies between AT and GC; the calculations using the electrostatic potential derived charges and our earlier Mulliken charges³² closely reproduce the experimental ratio of ~ 60%, whereas the H-bond energies calculated using the charges reported by Ornstein and Rein,³⁴ Renugopalakrishnan, Sakshminarayanan, and Sasisekharan,³⁵ and Pullman and Pullman³³ are not as accurate in

Table XX. Charges for 9-CH₃ adenine.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM UA
Charges				
N1	-0.774	-0.760	-0.817	-0.859
C2	0.661	0.571	0.607	0.700
N3	-0.728	-0.717	-0.787	-0.776
C4	0.546	0.695	0.587	0.580
C5	-0.097	-0.151	-0.019	-0.104
C6	0.769	0.813	0.855	0.933
N7	-0.543	-0.599	-0.609	-0.713
C8	0.263	0.488	0.180	0.575
N9	-0.063	-0.451	-0.191	-0.340
N6	-0.768	-0.793	-0.971	-1.004
H2	-0.032	-	0.043	-
H8	0.062	-	0.143	-
C9M	-0.431	0.230	0.298	0.185
HN6A	0.335	0.335	0.410	0.411
HN6B	0.324	0.340	0.386	0.411
HC9A	0.158	-	-0.043	-
HC9B	0.159	-	-0.037	-
HC9C	0.159	-	-0.037	-
μ_x	2.274	2.227	2.357	2.358
μ_y	0.264	0.204	0.302	0.321
μ_z	15.185	15.294	15.229	15.643
Q_a	-10.406	-10.512	-9.521	-10.304
Q_b	-4.779	-4.782	-5.708	-5.339
Q_c	0.79	1.23	1.20	1.50
s	8.8	11.05	12.1	14.8

^aSee Table XII for notation.**Table XXI.** Charges for adenine.^a

Property	STO-3G-AA	Basis Set/Model		
		STO-3G-UA	CLEM-AA	CLEM UA
Charges				
N1	-0.769	-0.768	-0.824	-0.860
C2	0.648	0.592	0.607	0.720
N3	-0.757	-0.750	-0.828	-0.830
C4	0.680	0.746	0.744	0.763
C5	-0.022	-0.081	0.078	-0.002
C6	0.728	0.793	0.829	0.857
N7	-0.574	-0.616	-0.656	-0.740
C8	0.373	0.528	0.329	0.687
N9	-0.568	-0.685	-0.693	-0.897
N6	-0.777	-0.819	-1.008	-0.994
H2	-0.023	-	0.051	-
H8	0.049	-	0.128	-
H9	0.338	0.365	0.415	0.467
HN6A	0.345	0.349	0.428	0.420
HN6B	0.328	0.346	0.400	0.407
μ_x	2.166	2.140	2.298	2.294
μ_y	-0.147	-0.142	0.207	0.183
μ_z	9.772	11.461	12.732	13.862
Q_a	-5.200	-6.936	-5.809	-7.556
Q_b	-4.572	-4.525	-6.922	-6.306
Q_c	0.98	1.04	1.31	1.60
s	9.9	10.5	11.8	14.5

^aSee Table XII for notation.

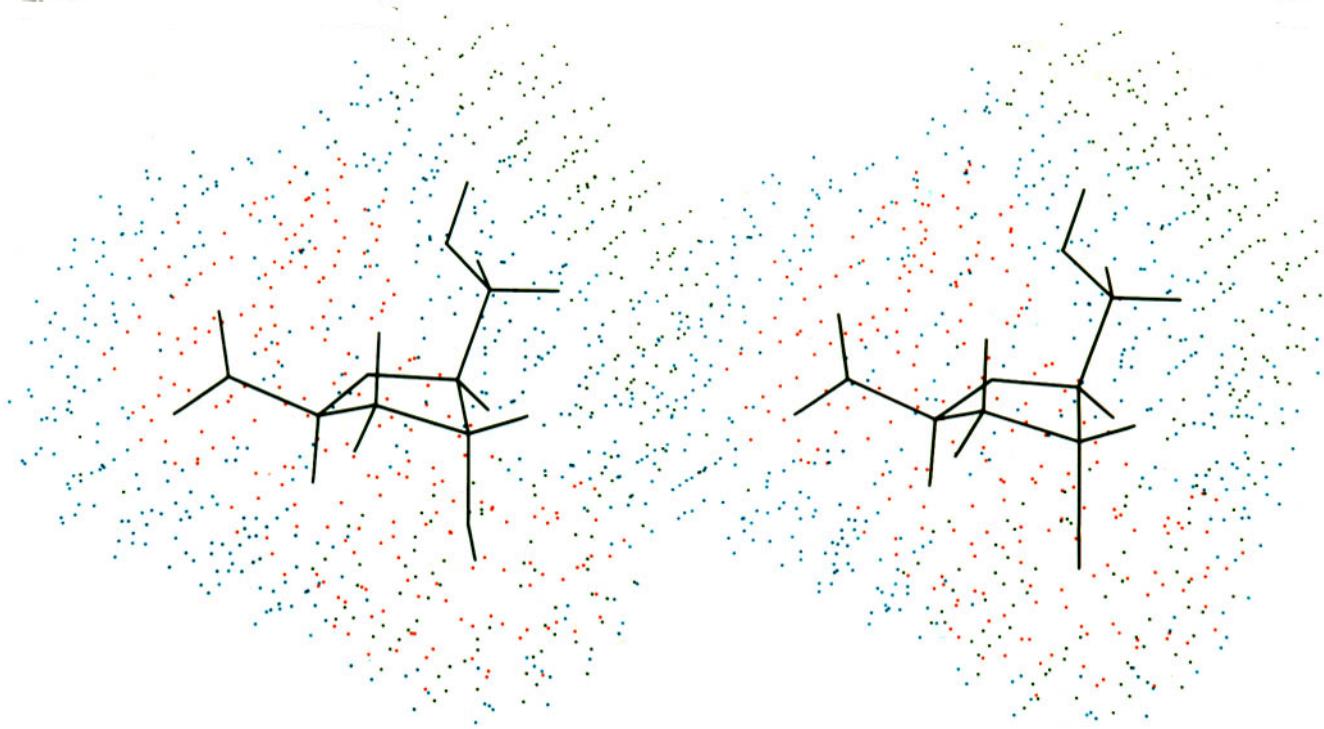


Figure 5. Quantum mechanical potential for 1-NH₂ deoxyribose. See Figure 1 for color code.

this regard. The charges reported by Ornstein and Rein,³⁴ and Renugopalakrishnan, Sakshminarayanan, and Sasisekharan³⁵ do not reproduce the dipole moments²⁸ as well for the isolated monomers,³⁸ but the charges reported in ref. 35 give qualitatively reasonable H-bond energies by comparison with experiment, although the GC H bond energy is significantly too strong in comparison with the AT structure.

Langlet, Claverie, and Caron³⁶ have carried out more precise calculations of the Watson-Crick H bonding in nucleic acid bases, and the agreement between our simple model and theirs is encouraging; both give quite good agreement with experiment. Given their finding that the Hoogsteen base pair of AT is slightly more stable than Watson-Crick, we examined this dimer as well and found results qualitatively similar to those of ref. 35 (see parentheses in Table XXV).

In Table XXV we have carried out the calculations with $\epsilon = 1$ and a distance-dependent dielectric $\epsilon = R$. Elsewhere we discuss³⁷ why we prefer the latter approach for solution-phase simulations, but it is clear that the former model is more appropriate for comparison with the gas-phase data of Yanson, Teplitsky, and Sukhodub.³⁷ On the other hand, the $\epsilon = R$ model does indirectly include polarization effects for attractive interactions in that it weighs more

heavily closer atom-atom interactions. The results in Table XXV show that the H-bond energies we calculate are in satisfactory agreement with experiment with either model.

Table XXII. Charges of 1-NH₂ ribose.^a

Property	Basis Set/Model			
	STO-3G AA	STO-3G UA	CLEM-AA	CLEM-UA
Charges				
O1'	-0.343	-0.413	-0.452	-0.503
C1'	0.373	0.537	0.543	0.726
C4'	0.100	0.197	0.098	0.276
C2'	0.101	0.083	0.181	0.094
C3'	0.303	0.271	0.446	0.315
N	-0.828	-0.894	-1.074	-1.145
O3'	-0.520	-0.542	-0.683	-0.683
C5'	0.180	0.175	0.271	0.174
H1'	0.054	-	0.057	-
HNA	0.312	0.325	0.372	0.386
HNB	0.308	0.317	0.376	0.387
H2'	0.008	-	-0.027	-
O2'	-0.546	-0.512	-0.682	-0.669
H3'	0.007	-	-0.037	-
H4'	0.061	-	0.043	-
HO3'	0.303	0.316	0.391	0.410
O5'	-0.463	-0.488	-0.596	-0.600
H5'	0.001	-	-0.036	-
H5'	0.016	-	-0.006	-
HO5'	0.289	0.300	0.379	0.390
HO2'	0.324	0.329	0.436	0.442
<i>ux</i>	1.723	1.749	2.452	2.497
<i>uy</i>	-1.740	-1.754	-1.917	-1.930
<i>uz</i>	0.836	0.848	0.863	0.868
<i>q_x</i>	5.375	4.949	7.536	7.254
<i>q_y</i>	-3.658	-3.253	-5.344	-5.128
<i>q_z</i>	-1.717	-1.696	-2.192	-2.126
<i>s</i>	0.69	0.75	1.02	1.10
	6.9	7.6	8.2	8.9

^a Notation as in Table XII; AMBER (ref. 24) optimized geometry was used, with C(3') endo sugar pucker; O1—C1' bond is along the Z axis and O1'—C1—N are in the XZ plane, with N in the +X direction. We used a model where the base is replaced by —NH₂.

Table XXIII. Charges of 1-NH₂ deoxy ribose.^a

Property	Basis Set/Model			
	STO-3G-AA	STO-3G-UA	CLEM-AA	CLEM-UA
Charges				
O1'	-0.368 (-0.371) ^b	-0.386	-0.494	-0.475
C1'	0.558 (0.470)	0.502	0.861	0.730
C4'	0.036 (0.029)	0.185	0.122	0.253
C2'	-0.307 (-0.351)	-0.047	-0.373	-0.103
C3'	0.233 (0.417)	0.172	0.276	0.190
N	-0.869 (-0.811)	-0.876	-1.172	-1.151
O3'	-0.508 (-0.541)	-0.514	-0.665	-0.666
C5'	0.118 (0.171)	0.153	0.206	0.144
H1'	0.009 (0.017)	-	-0.015	-
HNA	0.305 (0.298)	0.313	0.374	0.376
HNB	0.329 (0.309)	0.327	0.415	0.409
H2'	0.081 (0.071)	-	0.071	-
H2'	0.081 (0.102)	-	0.080	-
H3'	0.025 (-0.001)	-	0.016	-
H4'	0.056 (0.047)	-	0.043	-
HO3'	0.306 (0.306)	0.313	0.412	0.423
OS'	-0.404 (-0.477)	-0.425	-0.488	-0.489
HS'	0.003 (-0.004)	-	-0.039	-
HS'	0.039 (0.023)	-	0.017	-
HO5'	0.279 (0.296)	0.285	0.354	0.360
	1.783	1.799	2.389	2.426
H _x	-1.619	-1.619	-1.899	-1.901
H _y	0.215	0.234	0.121	0.131
H _z	8.843	9.081	9.957	10.271
Q _d	-1.594	-1.657	-1.372	-1.508
Q _d	-7.249	-7.423	-8.585	-8.763
Q _e	0.68	0.85	1.08	1.25
Q _f	7.4	9.4	9.8	11.2

^a Notation as in Table XII and XXII, except sugar pucker is C(2') endo.

^b Charges derived using a C(3') endo conformation.

DISCUSSIONS AND CONCLUSIONS

We have presented applications on a variety of molecules of a method for generating point-charge models for molecules, which can then be used in simulations. It is clear from the results of Watson-Crick hydrogen-bonding studies that this method, when used with a suitable *ab initio* basis set, can give very good agreement with experimental enthalpies of H-bond formation for nucleic acid base pairs. It is superior to Mulliken populations in this regard and is, we believe, a preferred method of approach for generating such charges. However, even given a set of atomic charges that reproduce well the molecular electrostatic potentials, one still faces the task of determining a set of nonbonded parameters that can be used with the charges to give good geometries and energies for complex formation. It should also be kept in mind that in most typical simulations, the intermolecular potential contains electrostatic, dispersion, and repulsion components, but no polarization or charge transfer. Thus the parameters used must compensate for the absence of some components. For example, one often uses electrostatic charges that lead to larger dipole moments than the gas-phase value for the molecule, in order to simulate polarization effects. In any case, the charges and nonbonded parameters must be tested in some model cases to ensure that they lead to results in good agreement with experiment or accurate calculations, and we have tried to do that here and elsewhere.

We have also compared, in the case of 1-NH₂ deoxyribose, the difference in charges when one uses a C3' endo rather than a C2' endo geometry (Table XXIII, all-atom approximation). The H-bonding-site charges change rather little, with the exception of O5', which changes from -0.404 to -0.477. However, the atoms to which it is bonded, C5' and HO5', become 0.070 more positive, almost exactly compensating for this charge shift. There are larger polarizations in the C2', C3', C4' charges, but since these are not likely to form strong electrostatic interactions, it is not clear how significant these shifts are. More studies are required to more precisely delineate the validity of the approximation of using a set of charges derived from one conformation to represent others.

The question of whether one should include lone pairs in partial-charge representations of molecules has been addressed here, but no single definitive answer to this question has been given. Obviously, inclusion of lone pairs improves the fit of the model to the electrostatic potential. In the case of (H₂O)₂, the directionality of the hydrogen bond, the angle the OH bond of the proton donor makes with the bisector of the HOH angle of the acceptor, goes from $\theta = 25$ to $\theta = 65$ upon addition of lone pairs, and the latter is closer to the experimental angle. The lone pairs in the optimum model, interestingly, are inverted from the tetrahedral direction and point in the direction of the hydrogens. This seemingly strange result is consistent with point-charge models for water based on *ab initio* dimer surfaces³⁹ and with such models designed to fit water liquid properties.⁴⁰ Such models, unlike those in which the lone pairs are in the "traditional" direction, can give excellent agreement

Table XXIV. Charges for dimethyl phosphate.^a

Charges	Basis Set/Model			
	STO-3G-AA	STO-3G-UA	STO-3G*-AA ^b	STO-3G*-UA
P	1.560	1.596	0.867	0.913
OCA	-0.509	-0.534	-0.377	-0.409
OCB	-0.508	-0.535	-0.375	-0.410
CA	0.032	0.086	0.034	0.108
CB	0.037	0.088	0.037	0.110
OFA	-0.846	-0.849	-0.650	-0.655
OFB	-0.848	-0.851	-0.652	-0.657
HCA1	0.003	-	0.005	-
HCA2	0.025	-	0.032	-
HCA3	0.014	-	0.023	-
HCB1	0.002	-	0.004	-
HCB2	0.023	-	0.030	-
HCB3	0.015	-	0.024	-
O	1.12	1.17	1.08	1.17
S	1.2	1.45	1.2	1.3

^a Notation as in Table XII; geometry from ref. 31.

^b STO-3G* basis set was used; see ref. 32.

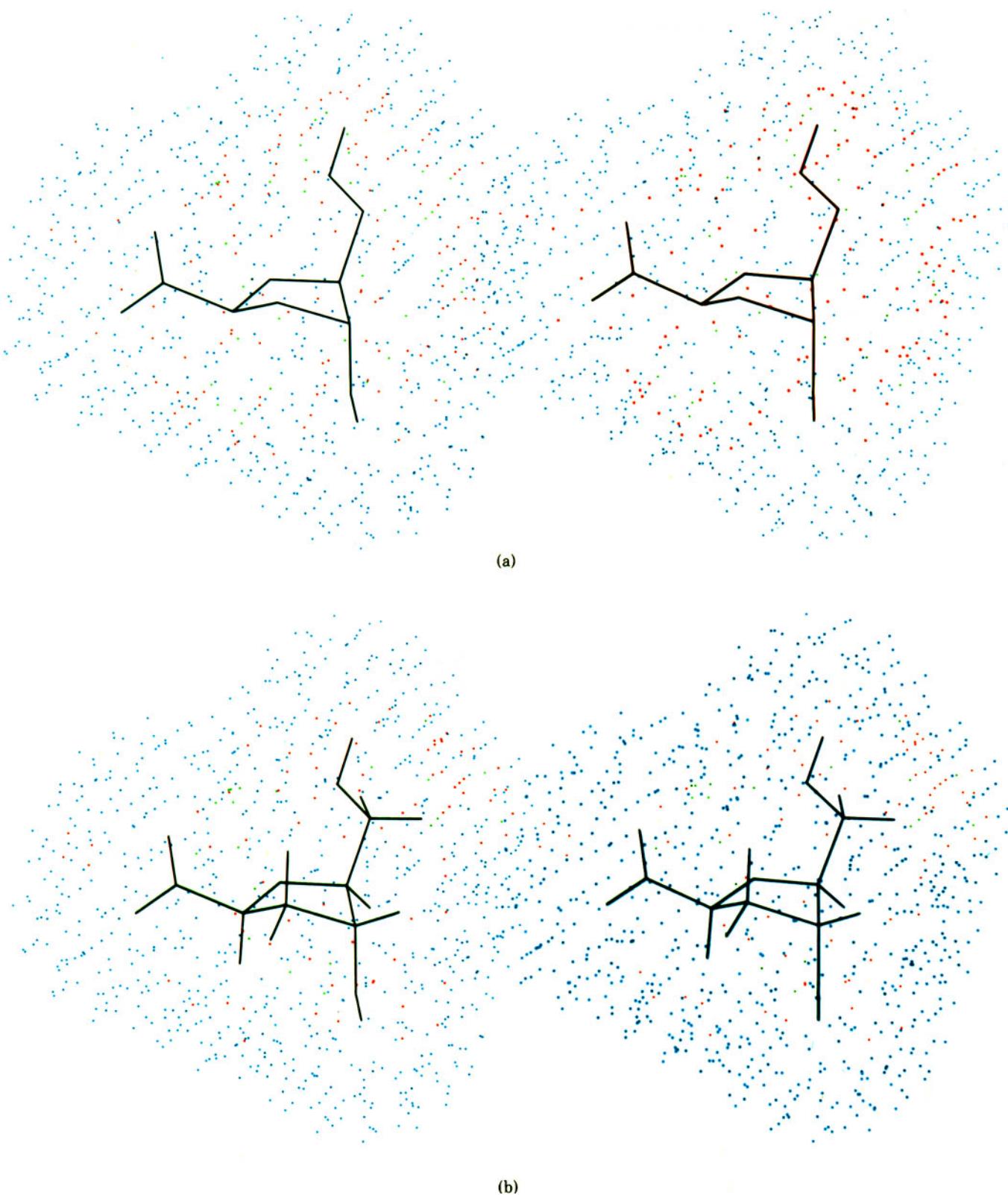


Figure 6. (a) Difference between united-atom model and quantum mechanical potential for 1-NH₂ deoxyribose. Color code as in Figure 2(a). (b) Difference between all-atom model and quantum mechanical potential for 1-NH₂ deoxyribose. Color code as in Figure 2(a).

with both the dipole and quadrupole moments for water.

In the case of CH₃OH, (CH₃)₂O, and CH₂O, we have found numerical problems in deriving

lone-pair-containing partial-charge models. These charges may collapse into the atom, extend too far from the atom, or become unrealistically large in magnitude, unless appropriate restraints are

Table XXV. Hydrogen-bond energies calculated with different charge models.

Model		$\Delta E(\text{AT})^{\text{d}}$	$R(\text{H}\cdots\text{B}), \text{AT}^{\text{e}}$	$\Delta E(\text{GC})^{\text{f}}$	$R(\text{H}\cdots\text{B}), \text{GC}^{\text{g}}$
Charges ^a	ϵ^{b}	$R^{\star\text{c}}$			
STO-UA ^h	1	2.0	11.3 (11.8) ^g	1.92, 1.86	21.2
STO-NUA ⁱ	1	2.0	11.9(13.5) ^g	1.91, 1.91	21.1
STO-UA ^h	R	2.0	13.4	1.86, 1.82	21.7
STO-NUA ⁱ	R	2.0	13.7	1.87, 1.81	22.1
Clem ^j	1	2.1	13.1	1.91, 1.98	22.6
Clem	R	2.1	15.2	1.87, 1.93	23.2
Pullman ^k	1	2.0	8.0	1.93, 1.94	18.6
Pullman	R	2.0	8.5	1.90, 1.91	19.9
Rein ^l	1	-	5.3	1.89, 1.92	11.1
Rein	R	-	5.6	1.82, 1.88	12.0
Sasi ^m	1	2.0	9.1	1.91, 1.91	23.6
Sasi	R	2.0	10.0	1.88, 1.89	25.8
PAK ⁿ	1	-	6.8	1.86, 1.89	13.1
PAK	R	-	6.6	1.78, 1.84	11.7
Claverie ^o			12.9 (13.6) ^g	-	23.7
Yanson (exp) ^p			13.0	-	21.0

^aSet of charges used.^bDielectric model.^c10–12 potential minimum energy R ; all well depths 0.5 kcal/mol.^dEnergy of AT Watson–Crick H bond in kcal/mol.^eH6 ··· O4 H1—H3 H-bond distances in AT in Å.^fEnergy of GC hydrogen bond in kcal/mol.^gO6 ··· H4, H1 ··· H3, and H2 ··· O2 H-bond distances in GC (Å).^hThis study; united atom charges STO-3G basis set.ⁱThis study; all atom charges; STO-3G basis set.^jThis study; Clementi basis set (ref. 17).^kCharges from Pullman and Pullman (ref. 33).^lCharges from Ornstein and Rein (ref. 34).^mCharges from Renugopalakrishnan, Sakshminarayanan, and Sasisekharan (ref. 35).ⁿCharges from Singh and Kollman (ref. 29) based on Mulliken population.^oSemiempirical analysis of Langlet, Claverie, and Caron (ref. 36).^pExperimental data of Yanson, Teplitsky, and Sukhodub (ref. 37).^qCalculated energy for Hoogsteen base pair.

placed on the fitting procedure. We note that our software allows such restraints to be imposed. Often, the “physically realistic” set of charges differs from those that are not so only slightly in quality of fit to the quantum mechanical potential.

For H₂CO we have also compared the lone-pair models with those without lone pairs for H₂CO ··· HOH. Again, the qualitative features do not change upon inclusion of lone pairs, but the angle the C—O bond makes with the HO bond increases slightly from 50° to 60°.

In summary, there is no compelling qualitative reason to include lone pairs in simulations including N and O lone pairs (a stronger case can be made concerning the inclusion of lone pairs for second-row atoms, such as S).⁴¹ In the long term, as more accurate models develop, lone pairs or higher multipoles than monopoles in the electrostatic expansion can be employed, but for the current status of simulations, atom-centered point monopoles seem reasonable and capable of reproducing well many molecular properties and, as such, should be useful. In this article we have presented a simple and systematic approach to

generating such monopoles. This approach can be employed for either atom-centered or lone-pair models and for either all-atom or united-atom models. It is a very simple generalization to determine charge models with higher multipoles as well. Elsewhere we present a complete set of charges appropriate for simulations of proteins and nucleic acids.⁴¹

P.A.K. is pleased to acknowledge research support by the NSF (CHE-80-26560) and NIH (CA-25644). We are also very grateful to F. van Duijnveldt for his program to carry out the least square fitting to quantum mechanical potentials and to Kitty Ghio for the PISA electrostatic potential program.

References

- N. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977); O. Ermer and S. Lifson, *ibid.*, **95**, 4121 (1973).
- A. Pullman, C. Zakrewska, and D. Perahia, *Int. J. Quantum Chem.*, **16**, 393 (1979).
- L. Dunfield, A. Burgess, and H. Scheraga, *J. Phys. Chem.*, **82**, 2609 (1978).
- D. M. Hayes and P. Kollman, *J. Am. Chem. Soc.*, **98**, 3335, 7811 (1976).
- E. Scrocco and J. Tomasi, *Adv. Quantum Chem.*, **11**, 115 (1978).

6. P. H. Smit, J. L. Derissen, and F. B. van Duijneveldt, *Mol. Phys.*, **37**, 521 (1979).
7. S. R. Cox and D. E. Williams, *J. Comput. Chem.*, **2**, 304 (1981).
8. U. C. Singh and P. Kollman, "GAUSSIAN 80 UCSF," *Quantum Chemistry Exchange Program Bull.*, **2**, 17 (1982).
9. J. S. Binkley, R. A. Whiteside, R. Kirshnan, R. Seeger, D. J. Defrees, H. B. Schlegel, S. Topiol, L. R. Kahn, and J. A. Pople, "GAUSSIAN 80," *Quantum Chemistry Program Exchange*, (1980).
10. M. Dupuis, D. Spangler, and J. J. Wendoloski, "GAMESS," NRCC Software Catalog, 1980, Program No. QG01, Vol. I.
11. P. Marsh and D. E. Williams, "GAUSSIAN 79," *Quantum Chemistry Program Exchange*, (1980).
12. Subroutine courtesy of R. Bonaccosi from the University of Pisa.
13. M. Connolly, *Quantum Chemistry Program Exchange*, (1982).
14. R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1980, Vol. 1, Chap. 5.
15. K. Kitaura and K. Morokuma, *Int. J. Quantum Chem.*, **10**, 325 (1976).
16. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).
17. E. Clementi, J. M. Andre, M. C. Andre, D. Klint, and D. Hahn, *Acta Phys. Chem.*, **27**, 493 (1969).
18. W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).
19. P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta*, **28**, 213 (1973).
20. For O (9s5p2d/4s3p1d), T. H. Dunning, Jr., *J. Chem. Phys.*, **53**, 2823 (1970).
For H (6s2p/2s1p), E. Clementi and H. Popkie, *J. Chem. Phys.*, **57**, 1077 (1972).
21. W. S. Benedict and L. D. Kaplan, *J. Chem. Phys.*, **30**, 388 (1959).
22. W. H. Flygare, *J. Am. Chem. Soc.*, **94**, 330 (1972).
23. W. Flygare and R. C. Benson, *Mol. Phys.*, **20**, 225 (1971).
24. P. Weiner and P. Kollman, *J. Comput. Chem.*, **2**, 287 (1981).
25. P. Kollman, in *Modern Theoretical Chemistry*, H. F. Schaefer, Ed., Plenum, New York, 1977, Vol. 4.
26. T. R. Dyke and J. Muentner, *J. Chem. Phys.*, **60**, 2929 (1974).
27. H. Popkie, H. Kistenmacher, and E. Clementi, *J. Chem. Phys.*, **59**, 1325 (1973).
28. D. Hankins, J. Moskowitz, and F. Stillinger, *J. Chem. Phys.*, **53**, 4544 (1970).
29. P. Kollman, P. Weiner, and A. Dearing, *Biopolymers*, **20**, 2583 (1981).
30. V. Bloomfield, D. Crothers, and I. Tinoco, *Physical Chemistry of Nucleic Acids*, Harper and Row, New York, 1974.
31. M. Newton, *J. Am. Chem. Soc.*, **95**, 256 (1973).
32. J. B. Collins, P. Schleyer, J. Binkley, and J. Pople, *J. Chem. Phys.*, **64**, 5142 (1976).
33. B. Pullman and A. Pullman, *Prog. Nucleic Acid Res. Mol. Biol.*, **9**, 327 (1969).
34. R. Ornstein and R. Rein, *Biopolymers*, **18**, 1277, 2821 (1979).
35. V. Renugopalakrishnan, A. Sakshminarayanan, and V. Sasisekharan, *Biopolymers*, **10**, 1159 (1971).
36. J. Langlet, P. Claverie, and F. Caron, in *Intermolecular Forces*, Proc. 14th Jerusalem Symp., B. Pullman, Ed., Reidel, Dordrecht 1981.
37. I. Yanson, A. Teplitsky, and L. Sukhodub, *Biopolymers*, **18**, 1149 (1979).
38. The calculated dipole moments in debyes for the N—CH₃ bases are for A, T, C, G: 2.2, 3.3, 5.4, 6.3 [this study, UA model STO-3G basis (ref. 16)]; 2.4, 3.7, 5.7, 6.6 [this study UA model Clementi basis (ref. 17)]; 1.8, 2.2, 4.4, 4.7 (ref. 32); 1.1, 3.4, 4.4, 5.4 (ref. 33); 2.6, 3.8, 8.1, 7.8 (ref. 34); 3.0, 3.9, 6.5, 7.2 (ref. 35). Indirect experimental data in ref. 30 suggests that the dipole moments for 9-CH₃ adenine and 1-CH₃ thymine are ~ 3.0 and 3.9 D, respectively.
39. H. Popkie, H. Kistenmacher, and E. Clementi, *J. Chem. Phys.*, **59**, 1325 (1973).
40. W. Jorgensen, *J. Chem. Phys. Lett.*, **77**, 4156 (1982).
41. S. J. Weiner, U. C. Singh, K. Ghio, G. Alagona, D. Case, P. Weiner and P. Kollman, *J. Amer. Chem. Soc.* (In press).