J Chim Phys (1997) 94, 1448-1481

New methods for potential functions for simulating biological molecules

GD Hawkins, CJ Cramer, DG Truhlar*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, USA

* *Correspondence and reprints.*

Résumé: Nous employons CM1A, un modèle classe-IV pour les charges, et les modèles de solvatation SM5.4/A et SM5.4PD/A pour calculer les charges atomiques et les énergies de solvatation de 9-methyladenine, thymine, et les dipeptides d'alanine et serine. Le modèle quantum mécanique CM1A pourvoit les charges atomiques qui sont si précis ou plus précis que les autres charges populaires utilisé pour les simulations dynamiques. Les charges CM1A, cependant, sont très économiques à calculer; donc, ils sont prometteurs pour examiner les effets des changement conformationel, des substitutions, d'attachement, même des réactions. Les modèles de solvatation ont été parameterisé contre beaucoup de groupes fonctionels et ils sont bien adapté aux calculations rapide pour systèmes grands.

Abstract: We use the CM1A class IV charge model and the SM5.4/A and SM5.4PD/A solvation models to calculate atomic charges and solvation energies for 9-methyladenine and thymine and for alanine and serine dipeptides. The CM1A quantum mechanical charge model provides atomic charges as accurate as or more accurate than those used in popular molecular dynamics force fields but is very economical in both computer time and effort required to generate charges; thus it is very promising for examining effects of conformational changes, substituents, solvation, binding, and even reaction. The solvation models have been parameterized over multiple functionalities and are well suited to rapid calculations on large systems.

Mots clés: modèle charges, énergies de solvatation, adenine, thymine, dipeptides d'alanine et serine

Keywords: charge model, solvation energies, adenine, thymine, alanine and serine dipeptides

1. Introduction

The development of potential functions for biomolecular simulations can benefit from a variety of approaches. In this paper we will apply two methods developed in our group to biological molecules:

(i) class IV charges for molecules in the gas-phase and in aqueous solution;

(ii) the pairwise descreening approximation for the solvation energy.

2. Theory

2.1. CM1A charge model

We can classify methods used to obtain atomic charges as follows:

1. Class I charges are obtained without quantum mechanics, e.g., by dividing the dipole moment of a diatomic molecules by its bond length.

2. Class II charges are obtained by some reasonable partitioning of an electron density computed by quantum mechanics into atomic populations. Mulliken [1], Löwdin [2], Weinhold and coworkers [3], Bader [4], and others have developed schemes for accomplishing this.

3. Class III charges are obtained by fitting atomic charges to reproduce calculations of true observables (like electrostatic potentials or molecular multipole moments). Popular methods include the Singh-Kollman-Besler-Merz [5,6] (SKBM) and CHELPG [7,8] methods. Such methods have many advantages, but also one striking disadvantage, namely that the calculated observables are not necessarily well converged with respect to the treatment of electron correlation and the basis set. A second disadvantage for practical work is that the calculated charges of buried atoms, i.e., atoms not near the surface of the molecule, are numerically unstable and sometimes unphysical [9].

4. Class IV charges are charges from a model that is designed to reproduce or predict accurately either experimental observables or well converged quantum mechanical calculations.

Many popular force fields in use for molecular dynamics (MD)

calculations employ either class II or class III charges. Recently we have devised a scheme for calculating class IV charges, and we believe it has a number of advantages. In this article we will illustrate the use of our class IV CM1A (Charge Model 1 based on AM1) model to calculate atomic charges for biological molecules and to serve as a charge model in solvation calculations.

The CM1A charge model [10] is based on a two-step mapping. We begin with a semiempirical molecular orbital wave function obtained by the AM1 [11] (Austin Model 1) method, and we calculate the Mulliken charge $q_\alpha^{(0)}$,α) on atom α. Then we calculate an improved charge by

$$q_\alpha^{(1)} = q_\alpha^{(0)} + B_\alpha \Delta q_\alpha \tag{1}$$

where B_α is the sum of bond orders [12] at atom α (calculated from the same AM1 wave function as used to calculate $q_\alpha^{(0)}$),

$$\Delta q_\alpha = c_\alpha q_\alpha^{(0)} + d_\alpha, \tag{2}$$

and c_α and d_α are semiempirical parameters fitted to experimental molecular multipole moments for small molecules (one could also fit c_α and d_α to well converged quantum mechanical calculations for small molecules). Finally the CM1A charge is

$$q_\alpha^{(2)} = q_\alpha^{(1)} - \sum_{\alpha' \neq \alpha} B_{\alpha\alpha'} \Delta q_{\alpha'} \tag{3}$$

where $B_{\alpha\alpha'}$ is the bond order [12] of atom α to atom α'. Equation (3) renormalizes the charge, but *locally*.

Because c_α and d_α depend only on atomic number (i.e., 1 for H, 6 for C,...) in the CM1A method, one can obtain values of these parameters by calibration for small molecules and then apply the method to large molecules. Thus it may be considered to be a bootstrap technique. We view the mapping step in eq. (2) as a correction for systematic errors in the electronegativities of the individual atoms implied by the AM1 model chemistry. Such systematic errors are endemic in practical computational levels not only in semiempirical molecular orbital theory, but also in *ab initio* electronic structure levels such as HF/6-31G*. We obtained values of c_α and d_α for H, C, N, O, F, Si, S, Cl,

Br, and I from a training set of 195 small molecules [10]. The average $|c_\alpha|$ is 0.07, and the average $|d_\alpha|$ is 0.06 in the CM1A model. Thus the change in atomic charges is quantitatively small, but this change is very significant for molecular properties like dipole moments. The RMS (root-mean-square) errors in the dipole moments (in Debyes) of 23 test molecules are illustrated in the following table [10]:

class II charges	
HF/6-31G* Weinhold natural populations	1.05
HF/6-31G* Mulliken charges	0.93
AM1 Mulliken charges	0.89
class III charges	
HF/6-31G* Merz-Kollman charges	0.34
HF/6-31G* ChelpG charges	0.33
class IV charges	
CM1A	0.27

The use of CM1A charges for modeling biological molecules has several advantages, namely:

(i) the calculations are very efficient so large systems may be treated and extensive conformational exploration (e.g., multiple ϕ–ψ maps [13]) may be carried out on mid-sized systems;

(ii) buried atoms are described on an equal footing with atoms near the molecular surface;

(iii) the results are more accurate than extended-basis-set *ab initio* Hartree-Fock calculations even when such calculations are affordable and when ESP fitting to such calculations is stable.

Although the inaccuracy associated with truncating the charge distribution at each site at the first (monopole) term and including only atomic centers as sites are widely appreciated [14], the distributed monopole model is so intuitive and so computationally convenient that we think it will remain popular essentially indefinitely. Forcing the distributed partial charges to reproduce physical observables like electrostatic potentials or molecular dipole

moments, as class III charges do, eliminates many of the severe problems that may be encountered when the charges are determined more arbitrarily. Determination of class III charges requires carrying out a new fit to high-level calculations for every case. The class IV charge model described above provides a rapid and efficient method for determining charges of comparable quality or better. Our method, unlike fitting, is simple enough that the charges may be used in a variety of ways in strategies for creating potential functions. This is illustrated in the present paper by their use in the SM5.4PD method, but we hope they will be useful in a variety of contexts for potential function generation.

2.2. Free energy of solvation

In general, for simulations of processes in aqueous solution, we are interested in the potential of mean force for the solute. Let x denote a full set of coordinates for an N-body solute. If $V(x)$ is the potential energy of the solute in the absence of solvent, and $\Delta G_S(x)$ is the free energy of solvation, the potential of mean force is

$$W(x) = V(x) + \Delta G_S(x) \tag{4}$$

One way to calculate this quantity is by a statistical average of the classical potential energy $V(x,y)$, where y denotes a full set of solvent coordinates. The average is carried out over a canonical ensemble of solvent configurations y for a fixed set of solute coordinates x:

$$W(x) = k_B T \ln < \exp[V(x,y) / k_B T] >_{\text{average over } y} \tag{5}$$

However one may also model $\Delta G_S(x)$ without explicitly dealing with the individual solvent coordinates y. A model that does this is called a continuum solvation model.

We have been studying continuum solvation models for aqueous solutions in which for fixed x, we can approximate $\Delta G_S(x)$ by two terms:

$$\Delta G_S(x) = \Delta G_{EP}(x) + G_{CDS}(x) \tag{6}$$

where,

$$\Delta G_{EP}(x) = G_P(x) + \Delta E_E(x) \tag{7}$$

Here $G_P(x)$ is the free energy of electric polarization of the solvent medium. It equals the solute-solvent electrostatic interaction with equilibrated (i.e., polarized) solvent minus the work required to polarize the solvent. Under the assumption of linear response, it is easily shown that the latter equals one half the former [15]. Thus $\Delta G_P(x)$ equals one half the solute-solvent electrostatic interaction energy. $\Delta E_E(x)$ is the increase in internal energy of the solute when it is polarized by the polarized solvent. In order to calculate the extent of mutual polarization of the solute and the solvent, we use the self-consistent reaction field [16-19] (SCRF) method. (In more complete treatments one also optimizes the geometry of the solute in solution [20], but that is not considered in the examples in this paper.)

The first term in eq. (6), which we have just discussed, accounts for the bulk electrostatic effect of the solvent and under many conditions it should provide a reasonable estimate of the effect of the bulk of the solvent. However the solvent has different properties in the first shell around the solute, and one must also consider dispersion interactions and specific interactions such as the non-electrostatic aspects of hydrogen bonding for solvent molecules in this first shell. Furthermore, one must account for the free energy cost of solvent structural changes in the first hydration shell. Of course these effects extend to a lesser extent into the second solvation shell, but it is well known that the major effects of solvation, except for long-range classical electrostatic effects, are primarily localized in the first shell of solvent. We account for all these effects by the second term, $G_{CDS}(x)$, in eq. (6). This term has the form

$$G_{CDS}(x) = \sum_{\alpha} \sigma_{\alpha}(x) A_{\alpha}(x) \qquad (8)$$

where $\sigma_{\alpha}(x)$ is a microscopic surface tension, and $A_{\alpha}(x)$ is the solvent-accessible surface area [21] of atom α. Notice that $\sigma_{\alpha}(x)$ is assumed to depend on the solute geometry x. We have developed a reasonable set of functional forms for this dependence, and these forms are collectively referred to as the SM5 solvation model. For a given treatment of the

electrostatics ($\Delta G_{EP}(x)$ in eq. (6)), we optimize the parameters in these functional forms to experimental data on free energies of solvation to account for first-solvation effects and also, inevitably, to make up in a semiempirical sense for any systematic deficiencies in the whole model that are capturable by the assumed functional forms [22,23]. The focus of this article, though, is on the electrostatics, in particular on SM5.4 models for aqueous solutions, where the descriptor SM5.4 denotes that we combine the SM5 functional forms for microscopic surface tensions with class IV charges. In the present paper we present two different versions of this kind of solvation model, the first denoted simply SM5.4 and the second denoted SM5.4PD. Our general experience with biological molecules suggests that parameterizations based on AM1 are more realistic than those based on PM3. Thus we only present the AM1 results in this paper. In papers where we apply both AM1 and PM3, results based on solvation models parameterized for AM1 have the suffix /A, but we do not need to carry along that distinction in the notation here.

Both the SM5.4 and SM5.4PD models are particularly relevant to the problem of creating potential functions for simulating complete systems since, by making reasonable approximations, they reduce the electrostatic problem almost (SM5.4) or completely (SM5.4PD) to a sequence of analytic calculations; the latter method is particularly intriguing in that the free energy effect of bulk solvation is calculated by a series of analytic steps each involving only individual pairs of atoms. Thus large systems may be treated efficiently. If solute electronic polarization were neglected, these analytic treatments would reduce the solvent polarization contribution to analytic functions, and in the PD model the functions involve only the distances between pairs of atoms. However we do not neglect solute polarization here. Rather we incorporate either of the analytic results for the electrostatic effect into a Fock operator [15-20,22,23] which allows us to calculate polarized molecular orbitals for the solute. The molecular orbital calculations are performed by semiempirical methods that are economical for very large systems. In fact other groups have recently performed such calculations for

whole proteins [24,25].

It is widely recognized that environmental effects can significantly change atomic charges and dipoles. These environmental effects may be classified as intramolecular, intermolecular, and long-range. When a system of point charges is placed in a bulk dielectric (e.g., an aqueous solvent) it polarizes the dielectric, producing a so called reaction field that acts back on the original system, changing its charge distribution. The first molecular layer of the dielectric medium requires a quantum mechanical atomic-level description to account for specific intermolecular interactions (we call these first-solvation-shell effects), but there is also a long-range contribution which is probably adequately treated by considering the solvent to be a classical dielectric medium. The reaction field of this dielectric interacts with the solute, whose own reaction must ultimately be modeled by quantum mechanics. We have embedded the mapping procedure that produces class IV charges into a quantum mechanical solvation model that accomplishes this. The solvation model also includes intramolecular charge redistribution and first-solvation effects; thus it includes all three types of environmental effects identified earlier in this paragraph.

Our treatment involves two solute-solvent boundaries. The first boundary is placed at the van der Waals surface, which is the surface of the overlapping set of spheres with standard van der Waals radii. The solvent-accessible surface area (SASA) used in the $G_{CDS}(x)$ calculation is the area of a surface which is taken to be 1.7 Å beyond this van der Waals boundary. This surface passes through the first solvation shell, and thus the SASA is a continuum measure of the average number of solvent molecules in this shell. The second boundary is the interface between the dielectric medium and the solute, and it is the surface of a set of overlapping spheres with empirically determined "Coulomb radii." It is this second boundary, the Coulomb boundary, that is important for the electrostatic part of the calculation. The radii used for the molecules in the present study are (in Å) [22,23]:

		Coulomb	
	van der Waals	*SM5.4*	*SM5.4PD*
H	1.20	ρ^*(H)	1.17
C	1.70	1.78	1.89
N	1.55	1.60	1.66
O	1.52	1.92	1.94

where ρ^*(H) is an analytic function of H–O and H–N distances that is 1.28 Å when H is more than 2.0 Å from any O or N and then decreases to a plateau value of about 0.59Å. In fact it has the value 0.59 ± 0.02 Å whenever any H–O or H–N distance is in the interval 1.03 ± 0.19 Å.

In the SM5.4 method the electrostatics are calculated by the generalized Born approximation [19,20,22-28], using the dielectric descreening algorithm [29,30] of Still and coworkers. In this approach one writes

$$G_P(s) = \frac{1}{2}\left(\frac{1}{\varepsilon} - 1\right) \sum_{\alpha,\alpha'} \gamma_{\alpha\alpha'} q_\alpha q_{\alpha'} \qquad (9)$$

where ε is the solvent dielectric constant, q_α is the charge on atom α, and $\gamma_{\alpha\alpha'}$ is a Coulomb integral. Coulomb integrals for $\alpha \neq \alpha'$ are evaluated from one-center Coulomb integrals $\gamma_{\alpha\alpha}$ by a generalized Ohno-Klopman formula [31,32], and the one-center terms are evaluated by a modified Born approximation, which is the crux of the method. In the modified Born approximation [29], the polarization free energy density in the space around a particular solute atom is approximated by the Coulomb free energy density (i.e., the free energy density predicted by Coulomb's law around a spherical conducting solute immersed in a dielectric medium) if the space is occupied by solvent and by zero if the space is occupied by solute. Then the total free energy G_P is then obtained by integrating over the space occupied by solvent. The latter is calculated by assuming that the solute consists of overlapping spheres. Thus the solute spheres descreen the atom under consideration from some of the solvent.

In the SM5.4PD method [23] the space occupied by the solvent is

calculated approximately in a two-by-two fashion rather than by simultaneously considering the location and mutual overlap of all the solute spheres. Each sphere α' is considered to descreen atom α independently and the contributions are assumed pairwise additive [23,32,33]. Empirical scale factors are employed for the sizes of the descreening spheres to make this pairwise descreening approximation as adequate as possible across a wide range of solute functionalities [23,34].

The volume integrals required for the electrostatic part of SM5.4 calculations may be carried out in spherical coordinates and the angular integrations may be carried out analytically [35], leaving only a one-dimensional radial quadrature to be carried out numerically. The same analytic techniques may also be used for the SASA calculations [35]. In the SM5.4PD calculations, everything is analytic.

We note that a coulomb potential that is screened by solvent in a solute-shape-sensitive fashion is not isotropic and hence is much more suitable for detailed modeling than is an isotropic distance-dependent function. The extent to which hydrogen bonding is simply an electrostatic effect and the extent to which it is not are very open questions. Since our treatment of electrostatics includes the first solvation shell, it clearly contains some part of the hydrogen bonding effects. But the additional first-solvation-shell terms, i.e., the surface-tension terms, must make up for the nonelectrostatic part as well as for the treatment of the electrostatic part with the bulk dielectric constant right up to the solute Coulomb boundary. One can obtain similar accuracy in a parameterized model for various choices of the Coulomb boundary (as long as one keeps it in the vicinity of the first solvation shell) by adjusting the surface tensions to compensate for the shell that is included or excluded in the dielectric region by the change in Coulomb boundary. Therefore the separation of the free energy of solvation into electrostatic and nonelectrostatic parts is more arbitrary than the separation of potential energy into coulombic and noncoulombic parts.

We have parameterized the SM5.4 model against a data set of 252 free

energies of solvation for neutral and ionic solutes containing H, C, N, O, F, S, Cl, Br, and I [22]. We obtained a mean unsigned error in the free energy of solvation of 0.50 kcal/mol for 215 neutrals and 4.3 kcal/mol for 34 ions [23]. (Note that the mean $|\Delta G_S|$ is 76 kcal for the ions, but only 3 kcal for the neutrals). Restricting attention to neutral solutes containing only C, H, O, and N, which are the only atoms relevant to the examples in the present paper, we achieve a mean unsigned error of 0.56 kcal/mol for 150 molecules. Encouragingly, we also obtain good results with the SM5.4PD model. In particular, we obtained mean unsigned errors of 0.46 kcal/mol for 215 neutrals and 3.6 kcal/mol for 34 ions. Again restricting attention to neutral solutes containing only C, H, O, and N yields a mean unsigned error of 0.50 kcal/mol with the primary improvements occurring for aliphatic amines and amides.

3. Applications

Atomic charges. In order to illustrate the use of the new methods for biological molecules, we report here applications to four biological systems:

MeAde:	9-methyladenine
Thy:	free base thymine
Ac-Ala-NHMe:	alanine dipeptide (C_7 equatorial conformation)
Ac-Ser-NHMe:	serine dipeptide (C_7 equatorial conformation)

These molecules are illustrated, with their atomic numbering, in Figures 1 and 2. Notice that we use peptide numbering rather than reside numbering. Alanine dipeptide is the widely used common name for 1-(acetylamino)-N-methylpropanamide, also known as N^{α}-acetyl-*N*-methylalanamide, *N*-methylalanine acetamide, *N*-methylalanylacetamide, and *N*-acetylalanine-*N*′methylamide. Similarly, serine dipeptide denotes *N*-acetylserine-*N*′-methylamide.

We will center attention on the atomic charges in the gas phase and solution. The present charges may be compared to those used in some previous parameterized force fields that include charges for intramolecular

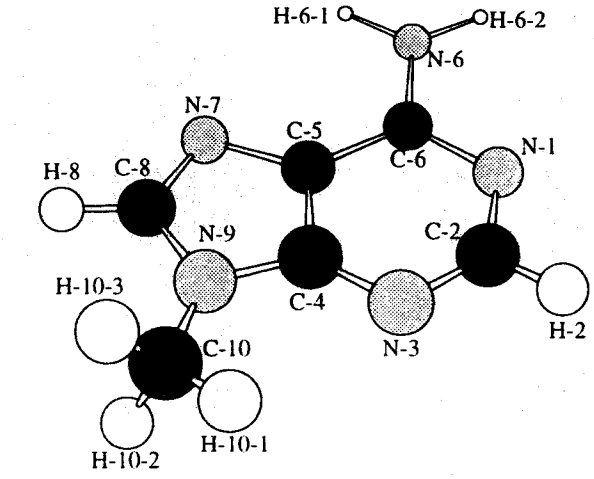

Fig. 1a

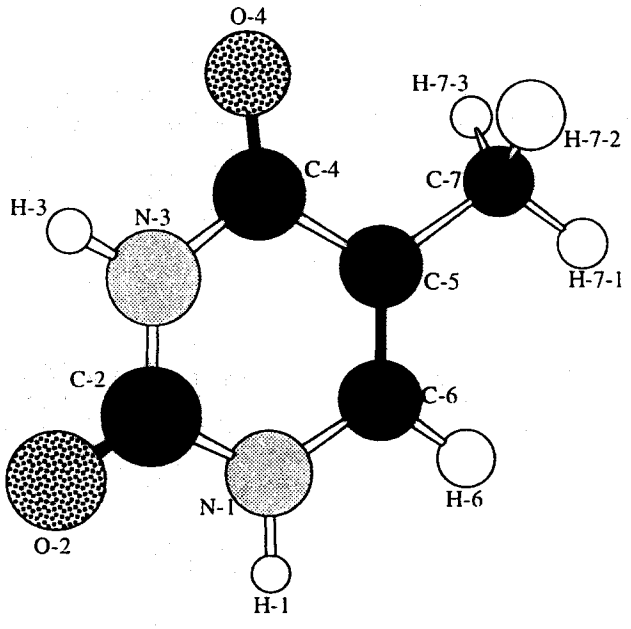

Fig. 1b

Figure 1. Atomic numbering schemes. (a) 9-methyladenine. (b) thymine.

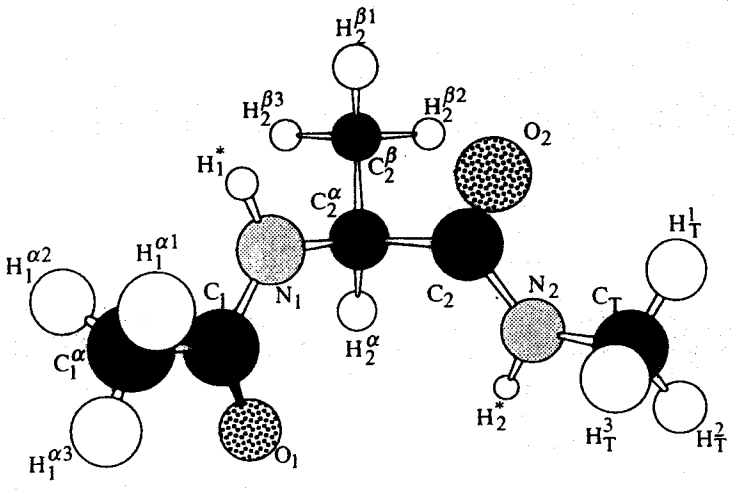

Fig. 2a

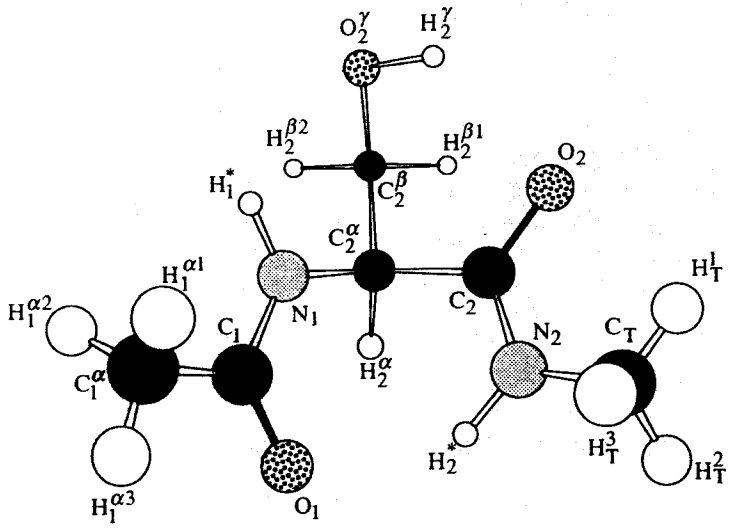

Fig. 2b

Figure 2. Atomic numbering schemes. (a) Ac-Ala-NHMe. (b) Ac-Ser-NHMe.

and/or intermolecular interactions.

Jorgensen and coworkers have developed the method of Optimized Potentials for Liquid Simulations (OPLS) [36-40]. They determined atomic charges by fitting the interaction energies of model compounds with water, using the Hartree-Fock method with a 6-31G* basis set (HF/6-31G*) [41]. In particular the charges were chosen so that interaction energies with TIP4P water agreed with the Hartree-Fock results, and the calculated interaction distances were decreased by 0.2–0.3 Å, the latter feature being necessary to obtain correct liquid densities. A similar procedure was used to obtain charges for the CHARMM-22 nucleic acid force field by MacKerell *et al.* [42]. However these workers parameterized interaction energies for nucleic acid bases with TIP3P water to HF/6-31G* interaction energies, and they scaled by a factor of 1.16 as well as shortening bond distances by approximately 0.2 Å. The factor of 1.16 and the reduced bond distances are used to account for the lack of polarization in the potential function and the absence of dispersion in the Hartree-Fock calculations. The experimental base pairing energies, dipole moment, and heats of sublimation were used as additional input in the charge parameterization. The approach of using water interaction energies for determining charges was originally introduced by Reiher and Karplus [43]. This method does not readily lend itself to estimating the effect on the charges of changing geometries, substituents, or solvation.

An alternative procedure has been used by the Scheraga and Kollman groups, namely derivation of the charges from quantum mechanical wave functions. The Scheraga group used overlap-normalized Mulliken analysis [1,44] of CNDO/2 [45] semiempirical molecular orbital wave functions to obtain atomic charges for the ECEPP [46] method, and these are retained without change in their more recent ECEPP/2 [47] method. These workers commented that the resulting charges are underestimated but that this could be compensated by adjusting the dielectric constant [46]. The Kollman group charges are incorporated in the AMBER program. The original AMBER-type charges were obtained from electrostatic potential (ESP) fitting based on

HF/STO-3G [41] wave functions (class III charges) [5]. Later work, which was formalized as the Weiner et al. (WEA) force field in the AMBER (Assisted Model Building and Energy Refinement) program [48], resulted in slightly different charges by the same basic procedure. More recent work by this group involves ESP fitting with an HF/6-31G* wave function and also a two-stage restrained electrostatic potential (RESP) fitting method [49,50], again based on HF/6-31G*. The restrained method forces charges on equivalent atoms and methyl and methylene hydrogens to be equivalent, and it was used to obtain new standardized charges for the Cornell *et al* (CEA) force field in AMBER [51]. Unlike the OPLS and CHARMM fitting procedures the methods used for the AMBER force fields do not modify the *ab initio* gas-phase calculations to account for solvent polarization effects. Nevertheless, like the OPLS and CHARMM force fields, the AMBER force field is intended for use in liquid-phase simulations. In fact, Weiner *et al.* attempted to derive a force field which can be used either with $\varepsilon = 1$ and full inclusion of solvent or with a distance-dependent dielectric constant without explicit solvent. This requires a careful balance of local and long-range interactions and also that the charges be representative of the peptide itself in the gas phase [52].

We will compare our calculations to all these methods, as well as to several other sets of results. In particular we consider the following other methods:

EFF	Empirical Force Field of Dauber and Hagler [53]. The changes are based in part on molecular orbital calculations [54].
CFF	Complete Force Field of Dauber-Osguthorpe *et al.* [55]
ECEPP	Empirical Conformational Energy Program for Peptides [46,47]
AMBER-WEA	Weiner *et al.* 1986 force field used in the AMBER program [48,56]. These are standardized values of charges obtained by electrostatic potential fitting to gas-phase calculations at the HF/STO-3G level [5].

ESP/6-31G*	Electrostatic fitting at the Hartree-Fock level without charge restraints [49].
AMBER-CEA	the second generation Cornell *et al.* force field in AMBER [51]. These are standardized values corresponding to RESP fitting at the HF/6-31G* level [49-51].
OPLS	the original OPLS force field [36-38] (Note: in the original OPLS force field, alkyl groups are treated as united atoms. Thus for alkyl groups we use OPLS-AA (all atom) charges, which do equal the united-atom alkyl charges when summed over all atoms in the alkyl group.)
OPLS-AA	the all-atom OPLS method [39,40]
CHARMM	original CHARMM (Chemistry at Harvard Macromolecular Mechanics) force field (sometimes called PARAM18) [57]
CHARMM-22	recent standardized charges for the all-atom CHARMM force field [42].
X-ray	charges derived from single-crystal X-ray diffraction data of nucleosides by partitioning electron density among the atoms during the refinement [58].

From our own work we present

CM1A	Class IV charges by the AM1-CM1A method
SKBM/MP2/pDZ	Charges obtained by the SKBM electrostatic potential fitting method from wave functions computed at the Møller-Plesset 2nd order perturbation theory [41] level (MP2) with the correlation-consistent polarized valence double zeta [59] (cc-pVDZ) basis set. The pDZ basis has 156 basis functions for Thy, 189 for MeAde, and 200 for Ac-Ala-NHMe.
CHELPG/MP2/pDZ	same as SKBM/MP2/pDZ except with the CHELPG

	electrostatic potential fitting method
SKBM/MP2/pTZ	same as SKBM/MP2/pDZ except with the cc-pVTZ [59] polarized triple zeta basis set. The pTZ basis set has 354 basis functions for Thy and 428 for MeAde.
CHELPG/MP2/pTZ	same as SKBM/MP2/pTZ except CHELPG
CHELPG/HF/pDZ	same as CHELPG/MP2/pDZ except Hartree-Fock
SKBM/HF/pDZ	same as SKBM/MP2/pDZ except Hartree-Fock
CHELPG/HF/MIDI!	same as CHELPG/HF/pDZ except with the MIDI! basis set
SKBM/HF/MIDI!	same as CHELPG/HF/MIDI! except SKBM

The MP2 results should be more reliable than the HF ones because MP2 includes electron correlation and HF does not.

All the present results for MeAde and Thy were computed using geometries optimized by the Hartree-Fock method with the MIDI! basis set, a partially polarized double zeta basis that was developed specifically to give accurate geometries and reasonable atomic charges at economical cost [60]. All the present results for the dipeptides were obtained with geometries optimized by the AM1 method, except for OPLS dipole moments which were computed at HF/MIDI! geometries. The HF/MIDI! and AM1 geometries for the alanine dipeptide are quite respectable, as illustrated by the critical ϕ, ψ angles. The best previous *ab initio* optimizations yielded ϕ, ψ = –86, +69 (HF/3-21G) [61], –79, +55 (HF/6-31+G*) [61], and –86, +79 (HF/6-31G**) [62]. CHARMM-19 gives –77.5, +90, and AMBER-86 gives –75, +68 [63]. AM1 and HF/MIDI! give –85, +64 and –83, +79, respectively.

In most cases where we carried out electrostatic fitting by the CHELPG and SKBM methods we also calculated charges by Mulliken analysis of the same wave functions. These results are not shown in the tables because they differ considerably from the electrostatic potential fitting results for the same wave functions, with deviations as large as 0.45 for MeAde, 0.60 for Thy, 0.51 for Ac-Ala-NHMe C_7^{eq}, and 0.48 for Ac-Ala-NHMe C_7^{ax}. It is well known that Mulliken analysis does not yield reliable charges, especially for

extended basis sets.

The new results for MeAde and Thy are compared to each other, to the atomic charges in popular MD force fields, and to the charges obtained by X-ray analysis in the crystal in Tables 1a, 1b, 2a, and 2b. Mean unsigned deviations from the X-ray results are 0.11–0.16 for the correlated calculations, 0.13–0.21 for the Hartree-Fock calculations, 0.10–0.18 for the popular MD force fields, 0.11–0.16 for CM1A gas phase charges, and 0.12–0.15 for CM1A charges polarized by the solvent. The CM1A gas-phase dipole moments are remarkably close, 0.23–0.28 D for MeAde and 0.03–0.04 D for Thy, to the highest-level ESP charges. The CM1A gas-phase charges are more generally similar to the OPLS ones than to those in the other force fields. The force field results are, on average, about equally similar to our gas-phase and aqueous ones. The excellent accuracy of the CM1A dipole moments is particularly striking.

The alanine dipeptide has been widely studied as a model for peptide conformation [52,61-70], and it can also be used as a test case for peptide charges [65]. There is general consensus that the C_7^{eq} conformation is the lowest-energy one in both the gas phase and solution, and we restrict attention to this case. The two electrostatic fitting methods in Tables 3a and 3b yield quite similar results, although the charge on C_2^{β} varies by 0.14 between the two methods. For a given fitting method, the Hartree-Fock charges show deviations from the MP2 ones as large as 0.16, with mean unsigned deviation of 0.07 (CHELPG) and 0.04 (SKBM). The OPLS and CM1A charges are in remarkably good agreement with each other, with only the charge on nitrogen being considerably different (the OPLS value being 0.13 less negative than the best *ab initio* result and the CM1A values being 0.24 more negative than this benchmark). The OPLS-AA value for the charge on nitrogen is farther than the original OPLS value from our best *ab initio* value and from CM1A but

Table 1a - Gas phase atomic charges and dipole moments of 9-methyladenine[a]

	gas phase								
	MP2/cc-pVTZ		MP2/cc-pVDZ		HF/cc-pVDZ		HF/MIDI!		HF/6-31G*
	CHELPG	SKBM	CHELPG	SKBM	CHELPG	SKBM	CHELPG	SKBM	ESP
N-1	–0.67	–0.66	–0.66	–0.64	–0.78	–0.76	–0.77	–0.74	–0.85
C-2	0.53	0.49	0.52	0.48	0.63	0.59	0.61	0.56	0.61
H-2	0.02	0.05	0.01	0.04	0.03	0.06	0.03	0.06	0.07
N-3	–0.66	–0.68	–0.63	–0.64	–0.74	–0.73	–0.73	–0.72	–0.79
C-4	0.42	0.42	0.41	0.40	0.48	0.44	0.53	0.50	0.54
C-5	0.01	0.07	–0.01	0.03	–0.09	–0.03	–0.13	–0.07	–0.10
C-6	0.57	0.49	0.57	0.51	0.71	0.64	0.75	0.70	0.89
N-6	–0.67	–0.67	–0.68	–0.68	–0.75	–0.75	–0.81	–0.81	–1.06
H-6-1	0.32	0.34	0.32	0.33	0.34	0.35	0.36	0.36	0.46
H-6-2	0.30	0.31	0.31	0.32	0.33	0.34	0.35	0.36	0.46
N-7	–0.51	–0.50	–0.50	–0.48	–0.53	–0.51	–0.52	–0.51	–0.58
C-8	0.18	0.06	0.19	0.07	0.20	0.08	0.24	0.15	0.20
H-8	0.11	0.16	0.10	0.14	0.12	0.17	0.11	0.15	0.15
MUD–MP2[b]	0.00	0.03	0.01	0.03	0.06	0.05	0.07	0.06	0.14
MUD–CM1A[c]	0.14	0.13	0.14	0.12	0.16	0.14	0.16	0.14	0.20
MUD–SM5.4[d]	0.14	0.13	0.13	0.12	0.16	0.14	0.16	0.14	0.20
MUD–X-ray[e]	0.11	0.13	0.11	0.11	0.14	0.14	0.14	0.15	0.21
μ (D)[f]	2.97	3.02	2.89	2.93	2.97	3.01	2.92	2.95	2.69

[a]The N-9 and methyl charges are omitted since these differ significantly in 9-methyladenine and free base adenine

[b]Mean unsigned deviation of 13 atomic charges from CHELPG/MP2/cc-pVTZ gas-phase charges

[c]Mean unsigned deviation of 13 atomic charges from CM1A gas-phase charges

[d]Mean unsigned deviation of 13 atomic charges from SM5.4 aqueous charges

[e]Mean unsigned deviation of 13 atomic charges from X-ray crystal charges

[f]dipole moment of 9-methyladenine calculated from atomic charges

Table 1b - Atomic charges and dipole moments of 9-methyladenine[a]

	MD force fields				gas	aqueous		crystal
		AMBER-		CHARMM-				
	OPLS	WEA	CEA	22	CM1A	SM5.4	SM5.4PD	X-ray
N-1	–0.53	–0.77	–0.76	–0.74	–0.57	–0.64	–0.62	–0.58
C-2	0.22	0.66	0.57	0.53	0.29	0.30	0.29	0.58
H-2	0.20	0.03	0.06	0.16	0.19	0.22	0.23	0.11
N-3	–0.55	–0.73	–0.74	–0.69	–0.43	–0.47	–0.45	–0.43
C-4	0.38	0.55	0.38	0.31	0.25	0.22	0.21	0.28
C-5	0.15	–0.10	0.07	0.23	–0.17	–0.18	–0.19	–0.18
C-6	0.44	0.77	0.69	0.43	0.50	0.50	0.48	0.44
N-6	–0.81	–0.77	–0.91	–0.80	–0.81	–0.79	–0.79	–0.53
H-6-1	0.36	0.34	0.42	0.40	0.39	0.41	0.41	0.32
H-6-2	0.39	0.34	0.42	0.40	0.40	0.42	0.42	0.32
N-7	–0.49	–0.54	–0.62	–0.63	–0.25	–0.32	–0.31	–0.38
C-8	0.20	0.26	0.16	0.38	0.12	0.15	0.16	0.33
H-8	0.20	0.06	0.19	0.18	0.20	0.25	0.26	0.01
MUD–MP2[b]	0.11	0.08	0.09	0.11	0.14	0.14	0.15	0.11
MUD–CM1A[c]	0.09	0.18	0.16	0.15	0.00	0.03	0.03	0.11
MUD-SM5.4[d]	0.09	0.18	0.15	0.14	0.03	0.00	0.01	0.12
MUD–X-ray[e]	0.14	0.15	0.18	0.14	0.11	0.12	0.12	0.00
μ (D)[f]	4.14	n.a.[g]	n.a.[g]	n.a.[g]	2.74	3.99	3.94	n.a.[g]

[a]The N-9 and methyl charges are omitted since these differ significantly in 9-methyladenine and free base adenine

[b]Mean unsigned deviation of 13 atomic charges from CHELPG/MP2/cc-pVTZ gas-phase charges

[c]Mean unsigned deviation of 13 atomic charges from CM1A gas-phase charges

[d]Mean unsigned deviation of 13 atomic charges from SM5.4 aqueous charges

[e]Mean unsigned deviation of 13 atomic charges from X-ray crystal charges

[f]dipole moment of 9-methyladenine calculated from atomic charges

[g]not available for 9-methyladenine

Table 2a - Gas phase atomic charges and dipole moments of thymine[a]

	gas phase								
	MP2/cc-pVTZ		MP2/cc-pVDZ		HF/cc-pVDZ		HF/MIDI!		HF/6-31G*
	CHELPG	SKBM	CHELPG	SKBM	CHELPG	SKBM	CHELPG	SKBM	ESP
C-2	0.74	0.69	0.71	0.67	0.88	0.81	0.85	0.78	0.81
O-2	–0.60	–0.54	–0.53	–0.52	–0.64	–0.62	–0.60	–0.57	–0.62
N-3	–0.60	–0.57	–0.58	–0.56	–0.69	–0.65	–0.71	–0.65	–0.79
H-3	0.35	0.35	0.34	0.34	0.37	0.37	0.38	0.37	0.42
C-4	0.67	0.63	0.64	0.61	0.81	0.77	0.77	0.71	0.79
O-4	–0.51	–0.51	–0.49	–0.48	–0.59	–0.58	–0.55	–0.53	–0.59
C-5	–0.10	–0.01	–0.13	–0.05	–0.21	–0.11	–0.18	–0.09	–0.10
C-7	–0.23	–0.50	–0.18	–0.44	–0.17	–0.46	–0.32	–0.58	–0.37
H-7-1	0.07	0.14	0.05	0.12	0.05	0.13	0.10	0.16	0.12
H-7-2	0.09	0.16	0.07	0.14	0.08	0.16	0.11	0.18	0.12
H-7-3	0.09	0.16	0.07	0.14	0.08	0.16	0.11	0.18	0.12
C-6	0.00	–0.05	0.01	–0.04	0.06	0.01	0.06	0.03	–0.12
H-6	0.15	0.17	0.14	0.16	0.17	0.18	0.18	0.18	0.21
MUD–MP2[b]	0.00	0.06	0.03	0.06	0.06	0.06	0.06	0.07	0.08
MUD–CM1A[c]	0.11	0.16	0.11	0.14	0.12	0.15	0.12	0.15	0.14
MUD-SM5.4[d]	0.12	0.16	0.11	0.14	0.12	0.15	0.12	0.15	0.13
MUD–X-ray[e]	0.12	0.16	0.12	0.15	0.13	0.16	0.14	0.17	0.18
μ (D)[f]	3.99	4.00	3.87	3.76	4.56	4.55	4.50	4.48	5.29

[a]The N-1 and H-1 charges are omitted since these differ significantly in deoxythymidine and free base thymine

[b]Mean unsigned deviation of 13 atomic charges from CHELPG/MP2/cc-pVTZ gas-phase charges

[c]Mean unsigned deviation of 13 atomic charges from CM1A gas-phase charges

[d]Mean unsigned deviation of 13 atomic charges from SM5.4 aqueous charges

[e]Mean unsigned deviation of 13 atomic charges from X-ray crystal charges

[f]dipole moment of thymine calculated from 15 atomic charges

Table 2b - Atomic charges and dipole moments of thymine[a]

	MD force fields				gas	aqueous		crystal
		AMBER-		CHARMM-				
	OPLS	WEA	CEA	22	CM1A	SM5.4	SM5.4PD	X-ray
C-2	0.50	0.85	0.57	0.57	0.78	0.78	0.78	0.61
O-2	–0.40	–0.49	–0.59	–0.47	–0.37	–0.41	–0.42	–0.65
N-3	–0.51	–0.85	–0.43	–0.46	–1.00	–1.02	–1.01	–0.51
H-3	0.36	0.36	0.34	0.36	0.48	0.48	0.48	0.39
C-4	0.45	0.81	0.52	0.54	0.59	0.59	0.59	0.66
O-4	–0.42	–0.46	–0.56	–0.49	–0.35	–0.41	–0.42	–0.47
C-5	–0.07	–0.18	0.00	–0.15	–0.22	–0.27	–0.26	–0.50
C-7	–0.14	–0.38	–0.23	–0.11	–0.14	–0.14	–0.14	0.32
H-7-1	0.08	0.11	0.08	0.07	0.07	0.11	0.11	0.09
H-7-2	0.08	0.11	0.08	0.07	0.10	0.09	0.09	0.09
H-7-3	0.08	0.11	0.08	0.07	0.10	0.09	0.09	0.09
C-6	0.08	0.03	–0.22	0.17	0.20	0.23	0.23	0.18
H-6	0.10	0.13	0.26	0.13	0.16	0.22	0.22	0.17
MUD–MP2[b]	0.09	0.08	0.08	0.08	0.11	0.12	0.12	0.12
MUD–CM1A[c]	0.12	0.10	0.18	0.11	0.00	0.02	0.02	0.16
MUD-SM5.4[d]	0.12	0.11	0.17	0.11	0.02	0.00	0.00	0.15
MUD–X-ray[e]	0.13	0.17	0.16	0.10	0.16	0.15	0.15	0.00
μ (D)[f]	4.14	n.a.[g]	n.a.[g]	4.51	4.03	6.18	6.21	n.a.[g]

[a]The N-1 and H-1 charges are omitted since these differ significantly in deoxythymidine and free base thymine

[b]Mean unsigned deviation of 13 atomic charges from CHELPG/MP2/cc-pVTZ gas-phase charges

[c]Mean unsigned deviation of 13 atomic charges from CM1A gas-phase charges

[d]Mean unsigned deviation of 13 atomic charges from SM5.4 aqueous charges

[e]Mean unsigned deviation of 13 atomic charges from X-ray crystal charges

[f]dipole moment of thymine calculated from 15 atomic charges

[g]not available for free base thymine

Table 3a - Atomic charges on alanine residue and dipole moments of alanine dipeptide in C_7 equatorial conformation

	MD force fields							
					OPLS-	AMBER-		
	EFF	CFF	ECPP	OPLS	AA	WEA	CEA	CHARMM
N_1	–0.26	–0.50	–0.34	–0.57	–0.50	–0.46	–0.42	–0.36
H_1^*	0.26	0.28	0.18	0.37	0.30	0.25	0.27	0.26
C_2^α	–0.11	0.12	–0.01	0.14	0.14	0.03	0.03	0.10
H_2^α	0.11	0.10	0.05	0.06	0.06	0.05	0.08	0.10
C_2^β	–0.33	–0.30	–0.09	–0.18	–0.18	–0.10	–0.18	–0.30
$H_2^{\beta 1}$	0.11	0.10	0.04	0.06	0.06	0.04	0.06	0.10
$H_2^{\beta 2}$	0.11	0.10	0.04	0.06	0.06	0.04	0.06	0.10
$H_2^{\beta 3}$	0.11	0.10	0.04	0.06	0.06	0.04	0.06	0.10
C_2	0.46	0.38	0.45	0.50	0.50	0.62	0.60	0.48
O_2	–0.46	–0.38	–0.38	–0.50	–0.50	–0.50	–0.57	–0.48
MUD–MP2[a]	0.13	0.08	0.14	0.08	0.09	0.13	0.13	0.10
MUD–CM1A[c]	0.14	0.09	0.14	0.07	0.09	0.13	0.12	0.11
MUD-SM5.4[d]	0.14	0.10	0.15	0.07	0.08	0.13	0.12	0.10
μ (D)[d]	n.a.[e]	n.a.[e]	n.a.[e]	2.79[f]	2.44[f]	n.a.[e]	n.a.[e]	4.52

[a]Mean unsigned deviation of 10 atomic charges from CHELPG/MP2/cc-pVDZ gas-phase charges

[b]Mean unsigned deviation of 10 atomic charges from CM1A gas-phase charges

[c]Mean unsigned deviation of 10 atomic charges from SM5.4 aqueous charges

[d]dipole moment in Debyes of entire dipeptide, calculated from atomic charges, in every case at HF/MIDI! geometry

[e]not available

[f]calculated at HF/MIDI! geometry (not AMBER geometry)

Table 3b - Atomic charges on alanine residue and dipole moments of alanine dipeptide in C_7 equatorial conformation

	gas phase					aqueous	
	MP2/cc-pVDZ		HF/cc-pVDZ				
	CHELPG	SKBM	CHELPG	SKBM	CM1A	SM5.4	SM5.4PD
N_1	–0.70	–0.70	–0.76	–0.75	–0.94	–0.93	–0.94
H_1^*	0.35	0.36	0.35	0.37	0.45	0.46	0.46
C_2^α	0.39	0.43	0.40	0.44	0.16	0.17	0.15
H_2^α	–0.01	0.00	0.00	0.01	0.13	0.13	0.13
C_2^β	–0.29	–0.43	–0.31	–0.45	–0.21	–0.20	–0.22
$H_2^{\beta 1}$	0.07	0.10	0.08	0.11	0.11	0.10	0.11
$H_2^{\beta 2}$	0.07	0.11	0.07	0.12	0.08	0.11	0.12
$H_2^{\beta 3}$	0.07	0.11	0.09	0.12	0.10	0.09	0.10
C_2	0.41	0.31	0.57	0.45	0.50	0.50	0.52
O_2	–0.45	–0.42	–0.57	–0.54	–0.41	–0.47	–0.47
MUD–MP2[a]	0.00	0.04	0.04	0.06	0.10	0.10	0.10
MUD–CM1A[c]	0.10	0.12	0.10	0.12	0.000	0.014	0.015
MUD-SM5.4[d]	0.10	0.12	0.10	0.11	0.014	0.000	0.009
μ (D)[d]	2.74	2.74	3.67	3.67	2.95	4.45	4.44

[a]Mean unsigned deviation of 10 atomic charges from CHELPG/MP2/cc-pVDZ gas-phase charges

[b]Mean unsigned deviation of 10 atomic charges from CM1A gas-phase charges

[c]Mean unsigned deviation of 10 atomic charges from SM5.4 aqueous charges

[d]dipole moment in Debyes of entire dipeptide, calculated from atomic charges, in every case at HF/MIDI! geometry

closer to the AMBER-CEA value. (The CM1P [11] value at the same AM1 geometry used the calculations in Table 3b is –0.45; clearly the charge on nitrogen is difficult to predict.) The AMBER value appears slightly less accurate than the OPLS, OPLS-AA, and CM1A charges.

When our CM1A charge model is incorporated in the SM5.4 and SM5.4PD solvation models, it allows for the effect of solute polarization by the reaction field produced by the solvent. The last two columns of Table 3b show only very small shifts in the atomic charge upon placing the dipeptide in aqueous solution. However these small changes have a huge effect on the dipole moment, changing it by 1.5 D. The SM5.4 and SM5.4PD models are in excellent agreement with each other for this change in dipole moment.

Finally in Tables 4a and 4b, we consider the serine dipeptide, again in the C_7 equatorial conformation. The only difference from the previous case is the replacement of the $-CH_3$ side chain by $-CH_2OH$, which participates in an intramolecular hydrogen bond with the terminating carbonyl. However this causes interesting differences in the various charges. The C^{α} charges obtained by ESP fitting to *ab initio* wave functions show much greater dependence on side chain than do either the standard MD changes or the CM1A changes. Since the CM1A changes do not suffer from buried-atom problems they well may be the most realistic values.

Solvation energies. The solvation energies are also of interest and can be used to assess the effect of variations in atomic charges on calculated observables. In addition we can compare SM5.4PD calculations to SM5.4 calculations to see if using the pairwise descreening to simplify the volume integral over the Coulomb free energy density is successful. Table 5 shows results for the three systems for which previous results are available [62,71-78] and compares to those results.

Table 5 shows that a wide range of values have been reported for all three cases. The various calculations differ in geometries, charges, method used to evaluate the free energy, etc. The energetics are particularly sensitive

Table 4a - Atomic charges on serine residue and dipole moments of N-acetylserine-N´-methylamide in C_7 equatorial conformation

	MD force fields					
					AMBER-	
	CFF	ECPP	OPLS	OPLS-AA	WEA	CEA
N_1	–0.50	–0.34	–0.57	–0.50	–0.46	–0.42
H_1^*	0.28	0.18	0.37	0.30	0.25	0.27
C_2^α	0.12	–0.01	0.14	0.14	0.03	–0.02
H_2^α	0.10	0.05	0.06	0.06	0.05	0.08
C_2^β	–0.17	0.13	0.14	0.14	0.02	0.21
$H_2^{\beta 1}$	0.10	0.02	0.06	0.06	0.12	0.04
$H_2^{\beta 2}$	0.10	0.02	0.06	0.06	0.12	0.04
O_2^γ	–0.38	–0.31	–0.70	–0.68	–0.55	–0.65
H_2^γ	0.38	0.17	0.43	0.42	0.31	0.43
C_2	0.38	0.45	0.50	0.50	0.62	0.60
O_2	–0.38	–0.38	–0.50	–0.50	–0.50	–0.57
MUD–MP2[a]	0.10	0.11	0.05	0.05	0.08	0.07
MUD–CM1A[b]	0.11	0.17	0.10	0.11	0.11	0.15
μ (D)[c]	n.a.[d]	n.a.[d]	1.86	1.51	n.a.[d]	n.a.[d]

[a]Mean unsigned deviation of 11 atomic charges from CHELPG/MP2/cc-pVDZ gas-phase charges

[b]Mean unsigned deviation of 11 atomic charges from CM1A gas-phase charges

[c]dipole moment in Debyes of entire N-acetylserine-N´-methylamide, calculated from atomic charges

[d]not available

Table 4b - Atomic charges on serine residue and dipole moments of N-acetylserine-N′-methylamide in C_7 equatorial conformation

	gas phase					aqueous	
	MP2/cc-pVDZ		HF/cc-pVDZ				
	CHELPG	SKBM	CHELPG	SKBM	CM1A	SM5.4	SM5.4PD
N_1	–0.56	–0.50	–0.61	–0.53	–0.93	–0.92	–0.92
H_1^*	0.32	0.32	0.33	0.32	0.46	0.47	0.47
C_2^α	0.08	0.00	0.07	–0.01	0.12	0.13	0.12
H_2^α	0.02	0.05	0.04	0.07	0.13	0.13	0.14
C_2^β	0.19	0.12	0.23	0.17	0.01	0.03	0.02
$H_2^{\beta 1}$	0.05	0.07	0.05	0.07	0.11	0.12	0.13
$H_2^{\beta 2}$	–0.01	0.02	–0.01	0.02	0.10	0.10	0.11
O_2^γ	–0.58	–0.56	–0.63	–0.61	–0.52	–0.54	–0.54
H_2^γ	0.37	0.38	0.39	0.40	0.39	0.39	0.39
C_2	0.48	0.43	0.66	0.59	0.51	0.51	0.52
O_2	–0.46	–0.44	–0.58	–0.56	–0.43	–0.47	–0.48
MUD–MP2[a]	0.00	0.04	0.04	0.05	0.10	0.10	0.11
MUD–CM1A[b]	0.10	0.10	0.13	0.12	0.00	0.01	0.01
μ (D)[c]	1.89	1.89	2.57	2.57	1.98	2.89	3.07

[a]Mean unsigned deviation of 11 atomic charges from CHELPG/MP2/cc-pVDZ gas-phase charges

[b]Mean unsigned deviation of 11 atomic charges from CM1A gas-phase charges

[c]dipole moment in Debyes of entire N-acetylserine-N′-methylamide, calculated from atomic charges

Table 5 - Free energies of solvation (kcal/mol)

	MeAde	Thy	alanine dipeptide
Present			
SM5.4	−15.0	−9.5	−7.6 [−8.2][a]
SM5.4PD	−13.4	−7.5	−9.1 [−9.3][a]
Other SCRF calculations			
SM2 [78]	−20.7	−15.9	−12.0
HF/6-31G** [62]			−17.8,−11.1
HF/6-31G* [73]	−6.5		
HF/6-31G* [75]	−8.5		
AM1 [74]	−9.5		
AM1 [75]	−10.8		
Quantum mechanical, polarizable solute + explicit solvent			
AM1 + TIP3P [77]		−8.5	
MD simulations, nonpolarizable solute			
AMBER-WEA [71]	−12.6		
AMBER-WEA [72]	−10.4		
AMBER-modified WEA [72]	−12.5, −14.9		
AMBER-CEA [51]	−16.3		
AMBER-CEA [76]	−12.0		
Experiment [72]	−13.6±1.1	n.a.	n.a.

[a]Values in brackets are for serine dipeptide, for which no previous results are known to us.

to the radii used in the electrostatics calculation, and these radii are intrinsically uncertain. As discussed above, we believe that it is particularly important to use surface tensions consistent with whatever choice is made. We believe that the use of this strategy with a geometry-dependent parameterization of the surface tensions in the SM5.4 methods makes these methods much less sensitive to the uncertainty in the choice of physical radii. Although both the SM5.4 and SM5.4PD methods give reasonable free energies of solvation, it is disappointing that these differ by 10–20% from each other.

In an attempt to better understand the differences, Table 6 compares the

Table 6 - G_P (kcal/mol)

	MeAde	Thy	alanine dipeptide
SM5.4 formulation			
CM1A charges (*g*)	–8.2	–6.7	–6.6
SM5.4 charges (*aq*)	–12.1	–11.2	–10.4
SM5.4PD charges (*aq*)	–12.4	–11.2	–11.4
CHELPG/MP2/pTZ charges (*g*)	–5.3	–9.0	–6.3
SKBM/MP2/pTZ charges (*g*)	–5.9	–9.1	–6.3
RESP charges (*g*)	–7.3		
OPLS charges			–7.7
OPLS-AA charges	-7.7		-7.0
CHARMM-22 charges		-9.6	
Present SM5.4PD formulation			
CM1A charges (*g*)	-8.5	-6.3	-8.6
SM5.4 charges (*aq*)	-12.5	-10.8	-13.1
SM5.4PD charges (*aq*)	-13.0	-10.9	-14.3
CHELPG/MP2/pTZ charges (*g*)	-5.4	-9.1	-7.1
SKBM/MP2/pTZ charges (*g*)	-5.8	-9.2	-7.3
RESP charges (*g*)	-7.2	-7.7	
OPLS charges			-8.9
OPLS-AA charges	-8.1		-8.2
CHARMM-22 charges		-9.6	
Poisson equation [79]			
OPLS charges	-10.8		
AMBER-WEA charges	-10.7		
CHARMM charges	-10.3		

polarization free energies calculated with several different sets of atomic charges. The polarization free energy is calculated first with the SM5.4 formalism (density descreening) and atomic radii (see above) and then with the SM5.4PD formalism and atomic radii. We also compare to some calculations in the literature [79]. The comparisons in each of the three

sections of this table differ from many similar comparisons in the literature in that for each molecule and section, all calculations are based on the same geometry, atomic radii, and method to evaluate the free energy. Thus the comparisons are a very direct measure of the effect of the charges on the free energy of solvation. We note that for the nucleic acid bases the density descreening and pairwise descreening formalisms agree quite well, but for the dipeptide they do not, perhaps due to the very polar carbonyl. Since the SM5.4 and SM5.4PD formalisms predict similar polarization free energies for the nucleic acid bases, it is clear that the differences in the first two rows of Table 5 are due primarily to different CDS contributions in the two parameterizations.

In comparing calculations with gas-phase charges and polarized charges in Table 6 we see differences in $G_P(x)$ of 3.8–4.5 kcal/mol. Such differences are offset by 2.0–3.2 kcal/mol of distortion cost, $\Delta E_E(x)$, in a self-consistent calculation, but this is not included in the unpolarized MD calculations. Thus such calculations can give the correct result for the resultant $\Delta G_{EP}(x)$ only if they have charges intermediate between the true the true gas-phase charges and the true aqueous charges, a situation which is perhaps not widely appreciated.

Table 7 provides another measure of the polarization of the solute by comparing dipole moments computed from the gas-phase and aqueous charges by various methods. Upon immersing the solutes in water, the SM5.4 dipoles increase by 1.3–2.2 D, and the SM5.4PD dipoles increase by 1.2–2.2 D. The pairwise descreening model is in quite good agreement with the more expensive calculation for this quantity.

Summary

In the first half of the paper we overviewed a new method for calculating partial atomic charges on large molecules both in the gas phase and in solution. The method yields class IV charges by a linear mapping of zero-overlap Mulliken charges obtained by semiempirical molecular orbital theory.

Table 7 - Dipole moments (Debyes) tabulated as [in the gas phase : in aqueous solution]

	MeAde	Thy	alanine dipeptide
AM1 : SM2 [30]	2.4 : 3.0	4.2 : 6.2	3.1 : 3.8
AM1 : SCRF [75]	2.3 : 3.1	4.2 : 5.8	
AM1 : SCRF + TIP3P [77]		4.2 : 5.9	
CM1A : SM5.4	2.7 : 4.0	4.0 : 6.2	2.9 : 4.4
CM1A : SM5.4PD	2.7 : 3.9	4.0 : 6.2	2.9 : 4.6

One can obtain partial atomic charges either in the gas phase or in aqueous solution.

In the second half of the paper we summarized an approximation scheme that reduces the problem of an N-body system in a solvent bath to a new N-body problem with a different potential function. The change in the potential function can be evaluated analytically.

The two ideas, class IV charges and pairwise descreening, are combined in a new semiempirical solvation method, denoted Solvation Model 5.4PD. Its use is illustrated by calculating free energies of solvation.

Acknowledgments

We are grateful to David Ferguson for helpful comments. This work was supported in part by the National Science Foundation under grant no. CHE94-23927 and by the National Institute of Standards and Technology Advanced Technology Program subcontract to Phillips Petroleum.

References

1. Mulliken RS (1955) *J. Chem. Phys.* 23, 1833-1840.
2. Löwdin P-O (1950) *J. Chem. Phys.* 18, 365-375.
3. Reed AE, Weinstock RB, Weinhold F (1985) *J. Chem. Phys.* 83, 735-746.
4. Bader RF (1985) *Accounts of Chemical Research* 18, 9-15.
5. Singh UC, Kollman PA (1984) *J. Comp. Chem.* 5, 129-145.
6. Besler BH, Merz KM, Kollman PA (1990) *J. Comp. Chem.* 11, 431-439.

7. Chirlian LE, Francl MM (1987) *J. Comp. Chem.* 8, 894-905.
8. Breneman CM, Wiberg KB (1990) *J. Comp. Chem.* 11, 361-373.
9. Francl MM, Carey C, Chirlian LE, Gange DM (1996) *J. Comp. Chem.* 17, 367-383.
10. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) *J. Computer-Aided Mol. Design* 9, 87-110.
11. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J. Am. Chem. Soc.* 107, 3902-3909.
12. Armstrong DR, Perkins PG, Stewart JJP (1973) *Journal of the Chemical Society, Dalton Transactions* 8, 838-840.
13. Ramachandran GN, Sasisekharan V (1968) *Advances in Protein Chemistry* 23, 283-438.
14. Price SL (1988) *Mol. Simul.* 1, 135-156.
15. Cramer CJ, Truhlar DG (1996) In *Solvent Effects and Chemical Reactivity*, Tapia O, Bertrán J, Eds., Kluwer, Dordrecht, 1-80.
16. Rivail J-L, Rinaldi D (1976) *Chem. Phys.* 18, 233-242.
17. Tapia O, Goscinski O (1975) *Mol. Phys.* 29, 1653-1661.
18. Yomosa S (1973) *J. Phys. Soc. of Japan* 35, 1738-1746.
19. Tapia O (1980) In *Quantum Theory of Chemical Reactions*, Daudel R, Pullman A, Salem L, Viellard A, Eds., Reidel, Dordrecht, Vol. 2, 25-72.
20. Cramer CJ, Truhlar DG (1991) *J. Am. Chem. Soc.* 113, 8305-8311, 9901(E).
21. Lee B, Richards FM (1971) *J. Mol. Biol.* 55, 379-400.
22. Chambers CC, Hawkins GD, Cramer CJ, Truhlar DG (1996) *J. Phys. Chem.* 100, 16385-16398.
23. Hawkins GD, Cramer CJ, Truhlar DG (1996) *J. Phys. Chem.* 100, 19824-19839.
24. Stewart JJP (1996) *Int. J. Quantum Chem.* 58, 133-146.
25. York DM, Lee T-S, Yang W (1996) *J. Am. Chem. Soc.* 118, 10940-10941.
26. Hoijtink GJ, de Boer E, Van der Meij PH, Weijland WP (1956) *Recueil de Travaux Chimique des Pays-Bas* 75, 487-503.
27. Peradejordi F (1963) *Cahiers de Physique* 17, 393-447.
28. Jano I (1965) Sur l'énergie de solvatation *Comptes Rendus Hebdomadaires des séances de L'Académie des Sciences, Paris* 261, 103-105.
29. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) *J. Am. Chem. Soc.* 112, 6127-6129.
30. Cramer CJ, Truhlar DG (1992) *J. Computer-Aided Mol. Design* 6, 629-666.
31. Ohno K (1964) *Theor. Chim. Acta* 2, 219-227.

32. Klopman G (1964) *J. Am. Chem. Soc.* 86, 4550-4557.
33. Schaefer M, Froemmel C (1990) *J. Mol. Biol.* 216, 1045-1066.
34. Hawkins GD, Cramer CJ, Truhlar DG (1995) *Chem. Phys. Lett.* 246, 122-129.
35. Liotard DA, Hawkins GD, Lynch GC, Cramer CJ, Truhlar DG (1995) *J. Comp. Chem.* 16, 422-440.
36. Jorgensen WL, Swenson CJ (1985) *J. Am. Chem. Soc.* 107, 569-578.
37. Jorgensen WL, Tirado-Rives J (1988) *J. Am. Chem. Soc.* 110, 1657-1666.
38. Pranata J, Wierschke SG, Jorgensen WL (1991) *J. Am. Chem. Soc.* 113, 2810-2819.
39. Kaminski G, Duffy EM, Matsui T, Jorgensen WL (1994) *J. Phys. Chem.* 98, 13077-13082.
40. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) *J. Am. Chem. Soc.* 118, 11225-11236.
41. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) *Ab initio molecular orbital theory*. Wiley, New York, p 245.
42. MacKerell Jr. AD, Wiórkiewicz-Kuczera J, Karplus M (1995) *J. Am. Chem. Soc.* 117, 11946-11975.
43. Reiher WE, Karplus M (1985) unpublished.
44. Yan JF, Momany FA, Hoffmann R, Scheraga HA (1970) *J. Phys. Chem.* 74, 420-421; erratum: (1970) 1974, 4611.
45. Pople JA, Segal GA (1966) *J. Chem. Phys.* 44, 3289-3296.
46. Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) *J. Phys. Chem.* 79, 2361-2381.
47. Neméthy G, Pottle MS, Scheraga HA (1983) *J. Phys. Chem.* 87, 1883-1887.
48. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) *J. Comp. Chem.* 7, 230-248.
49. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) *J. Am. Chem. Soc.* 115, 9620-9631.
50. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) *J. Phys. Chem.* 97, 269-280.
51. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz Jr. KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) *J. Am. Chem. Soc.* 117, 5179-5197.
52. Kollman PA, Dill KA (1991) *Journal of Biomolecular Structural Dynamics* 8, 1103-1107.
53. Dauber P, Hagler AT (1980) *Acc. Chem. Res.* 13, 105-112.
54. Hagler AT, Lapiccirella A (1976) *Biopolymers* 15, 1167-1200.
55. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M,

Hagler AT (1988) *Proteins* 4, 31-47.
56. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta Jr. S, Weiner P (1984) *J. Am. Chem. Soc.* 106, 765-784.
57. Brooks CR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) *J. Comp. Chem.* 4, 187-217.
58. Pearlman DA, Kim S-H (1990) *J. Mol. Biol.* 211, 171-187.
59. Dunning Jr. TH (1989) *J. Chem. Phys.* 90, 1007-1023.
60. Easton RE, Giesen DJ, Welch A, Cramer CJ, Truhlar DG (1996) *Theor. Chim. Acta* 93, 281-301.
61. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks Jr. CL, Pople J (1989) *Int. J. Quantum Chem. Quantum Biol. Symp.* 16, 311-322.
62. Gould IR, Cornell WD, Hillier IH (1994) *J. Am. Chem. Soc.* 116, 9250-9256.
63. Roterman IK, Lambert MH, Gibson KD, Scheraga HA (1989) *J. Biomol. Struct. Dyn.* 7, 421-453.
64. Momany FA, McGuire RF, Yan JF, Scheraga HA (1971 *J. Phys. Chem.* 75, 2286-2297.
65. Rossky PJ, Karplus M (1979 *J. Am. Chem. Soc.* 101, 1913-1937.
66. Weiner SJ, Singh UC, O'Donnell TJ, Kollman PA (1984) *J. Am. Chem. Soc.* 106, 6243-6245.
67. Pettitt BM, Karplus M (1985) *Chem. Phys. Lett.* 136, 383-386.
68. Böhm HJ, Brode S (1991) *J. Am. Chem. Soc.* 113, 7129-7135.
69. Gould IR, Kollman PA (1992) *J. Phys. Chem.* 96, 9255-9258.
70. Susnow R, Schutt C, Rabitz H, Subramaniam S (1994) *J. Comp. Chem.* 15, 947-962.
71. Bash P, Singh UC, Langridge R, Kollman PA (1987) *Science (Washington DC)* 236, 564-568.
72. Ferguson DM, Gould IR, Glauser WA, Schroeder S, Kollman PA (1992) *J. Comp. Chem.* 13, 525-532.
73. Young PE, Hillier IH (1993) *Chem. Phys. Lett.* 215, 405-408.
74. Orozco M, Luque FJ (1993) *Biopolymers* 33, 1851-1869.
75. Orozco M, Colominas C, Luque FJ (1996) *Chem. Phys.* 209, 19-29.
76. Miller JL, Kollman PA (1996) *J. Phys. Chem.* 100, 8587-8594.
77. Gao J (1996) *Reviews in Computational Chemistry*, Lipkowitz KB, Boyd DB, Eds., VCH, New York, Vol. 7, 119-186.
78. Cramer CJ, Truhlar DG (1992) *Chem. Phys. Lett.* 198, 74-80; erratum: (1993) 1202, 1567.
79. Mohan V, Davis ME, McCammon JA, Pettitt BM (1992) *J. Phys. Chem.* 96, 6428.

Corrections for the paper "New Methods for Potential Functions for Simulating Biological Molecules"
Journal de Chimie Physique 1997, **94**, 1448-1481

Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar
Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431

Some of the results using the SM5.4/AM1 and SM5.4PD/AM1 models for aqueous solvation are affected by a computer coding error in the Fock martix. These errors occur only in undistributed versions of AMSOL. Corrections are given below.

Page 1458. 0.50, 4.3, 0.56, 0.46, 3.6, and 0.50 should be 0.52, 4.1, 0.59, 0.49, 5.5, and 0.55, respectively.

Page 1465. 0.12–0.15 should be 0.07–0.16.

Page 1466 and 1468. The MUD-SM5.4 rows of Table 1a and Table 2a should be:

Table 1a: 0.12 0.12 0.11 0.11 0.12 0.11 0.13 0.12 0.17

Table 2a: 0.12 0.16 0.11 0.14 0.12 0.15 0.12 0.15 0.13

Page 1467 and 1469. The corrected portions of Table 1b and Table 2b are:

Table 1b	SM5.4	SM5.4PD	Table 2b	SM5.4	SM5.4PD
N-1	-0.78	-0.76	C-2	0.77	0.76
C-2	0.37	0.31	O-2	-0.41	-0.42
H-2	0.22	0.23	N-3	-1.04	-1.01
N-3	-0.49	-0.36	H-3	0.49	0.48
C-4	0.23	0.18	C-4	0.59	0.59
C-5	-0.14	-0.14	O-4	-0.42	-0.42
C-6	0.54	0.52	C-5	-0.26	-0.26
N-6	-0.74	-0.76	C-7	-0.14	-0.14
H-6-1	0.40	0.40	H-7-1	0.11	0.11
H-6-2	0.41	0.41	H-7-2	0.09	0.09
N-7	-0.43	-0.49	H-7-3	0.09	0.09
C-8	0.21	0.24	C-6	0.21	0.20
H-8	0.25	0.26	H-6	0.22	0.22
MUD-MP2	0.12	0.14	MUD-MP2	0.12	0.11
MUD-CM1A	0.07	0.07	MUD-CM1A	0.03	0.02
MUD-SM5.4	0.00	0.03	MUD-SM5.4	0.00	0.01
X-ray	0.11	0.07	X-ray	0.16	0.15
μ (D)	4.61	5.13	μ (D)	6.34	6.33

Corrections for the paper "New Methods for Potential Functions for Simulating Biological Molecules"
Journal de Chimie Physique 1997, **94**, 1448-1481

Gregory D. Hawkins, Christopher J. Cramer, and Donald G. Truhlar
Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431

Some of the results using the SM5.4/AM1 and SM5.4PD/AM1 models for aqueous solvation are affected by a computer coding error in the Fock martix. These errors occur only in undistributed versions of AMSOL. Corrections are given below.

Page 1458. 0.50, 4.3, 0.56, 0.46, 3.6, and 0.50 should be 0.52, 4.1, 0.59, 0.49, 5.5, and 0.55, respectively.

Page 1465. 0.12–0.15 should be 0.07–0.16.

Page 1466 and 1468. The MUD-SM5.4 rows of Table 1a and Table 2a should be:

Table 1a: 0.12 0.12 0.11 0.11 0.12 0.11 0.13 0.12 0.17

Table 2a: 0.12 0.16 0.11 0.14 0.12 0.15 0.12 0.15 0.13

Page 1467 and 1469. The corrected portions of Table 1b and Table 2b are:

Table 1b	SM5.4	SM5.4PD	Table 2b	SM5.4	SM5.4PD
N-1	-0.78	-0.76	C-2	0.77	0.76
C-2	0.37	0.31	O-2	-0.41	-0.42
H-2	0.22	0.23	N-3	-1.04	-1.01
N-3	-0.49	-0.36	H-3	0.49	0.48
C-4	0.23	0.18	C-4	0.59	0.59
C-5	-0.14	-0.14	O-4	-0.42	-0.42
C-6	0.54	0.52	C-5	-0.26	-0.26
N-6	-0.74	-0.76	C-7	-0.14	-0.14
H-6-1	0.40	0.40	H-7-1	0.11	0.11
H-6-2	0.41	0.41	H-7-2	0.09	0.09
N-7	-0.43	-0.49	H-7-3	0.09	0.09
C-8	0.21	0.24	C-6	0.21	0.20
H-8	0.25	0.26	H-6	0.22	0.22
MUD-MP2	0.12	0.14	MUD-MP2	0.12	0.11
MUD-CM1A	0.07	0.07	MUD-CM1A	0.03	0.02
MUD-SM5.4	0.00	0.03	MUD-SM5.4	0.00	0.01
X-ray	0.11	0.07	X-ray	0.16	0.15
μ (D)	4.61	5.13	μ (D)	6.34	6.33