

# Multipole electrostatic potential derived atomic charges in NDDO-methods with *spd*-basis sets

Anselm H. C. Horn · Timothy Clark

Received: 26 January 2006 / Accepted: 23 June 2006 / Published online: 22 August 2006  
© Springer-Verlag 2006

**Abstract** The recently introduced multipole approach for computing the molecular electrostatic potential (MEP) within the semiempirical neglect of diatomic differential overlap (NDDO) framework [Horn AHC, Lin Jr-H., Clark T (2005) *Theor Chem Acc* 114:159–168] has been used to obtain atomic charges of nearly ab initio quality by scaling the semiempirical MEP. The parameterization set comprised a total of 797 compounds and included not only the newly parameterized AM1\* elements Al, Si, P, S, Cl, Ti, Zr, and Mo but also the standard AM1 elements H, C, N, O and F. For comparison, the ZDO-approximated MEP was also calculated analytically in the *spd*-basis. For the AM1\*-optimized structures, single-point calculations at the B3LYP, HF and MP2 levels with the 6-31G(d) and LanL2DZP basis sets were performed to obtain the MEP. The regression analysis of all 12 combinations of semiempirical and ab initio MEP data yielded correlation coefficients of at least 0.99 in all cases. Scaling the analytical and multipole-derived semiempirical MEP by the regression coefficients yielded mean unsigned errors below 2.6 and 1.9 kcal mol<sup>-1</sup>, respectively. Subsequently, for 22 drug molecules from the World Drug Index, atomic charges were computed according to the RESP procedure using XX/6-31G(d) (XX=B3LYP, HF, MP2) and scaled AM1\* multipole

MEP; the correlation coefficients obtained are 0.83, 0.85 and 0.83, respectively. **Figure:** Schematic representation of the atomic charge generation: The molecular electrostatic potential (MEP) is calculated using the AM1\* Hamiltonian; then the semiempirical MEP is scaled to DFT or ab initio level, and atomic charges are generated subsequently by the restraint electrostatic potential (RESP) fit method.

**Keywords** AM1\* · Semiempirical MO theory · Atomic charges · Molecular electrostatic potential · Multipole

## Introduction

Recent improvements in computer hardware have enabled computational chemists to perform molecular-orbital calculations both for very large molecules and for large datasets of smaller ones (see e.g. [1]). The electron density obtained from quantum-mechanical calculations gives complete access to the ground-state properties of the molecules studied [2]. However, the difficulty of understanding its physical and chemical significance calls for simpler chemical concepts, such as net atomic charges, that have been used extensively as derivatives of the electron density.

The molecular electrostatic potential (MEP), rigorously defined in Eq. 1, can be considered a first level of simplification. It represents the fingerprint of a molecule's electrostatic behavior and has gained increasing importance in QSAR/QSPR (quantitative structure–activity and structure–property relationships) in recent years [3, 4]

$$\text{MEP}(\mathbf{r}) = \sum_{i=1}^n \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} - \int_{-\infty}^{+\infty} \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \quad (1)$$

Proceedings of “*Modeling Interactions in Biomolecules II*”, Prague, September 5th–9th, 2005.

**Electronic supplementary material** Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00894-006-0137-8> and is accessible to authorized users.

A. H. C. Horn · T. Clark (✉)  
Computer Chemie Centrum,  
Friedrich-Alexander-Universität Erlangen-Nürnberg,  
Nägelsbachstraße 25,  
91052 Erlangen, Germany  
e-mail: [clark@chemie.uni-erlangen.de](mailto:clark@chemie.uni-erlangen.de)

where  $i$  denotes the  $i$ th atom,  $Z_i$  are the nuclear charges,  $R_i$  the nuclear coordinates and  $\rho(r)$  the electron density.

However, in many cases the MEP is approximated as a classical Coulomb interaction using atomic monopoles, arrays of point charges around the atoms [5, 6] or distributed multipoles, [7] which, although they have found some interest, [8–12] have not been used widely in modeling or in silico screening applications. However, we recently defined an atom-based multipole model [13] within the framework of MNDO-like molecular orbital methods for minimal valence  $spd$ -basis sets. This model provides a natural description of the semiempirical electron density at negligible computational cost because it is derived directly from the multipole formalism used to approximate the two-electron integrals, [14, 15] and is therefore well suited for applications in cheminformatics.

However, multipole descriptions may still be too computationally demanding for large scale virtual high-throughput screening and classical molecular-dynamics simulations (e.g. AMBER [16]). This is where atomic partial charges come into play. It is often argued that a good choice of atomic charges is one that reflects the electrostatic properties of the molecule as closely as possible. This approach was first investigated by Momany [17] and Smit et al. [18] and further refined by Cox and Williams, [19] and other groups, [20–23] who calculated the MEP at ab initio levels of theory. However, the large computational effort required for such techniques makes the use of semiempirical neglect-of-diatomic-differential-overlap (NDDO) techniques particularly attractive, especially for large molecules or huge data sets [1].

Unfortunately, no single method has been adopted for calculating the MEP within NDDO theory, so that several methods have been developed and subsequently used to generate partial atomic charges. Most common is the quasi ab initio approach, [24, 25] in which all two-electron integrals present in the NDDO approximation are considered, but which requires a deorthogonalization of the basis. However, despite further research, [26] this approach still requires a computational effort larger than the original NDDO-calculation. Reynolds et al. [27] pointed out that applying the zero-differential overlap (ZDO) approximation strictly by ignoring the two-atom blocks of the density matrix, and calculating the remaining one-electron integrals for an  $sp$ -basis set explicitly leads to good results at a far lower computational cost. A similar ZDO approach was taken by Ford and Wang [28] and enhanced later by Bakowies and Thiel for partial charges in quantum mechanical/molecular mechanical (QM/MM) calculations [29]. Their technique used the multipole approximation for two-electron integrals [15] to evaluate the one-electron integrals necessary for the MEP (see Theory), but relies on additional atomic parameters designed to fit the results to

ab initio reference values as well as possible. In the same vein, we developed a fast model for NDDO wavefunctions that approximates the electrostatic potential (MEP) by point charges at the centers of charge density given by the natural atomic orbitals [5, 6, 30]. This technique is computationally very efficient but is limited to an  $sp^n$ -hybrid atomic orbitals to determine the positions and magnitudes of the charges without resorting to an integration step.

In our previous work, [13] we found that the MEP values from the atom-based multipole model showed a good linear correlation to those from ab initio calculations, although only a limited number of sample molecules were used. In this study, we examine whether these promising results can be extended to a larger number of molecules, including all elements parameterized for the AM1\* (the AM1 Hamiltonian for H, C, N, O and F extended with  $d$ -orbitals for the heavier elements) [31, 32] Hamiltonian. For comparison and verification of our results, we have calculated analytical integrals (for a Slater  $spd$ -basis) for the MEP calculation according to the ZDO approach [27]. In the second part of this work, partial atomic charges derived from the scaled multipole MEP according to the procedure recommended in the AMBER tutorial, [16] are compared to their pure ab initio counterparts.

## Theory

### Two-electron integrals in MNDO-like methods

Semiempirical methods of the MNDO type use Slater–Zener orbitals with quantum numbers  $n, l, m$  as their basis set (defined in the usual manner [33]). Because of the complexity of the mathematical expressions, these are not used directly for integral evaluation in current semiempirical techniques. In particular, for the two-electron integrals  $(\mu\nu|\lambda\sigma)$  (Eq. (2);  $e$  is the electronic charge,  $r_{12}$  is the inter-electronic distance and  $d\tau_1$  and  $d\tau_2$  denote the integration volume elements over the coordinates of electron 1 and 2) an elegant simplification [15] is used. In the NDDO-approximation [33] the three- and four-center two-electron integrals are neglected; the remaining two-center two-electron integrals  $(\mu^A\nu^A|\lambda^B\sigma^B)$  describe the electrostatic interactions between the charge distributions of electron 1,  $\rho^{\mu\nu}=\mu\nu$  on Atom A, and electron 2,  $\rho^{\lambda\sigma}=\lambda\sigma$  on Atom B.

$$(\mu\nu|\lambda\sigma) = \int \int \varphi_\mu(1)\varphi_\nu(1) \frac{e^2}{r_{12}} \varphi_\lambda(2)\varphi_\sigma(2) d\tau_1 d\tau_2 \quad (2)$$

Such a charge distribution  $\rho$  can be expressed in terms of a finite linear combination of normalized real spherical harmonics, [15] which finally leads to the result that the

interaction of two such charge distributions, and therefore the two-center two-electron integral  $(\mu\nu|\lambda\sigma)$ , can be evaluated as a sum over classical multipole interactions (Eq. 3) [15]

$$(\mu\nu|\lambda\sigma) = \sum_{l_1 l_2} \sum_{m=-|l_{\min}|}^{|l_{\min}|} [M_{l_1 m}^{\mu\nu} M_{l_2 m}^{\lambda\sigma}] \quad (3)$$

with  $l_{\min}$  being the smaller of  $l_1$  and  $l_2$ .

### One-electron integrals in MNDO-like methods

In MNDO-like methods the one-electron integrals describing the electron-core interaction are also calculated using this approach by mimicking the core charge of the nucleus by a spherical symmetric charge distribution [34]. Thus, the interaction of an electron occupying the atomic orbitals  $\phi_\mu$  and  $\phi_\nu$  of atom A with the atomic nucleus of atom B can be approximated by

$$Z^B \int_{-\infty}^{+\infty} \varphi_\mu \frac{1}{|\mathbf{r}_B - \mathbf{r}|} \varphi_\nu d\mathbf{r} \approx Z^B(\mu\nu|ss) \quad (4)$$

where  $Z^B$  is the nuclear charge of atom B.

### Atom-based multipole moments

A minimal *spd*-basis yields 45 unique charge distributions  $\mu\nu$  with 96 non-vanishing multipoles  $M_{\lambda\mu}^{\mu\nu}$ . Of those, all

octupole and hexadecapole moments are neglected, [35] so that only the remaining 52 monopoles, dipoles, and quadrupoles are retained. This treatment is identical to that introduced for MNDO/d [35]. Figure 1 shows the point-charge configurations arising from the multipoles up to  $l=2$  schematically; the separation between two adjacent point charges of opposite sign is  $2D$  (see [35, 36]). The atomic monopole, dipole and quadrupole moments can be calculated easily from this array of multipole point charges according to standard formulae [7]. Thus, the anisotropy of the electron density around any atom can be stored in ten floating point numbers.

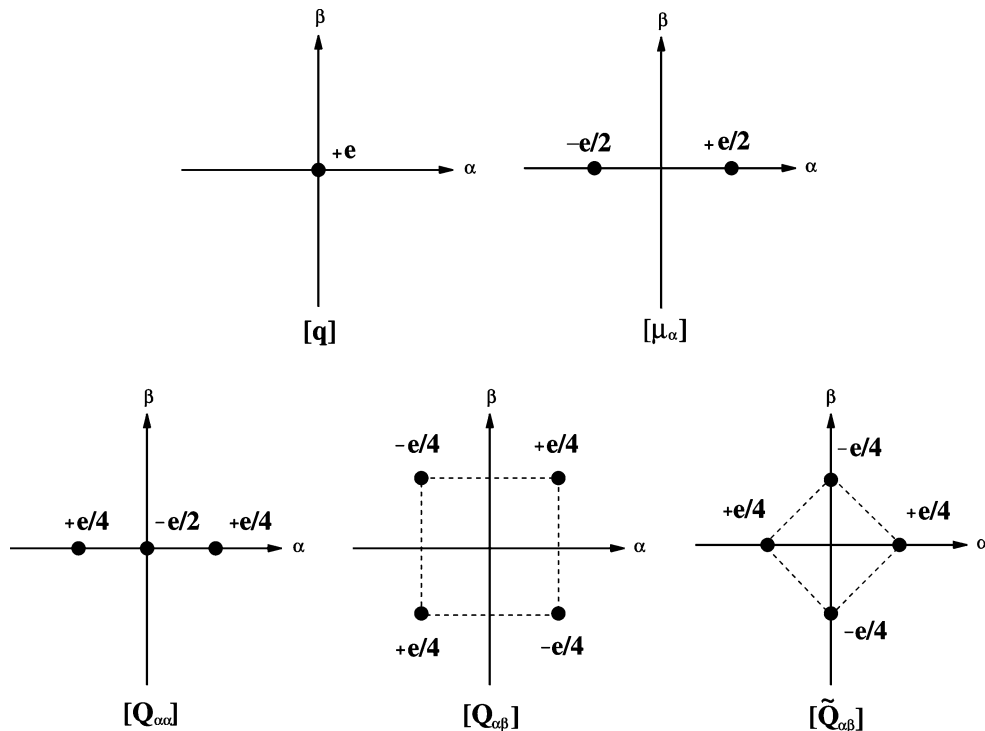
### Molecular electrostatic potential

Having defined the atom-based multipoles in the framework of semiempirical MO theory, we can now proceed to calculate the MEP using them. In terms of multipoles, the MEP is given according to

$$\text{MEP}(\mathbf{r}) = \hat{q} \left( \frac{1}{R} \right) - \hat{\mu}_\alpha \nabla_\alpha \left( \frac{1}{R} \right) + \frac{1}{3} \hat{\Theta}_{\alpha\beta} \nabla_\alpha \nabla_\beta \left( \frac{1}{R} \right) \quad (5)$$

where  $\hat{q}$ ,  $\hat{\mu}_\alpha$ , and  $\hat{\Theta}_{\alpha\beta}$  are the operators for monopole, dipole and quadrupole, respectively,  $R$  is the distance between the multipole center and the MEP point, and  $\nabla$  is the Nabla operator. Explicit formulae are given elsewhere

**Fig. 1** Point charge configurations ( $\alpha, \beta=x, y, z$ ). The separation of two neighboring point charges of opposite sign is  $2D$ . The details of the multipole formalism used in MNDO/d and AM1\* are given in References [15, 35, 36]



[7, 13]. With the atom-based multipoles, the MEP at a given point in space can be calculated easily by summing up the charge, dipole and quadrupole contributions of all atoms [7].

Within the SCF formalism, the electrostatic potential defined in Eq. 1 can be rewritten in terms of atomic orbitals (AOs):

$$\text{MEP}(\mathbf{r}) = \sum_{i=1}^n \frac{Z_i}{|\mathbf{r}_i - \mathbf{r}|} - \sum_{\mu\nu} P_{\mu\nu} \int_{-\infty}^{+\infty} \frac{\varphi_{\mu}(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} \varphi_{\nu}(\mathbf{r}') d\mathbf{r}' \quad (6)$$

The first term again represents the classical Coulomb interaction of the nuclei; the second describes the electronic contribution, with  $P_{\mu\nu}$  being an element of the density matrix. Note that the first term does not use the usual modifications [34, 37, 38] to the core-core potential in our treatment.

Equations 4–6 show that in NDDO methods the influence of an electron distribution on the MEP at a given point in space can be calculated using the appropriate two-electron integral expressions described by multipole interactions. Note that this approach uses the ZDO approximation implicitly. However, such a technique yields MEP values that are far too negative in the vicinities of the atomic nuclei, and therefore the first term of Eq. 6 was substituted by a new semiempirical function, in analogy to the MNDO core-repulsion functions [34], that contains two additional atom-specific parameters [28, 29].

A third way of calculating the MEP within NDDO methods was suggested by Reynolds et al. [27]. They apply the ZDO approximation to Eq. 5 and compute the remaining integrals analytically. The authors report an implementation for standard MNDO [34], AM1 [37] and MINDO/3 [39], with an *sp*-basis set. In order to compare the results of our multipole MEP, we also implemented this analytical approach including *d*-orbitals for AM1\* [31, 32]. Formulae are given in detail elsewhere [40].

#### Fit of atomic charges to MEP

The best least-squares fit of the charges to the electrostatic potential is obtained by the method of Lagrangian multipliers. The task is to minimize the difference of the quantum mechanically computed MEP and that caused by the atomic charges at a large number of points  $j$  around the molecule, given by

$$t(q_1, \dots, q_n) = \sum_{j=1}^{n_{\text{Points}}} \left( \text{MEP}_j^{\text{QM}} - \text{MEP}_j^{\text{classical}}(q_1, \dots, q_n) \right)^2 \quad (7)$$

We now define a constraint function that requires the sum of all atomic charges  $q_i$  to be equal to the total molecular charge

$$u(q_1, \dots, q_n) = \left( \sum_{i=1}^n q_i \right) - q_{\text{tot}} = 0 \quad (8)$$

The combined function containing the Lagrangian multiplier  $\lambda$  then reads

$$z(q_1, \dots, q_n) = t(q_1, \dots, q_n) + \lambda u(q_1, \dots, q_n) \quad (9)$$

Limiting points of this function are then found by setting its first derivative to zero, which yields a set of  $n+1$  equations in  $n+1$  unknowns. The solution of the corresponding matrix equation represents the atomic charges. Three common MEP charge methods, Merz–Kollman, [21] Chelp, [23] and ChelpG, [22] use this fitting scheme directly, differing only in the way the surface MEP points are created. The restrained electrostatic potential fit (RESP) method [41, 42] adds additional hyperbolic restraints to Eq. 8 that reduce the overall magnitude of the fitted charges, most often the statistically poorly determined charges of hidden atoms. These matrix equations, however, must then be solved iteratively.

#### Computational details

A set of 797 compounds (neutral and ionic species) containing the elements H, C, N, O, F, Al, Si, P, S, Cl, Ti, Zr, and Mo was retrieved from our parameterization data base. [31, 32] Additionally, 22 structures from a subset of the World Drug Index (WDI) [43] containing second-row elements were selected. The final dataset of 775 structures was then subjected to geometry optimization using the AM1\* Hamiltonian as implemented in the semiempirical program package VAMP [44]. The true nature of the minima was ensured by subsequent frequency calculations.

Following our established procedure [30], each molecule was then surrounded by a large number of points outside the van-der-Waals surface, at which the MEP was calculated semiempirically with AM1\* with both the multipole method and the analytical approaches. The same arrays of points were then used for MEP calculations at the B3LYP [45–47], HF, and MP2 [48] levels of theory combined with the 6-31G(d) [49] and LanL2DZP [50–54] basis sets using GAUSSIAN03 [55]. The choice for these methods and the 6-31G(d) basis originated from their traditional popularity in many areas of computational chemistry; LanL2DZP was selected in order to include a further basis set of similar size, and well documented performance, especially for the B3LYP functional (eg. [56]). Linear regression analyses of

the 6,683,318 MEP points were carried out for all 12 combinations of semiempirical and ab initio/DFT methods.

For the 22 WDI molecules, RESP charges were determined using the parameterization tool *antechamber* from the AMBER8 suite of programs [16] in the standard recommended two-step process [42], firstly with the three sets of 6-31G(d) MEP values, and secondly with the AM1\* values scaled by the appropriate linear regression coefficients. Thus, six sets of atomic charges were finally obtained.

## Results and discussion

### Regression analyses

The results of the linear regression analysis for all 12 combinations with over 6.6 million data points each are summarized in Table 1. Slope ( $m$ ) and intercept ( $t$ ) of the regression equations are of the same order of magnitude and share the same sign, as expected; for HF the slope is marginally steeper and the intercept smaller than for DFT and MP2. Both values are systematically larger for all multipole MEP combinations than for the analytical ones. Alhambra et al. [25], whose data set comprised 45 molecules with H, C, N, O, and F, correlated only the MEP minima energies of AM1 calculations to HF/6-31G(d) and obtained a slope of 0.6 for the Ford and Wang-approach and of 0.8 for the quasi-ab initio approach.

Mean unsigned errors (MUE) are  $2.6 \text{ kcal mol}^{-1} \text{ e}^{-1}$  for the analytical MEP approach and  $1.9 \text{ kcal mol}^{-1} \text{ e}^{-1}$  for the multipole ansatz, showing a systematic difference of about

$0.7 \text{ kcal mol}^{-1} \text{ e}^{-1}$ . For their data set of 24 molecules containing atoms H, C, N, and O, Bakowies and Thiel [29] used distinct shells of surface points for the calculation of their parameterized MEP; the radii of the shells were determined by scaling the van-der-Waals radii by a parameter  $f$ . For three areas of scaling ( $f=0.6-1.2, 1.4-2.0, 3.0$ ) with a total of roughly 20,000 surface points they obtained MUEs for AM1 of 5.8, 2.0, and  $0.6 \text{ kcal mol}^{-1} \text{ e}^{-1}$  and for MNDO 6.9, 2.4,  $0.7 \text{ kcal mol}^{-1} \text{ e}^{-1}$  relative to HF/6-31G(d). However, their dataset did not contain transition-metal elements and was parameterized. The standard deviation between the semiempirical and ab initio MEPs shown in Table 1 is  $0.7-0.9 \text{ kcal mol}^{-1} \text{ e}^{-1}$  larger for the analytical than for the multipole MEP, but rather constant within one semiempirical method.

The last two columns of Table 1, however, reveal some inconsistencies of both semiempirical MEP methods investigated. The largest positive and negative errors can be as high as  $100 \text{ kcal mol}^{-1}$ . In Table S1 of the Supplementary Material the MUE values are given for each molecule. From that one can see that the large error contributions stem from rather small molecules. The diatomic ZrC, for instance, has MUE values with respect to HF of more than  $10 \text{ kcal mol}^{-1} \text{ e}^{-1}$  and PO even higher errors of  $20 \text{ kcal mol}^{-1} \text{ e}^{-1}$  and more with respect to MP2. Most striking, the two isomers of  $\text{S}_2\text{F}_2$ , linear and trigonal, show average MUEs for the linear form ( $1.2-2.3 \text{ kcal mol}^{-1} \text{ e}^{-1}$ ), but large ones for the other ( $12.2 \text{ kcal mol}^{-1} \text{ e}^{-1}$  at worst).

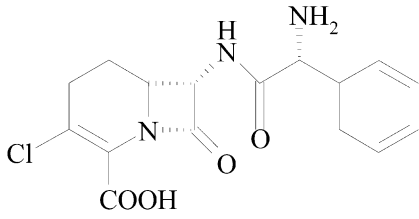
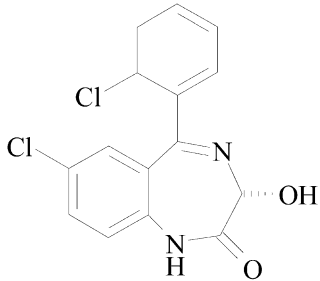
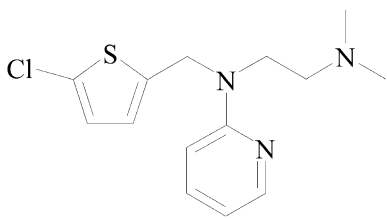
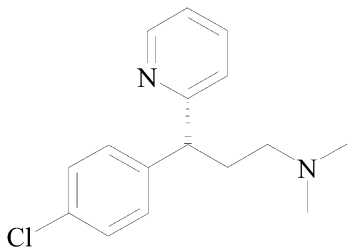
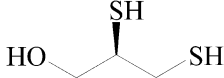
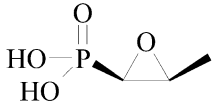
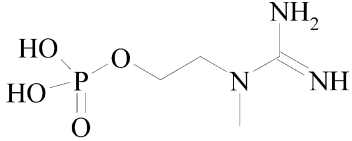
The problems have their origin in the semiempirical description of the molecule per se. The NDDO-approximation is most severe for small molecules and becomes better as the size of the molecule increases (because two-electron integrals over large distances can also be neglected in

**Table 1** Regression coefficients and mean unsigned errors (MUE), standard deviation ( $\sigma$ ), and most positive and negative errors ( $\text{Err}^+$ ,  $\text{Err}^-$ ) of the scaled AM1\* electrostatic potential in comparison to ab initio/DFT values

AM1*	Ab initio/DFT	$m^a$	$t^a$	MUE ( $\text{kcal mol}^{-1} \text{ e}^{-1}$ )	$\sigma$	$\text{Err}^+$	$\text{Err}^-$
Analytical	B3LYP/6-31G(d)	0.97764	-0.04425	2.52	3.90	60.5	-55.2
	B3LYP/LanL2DZP	0.97828	-0.04350	2.55	3.92	60.2	-56.7
	HF/6-31G(d)	0.98468	-0.00717	2.54	3.93	61.5	-63.4
	HF/LanL2DZP	0.98453	-0.00258	2.60	4.00	66.9	-64.0
	MP2/6-31G(d)	0.97989	-0.04272	2.54	3.95	61.7	-79.0
	MP2/LanL2DZP	0.98046	-0.03826	2.56	3.98	87.1	-104.0
Multipole	B3LYP/6-31G(d)	0.98260	-0.06241	1.87	3.08	53.1	-61.8
	B3LYP/LanL2DZP	0.98346	-0.06146	1.90	3.07	51.0	-63.3
	HF/6-31G(d)	0.98976	-0.02537	1.90	3.09	56.3	-70.6
	HF/LanL2DZP	0.98993	-0.02048	1.94	3.13	57.6	-70.6
	MP2/6-31G(d)	0.98485	-0.06091	1.90	3.14	65.0	-82.0
	MP2/LanL2DZP	0.98561	-0.05629	1.92	3.15	88.8	-91.2

$$^a \text{MEP}^{\text{abinitio/DFT}} = m \text{MEP}^{\text{AM1*}} + t$$

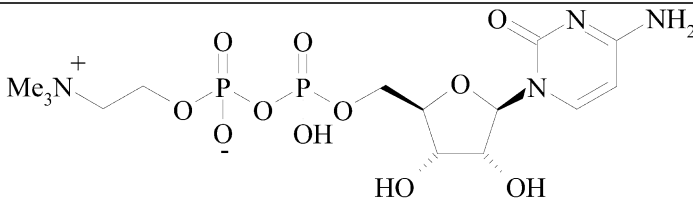
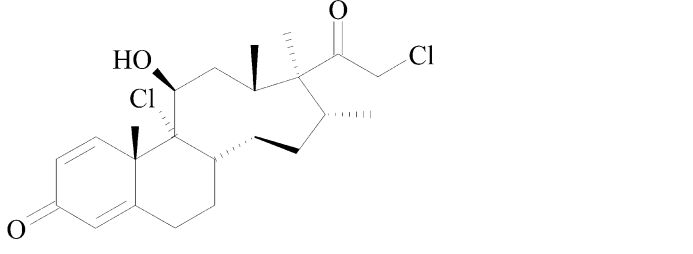
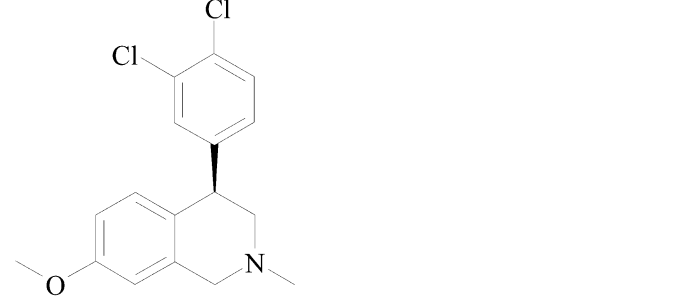
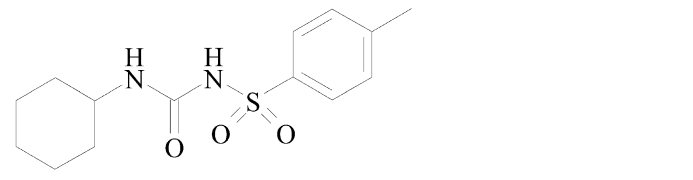
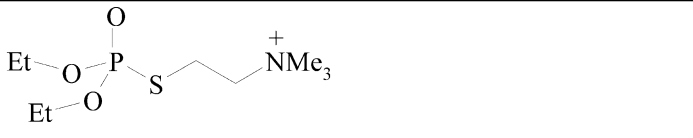
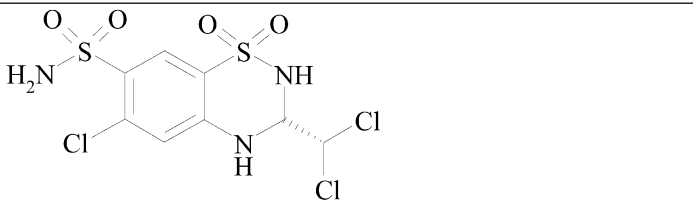
**Table 2** Overview of the 22 WDI compounds used for the charge derivation

n	Name	Structure
1	Loracarbef	
2	Lorazepam	
3	Chloropyrilene	
4	Chlorophenamine	
5	Dimercaprol	
6	Fosfomicin	
7	Creatinolphosphate	

accurate techniques). Structures with uncommon binding patterns are also more prone to errors in parameterized methods, as they are unlikely to be included in the parameterization data set. This situation shows once more

that even modern semiempirical methods cannot be used as black box methodologies, but that critical assessment of the results is crucial. However, the large errors occur for molecules that would not normally be considered using

**Table 2** (continued)

8	Citicoline	
9	Mometasone	
10	Diclofenine	
11	Glycyclamide	
12	Ecothiopate	
13	Trichlormethiazide	

semiempirical techniques and are therefore of little practical importance.

#### Atomic charges

Because all MEP linear regressions yielded correlations of similar quality, we decided to use just three of the 12 possible combinations for generating atomic charges. The

XX/6–31G(d)-AM1\* multipole was therefore compared with XX/6-31G(d) (XX=B3LYP, HF, MP2) because these are still the methods of choice in most force-field parameterizations (e.g. [57, 58]). Because the Hartree-Fock MEP is usually “more polar” than that obtained from methods that consider dynamic correlation; it is often preferred for force fields that are designed to reproduce results in condensed phases.

Table 2 (continued)

14	Fosinopril	
15	Thiamazole	
16	Fosfosal	
17	Psilocybine	
18	Fosfestrol	
19	Clothiapine	

Table 2 gives names and structures of the 22 compounds selected from a subset of the WDI [43]. The molecules were selected to provide as many *d*-elements as possible in the WDI-subset.

Note that we do not intend to test the quality of the atomic charges themselves, which are in any case neither uniquely defined nor measurable. Our aim is simply to test the ability of the fast semiempirical technique to provide



Table 2 (continued)

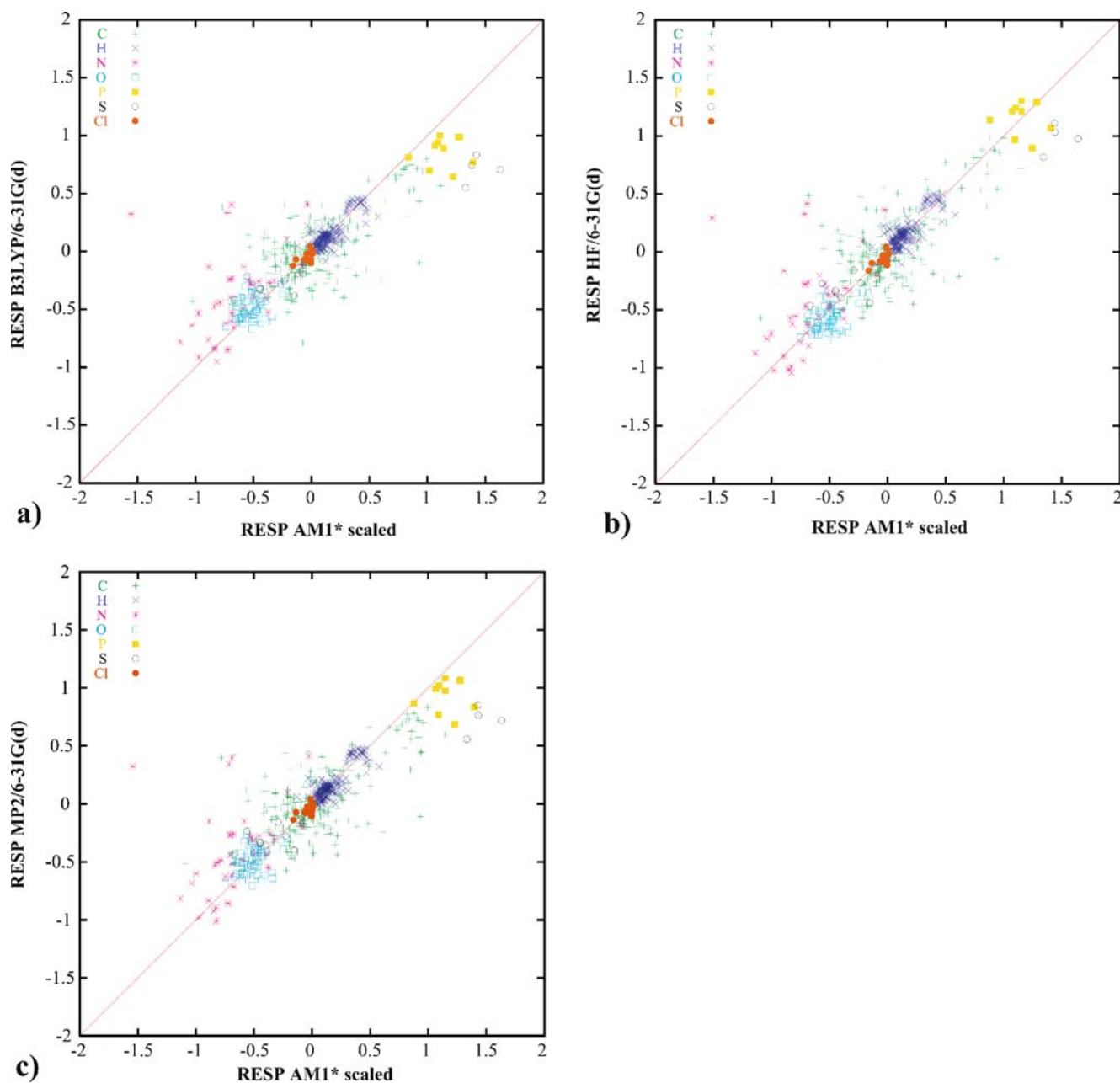
20	Triclofos	
21	Cefonicid	
22	Butoconazole	

MEP-derived charges comparable to those obtained from XX/6-31G(d). The correlation results and MUE values from Table 1, however, tell us that the agreement between the two sets of MEPs is satisfactory.

The ability of the RESP charges to reproduce the MEP used as input for the fit is measured by the root mean square (RMS) error and the relative RMS (RRMS), both listed in Table 3 for all 22 molecules of the WDI subset

**Table 3** RMS (in kcal mol<sup>-1</sup> e<sup>-1</sup>) and RRMS values for the RESP fit of the WDI subset (see text for details). AM1\* denotes multipole MEP calculated with the AM1\* Hamiltonian and scaled afterwards to respective ab initio/DFT values according to the regression formula (cf Table 1)

<i>n</i>	B3LYP/6-31G(d)		AM1*		HF/6-31G(d)		AM1*		MP2/6-31G(d)		AM1*	
	RMS	RRMS	RMS	RRMS	RMS	RRMS	RMS	RRMS	RMS	RRMS	RMS	RRMS
1	1.10	0.10	1.42	0.14	1.27	0.10	1.44	0.14	1.15	0.11	1.44	0.14
2	1.15	0.12	1.26	0.14	1.29	0.12	1.27	0.14	1.23	0.13	1.27	0.14
3	1.27	0.27	1.52	0.25	1.41	0.26	1.51	0.25	1.34	0.27	1.50	0.25
4	1.23	0.19	1.28	0.17	1.30	0.18	1.27	0.16	1.29	0.19	1.27	0.16
5	2.05	0.27	2.37	0.30	2.16	0.27	2.38	0.29	2.11	0.27	2.37	0.30
6	1.25	0.12	1.58	0.15	1.44	0.13	1.59	0.15	1.31	0.13	1.59	0.15
7	1.66	0.13	1.70	0.14	1.75	0.13	1.71	0.14	1.72	0.13	1.71	0.14
8	1.15	0.06	2.09	0.10	1.36	0.07	2.16	0.10	1.19	0.06	2.15	0.10
9	1.00	0.09	1.42	0.15	1.17	0.10	1.49	0.15	1.07	0.11	1.49	0.15
10	1.40	0.26	1.37	0.20	1.47	0.23	1.35	0.19	1.46	0.25	1.34	0.19
11	1.11	0.09	1.37	0.13	1.22	0.09	1.38	0.13	1.15	0.10	1.38	0.13
12	1.35	0.02	2.21	0.04	1.52	0.02	2.10	0.03	1.39	0.02	2.17	0.04
13	1.61	0.11	1.49	0.12	1.81	0.12	1.50	0.12	1.66	0.12	1.49	0.12
14	0.93	0.08	1.47	0.13	1.05	0.08	1.65	0.14	1.00	0.09	1.66	0.14
15	1.61	0.11	1.89	0.11	1.95	0.12	1.91	0.11	1.63	0.11	1.90	0.11
16	1.00	0.08	1.32	0.10	1.17	0.08	1.34	0.10	1.07	0.08	1.33	0.10
17	1.27	0.11	1.44	0.10	1.31	0.10	1.45	0.10	1.31	0.11	1.44	0.10
18	1.16	0.10	1.72	0.14	1.27	0.09	1.82	0.14	1.20	0.10	1.82	0.14
19	1.36	0.27	1.84	0.27	1.51	0.26	1.83	0.26	1.44	0.27	1.82	0.26
20	1.46	0.14	1.19	0.11	1.66	0.14	1.20	0.11	1.54	0.14	1.19	0.11
21	1.68	0.13	2.51	0.20	1.82	0.12	2.76	0.21	1.79	0.13	2.77	0.21
22	1.49	0.17	1.24	0.13	1.74	0.18	1.28	0.13	1.58	0.17	1.27	0.13



**Fig. 2** Correlation of RESP charges derived from XX/6-31G(d) (XX=B3LYP, HF, MP2) results with those from scaled AM1\* multipole MEP calculations. The WDI subset used comprised 367

H, 283 C, 40 N, 78 O, 11 P, 13 S, and 19 Cl atoms (depicted using the colors and shapes given in the legends). Correlation coefficient: **a** 0.83, **b** 0.85, **c** 0.83

(output by the program *antechamber* from the AMBER suite [16]). The RMS values for the original *ab initio*/DFT and scaled semiempirical MEP are comparable and range from 0.9 to 2.1 kcal mol<sup>-1</sup> e<sup>-1</sup> and from 1.2 to 2.8 kcal mol<sup>-1</sup> e<sup>-1</sup>, whereas the RRMS values lie in the range of 2–24 % and 4–30 %, respectively.

Bakowies and Thiel [29] found mean RMS and RRMS errors of parameterizing the HF/6-31G(d) MEP directly by means of an auxiliary function to be 2.6 kcal mol<sup>-1</sup> e<sup>-1</sup> and 25%, respectively. However, because we performed the

RESP fit to a scaled semiempirical calculation, the numbers cannot be compared directly, but rather indicate the order of magnitude of such error values.

The correlation between the three pairs of RESP charge sets (XX/6-31G(d) vs. AM1\* scaled) is shown in Fig. 2. The atomic-charge values cover the complete range from -1 to +1, some even fall beyond it. This is a well known shortcoming of MEP-derived charges, especially for buried atoms, and occurs here for both the scaled multipole and the *ab initio*/DFT MEP derived charges, although the latter

have generally smaller values. In particular, B3LYP/6-31G(d) RESP charges are found to be less polar than their HF and MP2 counterparts at the same basis set level, in accord with the literature [59].

The striking outlier in all three plots of Fig. 2 at approximately  $-1.5$ ;  $+0.4$  corresponds to the nitrogen of the  $\text{N}(\text{Me})_3^+$  group in compound #12. However, this large discrepancy is put into perspective by adding together all atomic charges around this buried atom ( $-\text{CH}_2-\text{N}(\text{Me})_3$ ), which yields a group charge of  $+0.9$  for XX/6-31G(d) and  $+0.7$  for scaled AM1\*. This charge variance originates from the fact that *ab initio*/DFT and AM1\* electron densities are similar but not identical, leading to potentially large differences in fitted charges that are poorly defined numerically. This conclusion is supported by the fact that the AM1\*-derived nitrogen charges of the WDI subset are all negative, whereas some are positive for all XX/6-31G(d) cases, although this contradicts chemical intuition.

A closer inspection of the RESP charges reveals that hydrogen and carbon atoms often differ in the sign of their charges, which again may just be an artefact of the fitting procedure: As discussed above for the trimethylammonium group, a similar group total charge and thus a surrounding MEP may be obtained by different combinations of atomic charges. All oxygen and sulfur charges, however, show the same polarity, although in the latter case the semiempirical ones are generally found to be larger. All methods assign charges of around  $+1$  to the phosphorous atoms. Compared to B3LYP and MP2, the AM1\* counterparts are again more polar; no such trend can be observed for the HF results. Finally, the RESP charges of chlorine are very small for all six sets, they lie in the range approximately  $-0.1$  to  $+0.1$ . Here, the XX/6-31G(d) methods yield positive charges for the three Cl atoms of compound #20 ( $+0.04$ ), whereas the scaled semiempirical approach produces just one positively charged chlorine in compound #1 ( $+0.01$ ).

## Summary and outlook

In this work we have demonstrated that our recently introduced NDDO multipole ansatz [13] is well suited for obtaining electrostatic potentials of nearly *ab initio*/DFT quality from semiempirical calculations on the AM1\* level [31] *via* simple linear scaling, and even outperforms the analytical integral approach (cf. Table 1). Because our multipole ansatz is derived from NDDO theory directly, no additional parameters are needed, in contrast to common semiempirical MEP approaches [28, 29]. Therefore, as soon as new NDDO parameters for an element become available, the MEP may be calculated readily, especially for transition metals [32].

We emphasize here that we have used a completely unmodified ZDO-procedure (i.e. the contributions of the nuclei to the MEP are treated by a simple point-charge Coulomb relationship and those of the electron density by the multipole treatment used for one- and two-electron integrals in MNDO-like methods) and have merely scaled these MEPs to obtain good agreement with *ab initio* and DFT-results.

In the second part of this work we showed that appropriately scaled AM1\* MEP derived RESP [42] charges resemble their XX/6-31G(d) (XX=B3LYP, HF, MP2) analogues reasonably well (cf Fig. 2). Of course, there is a trade-off between accuracy and speed: Despite some difference in electron density between NDDO and *ab initio*/DFT methods, the overall similarity is preserved; on the other hand, semiempirical calculations enjoy a speed-up of  $\sim$ at least a factor of  $10^3$ . This makes them and our charge-generation method ideally suited for applications with either a large number of compounds (e.g. high-throughput screening) or huge compounds (e.g. force-field parameter generation for enzyme inhibitors).

**Acknowledgements** The authors thank Dr. G. Wellein, G. Hager, and T. Zeiser from the Regionales Rechenzentrum Erlangen (RRZE) for providing a specially optimized AMBER compilation, and the Bioinformatics group (Prof. Dr. H. Sticht) for support with AMBER 8.

## References

1. Beck B, Horn A, Carpenter J, Clark T (1998) *J Chem Inf Comput Sci* 38:1214–1217
2. Hohenberg P, Kohn W (1964) *Phys Rev* 136:B864–B871
3. Murray JS, Lane P, Brinck T, Paulsen K, Grince ME, Politzer P (1993) *J Phys Chem* 97:9369–9373
4. Murray JS, Politzer P (1998) *J Mol Struct (Theochem)* 425:107–114
5. Rauhut G, Clark T (1993) *J Comput Chem* 14:503–509
6. Beck B, Rauhut G, Clark T (1994) *J Comput Chem* 15:1064–1073
7. Stone AJ (2002) *The Theory of Intermolecular Forces*. Clarendon, Oxford
8. Chipot C, Ángyán JG, Ferenczy GG, Scheraga HA (1993) *J Phys Chem* 97:6628–6636
9. Ferenczy GG (1991) *J Comput Chem* 12:913–917
10. Ferenczy GG, Winn PJ, Reynolds CA (1997) *J Phys Chem A* 101:5446–5455
11. Winn PJ, Ferenczy GG, Reynolds CA (1997) *J Phys Chem A* 101:5437–5445
12. Winn PJ, Ferenczy GG, Reynolds CA (1999) *J Comput Chem* 20:704–712
13. Horn AHC, Lin J-H, Clark T (2005) *Theor Chem Acc* 114:159–168
14. Thiel W (1988) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of Computational Chemistry*. Wiley, Chichester
15. Dewar MJS, Thiel W (1977) *Theor Chim Acta* 46:89–104
16. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM Jr, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G,

- Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) AMBER 8. University of California, San Francisco
17. Momany FA (1978) *J Phys Chem* 82:592–601
  18. Smit PH, Derissen JL, Duijneveldt JL (1979) *Mol Phys* 37:521–539
  19. Cox SR, Williams DE (1981) *J Comput Chem* 2:304–323
  20. Singh UC, Kollman PA (1984) *J Comput Chem* 5:129–145
  21. Besler BH, Merz KM Jr, Kollman PA (1990) *J Comput Chem* 11:431–439
  22. Breneman CM, Wiberg KB (1990) *J Comput Chem* 11:361–371
  23. Chirlian LE, Francl MM (1987) *J Comput Chem* 8:894–905
  24. Merz KM Jr, Besler BH (1990) *QCPE Bull* 10:15
  25. Alhambra C, Luque FJ, Orozco M (1994) *J Comput Chem* 15:12–22
  26. Cummings PL, Gready JE (1990) *Chem Phys Lett* 174:355–360
  27. Reynolds CA, Ferenczy GG, Richards WG (1992) *J Mol Struct (Theochem)* 256:249–269
  28. Ford GP, Wang B (1993) *J Comput Chem* 14:1101–1111
  29. Bakowies D, Thiel W (1996) *J Comput Chem* 17:87–108
  30. Beck B, Glenn RC, Clark T (1997) *J Comput Chem* 18:744–756
  31. Winget P, Horn AHC, Selçuki C, Martin B, Clark T (2003) *J Mol Model* 9:408–414
  32. Winget P, Clark T (2005) *J Mol Model* 11:439–456
  33. Pople JA, Beveridge DL (1970) *Approximate Molecular Orbital Theory*. McGraw-Hill, New York
  34. Dewar MJS, Thiel W (1977) *J Am Chem Soc* 99:4899–4907
  35. Thiel W, Voityuk AA (1992) *Theor Chim Acta* 81:391–404
  36. Thiel W, Voityuk AA (1996) *Theor Chim Acta* 93:315
  37. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902–3909
  38. Stewart JJP (1989) *J Comput Chem* 10:209–220
  39. Bingham RC, Dewar MJS, Lo DH (1975) *J Am Chem Soc* 97:1285–1293
  40. Kuznetsov A, Reinhold J (1984) *Wiss Z Karl-Marx-Univ Leipzig, Math-Naturwiss R* 33:397–403
  41. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) *J Phys Chem* 97:10269–10280
  42. Cornell WD, Cieplak P, Bayly CI, Kollman PA (1993) *J Am Chem Soc* 115:9620–9631
  43. Brüstle M, Beck B, Schindler T, King W, Mitchell T, Clark T (2002) *J Med Chem* 45:3345–3355
  44. Clark T, Alex A, Beck B, Burkhardt F, Chandrasekhar J, Gedeck P, Horn AHC, Hutter M, Martin B, Rauhut G, Sauer W, Schindler T, Steinke T (2003–2005) VAMP 8. Accelrys Inc Erlangen, San Diego. <http://www.accelrys.com>
  45. Becke AD (1993) *J Chem Phys* 98:5648–5652
  46. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785–789
  47. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) *J Phys Chem* 98:11623–11627
  48. Moller C, Plesset MS (1934) *Phys Rev* 46:618–622
  49. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1996) *Ab Initio Molecular Orbital Theory*. Wiley, New York
  50. Huzinaga S (1984) *Gaussian Basis Sets for Molecular Calculations*. Elsevier, Amsterdam
  51. Dunning TH Jr, Hay PJ (1976) In: Schaefer HF III (eds) *Modern Theoretical Chemistry*. Plenum, New York
  52. Hay PJ, Wadt WR (1985) *J Chem Phys* 82:270–283
  53. Wadt WR, Hay PJ (1985) *J Chem Phys* 82:284–298
  54. Hay PJ, Wadt WR (1985) *J Chem Phys* 82:299–310
  55. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniel AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) *Gaussian 03, Revision C.02*. Gaussian, Wallingford CT
  56. Weber CF, Puchta R, van Eikema Hommes NJR, Wasserscheid P, van Eldik R (2005) *Angew Chem Int Ed* 44:6033–6038
  57. Homeyer N, Horn AHC, Lanig H, Sticht H (2006) *J Mol Model* 11:281–289
  58. Wartha F, Horn AHC, Meiselbach H, Sticht H (2005) *J Chem Theory Comput* 1:315–324
  59. St-Amant A, Cornell WD, Kollman PA, Halgren TA (1995) *J Comput Chem* 16:1483–1506