

Tracking molecular charge distribution along reaction paths with atomic multipole moments

Karol M. Langner¹ · Wiktor L. Beker² · Edyta Dyguda-Kazimierowicz² · W. Andrzej Sokalski²

Received: 18 November 2015 / Accepted: 25 January 2016 / Published online: 10 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We explore the idea of supplementing partial atomic charges with cumulative multipole moments for modeling electrostatic effects during chemical reactions. To this end, we investigate the first stage of alkaline hydrolysis of *O,O*-dimethyl phosphorofluoridate and show how changes in atomic moments provide a more detailed description of charge redistribution during the reaction than is possible using charges alone. Furthermore, the electrostatic potential on the solvent-excluded surface for this reaction roughly converges at the quadrupolar level, with a root-mean-square deviation of ~ 1 kcal/mol compared to the ab initio Hartree–Fock expectation value. We arrive at similar conclusions for four other reactions, namely the alkaline hydrolysis of demeton-*S* and phosalone, carbon dioxide hydration, and hydrogen cyanide isomerization. Employing multipole moments on atoms therefore appears to be a feasible and compact way to derive catalytic fields defining the optimal catalytic environment for chemical reactions.

Keywords Chemical reaction · Charge redistribution · Atomic multipole moment · Reaction path

Introduction

An accurate representation of molecular charge distribution is important for modeling chemical reactions, especially catalytic processes which are often dominated by electrostatic effects. The electron redistribution around reactants defines catalytic fields which have been shown to be useful indicators for de novo catalyst design [1, 2]. In general, static catalytic fields are defined as molecular electrostatic potential changes during reaction progress [1, 2].

Typically, the charge distribution is approximated by assigning partial charges to atoms—*atomic charges*—which are roughly proportional to the local electron density and sum up to the net molecular charge. Atomic charges can be calculated from a population analysis of an underlying orbital representation like Mulliken charges [3], or fit to optimally reproduce a physical quantity such as the electrostatic potential [4–6]. Each type has certain disadvantages; for example, Mulliken or alternative charges can change with the basis set used, whereas fitted charges suffer from overfitting and redundancies resulting in unphysical values for buried atoms. Overall, it is useful to remember that any such division is arbitrary since no unique definition of an observable atomic charge exists. Nevertheless, any useful definition will reflect the inherent ionicity differences between atoms in a molecule [7].

During a chemical reaction, the anisotropy of a charge distribution around atoms is exacerbated due to strained electronic structure around breaking and forming bonds. Atomic charges alone will generally not describe such changes adequately. One way to improve this it to employ atomic dipoles and higher moments to describe local deviations from an isotropic distribution. Again, this can be implemented as an extension of population analyses using schemes such as distributed multipole analysis (DMA) [8]

Dedicated to the memory of Prof. Oleg Valeryevich Shishkin.

✉ W. Andrzej Sokalski
sokalski@pwr.edu.pl

¹ Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22908, USA

² Department of Chemistry, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

and cumulative atomic multipole moments (CAMM) [9, 10], or fitted along with atomic charges (monopoles) to a surrounding electrostatic potential [11–13]. In the present contribution, we look at the ability of atomic multipole moments obtained using the CAMM scheme to reproduce the molecular electrostatic potentials on the solvent-excluded surface around five reactions involving both neutral and charged reactants.

Cumulative atomic multipole moments

The anisotropic properties of a molecular charge distribution can generally be described by molecular moments, and such a description will suffice at large distances. Closer to the molecule, this becomes harder to achieve and quickly requires uncomfortably high moments in the multipole expansion. However, given molecular orbitals built from an atomic basis set, these molecular moments can be naturally partitioned among atoms. One way to do so is to divide each density element for two basis functions I and J equally between the two atoms involved:

$$\begin{aligned} \langle x^k y^l z^m \rangle &= \sum_i Z_i x_i^k y_i^l z_i^m - \sum_I \sum_J P_{IJ} \langle I | x^k y^l z^m | J \rangle \\ &\equiv \sum_i \langle x^k y^l z^m \rangle_i = \sum_i \left(Z_i x_i^k y_i^l z_i^m - \sum_{I \in i} \sum_J P_{IJ} \langle I | x^k y^l z^m | J \rangle \right), \end{aligned} \quad (1)$$

where i spans all atoms, P_{IJ} is an element of the density matrix in which all off-diagonal elements are halved, and Z_i is the nuclear charge. So an atomic multipole moment of rank klm can be defined as

$$M_{klm,i} = \langle x^k y^l z^m \rangle_i = Z_i x_i^k y_i^l z_i^m - \sum_{I \in i} \sum_J P_{IJ} \langle I | x^k y^l z^m | J \rangle. \quad (2)$$

The Cartesian atomic moments defined in (2) are all calculated relative to the same origin and therefore can be directly summed into molecular moments that are also centered at that origin. It is usually beneficial to move the moments to their local atomic coordinate systems, which can be done through coordinate substitution and an iterative recombination of the moments as follows:

$$\begin{aligned} M_{klm,i}^{\text{CAMM}} &= M_{klm,i} - \sum_{k' \geq 0} \sum_{l' \geq 0} \sum_{m' \geq 0} \binom{k}{k'} \binom{l}{l'} \binom{m}{m'} \\ &\quad \times x_i^{k-k'} y_i^{l-l'} z_i^{m-m'} M_{k'l'm',i}. \end{aligned} \quad (3)$$

These atomic moments centered on atoms have been called *cumulative atomic multipole moments* or CAMMs [10] when

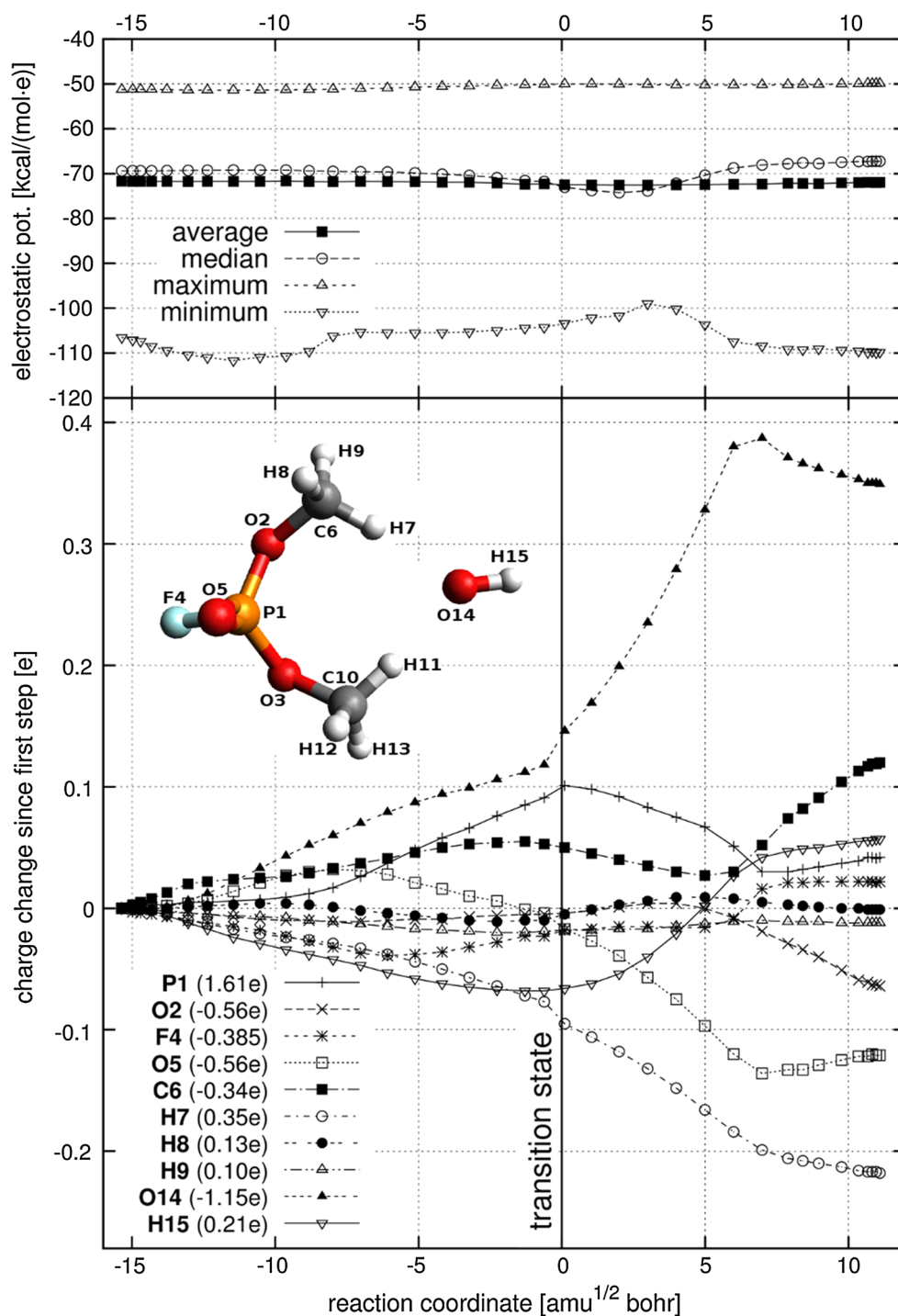
derived from the Mulliken population analysis, although it can also be applied to moments based on any other atomic charge definition as well as on any post-Hartree–Fock, ground state or excited state density matrix. Besides being invariant with respect to translation, *cumulative* atomic moments can be easily rotated while centered on atoms.

Charge redistribution along a reaction path

We take a close look at how individual atomic moments change during the first stage of the alkaline hydrolysis of *O,O*-dimethyl phosphorofluoridate (DMPF). This reaction step has been studied computationally in detail in the context of the hydrolytic degradation of several organophosphorus compounds [14], which are notable in that they are susceptible to enzymatic detoxification performed by phosphotriesterase. In general, basic hydrolysis of DMPF follows multistep addition–elimination mechanism, and the present analysis is limited to the first stage of the “A” path in DMPF degradation. The latter starts with nucleophilic attack of the hydroxyl ion on the phosphorus atom of DMPF, leading to the first pentavalent intermediate (designated by INT1 → TS1a → INT2a in Dyguda-Kazimierowicz et al. [14]). This particular reaction pathway features proton from the incoming hydroxide aligned with the phosphoryl oxygen atom. Results presented here refer to the geometries characterizing the corresponding reaction path, as obtained from intrinsic reaction coordinate simulation [14]. Accordingly, reaction coordinate “0” in Fig. 1 and subsequent plots corresponds to the transition state (TS) structure. Electrostatic potentials and CAMMs were calculated at the RHF/6-311++G(d,p) level using GAMESS-US version 11 APR 2008 (R1) [15].

Besides the evolution of atomic moments, Fig. 1 shows four measures of the ab initio molecular electrostatic potential (MEP) on the Connolly surface, namely its average, median, minimum, and maximum values (the average value is approximately constant, in accordance with Gauss’s law). The Connolly or solvent-excluded surface [16, 17] was chosen as it represents a typical region in space at which other molecules could interact. The probing distance was set to the van der Waals radii according to Pauling and Bondi (for carbon and fluorine), extended by the radius of the water molecule (1.4 Å). Further details on the implementation can be found in [18] and references quoted therein. These quantities do not change in an obviously concerted way, and the maximum or weakest potential on the surface does not exhibit any significant change at all. Meanwhile, the minimum or strongest potential has its largest absolute value at reaction coordinates of approximately $-10 \text{ amu}^{1/2} \text{ Bohr}$, where the other

Fig. 1 Evolution of atomic Mulliken charges and ab initio molecular electrostatic potential on the Connolly surface along the reaction path of the first stage of the alkaline hydrolysis of DMPF (INT1 \rightarrow TS1a \rightarrow INT2a). Atomic charges are plotted relative to their value at the first step (left hand side), with the embedded molecular structure defining the names and numbering of atoms used in the legend. The upper plot shows the corresponding evolution of the average, median and maximum values of molecular electrostatic potential using the same reaction coordinate



measures remain constant. The fact that the minimum (most negative) potential changes the most is not surprising, since the reactants are charged negatively, thus emphasizing negative potentials, and these are the most representative for monitoring charge redistribution during the reaction.

Already from these crude characteristics of the MEP, it is evident that the largest reorganization takes place before

and after the TS, the second region being just before a reaction coordinate of $+5 \text{ amu}^{1/2} \text{ Bohr}$. The circumstances in this second region are different, since along with a rise in the minimum value, there is a visible dip in the median. This means that the negative potential spreads on the surface, which might be caused by a conformational change in the reactants relative to the surface. This pattern is reflected in the evolution of atomic charges (Fig. 1), where the

largest changes also take place after the transition state, and can be identified in the evolution of a number of atomic moments (Fig. 2). The first region (around reaction coordinate $-10 \text{ amu}^{1/2}\text{Bohr}$) involves local charge redistribution within the hydroxyl ion and nearby methyl groups (electron transfer from C6 to H7). After that, redistribution gradually intensifies with little variation in the minimum value, and even the charge on the central phosphorus (P1) changes by $0.1e$ before the transition state.

Much of the charge transfer takes place following the TS, which can be read from the atomic charge evolution in Fig. 1. After the transition state, the approaching oxygen atom is 2.6 \AA from the phosphorus situated between the methyl groups, with H7–O14–H11 being roughly linear. The O14 oxygen proceeds to give away almost $0.3e$, and it is not surprising that much of the charge donated by O14 ends up on the other oxygen atoms bonded to the phosphorus atom. However, the second largest change is found for the H7 and H11 atoms (above $0.2e$), a drop that returns their charge to “standard” values, similar to the other hydrogen atoms of the methyl groups. At first sight, it may seem surprising that much of the charge redistribution takes place after the transition state; nonetheless, this agrees with energetic considerations. Since the transition state is essentially a stationary point on the potential energy surface, its derivative there with respect to the reaction coordinates is zero and the energy should not change very much in the vicinity. As the energy is a function of the charge distribution, it follows the latter should also not change significantly.

Mulliken charges and the associated CAMM atomic moments are arbitrary and often strongly basis set dependent [9], but the plotted *changes* in atomic charges illustrate the magnitude of charge redistribution and should be less sensitive (Löwdin charges were compared in this case). If anything, these changes pinpoint which atoms participate in the reaction and in which direction the flow of charge takes place. The same is true for atomic moments, some of which are plotted in Fig. 2. They describe the finer effects of charge redistribution, also within the bounds of individual atoms. It is hard to draw any final conclusions from them, but some general observations about the role of certain atoms are possible. The largest dipole moments in the system (O2, P1, C6) are all reinforced during the reaction; however, their directions do not change significantly. The direction of dipoles in the hydroxyl group changes by about a right angle, which is expected when the methyl groups stop interacting and the oxygen atom enters the new bond. Surprisingly, the dipole moment on the fluorine atom changes its direction by almost 180° , but its magnitude is small so the reversal is not very meaningful. For the higher moments, the central phosphorus atom

exhibits by far the largest values and most rapid changes, which is understandable since the charge distribution around its nucleus is the most anisotropic.

Convergence of moment-derived molecular electrostatic potentials

Not all of the changes in atomic moments described in the preceding section will have an effect on the surroundings, especially the changes observed for many of the higher atomic moments. This leads to the question of how the moment-derived molecular potential converges when increasing the rank L of the multipole expansion. For example, the hexadecapole (denoted in Fig. 2 by $M^{(4)}$) changes drastically after the transition state, while the difference in the multipole-derived electrostatic potential on the Connolly surface between ranks $L = 3$ and $L = 4$ (Fig. 3) is very small. Fortunately, the redistribution of molecular charge density can also be monitored more directly through the multipole expansion of any well-defined molecular property, such as the MEP or electric fields. In their case, the arbitrary character of a particular atomic charge definition is, to a large degree, eliminated, yielding static or dynamic catalytic fields [2] and aiding *de novo* catalyst design.

In addition to DMPF hydrolysis (Fig. 3), we consider the convergence of the MEP for the alkaline hydrolysis of phosalone and demeton-S, as well as HCN–HNC isomerization and carbonic acid synthesis, with corresponding results in Figs. 4, 5, 6 and 7. In all cases, the basic question is how well atomic multipole-derived potentials describe the molecular electrostatic potential around reactants and how this representation converges with the expansion rank. To this end, we show the root-mean-square deviation (RMSD) of the MEP estimated from atomic multipole expansions relative to the *ab initio* restricted Hartree–Fock value on the Connolly surface, in the case of DMPF using the same reaction coordinates as in Figs. 1 and 2.

In the case of DMPF hydrolysis, using only atomic charges ($L = 0$) implies an error of around 5 kcal/mol , and adding charge–dipole interactions ($L = 1$) lowers this to below 3 kcal/mol . Moments up to octupoles ($L = 3$) are needed in order to bring the deviation below 1 kcal/mol , similar to the octupole-level convergence observed for the transferable atom equivalent method [19]. However, the lowest root-mean-square (RMS) that can be achieved on the Connolly surface in this case is slightly below 0.7 kcal/mol . The average value on this surface, plotted in Fig. 1, is about -70 kcal/mol , which means that the converged expansion carries an error below 1% . If the multipole expansion is in fact converged, then this value can be interpreted as an estimate of the average value of

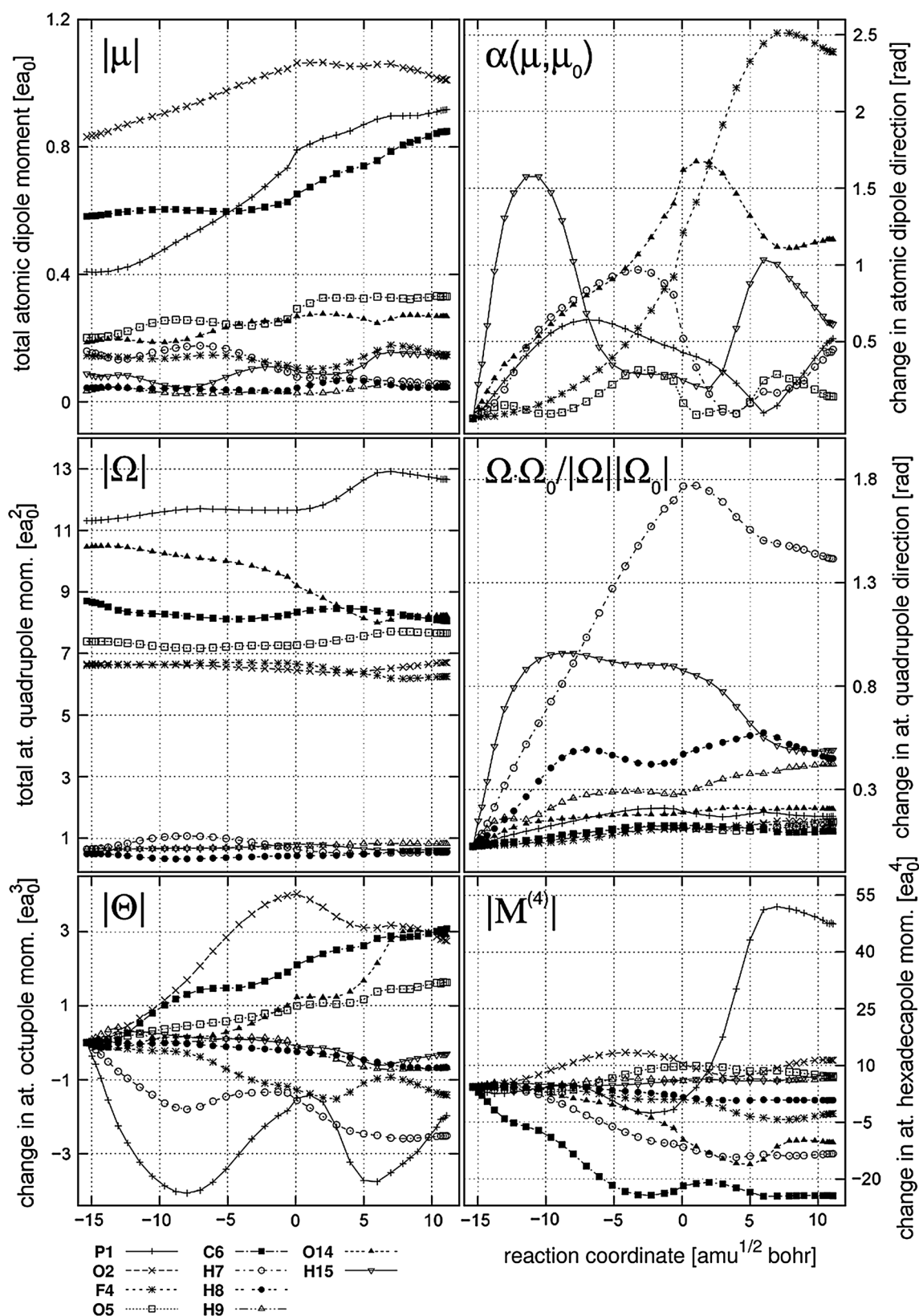


Fig. 2 Evolution of atomic multipole (CAMP) moments along the reaction path of the first stage of the alkaline hydrolysis of DMPF, INT1 → TS1a → INT2a, with the same path and atom definitions as in Fig. 1

Fig. 3 Root-mean-square deviation of the multipole-derived electrostatic potential compared to its ab initio value on the Connolly surface along the reaction path of the first stage of the alkaline hydrolysis of DMPF

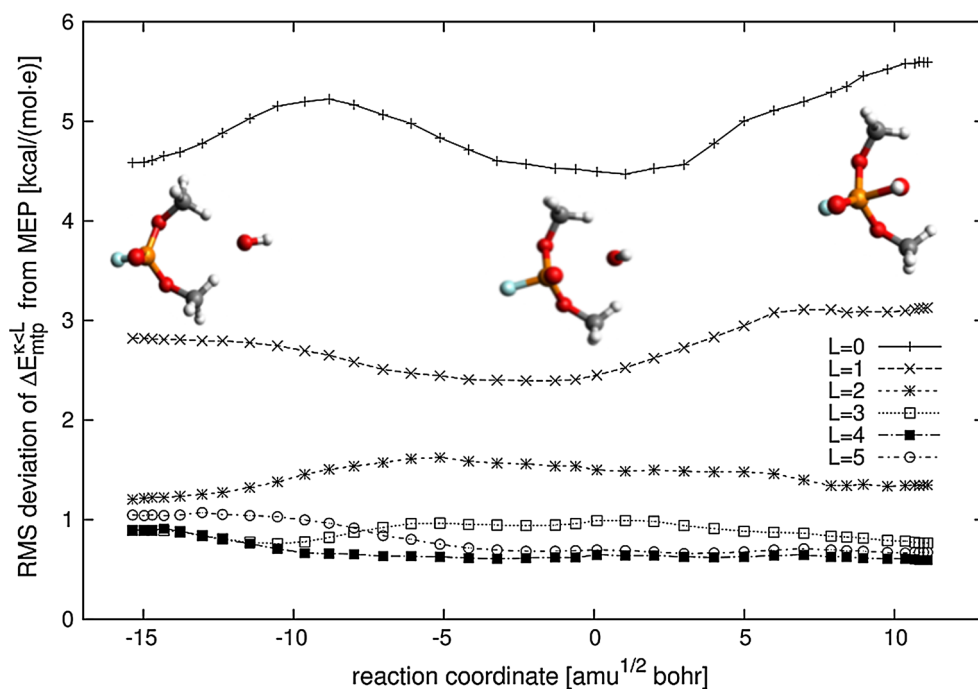
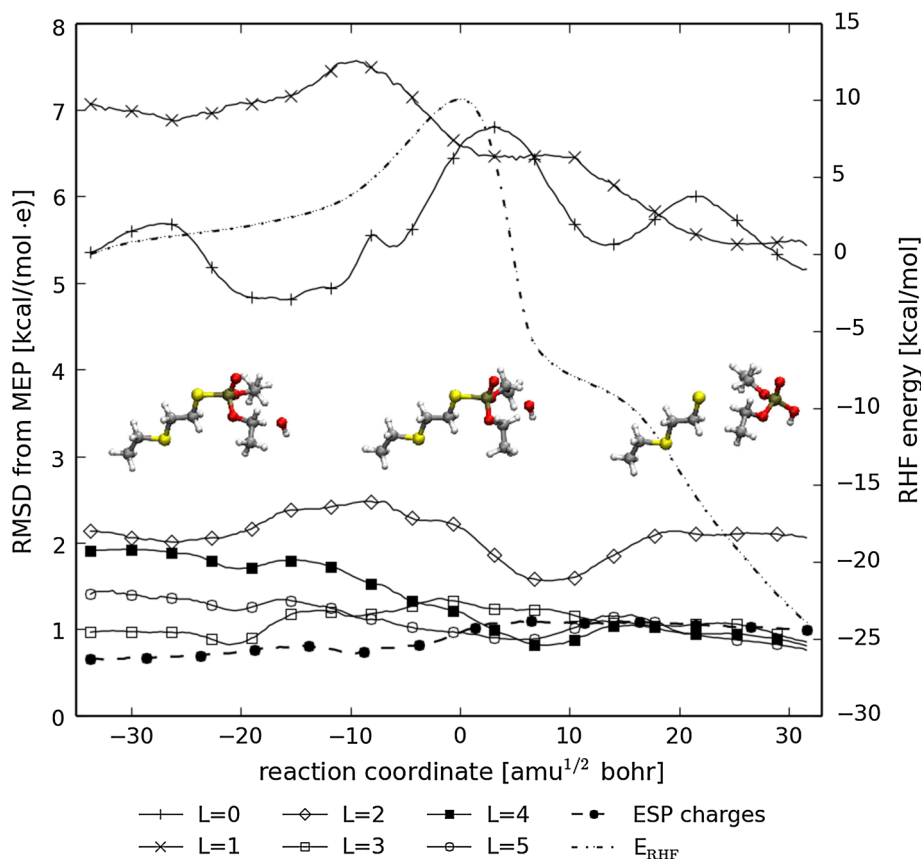


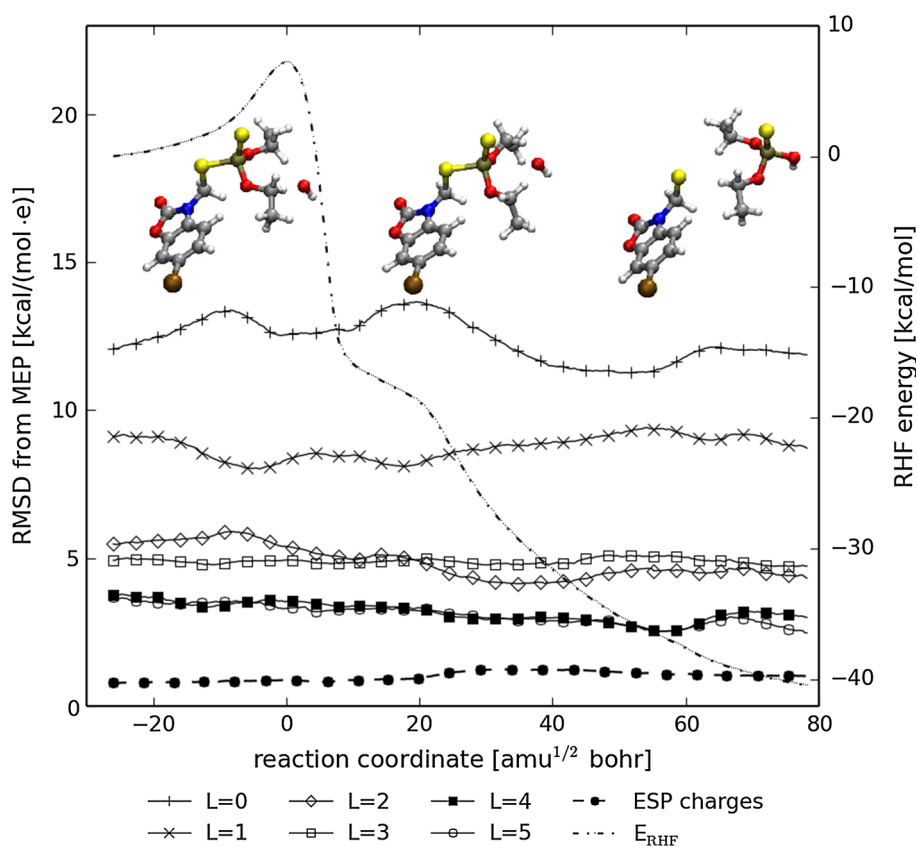
Fig. 4 Root-mean-square deviation of the multipole-derived electrostatic potential compared to its ab initio value on the Connolly surface along the reaction path of the alkaline hydrolysis of demeton-S. Data for atomic charges fitted to electrostatic potential (ESP) shown for comparison



penetration effects at this distance. To compare, using atomic charges implies an error of about 6 %. It should be mentioned, however, that in this case the multipole

expansion starts to diverge for higher ranks, and we have found that the RMS deviation consequently starts to increase for $L > 9$.

Fig. 5 Root-mean-square deviation of the multipole-derived electrostatic potential compared to its ab initio value on the Connolly surface along the reaction path of the alkaline hydrolysis of phosalone. Data for atomic charges fitted to electrostatic potential (ESP) shown for comparison



Presently, we look at alkaline hydrolysis of two other organophosphorus compounds, i.e., demeton-*S* (Fig. 4) and phosalone (Fig. 5). According to the recent study of the mechanism of these reactions [20], certain conformation of the incoming hydroxide is associated with a single-step direct-displacement reaction pathway. The results discussed in what follows refer to the structures along intrinsic reaction coordinate trajectory characterizing “B” paths of demeton-*S* and phosalone hydrolysis, as described in ref. [20]. As a reference, we also present RMSD values obtained from charges fit to the electrostatic potential according to the Merz–Singh–Kollman scheme [4] and we will refer to these as ESP charges (obtained using keywords MK or ESP in Gaussian 09 [21] at the RHF/6-31+G(d) level; the same parameters were used to calculate electrostatic potentials on the Connolly surface). Overall, the results look similar to those for DMPF. However, a comparison with ESP charges shows that the RMSD for the electrostatic potential obtained from atomic moments (CAMMs were calculated using GAMESS-US [15] at the RHF/6-31+G(d) basis set level) rarely does a better job, dropping below the reference only in the demeton reaction and only after the transition state.

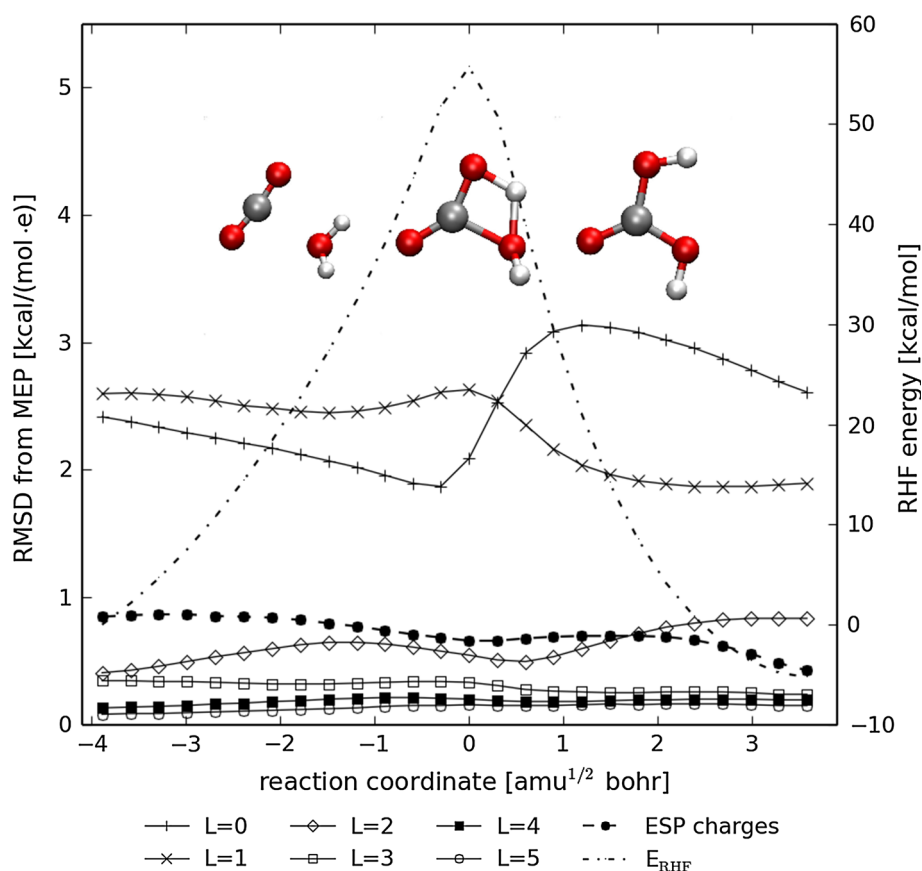
Finally, we turn to two reactions involving neutral reactants, namely the second phase of the reaction of CO₂

with water during which a CO double bond is formed (Fig. 6) and the isomerisation of HCN to HNC (Fig. 7). In both cases, the transition state geometry was obtained here at the RHF/6-31G(d,p) level, and potentials and atomic moments were generated using the 6-311++G(d,p) basis set. Using only atomic charges ($L = 0$) in these cases implies an error up to 3 kcal/mol, and adding charge–quadrupole interactions ($L = 2$) lowers it significantly in both cases to below 1 kcal/mol. Atomic moments up to octupoles ($L = 3$) are needed in order to bring the deviation into the sub-millihartree range. The RMS deviation for this reaction converges at higher multipole ranks to around 0.1 kcal/mol or 0.2 mH, and it appears to be stable up to $L = 16$.

It is important to emphasize that the decrease in RMSD observed in the last two reaction is more than 20-fold. Compared to the magnitude of the electrostatic potential, which oscillates around 20 mH, this corresponds to improving the deviation from 20 to 1 %. Again, using octupoles ($L = 3$) generally brings the MEP estimated from atomic moments close to the best approximation of the exact potential that is possible—with an estimate of 1 % for the electrostatic penetration effect.

One may find it surprising that in one system atomic multipoles reproduce the charge distribution more

Fig. 6 Root-mean-square deviation of the multipole-derived electrostatic potential compared to its ab initio value on the Connolly surface along the reaction path of carbon dioxide hydration



accurately than in others. In other words, the question is why are CAMMs so advantageous compared to ESP charges in the case of HCN isomerisation and carbonic acid synthesis, but not for the hydrolyses of demeton-*S* and phosalone? One possible explanation is the total charges of these systems. For neutral reactants, like HCN or $\text{CO}_2 + \text{H}_2\text{O}$, parts of electron density are considerably inhomogeneous and thus difficult to describe by point charges. In such cases, anisotropic features such as lone pairs may have major contributions to total MEP, leading to high RMSD values if higher multipoles are not included. If these parts undergo significant reorganisation during the reaction, which actually takes place, the deviation of MEP should vary notably along the path. This is indeed observed in the case of $\text{HCN} \rightarrow \text{HNC}$ reaction (Fig. 7). On the other hand, if a molecular system is charged, the Coulomb term becomes dominant, and thus, other multipole moments may be relatively negligible. Therefore, one may expect that in charged complexes ESP charges will often be sufficient, and this seems to be the case for the two $\text{S}_{\text{N}}2$ -like reactions we consider (Figs. 4 and 5).

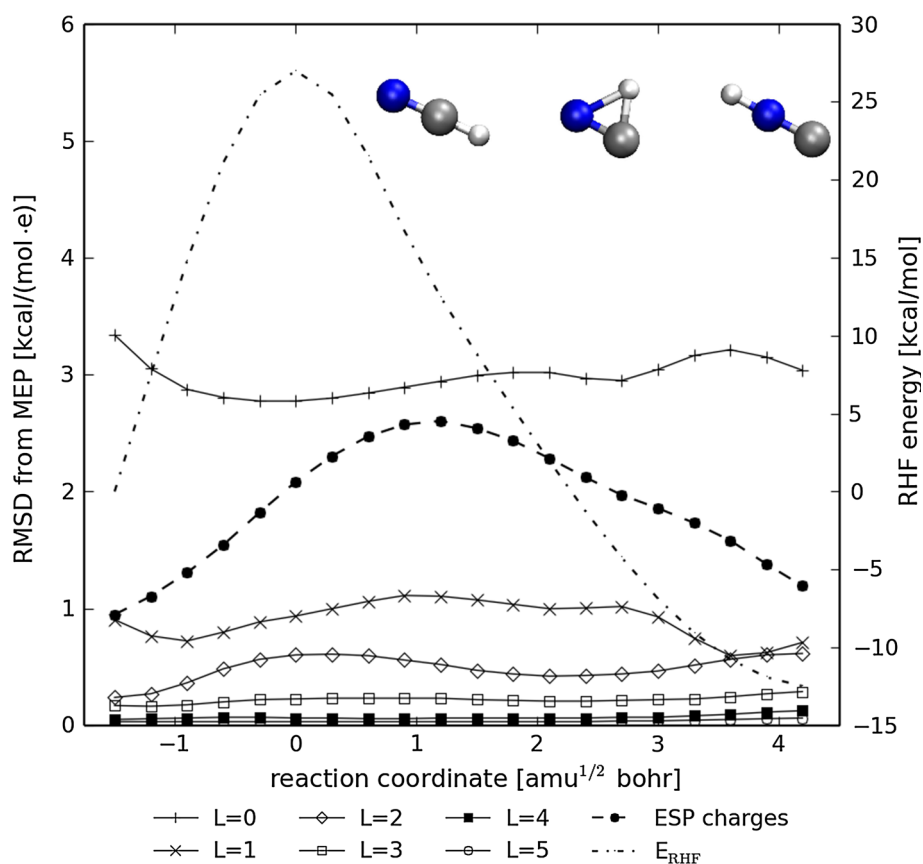
A logical way to improve the anisotropic capabilities of an ESP charge model is to also fit atomic moments within the same procedure, an approach that has recently gained popularity [11–13]. Another potential problem with fitted

models is that many charges, especially for buried atoms, become almost completely redundant with respect to the target electrostatic potential, resulting in unphysical values. This problem becomes only worse when multipole moments are added to the mix, and one can imagine that it will affect larger assemblies more drastically. If such buried atoms happen to be close to the place where the most interesting chemistry appears during a reaction, then the description will ultimately not be appropriate. Jakobsen and Jensen [12] have shown how to sidestep this issue by systematically removing atomic charges moments that are not necessary to reproduce the molecular electrostatic potential.

Conclusions

Some of the most interesting questions to be asked about bond formation and dissociation concern the changes that take place in the electron distributions around atoms. Since a representation in terms of multipole expanded atomic moments can describe the distribution of charge around molecules, it is natural to ask whether they can be used to characterize the changes that occur during reactions. We find this to be the case, based on cumulative atomic

Fig. 7 Root-mean-square deviation of the multipole-derived electrostatic potential compared to its ab initio value on the Connolly surface along the reaction path of hydrogen cyanide isomerisation



multipole moments calculated for five chemical reaction pathways. Changes in individual atomic moments provide additional insight into the redistribution of electron charge during reactions, although these changes can be difficult to interpret outside the context of the multipole expansion. In terms of the molecular electrostatic potential, we find that atomic moments obtained with the CAMM scheme are able to reproduce the ab initio potential on the excluded solvent surface, and generally do so to within 1 kcal/mol when at least quadrupole moments are included on atoms. Therefore, atomic multipoles appear to be a compact and versatile representation for charge distribution during chemical reactions which can aid the robust theoretical design of biocatalysts, especially within the catalytic field technique that relies on the molecular electrostatic potential differences.

Acknowledgments This work was supported by Wrocław Research Centre EIT+ within the project “Biotechnologies and advanced medical technologies” BIOMED (POIG.01.01.02-02-003/08) co-financed by European Regional Development Fund (Operational Programme Innovative Economy, 1.1.2). Partial financing by a statutory activity subsidy from Polish Ministry of Science and Higher Education for Faculty of Chemistry of Wrocław University of Technology is also acknowledged. Calculations were performed at supercomputer centers in Wrocław (WCSS), Poznań (PCSS) and Warsaw (ICM). KML thanks Paweł Kędzierski for providing tools to calculate the Connolly surface.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Sokalski WA (1985) *J Mol Catal* 30:395
- Szefczyk B, Mulholland A, Ranaghan K, Sokalski WA (2004) *J Am Chem Soc* 126(49):16148
- Mulliken RS (1955) *J Chem Phys* 23(10):1833
- Singh UC, Kollman PA (1984) *J Comput Chem* 5:129
- Bayly C, Cieplak P, Cornell W, Kollman P (1993) *J Phys Chem* 97:10369
- Beck B, Clark T, Glen R (1997) *J Comput Chem* 18:744
- Meister J, Schwarz WHE (1994) *J Phys Chem* 98:8245
- Stone AJ (1981) *Chem Phys Lett* 83(2):233
- Sokalski WA, Poirier RA (1983) *Chem Phys Lett* 98:86
- Sokalski WA, Sawaryn A (1987) *J Chem Phys* 87:526
- Kramer C, Gedeck P, Meuwly M (2012) *J Comput Chem* 33:1673
- Jakobsen S, Jensen F (2014) *J Chem Theory Comput* 10:54923
- Kramer K, Spinn A, Liedl K (2014) *J Chem Theory Comput* 10:4488
- Dyguda-Kazimierowicz E, Sokalski WA, Leszczyński J (2008) *J Phys Chem B* 112:9982
- Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus

- TL, Dupuis M, Montgomery JA (1993) *J Comput Chem* 14(11):1347. doi:[10.1002/jcc.540141112](https://doi.org/10.1002/jcc.540141112)
16. Connolly M (1983) *J Appl Crystallogr* 16:548
17. Connolly M (1983) *Science* 221(4612):709
18. Kedzierski P (2001) Study of the nature of interactions in the active sites of enzymes. Ph.D. thesis, Wroclaw University of Technology
19. Whitehead C, Breneman C, Sukumar N, Ryan M (2003) *J Comput Chem* 24:512
20. Dyguda-Kazimierowicz E, Roszak S, Sokalski WA (2014) *J Phys Chem B* 118:7277
21. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JN, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas AU, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) *Gaussian 09 Revision D.01*. Gaussian Inc., Wallingford, CT