

# Modeling Biophysical and Biological Properties from the Characteristics of the Molecular Electron Density, Electron Localization and Delocalization Matrices, and the Electrostatic Potential

Chérif F. Matta<sup>\*[a,b,c]</sup>

The electron density and the electrostatic potential are fundamentally related to the molecular hamiltonian, and hence are the ultimate source of all properties in the ground- and excited-states. The advantages of using molecular descriptors derived from these fundamental scalar fields, both accessible from theory and from experiment, in the formulation of quantitative structure-to-activity and structure-to-property relationships, collectively abbreviated as QSAR, are discussed. A few such descriptors encode for a wide variety of properties including, for example, electronic transition energies,  $pK_a$ 's, rates of ester hydrolysis, NMR chemical shifts, DNA dimers binding energies,  $\pi$ -stacking energies, toxicological indices, cytotoxicities, hepatotoxicities, carcinogenicities, partial molar volumes, partition coefficients ( $\log P$ ), hydrogen bond donor capacities, enzyme-substrate complementarities, bioisosterism, and regularities in the genetic code. Electronic fingerprinting from the topological analysis of the electron density is shown to be comparable and

possibly superior to Hammett constants and can be used in conjunction with traditional bulk and liposolubility descriptors to accurately predict biological activities. A new class of descriptors obtained from the quantum theory of atoms in molecules' (QTAIM) localization and delocalization indices and bond properties, cast in matrix format, is shown to quantify transferability and molecular similarity meaningfully. Properties such as "interacting quantum atoms (IQA)" energies which are expressible into an interaction matrix of two body terms (and diagonal one body "self" terms, as IQA energies) can be used in the same manner. The proposed QSAR-type studies based on similarity distances derived from such matrix representatives of molecular structure necessitate extensive investigation before their utility is unequivocally established. © 2014 The Author and the Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23608

## Introduction

Quantitative structure-to-activity (or -property) relationships (QSAR/QSPR) relate a set of molecular descriptors to biological, pharmacological, or physicochemical properties.<sup>[1–15]</sup> For simplicity, we will ignore the nuance between property and activity and use QSAR to designate both QSAR and QSPR in this article as the principles in modeling either are essentially the same. QSAR correlations are constructed to predict the activities of compounds (physicochemical, spectroscopic, biological, pharmacological, and so forth) from other more readily measurable experimental properties or from calculated properties regardless of the presence of a clear predictor–predicted causal link.

Not infrequently, QSAR models, despite of being predictive (and hence useful), provide little or no clue as to why they work. We can cite, for example, the chemical graph theoretical (connectivity) indices,<sup>[6]</sup> which useful as they may be, cannot be generally expected to provide, as such, much insight on the mode of action of a drug.

Alternatively, QSAR modeling based on descriptors<sup>[16]</sup> obtained directly from the electron density, electron (de)localization, or the molecular electrostatic potential (MESP), through the application of a topological analysis<sup>[17–19]</sup> can reach beyond the primarily utilitarian approach.<sup>[15]</sup> Descriptors derived from the electron density<sup>[16]</sup> (Fig. 1) have been used by various research groups in the construction of physically

insightful QSAR models.<sup>[15]</sup> The present review samples the relevant literature and expands the use of such descriptors in QSAR, especially those obtained from Bader's quantum theory of atoms in molecules (QTAIM),<sup>[17–19]</sup> in the construction of predictive and physically insightful QSAR models. The terminology of QTAIM will be used without definition since it is reviewed elsewhere.<sup>[17–24]</sup>

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

[a] C. F. Matta

Department of Chemistry and Physics, Mount Saint Vincent University, Halifax, Nova Scotia, Canada B3M 2J6

Fax: +1-(902)-457-6134 Email: cherif.matta@msvu.ca

[b] C. F. Matta

Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J3

[c] C. F. Matta

Department of Chemistry, Saint Mary's University, Halifax, Nova Scotia, Canada B3H 3C3

Contract grant sponsor: Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI), and Mount Saint Vincent University

© 2014 The Authors Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

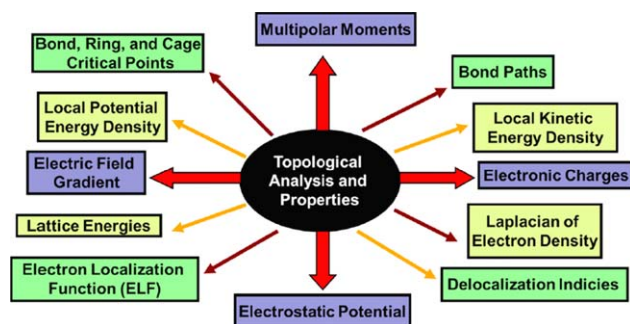


Figure 1. Examples of properties that can be used as molecular descriptors derived from the electron density. (Reproduced with permission from Ref. 16 © 2012 American Chemical Society). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

Electron density-derived descriptors have clear physical meanings that can be often related to the predicted property. In addition, these descriptors are often accessible from both experiment (X-ray diffraction)<sup>[19,25–29]</sup> and theoretical calculations except for properties that require the full density matrix obtainable only from quantum mechanical calculations.

The analysis of the electron density can be a rich source of raw atomic and bonding descriptors that can be processed through empirical models to predict (and possibly explain) physicochemical and biological properties of molecules.<sup>[15]</sup> The use of QTAIM descriptors in QSAR may possibly increase drastically in the near future due to the fast and ever-growing availability of computational resources and efficient computer programs.

A particular advantage of QTAIM analysis of the electron density is that it provides a complete set of consistent properties (not only atomic charges) within one coherent and quantum mechanically sound theory.<sup>[17]</sup> Further, QTAIM atomic and bond properties are well-known for their insensitivity to the underlying levels of *ab initio* electronic structure calculations, an assertion especially true when differences rather than absolute values are considered.<sup>[30–32]</sup> This is contrasted with, say, Hirshfeld or Mulliken atomic charges, which are arbitrarily defined and basis set unstable.

## Quantitative Structure–Activity Relationships (QSAR)

QSAR modeling<sup>[1–15]</sup> aims at the construction of empirical formulae capable of predicting the biological activity of an untested molecule given a set of its structural and electronic properties, often on the basis of some sort of multivariate regression.<sup>[8]</sup> This approach is widespread in the field of drug discovery and is especially useful when the mechanism of action of the drug is not known. The field is so vast that there are journals with names such as *SAR QSAR Environ. Res.* or *Mol. Inform.* (formerly *Quant. Struct.-Act. Relat.* and later *QSAR Comb. Sci.*) and even a UK-based society (*The QSAR Society*).

The properties used in the construction of a QSAR model can include experimentally determined quantities, results of (quantum chemical) calculations, simple counts of atoms or bond types, and connectivity information regardless of the

presence of an obvious causal link between them and the modeled activity. The basic assumption is expressed as

$$\begin{aligned} \text{biological activity} &\equiv \log\left(\frac{1}{C}\right) = f(x_1, x_2, \dots, x_n) \\ &= \text{constant} + \sum_{i=1}^n \sum_{j=1}^m a_{ij} x_i^j, \end{aligned} \quad (1)$$

in which  $C$  is the concentration of the compound necessary to reach a biological endpoint such as  $LD_{50}$  (lethal dose 50%),  $IC_{50}$  (inhibitory concentration 50%), or  $ED_{50}$  (effective dose 50%),  $x_i$  is the  $i$ th predictor (from a set of  $n$ ) raised to the  $j$ th power, and  $a_{ij}$  are the weighing coefficient obtained from the statistical fitting. The last equality of Eq. (1) gives the explicit form of the oft assumed function relating the predictors to the predicted response. The predictors can be experimental or calculated properties whether structural, physicochemical, or quantum chemical or any combinations of all the above. It is recommended that the number of compounds used to build the statistical model be at least  $5n$ , and the rule of thumb is that the greater the ratio of the number of compounds to the number of parameters, the better the model.<sup>[10]</sup>

It is impossible to review the immense QSAR literature but we cite a few examples here to illustrate the principally utilitarian use of QSAR in drug discovery. The HOMO–LUMO gap and the local density of electronic states, for example, have been used by Vendrame and coworkers<sup>[33]</sup> to construct a QSAR model capable of predicting the biological activity of steroids with 100% success rate. These workers also constructed a related QSAR model to identify the carcinogenic activity of a given polycyclic aromatic hydrocarbon, a model that exhibited an accuracy of over 80%.<sup>[33]</sup> In another study, it was found that the quantum mechanically calculated electron affinities of 270 nitroaromatic compounds provide a statistically significant basis for the discrimination between those that are mutagenic (Ames test positive) and those that are not (Ames test negative).<sup>[34]</sup> Hatch et al.<sup>[35]</sup> developed QSAR models for 80 amines with mutagenic activities spanning 10 orders of magnitude based on the total energy of the conjugated  $\pi$ -electrons and on the energy of the LUMO.

Clearly, there is no shortage of examples of QSAR models in the literature. A final example to help bringing the point to the fore is that of a strongly predictive (and hence useful) model in which the connections between the predictors and the response are not obvious. In this example, the effectiveness of a series of ketones with aromatic substituents against *Candida albicans*, measured by the negative logarithm of the minimum inhibitory concentration ( $pMIC$ ), was found to be robustly predicted by the following model:<sup>[36]</sup>

$$\begin{aligned} pMIC &= 17.4735 + 0.3393 [Lx] - 1.7829 [IP] \\ &\quad - 0.0708 [MR] + 0.8918 [B4 (A)], \end{aligned} \quad (2)$$

where the predictors include molecular length [Lx], ionization potential [IP], molar refractivity [MR], and the Verloop B4 steric parameter [B4 (A)] of the substituent in the aromatic ring A. The connection between these quantities and the MIC of the compound is not evident, but this does not reduce the importance of the QSAR model in the design of more potent inhibitors.

From the above few examples, one can say that the predictors in QSAR modeling can be selected to maximize the correlation with the response before other considerations such as causal link or mechanism of action. Such descriptor(s) selection can be automated by modern statistical methods implemented in numerous software packages but this black-box approach can lead to problems such as chance correlations<sup>[37]</sup> and even the inconsistency of the sign of the correlation coefficient when only part of the original dataset is analyzed.<sup>[38]</sup> It has been argued recently that the use of electron density-based molecular descriptors can eliminate a source of uncertainty in QSAR modeling since they are well-defined, physically meaningful, and fundamentally related to activity, which would constitute one way to remedy for the aforementioned possible limitations of traditional QSAR.<sup>[15]</sup>

## The Electron Density $\rho(\mathbf{r})$ and QSAR Modeling

The electron density  $\rho(\mathbf{r})$  at a point  $\mathbf{r}$  is the probability density of finding an electron of either spin in a volume element  $d\tau = dx dy dz$  at  $\mathbf{r}$  weighted by the total number of electrons ( $N$ ). The electron density can be calculated from the many-electron wavefunction  $\Psi$ :

$$\rho(\mathbf{r}) = N \int d\tau' \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \quad (3)$$

where  $\mathbf{x}_i$  is the set of spatial and spin coordinate of the electron  $i$ , and  $\int d\tau'$  is short for integration over the spatial coordinates of all electrons except one and summation over both spins.

The total electric charge density  $\rho_{\text{total}}(\mathbf{r})$  is the sum of the negative electronic charge density and the positive nuclear charge density (discrete distribution of point nuclear charges), which in atomic units (a.u.) ( $e = 1$ ) is given by:

$$\rho_{\text{total}}(\mathbf{r}) = -\rho(\mathbf{r}) + \sum_{\alpha} Z_{\alpha} \delta(\mathbf{R}_{\alpha} - \mathbf{r}), \quad (4)$$

where  $Z_{\alpha}$  and  $\mathbf{R}_{\alpha}$  are the charge and position of nucleus  $\alpha$ , and  $\delta(\mathbf{R}_{\alpha} - \mathbf{r})$  is a Dirac delta function.

Hohenberg and Kohn (HK)<sup>[39]</sup> proved the existence of a bijective mapping between the electron density  $\rho(\mathbf{r})$  of a many-electron nondegenerate ground state (in the absence of external magnetic fields), and the external potential  $v[\rho(\mathbf{r})]$ , that is, the nuclear potential augmented by scalar potentials external to the system of electrons. In other words, the electron density is bijectively mapped into the nuclear charges (the atomic numbers in a.u.,  $Z_{\alpha}$ ) and positions  $[\delta(\mathbf{R}_{\alpha} - \mathbf{r})]$ , which, together, generate to the external potential. This last assertion is a consequence of Kato's cusp condition, which states that at the position of an atomic nucleus the derivative of the spherically averaged density ( $\bar{\rho}$ ) with respect to the distance from that nucleus ( $d\bar{\rho}/dr_{\alpha}$ ) equals  $-2Z_{\alpha}\bar{\rho}(\mathbf{R}_{\alpha})$ .<sup>[40]</sup> Further, the electron density also determines the total number of electrons in the system  $N$  via its integral over all space (Fig. 2). This mapping implies the interdeterminacy of the two terms of the R.H.S. of Eq. (4), and thus the electron density analyzed in this work, contains the same information as the total charge density.

$$\hat{H}_{\text{elec.}} = - \underbrace{\sum_{i=1}^N \frac{1}{2} \nabla_i^2}_{\text{electronic K.E. term}} - \underbrace{\sum_{i=1}^N \sum_{\alpha=1}^M \frac{Z_{\alpha}}{|\mathbf{R}_{\alpha} - \mathbf{r}_i|}}_{\text{electron - nuclei attraction term}} + \underbrace{\sum_{i=1}^N \sum_{i>j}^N \frac{1}{r_{ij}}}_{\text{electron - electron repulsion term}}$$

(i)  $\rho(\mathbf{r})$  exhibits cusps at the position of the nuclei  $\rightarrow \therefore \rho(\mathbf{r})$  fixes the positions  $\mathbf{R}_{\alpha}$  and the number  $M$  of the nuclei.

(ii) Kato's cusp condition  $\rightarrow \rho(\mathbf{r})$  determines nuclear charges  $Z_{\alpha} \rightarrow$  the external potential is fixed (and so is the nuclear-nuclear repulsion term, not shown).

(iii)  $\rho(\mathbf{r})$  determines  $N$  since  $\int \rho(\mathbf{r}) d\mathbf{r} = N$ . This completes the unique specification of the Hamiltonian.

(\*) Atomic units are used throughout.

Figure 2. The ground-state electron density,  $\rho(\mathbf{r})$ , determines the Born-Oppenheimer (BO) Hamiltonian uniquely. The full BO Hamiltonian includes, in addition to the electronic Hamiltonian displayed in the figure, the nuclear-nuclear repulsion term which is fixed from points (i) and (ii). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

As specifying  $v[\rho(\mathbf{r})]$  and  $N[\rho(\mathbf{r})]$  completely defines the Hamiltonian (Fig. 2), the ground-state wavefunction  $\Psi[\rho(\mathbf{r})]$  and all derived ground-state properties  $\{O\}$  including the total energy are uniquely specified as well, thus we may write:

$$\rho(\mathbf{r}) \rightarrow \left\{ \begin{array}{l} v[\rho(\mathbf{r})] \\ N[\rho(\mathbf{r})] \end{array} \right\} \rightarrow \hat{H}[\rho(\mathbf{r})] \rightarrow \Psi[\rho(\mathbf{r})] \rightarrow O[\rho(\mathbf{r})]. \quad (5)$$

The electron density is thus the scalar field par excellence for the construction of meaningful QSAR correlations since by necessity all molecular properties are uniquely mapped to it whether the exact functional relationships is known or not.

Riess and Münch (RM) demonstrated that the first HK theorem, expressed in the relation (5), applies not only to the total system but also that it applies to any finite bounded subsystem of any size and shape.<sup>[41]</sup> That is to say, an arbitrarily small open subsystem within a molecule enclosed by sharp but otherwise arbitrary boundaries can be mapped uniquely to all ground-state properties of the full system, including the total full electron density,<sup>[41]</sup> which can be written symbolically as:

$$\rho(\mathbf{r})_{\omega} \rightarrow \left\{ \begin{array}{l} v[\rho(\mathbf{r})] \\ N[\rho(\mathbf{r})] \end{array} \right\} \rightarrow \hat{H}[\rho(\mathbf{r})] \rightarrow \Psi[\rho(\mathbf{r})] \rightarrow O[\rho(\mathbf{r})], \quad (6)$$

↙ ↘  
 $\rho(\mathbf{r})$

where  $\omega$  is any open system with sharp boundaries delimiting it, in three-dimensional (3-D) space, from the rest of the total many-electron system.

Bader and Becker<sup>[42]</sup> applied this extension of the HK theorem to atoms in molecules (AIMs) bounded by surfaces of zero-flux in the gradient vector field associated with the electron density (when  $\omega = \Omega$ ). Mezey<sup>[43]</sup> illustrates the use of RM's extension of the HK theorem in conjunction with Carbó's similarity index in what he terms the "holographic electron density theorem". Mapping (6) constitutes fundamental grounds for

basing the empirical modeling of properties in QSAR-type studies on electron density-derived properties. This is particularly true when  $\omega$  itself is well-defined from quantum mechanics, that is, when  $\omega = \Omega$ .<sup>[17]</sup>

Bader and Zou<sup>[44]</sup> emphasized that Dirac's definition of a quantum mechanical "observable"<sup>[45]</sup> applies to  $\rho(\mathbf{r})$  as it fulfils the following conditions:

1.  $\rho(\mathbf{r})$  is a real dynamical variable, which is the expectation value of a linear Hermitian operator  $\hat{\rho}(\mathbf{r})$ , and
2. The eigenstates of  $\hat{\rho}(\mathbf{r})$  form a complete set of coordinate states  $|\mathbf{r}\rangle$ .

Dirac comments further that "[i]n practice it may be very awkward, or perhaps even beyond the ingenuity of the experimenter, to devise an apparatus which could measure some particular observable, but the theory always allows one to imagine that the measurement can be made."<sup>[45]</sup> While not every "observable" in the quantum mechanical sense is observable in practice, the electron density is. Thus, the electron density is not only the source of all ground-state properties but also it is a quantum mechanical observable, which is readily accessible from both theory and experiment (primarily from X-ray scattering experiments),<sup>[19,25–29]</sup> Figure 3. These qualities of the density when considered together with relation (6) lead one to conclude that atomic properties such as atomic charges themselves are quantum mechanical observables accessible experimentally and, contrary to the widely held view, are unique.<sup>[44,46]</sup>

## The Electron Density and Molecular Similarity

A minimum degree of similarity is often required to exist between the molecules used in the construction of a QSAR model, say a common molecular skeleton with different substituents such as the series of substituted benzoic acids.<sup>[47]</sup> Researchers have used a diversity of molecular features in quantifying similarity, some base similarity on a comparison of amino acid sequences,<sup>[48–50]</sup> others on the degree of mismatching of 2-D chemical graphs,<sup>[51]</sup> and others on 3-D molecular superpositions.<sup>[52]</sup> By relations (5) and (6), a similarity of  $\rho(\mathbf{r})$  necessarily leads to the similarity of all other ground-state properties, and hence the most fundamental molecular comparisons are those effected at the electron density level. This realization is credited to Carbó-Dorca and coworkers<sup>[9,53–55]</sup> who used the electron density and its associated MESP<sup>[56–68]</sup> to quantify molecular similarity. Carbó's (dimensionless) similarity index can be written in its general form as<sup>[9,53]</sup>

$$C_{M_i M_j}(\hat{O}; \Theta) = \frac{\max_{\Theta} \int \int \rho_{M_i}(\mathbf{r}_i) \hat{O}(\mathbf{r}_i, \mathbf{r}_j) \rho_{M_j}(\mathbf{r}_j; \Theta) d\mathbf{r}_i d\mathbf{r}_j}{\sqrt{\int \int \rho_{M_i}(\mathbf{r}_i) \hat{O}(\mathbf{r}_i, \mathbf{r}_j) \rho_{M_i}(\mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \int \int \rho_{M_j}(\mathbf{r}_j) \hat{O}(\mathbf{r}_i, \mathbf{r}_j) \rho_{M_j}(\mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j}}, \quad (7)$$

where  $\rho_{M_i}$  and  $\rho_{M_j}$  are the electron densities of  $i$ th and  $j$ th molecules,  $\Theta$  is an alignment parameter, and  $\hat{O}(\mathbf{r}_i, \mathbf{r}_j)$  is an operator

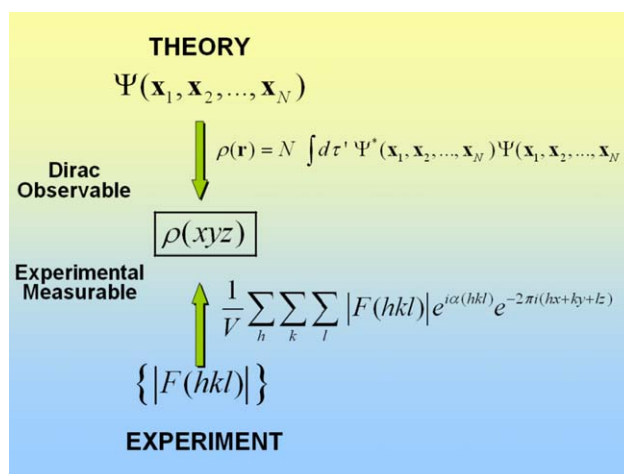


Figure 3. The electron density as theoretically accessible and experimentally measurable Dirac observable. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

that defines the similarity index. If  $\hat{O}(\mathbf{r}_i, \mathbf{r}_j) = \delta(\mathbf{r}_i - \mathbf{r}_j)$ , a Dirac delta function, and assuming maximal alignment the index is<sup>[9]</sup>

$$C_{M_i M_j} = \frac{\int \rho_{M_i} \rho_{M_j} d\mathbf{r}}{\sqrt{\int \rho_{M_i}^2 d\mathbf{r} \int \rho_{M_j}^2 d\mathbf{r}}}. \quad (8)$$

and when  $\hat{O}(\mathbf{r}_i, \mathbf{r}_j) = |\mathbf{r}_i - \mathbf{r}_j|^{-1}$ , the Coulomb operator,  $C_{M_i M_j}$  is a Coulomb similarity index.

The denominator in Eqs. (7) or (8) causes  $C_{M_i M_j}$  to become unity when a molecule is compared with itself ( $i = j$ ). Thus Carbó indexes range from a maximum of unity (identity) and approach but never reach zero, since there will always be some overlap of electron density no matter how dissimilar two molecules are (i.e.,  $0 < C_{M_i M_j} < 1, i \neq j$ ).

Cioslowski and Nanayakkara (CN)<sup>[69]</sup> extend the applicability of a Carbó-type indexes to pairs of AIMs as defined by QTAIM. In this approach, the comparison is between a pair of AIMs (AIMs  $A$  and  $B$ ) rather than between pairs of molecules (molecules  $M_i$  and  $M_j$ ). This is achieved by taking into account only the intersection of the spaces occupied by the two atoms delimited from the rest of their respective molecules by zero-flux surfaces. When the atom is not bound and extends to infinity, then a cutoff of 0.001 a.u. is taken as its outer isosurface. The atomic similarity index is defined as<sup>[69]</sup>

$$S_{A(M_i)B(M_j)} = \left[ \frac{\int \rho_{M_i} d\mathbf{r}}{\Omega_{A(M_i)}} \right] \times \left[ \frac{\int \rho_{M_j} d\mathbf{r}}{\Omega_{B(M_j)}} \right], \quad (9)$$

where

$$\Omega_{A(M_i)B(M_j)} = \Omega_{B(M_j)A(M_i)} \equiv \Omega_{A(M_i)} \cap \Omega_{B(M_j)}, \quad (10)$$

and where the intersection of the two basins is calculated after superposing the nuclei of the two atoms and maximizing the index through a rotation that maximizes their overlap. If one compares two atoms in the same molecule, then  $i = j$  in

Eqs. (9) and (10). CN provide several examples showing that their atomic similarity index yields results that conform with chemical observation, intuition, and expectation.<sup>[69]</sup>

Despite of their ground-breaking nature, Carbó's indexes can be impractical because (i) they necessitate molecular alignment, that is, maximizing  $C_{M_i M_j}$  through an exhaustive search of the superposition space<sup>[9,70,71]</sup> and (ii) can be dominated by nuclear maxima, which overshadows the subtle differences in the valence shells that govern chemistry. A solution to these problems has been proposed by Popelier in what he terms the quantum topological molecular similarity (QTMS) approach, briefly reviewed below.

## The Electron Density, Transferability, and Additivity

The contributions of a given atom (or of a group of atoms) to a set of molecular properties can often be predicted from its corresponding contributions to the molecular properties of a chemically different molecule. When such a prediction is possible with chemical accuracy the atom (or group) is said to be transferable. Transferability, when coupled with additivity, leads to additivity schemes such as those developed by Benson and coworkers.<sup>[72–74]</sup>

Bader and coworkers<sup>[17,24,75–82]</sup> stress that the QTAIM definition of AIMs maximizes their transferability. QTAIM partitions space exhaustively at the zero-flux surfaces<sup>[83–85]</sup> without any gaps unaccounted for into “nonoverlapping, bounded, space-filling open quantum systems”,<sup>[82]</sup> and hence AIMs defined in this manner also exhibit additivity since

$$\begin{aligned} \langle \hat{O} \rangle &= \sum_{\Omega} \left( \int_{\Omega} \left\{ N \int \frac{1}{2} [\Psi^* \hat{O} \Psi + (\hat{O} \Psi)^* \Psi] d\tau' \right\} d\mathbf{r} \right) \\ &= \sum_{\Omega} \left( \int_{\Omega} \rho_O d\mathbf{r} \right) = \sum_{\Omega} \langle \hat{O}(\Omega) \rangle, \end{aligned} \quad (11)$$

where the left hand side (L.H.S.) is the usual molecular expectation value of a linear Hermitian operator  $\hat{O}$ , while the three equivalent expressions on the R.H.S. yield the molecular value as a sum of atomic averages obtained from the integral of a real space dressed density  $\rho_O(\mathbf{r})$  over the volume of each atom  $\Omega$ .

The transferability and additivity of atomic properties has been used in the theoretical reconstruction of large molecules from fragments synthesized in molds. The goal of this type of reconstruction is to bypass or reduce the nonlinear scaling bottleneck as a function of the size of ab initio calculations. Examples of such reconstruction includes the theoretical synthesis of polypeptides from amino acid residues extracted from molds and melded together,<sup>[75,77–79]</sup> and the theoretical<sup>[86]</sup> and experimental<sup>[87]</sup> reconstruction of complicated alkaloids such as the opioids from their constituent molecular building blocks. Experimentally, the advantage of this approach is to circumvent the problem of obtaining electron densities of molecules that are difficult to crystallize from the crystallographic electron densities of more readily crystallizable smaller molecules. The transferability of larger molecular fragments saturated with

hydrogen atoms termed “kernels” has been extensively tested by Huang, Massa, and Karle and has been used to predict total energies, interactions energies, and binding energies of gigantic molecules with chemical accuracy at a fraction of the computational cost of direct calculations, the reconstruction method being termed the “kernel energy method.”<sup>[88–95]</sup> The transferability of the electron density obtained from theory and experiment has allowed several groups to build libraries of transferable X-ray crystallographic multipolar parameters that are used to assist in the refinement and in the modeling of the electron density of large biological macromolecules.<sup>[87,96–109]</sup>

It is an observational fact that properties of AIMs are often transferable within chemical precision, that is, to within  $\leq 1$  kcal mol<sup>-1</sup>, which is practically tantamount to perfect transferability. Such high transferability (and additivity) has been reported for QTAIM atomic energies<sup>[17,110]</sup> and for atomic response properties such as polarizabilities<sup>[111]</sup> and magnetic susceptibilities<sup>[112]</sup>—all closely paralleling experiment.

But why and how does transferability work? Actually there are two types of transferability: Direct transferability, which is the almost constancy of the atomic properties of a given atom or group from one molecule to another, and compensatory transferability, where the change in one atom or group cancels the opposite change of similar magnitude in its neighboring atom(s) or group(s).<sup>[76,113]</sup>

To understand Benson's additivity schemes, one must take both types of transferability into account.<sup>[113]</sup> We quote an example given by Cortés-Guzmán and Bader:<sup>[113]</sup>

The energy of CH<sub>2</sub>XCH<sub>3</sub> is the mean of the energies of CH<sub>2</sub>XCH<sub>2</sub>X and CH<sub>3</sub>CH<sub>3</sub> to within  $\leq 0.5$  kcal mol<sup>-1</sup> (1 kcal = 4.184 kJ), experimental<sup>[74]</sup> or theoretical,<sup>[77]</sup> for X = OH, Cl and I. Theoretical results for X = F give the energy of CH<sub>3</sub> in fluoroethane as being 15.5 kcal mol<sup>-1</sup> more stable, that is, more negative, than in ethane while the energy of CH<sub>2</sub>F is found to increase by 15.6 kcal mol<sup>-1</sup> relative to its value in CH<sub>2</sub>FCH<sub>2</sub>F, the energy changes being accompanied by a transfer of 0.061 e from CH<sub>3</sub> to CH<sub>2</sub>F. The same principle applies to systems with delocalized electrons, the energy of pyridine, experimental or theoretical, equaling the mean of the energies of pyrazine and benzene to within 0.1 kcal mol<sup>-1</sup> using the measured heats of formation.<sup>[74]</sup>

Bader and Cortés-Guzmán close their paper by noting the similarity of the net result of compensatory transferability and Le Châtelier principle:<sup>[113]</sup>

There appears to be a Le Châtelier principle at work—one that states that two open systems brought into contact respond in such a way as to minimize the overall changes in form and energy, resulting in many cases in a conservation of energy.

Taking advantage of the transferability of QTAIM atoms, Breneman invented the “transferable atom equivalent” (TAE)<sup>[114–117]</sup> method, which is designed to reconstruct accurate approximations to ab initio electron densities at a tiny fraction of the

computational effort. A TAE is a mononuclear atomic electron density region delimited by zero-flux surfaces extracted from a mold molecule and stored in an electronic database which incorporates a large number of element combinations, atom-types, and electronic environments. The electron density reconstruction is achieved with the RECON program by recalling the desired TAE from the database, and after the necessary rotation and translation, each newly added fragment is melded to the growing reconstructed molecule to maximize the matching of its zero-flux surface with that of the growing molecule. The automatic melding of zero-flux surfaces is affected by allowing the quantum stress tensor to exert pressure on the two surfaces until the condition of pressure equalization is satisfied, a point at which the two nearly matching surfaces become melded into one gapless approximate interatomic zero-flux surface. This melding procedure necessitates some adjustments to bond lengths.

In addition to the TAE themselves, the database also stores derivatives of their properties with respect to radial variations of the zero-flux surface. These derivatives are used to correct atomic properties for their variations accompanying the change in the interatomic surface resulting from the melding. Molecular properties/descriptors such as total energy, total volume, surface area of a given isodensity envelope, electrostatic potential, dipole moment, Fukui function, and bare nuclear potential are then calculated by combining the properties of the composing TAEs.<sup>[114–117]</sup>

Several robust predictive QSAR models have been built from TAE. Examples include a model capable of the accurate prediction of the acute toxicity of 375 organic molecules to fish in which 300 molecules were used as a training set to construct the model while 75 molecules were left as the test set for which a “leave one out” cross-validated  $r^2$ , usually given the symbol  $q^2$ , of 0.81 was obtained.<sup>[117]</sup> The speed with which ab initio quality results can be obtained using the TAE approach is relevant to high speed screening of large molecular sets ubiquitous in the in silico phase of drug design.

## Categories of QTAIM Descriptors

QTAIM defines a number of atomic and bonding properties simultaneously from the topological and the topographical analysis of the electron density that extract atomically-resolved and bond-resolved levels of description of a molecule from its many-electron wavefunction. We can classify these QTAIM properties broadly as follows:

- i. **Bond properties.** These can be grouped into: (a) local properties evaluated at the bond critical point (BCP) such as the electron density ( $\rho_{\text{BCP}}$ ) or the energy densities (e.g.,  $G_{\text{BCP}}$ ,  $V_{\text{BCP}}$  and  $H_{\text{BCP}}$ ); (b) properties integrated along the bond path such as the bond path length; and (c) properties integrated over the interatomic zero-flux surface such as the integrated electron density.<sup>[17]</sup>
- ii. **Atomic properties [Prop. ( $\Omega$ )].** These are properties integrated over the 3-D atomic basin bounded by the union of zero-flux interatomic surfaces and, when extending to infinity, a chosen outer isodensity envelope,

normally taken as the 0.001 a.u. envelope. The general form of this integration is given in Eq. (11), which necessarily leads to additivity of atomic properties. Examples include the atomic energy [ $E(\Omega)$ ], the electrostatic charge  $q(\Omega)$  (electrostatic monopole) and higher multipoles such as the dipole moment [ $\mu(\Omega)$ ].<sup>[17]</sup>

- iii. **Interatomic properties [Prop. ( $\Omega, \Omega'$ )].** These are generally integrated properties that depend on pairs of atomic basins. Examples include the delocalization index (DI)  $\delta(\Omega, \Omega')$ <sup>[118]</sup> and the interaction energy components of the interacting quantum atoms (IQA).<sup>[119–121]</sup> One can also calculate interatomic properties that do not involve any integration involving two basins such as interatomic distances, interatomic Coulomb repulsion energies taking only atomic monopoles, or nuclear–nuclear repulsion energies. Interatomic properties can be organized in matrix format that can be used in QSAR (See the section entitled “Localization/Delocalization Matrices (LDM), Delocalization Matrices (DM), and Density-Weighted Connectivity Matrices (DWCM) in QSAR” below).
- iv. **MESP (and field) [ $V(\mathbf{r})$ ,  $E(\mathbf{r})$ ].** These fields are determined entirely by the charge density distribution<sup>[56–68]</sup> and represent an additional source of descriptors that are directly related to reactivity that can be used in QSAR (See the section entitled “Molecular Electrostatic Potential (MESP) and Field (MESF)” below). While these are not QTAIM-derived properties, the MESP has been shown to be accurately reproduced from truncated multipolar expansion representations of the atomic densities within molecules using the QTAIM multipoles.<sup>[122]</sup> Numerical evidence has shown, thus, that the relation between QTAIM atomic electrostatic properties and the MESP is a bijective mapping, not only in principle but also in computational practice as well.

Many of these QTAIM properties have been used in QSAR studies and drug design, but one can probably say safely that this is a field in its infancy despite of the enormous literature that has already accumulated. The purpose of the remainder of this article is not an exhaustive review of this extensive literature, an overwhelming task for the present author, but rather to emphasize (i) the breadth, depth, and potential practical utility of such an endeavor, with a necessarily biased sampling of the literature and (ii) to show that QTAIM is not exclusive of other approaches, as we read sometimes in the literature, in fact—to the contrary—it gains and complements (and is complemented by) other approaches.

## QTAIM Bond Properties in QSAR

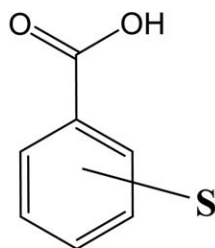
The predictive power of BCP properties-based descriptors is illustrated with the QTMS approach of Popelier and coworkers,<sup>[123–144]</sup> and the work of Buttingsrud et al.<sup>[145–147]</sup>

## Quantum topological molecular similarity (QTMS)

Substituent effects can be gauged by Hammett substituent constants,<sup>[1,5,47,148]</sup> that is, the change in the  $pK_a$  of benzoic

acid on substitution in the aromatic ring by the substituent S (Scheme 1):

$$\sigma = \log \frac{K_S}{K_H} = \rho K_{a,H} - \rho K_{a,S}. \quad (12)$$



Benzoic acid with a substituent S in the aromatic ring

Scheme 1.

Popelier developed a method termed “quantum topological molecular similarity” (QTMS) that has been tested extensively for its capability of reproducing the trends predicted from Hammett constants accurately.<sup>[123]</sup> The QTMS descriptors have been shown to convey essentially the same information contained in the Hammett substituent constants and may constitute an alternative in QSAR modeling since it is generally easier to compute QTMS descriptors than it is to determine the experimental values of new nontabulated Hammett constants.<sup>[124–126,129,131,140,142]</sup>

In the QTMS approach,  $\rho(\mathbf{r})$  is sampled at the BCPs, a sampling that entails several advantages such as the speed of the calculation and the emphasis on the chemically relevant bonding regions with only an indirect perturbing influence of the chemically irrelevant nuclear regions. Further, QTMS bypasses the molecular alignment problem, a considerable time-saver compared to methods requiring maximization with respect to the parameter  $\Theta$  in Eq. (7). We note in passing that unlike Eq. (7), however, QTMS cannot distinguish between enantiomers, which does not constitute a limitation as long as the collected experimental data include those obtained using the correct enantiomer(s). Recapping, the QTMS method has a number of advantages with respect to its applicability in QSAR-type drug design: (i) It is entirely based on the topology of the electron density in the chemically relevant bonding region and not dominated by nuclear maxima and (ii) its is considerably fast given that it does not require molecular prealignment and/or space integration as the electron density is only sampled at a set of point (the BCPs) rather than being considered in its entirety.

The QTMS method starts with the construction of a multidimensional mathematical space using bond properties determined at the BCPs. Each bond property is represented by a Cartesian axis after converting each to a dimensionless relative scale to avoid dimensional nonhomogeneity. The dimensionless Euclidian distance between two molecules in this mathematical space is then a measure of their dissimilarity, the larger the distance the less similar they are.

Examples of bond descriptors include, but are not limited to, the electron density at the BCP ( $\rho_{\text{BCP}}$ ), the Laplacian of  $\rho_{\text{BCP}}$

( $\nabla^2 \rho_{\text{BCP}}$ ), the bond ellipticity ( $\varepsilon$ ), the three principal curvatures of the electron density evaluated at the BCP ( $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ ), the kinetic energy density at the BCP ( $K_{\text{BCP}}$ ), the total energy density at the BCP ( $H_{\text{BCP}}$ ), and the equilibrium bond length ( $R_e$ ).

If we suppose the BCP space is constructed from  $\rho_{\text{BCP}}$ ,  $\nabla^2 \rho_{\text{BCP}}$  and  $\varepsilon$ , then the distance between molecules A and B is:

$$d(A, B) = \sum_{i \in A} \sum_{i \in B} \sqrt{[(\rho_{\text{BCP}}^A - \rho_{\text{BCP}}^B)^2 + (\nabla^2 \rho_{\text{BCP}}^A - \nabla^2 \rho_{\text{BCP}}^B)^2 + (\varepsilon^A - \varepsilon^B)^2]_i}, \quad (13)$$

where the sum runs over the  $i$  common bonds between the two molecules A and B, and where standardised BCP properties are used, which means that property  $x$  is replaced by  $\{[x - \text{mean}(x)]/\text{standard deviation of } x\}$ , to avoid dimensional inhomogeneity. In this way, the smaller  $d(A, B)$  the more similar are A and B with a lower bound of  $d = 0$  when A and B are one and the same molecule.

By construction, QTMS is an electronic structure fingerprinting tool, but it is devoid of a built-in capability to represent molecular size or hydrophobicity (as recognized by Popelier on a number of occasions). This is so since the sampling is of intensive properties (densities) rather than extensive properties (integrated densities). This feature can actually be a strength of QTMS rather than a limitation, because whenever QTMS alone (i.e., without the inclusion of extraneous steric or log  $P$  information) is capable of yielding a strongly predictive QSAR model, then this implies that the properties being modeled are determined primarily by electronic structure rather than by steric bulk, molecular size, or solvent partitioning tendencies. Conversely, failure of QTMS alone to predict activity is a diagnostic that steric bulk and/or hydrophobicity are important in determining activity. In the latter cases, QTMS can be used in conjunction with traditional steric bulk and/or hydrophobicity measures to yield powerful predictive combined models.

The exclusive sensitivity of QTMS to electronic structure parallels that of the Hammett  $\sigma$  constants. Both Hammett constants and QTMS must be supplemented with steric bulk and/or hydrophobicity descriptors when these are important in QSAR models. This parallelism between QTMS and Hammett constants lends merit to Popelier's proposal of the eventual replacement of the latter by the former given the relative ease with which QTMS descriptors may be obtained.<sup>[130,132]</sup>

A remarkable feature of QTMS is its apparent inherent capability to pinpoint the “active center” automatically. By this, we mean that the experimental Hammett constants ( $\sigma$ )-sequence relative to *p*-aminobenzoic acid (Table 1) is reproduced if the BCPs of the “active center,” and only the active center (here the bonds in the carboxylic group: O(–)H, C(=)O, and C(–)OH), are included in the distance calculation. The inclusion of any additional bond(s) destroys the agreement with the experimental sequence, and the more “irrelevant” bonds are incorporated in the Euclidian space the more the disagreement with experiment<sup>[123]</sup> (Table 1, especially the last column). As a corollary and by construction, the Euclidian QTMS molecular comparison space need not include (and probably is better not to include) all the common bonds of a series of

**Table 1.** Hammett  $\sigma$ -constants ranking of seven *p*-substituted benzoic acids relative to *p*-amino benzoic acid as the number of BCP defining the Euclidian distance Eq. (13) increases, and the number of ordering disagreement with the experimental sequence.<sup>[a]</sup>

Experimental sequence										
QTMS sequences	No. of BCPs	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	H	F	Cl	CN	NO <sub>2</sub>	# of disagreements/8
OH	1	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	H	F	Cl	CN	NO <sub>2</sub>	0
C—O	1	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	H	F	Cl	CN	NO <sub>2</sub>	0
COOH	3	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	H	F	Cl	CN	NO <sub>2</sub>	0
C—COOH	4	NH <sub>2</sub>	OCH <sub>3</sub>	F	CH <sub>3</sub>	Cl	H	CN	NO <sub>2</sub>	4
C <sub>6</sub>	6	NH <sub>2</sub>	CN	CH <sub>3</sub>	OCH <sub>3</sub>	H	Cl	NO <sub>2</sub>	F	5
C <sub>6</sub> H <sub>4</sub>	10	NH <sub>2</sub>	CH <sub>3</sub>	OCH <sub>3</sub>	CN	H	Cl	NO <sub>2</sub>	F	6
C <sub>6</sub> —COOH	10	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	Cl	CN	H	NO <sub>2</sub>	F	5
C <sub>6</sub> H <sub>4</sub> —COOH	14	NH <sub>2</sub>	OCH <sub>3</sub>	CH <sub>3</sub>	Cl	H	CN	NO <sub>2</sub>	F	5
S—C <sub>6</sub> H <sub>4</sub> —COOH <sup>[b]</sup>	15	NH <sub>2</sub>	CN	H	OCH <sub>3</sub>	CH <sub>3</sub>	NO <sub>2</sub>	F	Cl	7

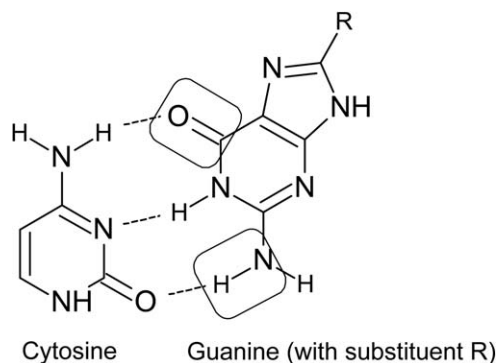
[a] Adapted from Ref. 123 with permission © 1999 American Chemical Society. [b] S denotes a substituent attached to the ring.

molecules. In other words, one can include in the QTMS distance calculation only those bonds existing in a common “subgraph” that maximizes the predictability of the model. That is to say, given a well-characterized active center in otherwise unrelated molecules (i.e., molecules differing in their molecular skeletons), the relative activity can be predicted from the distances computed from the properties of the bonds in this active center alone. Figure 4a supports this hypothesis as can be seen from the plot of  $pK_a$  values predicted on the basis of the Euclidian distance including only the bonds of the carboxylic groups of 40 unrelated organic acids<sup>[129]</sup> after a selection of the strongest predictors (from the set of BCP properties) on the basis of a partial least square (PLS) regression procedure. A less stringent test is presented in the plots of Figures 4b and 4c that demonstrate the ability of QTMS to predict the  $pK_a$ 's of related molecules with accuracy: Substituted anilines in 4b and substituted phenols in 4c. The possibility of chance correlations has been eliminated in this study through the “leave one out” cross validation<sup>[129]</sup> leading to final statistics that are of higher quality than those obtained from traditional descriptors such as Hammett constants.<sup>[129]</sup>

The versatility of QTMS has been demonstrated by its ability to predict a wide range of physicochemical and pharmacological properties. A few highlights from this work are given next for the sake of illustration.

Examples of predictive QSAR using QTMS only (without incorporating steric or log *P* information into the model) include the prediction of the proton affinities and the  $pK_a$ 's of 125 pyridine derivatives,<sup>[140]</sup> the relative bond dissociation enthalpies of phenolic antioxidants,<sup>[131]</sup> and the interaction energies of substituted cytosine–guanine (CG) Watson–Crick (WC) base pairs.<sup>[149]</sup> The predicted properties in this selection of examples are all primarily dependent on electronic (and much less on steric or solvent partitioning) properties. The latter study will now be briefly discussed.

Xue and Popelier (XP)<sup>[149]</sup> calculated the effect of 37 substituents at position 8 of guanine on the binding energy of the CG WC base pair. The position of the substituent R, attached to C8 of guanine, is indicated in the following structure with the most “predictive bonds” (see below) enclosed in boxes (Scheme 2):



Scheme 2.

The binding energies of the unsubstituted and substituted dimers were all calculated at the B3LYP/6-311+G(2d,p) level of density functional theory (DFT) with the inclusion of counterpoise basis set superposition error corrections.<sup>[150,151]</sup> If  $\Delta E$  represents the interaction energy of the unsubstituted GC base pair and  $\Delta E_S$  that of a guanine(C8)-substituted GC pair, then  $\Delta\Delta E \equiv \Delta E_S - \Delta E$  represents the change in  $\Delta E$  on substitution. Defined in this manner, a positive value of  $\Delta\Delta E$  implies a substitution that destabilizes the base pair and a negative value is a substitution that favors the hydrogen bonding of the WC pair. XP then compare a multiple linear regression model with a model constructed solely using Hammett  $\sigma_m$  constants. The QSAR models these authors report are:

$$\Delta\Delta E(\text{kJ mol}^{-1}) = 1.95(\pm 0.32) - 12.70(\pm 1.00)\sigma_m, \quad (14)$$

$$[r^2 = 0.82, \text{RMSD} = 1.25, n = 37]$$

where RMSD is root mean square deviation, using traditional Hammett constants, and:

$$\Delta\Delta E(\text{kJ mol}^{-1}) = -126.28 + 3865.81G_{\text{N-H}} - 705.74\varepsilon_{\text{C=O}}, \quad (15)$$

$$[\text{Training set : } r^2 = 0.96, \text{RMSD} = 0.58, n = 28]$$

$$[\text{Test set : } r^2 = 0.99, \text{RMSD} = 0.24, n = 9]$$

using the QTAIM properties,<sup>[17]</sup> where  $\varepsilon (\equiv \lambda_1/\lambda_2 - 1)$  is the bond ellipticity and  $G$  the gradient kinetic energy density ( $G = 1/2N \int (\nabla\Psi^* \cdot \nabla\Psi) d\tau'$ ) evaluated at the BCP, and where



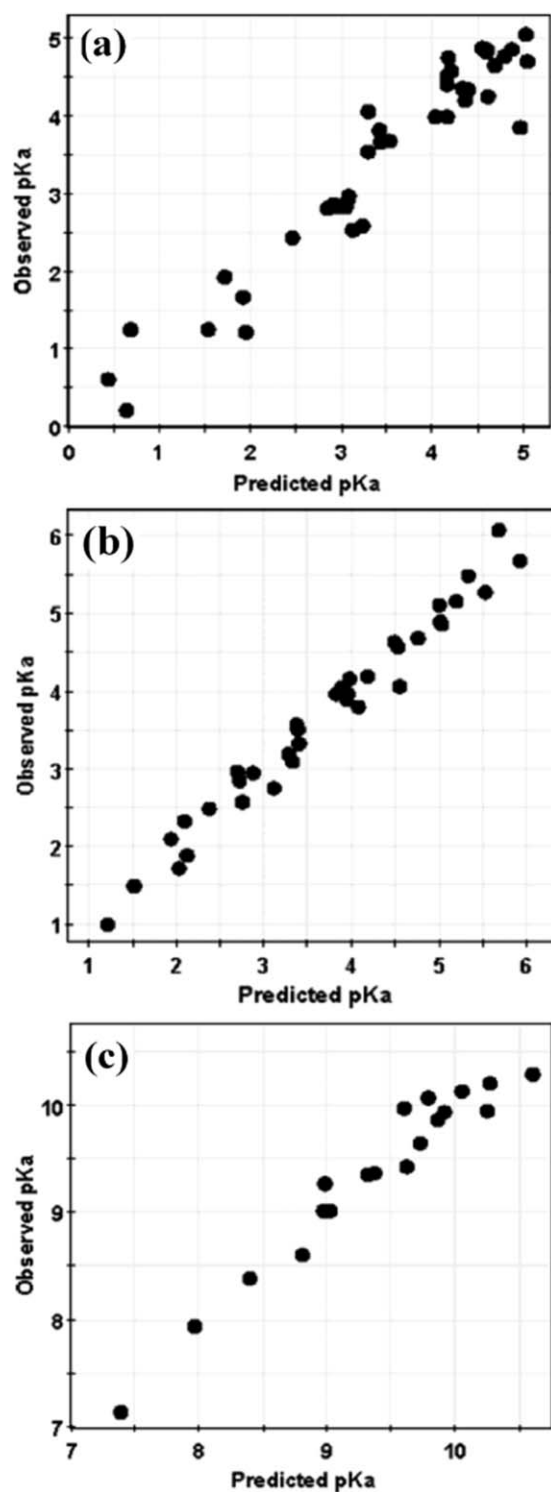


Figure 4. (a) Correlation of observed and predicted  $pK_a$ 's of a set of 40 generally unrelated small organic carboxylic acids. (b) Correlation of observed and predicted  $pK_a$ 's of 36 substituted anilines. (c) Correlation of observed and predicted  $pK_a$ 's of 19 substituted phenols. (Reproduced with permission from Ref. 129 © 2004 American Chemical Society).

the bonds are those enclosed in boxes in the chemical structure above.

As can be seen from the regression quality indicator lines below Eq. (15), the molecular set has been divided into two subsets: A training set used to construct the model and a test

external to the training set to test the performance of the constructed model outside of its bounds (this is necessary to ensure the predictivity of the QSAR model rather than just its ability to model the given data accurately). Clearly, the QSAR model based on QTAIM descriptors is characterized by significantly superior statistical quality indicators that the corresponding traditional Hammett constants-based model. This example, while not strictly QTMS since Eq. (13) has not been used, is nevertheless closely associated to QTMS since the descriptors used in Eq. (15) are typically incorporated into QTMS models. These results add weight to the proposal that QTMS can be an inexpensive alternative to the Hammett constants-based QSAR approach.

The utility of QTMS has been extended beyond the prediction of static molecular properties. For example, this approach has been shown capable of reproducing the rate constants ( $k$ ) of a model reaction, namely, the hydrolysis of esters:<sup>[127]</sup>



The alkaline hydrolysis rate constant at 25°C of 40 different esters (16 ethyl-containing and 24 other esters) were calculated from a QTMS QSAR model and compared with experimental values. The QSAR model yielded  $r^2 = 0.930$  and  $q^2 = 0.863$ . Further, an examination of the "variable importance in the projection (VIP)" obtained from the PLS regression has (correctly) identified the bonds in molecular fragment enclosed in square brackets,  $\text{R}[(\text{C}=\text{O})\text{O}]\text{R}'$ , as the most important in determining activity,<sup>[127]</sup> that is, essentially automatically locating the active center.

It is customary to regard the accurate prediction of  $pK_a$ 's as one of the initial validation steps of a new drug design approach. QTMS has moved beyond that stage and has been applied to the direct prediction of biological properties. For example, QTMS<sup>[152]</sup> has been shown to have comparable predictivity as the more traditional comparative molecular field analysis (CoMFA)<sup>[153]</sup> modeling of the toxicological indices of a series of health-hazardous lipid-soluble environmental pollutants (polyhalogenated dibenzo-*p*-dioxins).<sup>[154]</sup> A regression of  $r^2 = 0.91$  and  $q^2 = 0.86$  was obtained in a QTMS modeling of the cell-growth inhibitory activities of 15 substituted (*E*)-1-phenylbut-1-ene-3-ones.<sup>[155]</sup> The method has also delivered the automatic identification of the region in these congeners responsible for their biological activity, namely, the site of the Michael addition.<sup>[155]</sup> A study of the cytotoxicity of *o*-alkyl substituted 4-*X*-phenol (*X* being the substituent)<sup>[133]</sup> has shown that this biological activity is accurately predicted by QTMS alone (without the inclusion of any steric information), which indicates that cytotoxicity is primarily governed electronically, a result that does not support a previous proposal by the Hansch group that the latter is more important in determining activity.<sup>[156]</sup> The reverse situation is encountered in a study of the hepatotoxicity of a series of phenols, which demonstrates that, in this case, QTMS must be combined with external hydrophobicity descriptors such as  $\log P$ , (see the section entitled "Partitioning Between Immiscible Liquids and  $\log P$ " below) for a successful modeling,<sup>[134]</sup> liposolubility being a known key factor in determining

the hepatotoxicity of phenols. More recently and similarly, it has been shown that QSAR models for the toxicity of aldehydes, another class of environmental pollutants, to a model organism (*Tetrahymena pyriformis*) can be constructed from a proper mix of QTMS descriptors and  $\log K_{o/w}$ <sup>[141]</sup> [defined in Eq. (42) and discussed in the section entitled “Partitioning Between Immiscible Liquids and  $\log P$ ”]. An earlier study has also shown that  $\log K_{o/w}$  and  $E_{LUMO}$  are essential extraneous ingredients in the QTMS modeling of the toxicity of nitroaromatics, another class of pollutants, to *Saccharomyces cerevisiae*.<sup>[136]</sup>

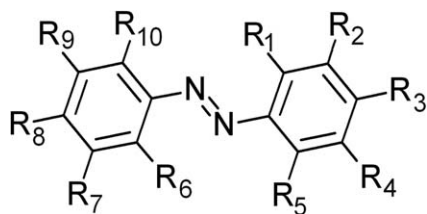
### Bond properties as predictors of spectroscopic transitions and NMR proton chemical shifts

The calculation of UV excitation spectra accurately remains a challenge for computational quantum chemistry to this day and usually necessitate an expensive high level of explicit treatment of electron correlation. The fast empirical modeling of such key analytical characteristics of molecules is, thus, of particular significance from the practical point of view.

The first HK theorem<sup>[39]</sup> [relation (5)] has been proven for nondegenerate *ground states*, as discussed above in the section entitled “The Electron Density  $\rho(r)$  and QSAR Modeling”. The mapping expressed in expression (5) shows, however, that the ground-state density,  $\rho(\mathbf{r})$ , does specify completely the Hamiltonian operator  $\hat{H}[\rho(\mathbf{r})]$  (Fig. 2). By specifying  $\hat{H}$  and through the time-independent many-particle Schrödinger equation, the ground-state density  $\rho(\mathbf{r})$  in principle also determines the *excited states* and their properties. In other words, *the energies of the excited states are also functionals of the ground-state density, even though they are not functionals of the corresponding excited state densities* since there exists different external potentials that can yield the same excited state density (i.e., there is no HK theorem for excited states).<sup>[157]</sup>

It is not entirely surprising, thus, to realize that (approximations to) *the ground-state density and ground-state geometry, both of which are mapped to the ground-state external potential, contain information about excited states*. In this section, we show that the ground-state bond properties do indeed encode information that can be used in the QSAR modeling of UV excitation energies.

About seven years ago, Buttingsrud, Alsberg, and Åstrand (BAA) have been able to accurately predict  $\lambda_{max}$  and the corresponding excitation energies ( $\Delta E_{h\nu}$ ) of 191 substituted azobenzene dyes from two sets of ground-state QSAR models: The first is based on optimized ground-state bond lengths, and the second on QTAIM BCP descriptors of the ground state.<sup>[146]</sup> The molecular set studied has the following general structure (Scheme 3):



Scheme 3.

where some of the R's may belong to condensed cyclic systems.

Several PLS models based only on optimized ground-state bond lengths (including different subsets of bonds in the molecules) have been tested in modeling the excitation properties. Even the simplest model, the one that includes only the N=N bond length, yields already an elevated  $r^2$  of 0.90 ( $\lambda_{max}$ ) and a corresponding  $r^2 = 0.86$  ( $\Delta E_{h\nu}$ ); indicating that this bond appears essential in describing the excitation<sup>[146]</sup> and hence represents the chromophore.

The cross validated statistics of the one-bond length model are  $q^2 = 0.89$  for the training set and  $q^2(\text{test}) = 0.91$  for the test set with a root mean square error of cross validation (RMSECV) of 11.2 nm, and a root mean square error of prediction (RMSEP) of 12.6 nm. The equivalent one-bond (N=N) QTAIM bond property model (based on two descriptors:  $\nabla^2 \rho_{BCP}$  and  $\varepsilon$ ) yields  $q^2 = 0.91$  for the training set and  $q^2(\text{test}) = 0.92$  for the test set with a RMSECV of 9.9 nm and a RMSEP of 12.3 nm.

The best QSAR of all in both sets of models are those that includes all bonds, in which case the statistics are (bond length model/QTAIM model):  $q^2 = 0.94/0.97$ ,  $q^2(\text{test}) = 0.96/0.98$ , RMSECV = 8.0/5.4 nm, and RMSEP = 7.6/5.4 nm. The above observations are consistent with the trends obtained from an inspection of Table 2 of Ref. 146 that (i) generally, QTAIM modeling yields more robust statistics than the modeling based on bond lengths only, and that (ii) there is an improvement, albeit modest, from the simplest QTAIM model (based only on the N=N bond) to the most complex based on all bonds, a small gain at the cost of a significant increase in the complexity of the model by the empirical fine-tuning of the perturbation of the chromophore by its surroundings beyond the perturbation captured in the mere change in the N=N bond length and properties.

In another study, BAA use QTAIM bond properties to predict 170 distinct NMR proton isotropic shieldings in 60 substituted benzenes obtained from the direct electronic structure calculation. The 60 molecules were grouped into four separate groups, with some overlap (some molecules are members of more than one group), according to the nature of the substituents: group I, II, III, and IV, which include 12, 11, 38, and 28 compounds, respectively.<sup>[145]</sup> The purpose of the study is not to calculate the NMR shifts themselves since these are calculated directly (and more accurately) by the authors, but rather to use these data as a test-case example of a local property to be predicted by QTAIM bond properties.

Models of increasing complexity ranging from the simplest model, which includes the QTAIM bond properties of only the bond shared by the proton of interest (model a) to the most complex model that includes the properties of all bonds (model e) are tested for their predictivity. The models are further classified in three categories: Those based on bond lengths only (models B); those based on three “standard” QTAIM bond properties, namely,  $\rho_{BCP}$ ,  $\nabla^2 \rho_{BCP}$  and  $\varepsilon$ , (model S); and finally those that include more QTAIM bond properties, a group of models termed “extended” (model E). An examination of Table 4 of Ref. 145 reveals a general and significant

improvement in the **S** models over the **B** models, but only a marginal improvement over the **S** models on passing to the **E** models. This observation is consistent across the table, hence we quote here only the extreme values for the simplest model **a** (one-bond only) and the most complex model **e** (all bonds) applied to group **I** of 12 molecules exhibiting 20 distinct proton shieldings in the order **B/S/E**: model **Ia**:  $q^2 = 0.6/0.98/0.98$ , RMSECV = 0.12/0.027/0.025 ppm, and model **Ie**:  $q^2 = 0.995/0.997/0.996$ , RMSECV = 0.014/0.011/0.012 ppm.<sup>[145]</sup>

In some cases, the simplest model **a** fails completely to capture the proton shielding (group **II** of chlorobenzenes). The modeling in this case fails presumably due to the size and relatively large electron population of chlorine (compared to fluorine, for example) and which may cause "through space" shielding that invalidates the BCP modeling. Again, as in QTMS, bond properties can be used in electronic fingerprinting but fail to capture size effects. Even in these failed cases, the accuracy of the modeling is quickly recovered in models of the complexity **c** (seven bonds) and beyond for this group of molecules. For a comparison (**B/S/E**): model **Ila**:  $q^2 = 0.1/0.5/0.0$ , RMSECV = 0.11/0.11/0.08 ppm, and model **Ile**:  $q^2 = 0.97/0.96/0.96$ , RMSECV = 0.019/0.022/0.020 ppm.<sup>[145]</sup>

## QTAIM Atomic Properties in QSAR

These properties are more expensive to obtain than bond properties since their calculation involves 3-D numerical integrations over atomic basins [Eq. (11)]. However, exploitation of transferability and additivity and the use of precalculated database libraries of transferable atomic and fragment properties can virtually eliminate computational time once these libraries have been constructed, as in Breneman's TAE approach outlined above in the section entitled "The Electron Density, Transferability, and Additivity".

### Atomic energy and charge of the acidic hydrogen as predictors of $pK_a$

Adam demonstrated that the  $pK_a$  of a weak acid is inversely proportional to the QTAIM energy of the dissociating (acidic) hydrogen in the parent undissociated acid molecule at a given set of thermodynamic conditions.<sup>[158]</sup> The milestones of his demonstration are summarized here.

The  $pK_a$  of an organic acid is related to the Gibbs energy of the acid dissociation reaction ( $HA + H_2O \rightleftharpoons H_3O^+ + A^-$ ) and is expressed:

$$pK_a \equiv -\ln K_a = \frac{\Delta G^0}{2.303 RT}. \quad (16)$$

Re-expressing  $\Delta G^0$  in terms of the molar dissociation energy ( $U^0$ ) and the standard molar partition functions [ $q^0$ ; not to be confused with atomic charge, which is given the symbol  $q(\Omega)$ ], we get:

$$pK_a = \frac{U_{A^-}^0 - U_{HA}^0}{2.303 RT} - \log \left( \frac{q_{A^-}^0}{q_{HA}^0} \right) + C', \quad (17)$$

where

$$C' \equiv \frac{U_{H_3O^+}^0 - U_{H_2O}^0}{2.303 RT} - \log \left( \frac{q_{H_3O^+}^0}{q_{H_2O}^0} \right), \quad (18)$$

and where

$$q_{A^-}^0 \approx q_{HA}^0 \rightarrow \log \left( \frac{q_{A^-}^0}{q_{HA}^0} \right) \approx 0. \quad (19)$$

The last term in Eq. (17), defined in Eq. (18), depends on the characteristics of  $H_2O$  and  $H_3O^+$  and the temperature and, hence, is independent of the nature of the acid. The assumption implicit in Eq. (19) is expected to be more valid the larger the acid molecule.<sup>[158]</sup> Substituting approximation (19) into Eq. (17):

$$pK_a = \frac{U_{A^-}^0 - U_{HA}^0}{2.303 RT} + C'. \quad (20)$$

The molecular dissociation energy is given by:

$$U^0 = E + E^{ZPE} - \sum_{i=1}^n E_i, \quad (21)$$

where  $E$  is the bottom-of-the-well energy of dissociation,  $E^{ZPE}$  the zero-point vibrational energy,  $E_i$  is the electronic energy of the separated  $i$ th ground-state atom. The numerator of the first term of the R.H.S. of Eq. (20) can then be re-expressed as:

$$U_{A^-}^0 - U_{HA}^0 = E_{A^-} - E_{HA} + E_H + E_{A^-}^{ZPE} - E_{HA}^{ZPE}, \quad (22)$$

in which  $E_{A^-}$  and  $E_{HA}$  are the vibrationless total energies of the anion and the undissociated acid, respectively. The difference  $E_{A^-}^{ZPE} - E_{HA}^{ZPE}$  is expected to be small and approximately constant for a homologous series, and  $E_H$  is the energy of free ground-state hydrogen atoms, which is another constant, leading to a re-expression of Eq. (20) as:

$$pK_a = \frac{E_{A^-} - E_{HA}}{2.303 RT} + C. \quad (23)$$

where  $C$  is a new constant.

Assuming that the atomic energy of the dissociating hydrogen atom can be approximated as the difference between the total energies of the anion and the neutral undissociated acid, especially for a larger acid molecule.<sup>[158]</sup>

$$E_{A^-} - E_{HA} \approx -E(H), \quad (24)$$

then,

$$pK_a = \frac{-E(H)}{2.303 RT} + C. \quad (25)$$

It is important to realize that this result is independent of the particular model chemistry, which also may or may not

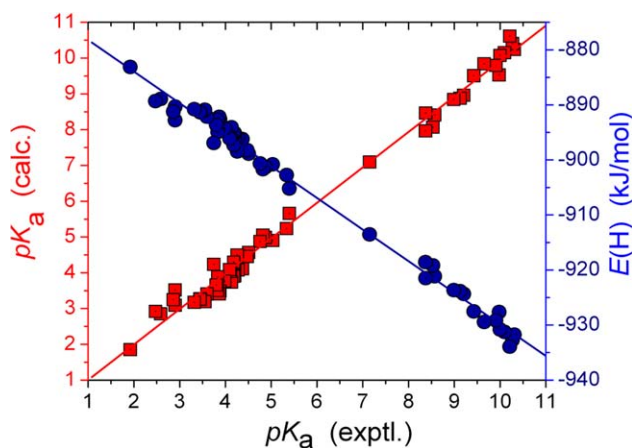


Figure 5. Experimental  $pK_a$  (abscissa) versus the values calculated from Eq. (26) (left ordinate, square data points) and the QAIM energy of the acidic hydrogen atom in the undissociated acid (right ordinate, circular data points), along with the corresponding least squares lines. (Plots drawn from the data listed in Table 13 of Ref. 158). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

incorporate solvation effects. The model embodied in Eq. (25) is validated by an excellent agreement with experiment. For instance, the following equation is the result of a statistical fitting of the  $pK_a$ 's of 53 unrelated organic acids to Eq. (25):<sup>[158]</sup>

$$pK_a(\text{exptl.}) = -150.486 - 0.1725 \times E(\text{H})(\text{kJ mol}^{-1}), \quad (26)$$

which yields an  $r^2$  of 0.991.

The correlation of the experimental and calculated  $pK_a$  values and the correlation of  $pK_a$ 's with the QAIM atomic energies of the acidic hydrogen are displayed in Figure 5. The statistical correlation indicators reported by Adam<sup>[158]</sup> based on Eqs. (25) and (26) for different sets of molecules are collected in Table 2.

Adam uses DFT [PW91/6-311+G(d,p)] model chemistry from which he obtains the atomic energies of the acidic hydrogen,<sup>[158]</sup> raising the questions of (i) the physical meaning of QAIM atomic energies obtained from DFT calculations and (ii) the effect of using a different level of theory with the possible result of obtaining significantly different atomic energies. The first question is addressed in detail in the Appendix of Ref. 159. As for the second question, QAIM atomic energies that are numerically obtained by the same integration procedure but from different underlying electronic structure theories (such as Hartree–Fock, Møller–Plesset perturbation theory, or DFT) are very strongly correlated statistically ( $r^2$  typically = 0.99) and

Table 3. Correlation of experimental  $pK_a$  values with the Hammett substituent constants  $\sigma$ , the charges of the acidic hydrogen  $q(\text{H})$ , and the total charges on the carboxylic group of monosubstituted benzoic acids.<sup>[a]</sup>

Substituent	$pK_a$	$\sigma$	$q(\text{H})$	$q(\text{COOH})$
H(benzoic acid)	4.19	0.00	0.5803	-0.1518
<i>m</i> -Bromo	3.81	0.39	0.5840	-0.1189
<i>m</i> -Chloro	3.83	0.37	0.5838	-0.1233
<i>m</i> -Cyano	3.60	0.56	0.5865	-0.1090
<i>m</i> -Fluoro	3.87	0.34	0.5830	-0.1279
<i>m</i> -Hydroxy	4.08	0.12	0.5806	-0.1439
<i>m</i> -Methoxy	4.09	0.12	0.5798	-0.1501
<i>m</i> -Methyl	4.27	-0.07	0.5796	-0.1571
<i>m</i> -Nitro	3.49	0.71	0.5868	-0.1008
<i>p</i> -Bromo	3.97	0.23	0.5843	-0.1316
<i>p</i> -Chloro	3.98	0.23	0.5826	-0.1388
<i>p</i> -Cyano	3.55	0.66	0.5859	-0.1216
<i>p</i> -Fluoro	4.14	0.06	0.5818	-0.1422
<i>p</i> -Hydroxy	4.57 <sup>[b]</sup>	-0.37	0.5790	-0.1642
<i>p</i> -Methoxy	4.47 <sup>[b]</sup>	-0.27	0.5783	-0.1628
<i>p</i> -Methyl	4.37	-0.17	0.5792	-0.1501
<i>p</i> -Nitro	3.43	0.78	0.5864	-0.1220
$\rho^{2[c]}$		0.9985	0.9134	0.8698

[a] Data were collected from Ref. 160. [b] The values entered for  $pK_a$  of these two acids in Table 1 of Ref. 160 are incorrect and were corrected in the present table. [c] The correlation coefficients are those of the respective single variable regression equations to reproduce the experimental  $pK_a$  values obtained from Ref. 160.

hence, from the point of view of statistical modeling, atomic energies contain essentially the same information no matter what underlying computational level of theory has been used and despite of their possibly significantly differing magnitudes.<sup>[32]</sup> References 30 and 31 contain additional discussions of the effects of underlying electronic structure calculation levels of theory on QAIM-derived properties.

Simultaneous to the appearance of Adam's paper, Hollingsworth, Seybold, and Hadad (HSH)<sup>[160]</sup> report QSAR models to predict the  $pK_a$  values of a series of substituted benzoic acids from models based on the atomic charges of the carboxylic group and of the acidic hydrogen, using several different definitions of atomic charges including that of QAIM. Various data from the HSH paper are assembled in Table 3, which lists the  $pK_a$ 's against the Hammett constants [defined in Eq. (12)], and the total QAIM charges of the acidic hydrogens,  $q(\text{H})$ , and of the carboxylic acid group,  $q(\text{COOH})$ . As can be seen from the table, Hammett constants are the better predictors in this case ( $r^2 = 1.000$ ), which is expected since they are constructed to deliver  $pK_a$ 's of substituted benzoic acids, but the QAIM charges of the acidic hydrogens are also excellent predictors of the respective acidities ( $r^2 = 0.913$ ):<sup>[160]</sup>

Table 2. Adam's statistics<sup>[158]</sup> obtained using Eq. (26) to calculate the  $pK_a$ 's of different families of organic compounds.

Compound families	$n^{[a]}$	$r^2$	Average deviation	Maximum absolute deviation
Aliphatic carboxylic acids	19	0.98	0.10	0.34
Substituted phenols	20	0.92	0.21	0.60
Substituted anilinium ions	12	0.96	0.14	0.27
3,4-Substituted pyridinium ions	11	0.97	0.33	0.68
Hydrated aliphatic acids, substituted benzoic acids, and substituted phenols	53	0.99	0.22	0.62

[a]  $n$  is the number of compounds in the series and  $r$  is Pearson's regression coefficient.

$$pK_a(\text{exptl.}) = 69.5(\pm 5.2) - 112.5(\pm 8.9) \times q(\text{H})(\text{a.u.}), \quad (27)$$

which is in line with the simple intuition that an acid with a more electropositive acidic hydrogen (greater positive charge,  $q(\text{H}) > 0$ ) is more acidic and, thus, has a lower  $pK_a$  value, as reflected in the negative coefficient of the second term.

There is a striking consistency and similarity in the form of Eqs. (27) and (26), which deserves a comment. Atomic energies are always negative ( $E(\Omega) < 0$ , always), and they usually become more negative with an increase in the electron population [ $N(\Omega) = Z_\Omega - q(\Omega)$ ] of a given atom. This is especially true for electropositive hydrogen atoms where the intra-atomic electron–electron repulsion is nonexistent. It is, thus, reasonable that the intercept in Eq. (26) is also negative leading to a lower  $pK_a$  for an acid with more electropositive acidic hydrogen since it has a higher (less negative) energy, just as HSH result, Eq. (27), suggests.

### QTAIM descriptors of bulk and charge-separation for the prediction of partial molar volumes (PMV)

The partial molar volume (PMV, or  $V^0$ ) is the rate of change of the volume of a solution with respect to the number of moles  $n$  of solute, extrapolated to infinite dilution. This quantity has the units ( $\text{L}^3 \text{mol}^{-1}$ ) and for a two-component system is defined:

$$V^0 = \left( \frac{\partial V}{\partial n_{\text{solute}}} \right)_{T,P,n_{\text{solvent}}}, \quad (28)$$

where  $V$  is the volume of the solution at a given temperature ( $T$ ), pressure ( $P$ ), and number ( $n$ ) of moles of solvent. Equivalently,  $V^0$  is the first derivative of the chemical potential of the solute with respect to the pressure.<sup>[161]</sup> Experimentally, the “apparent” PMV ( $V_{\text{app}}^0$ ) is measured, which only equals  $V^0$  at infinite dilution (to eliminate solute–solute association) since<sup>[162]</sup>

$$V^0 = V_{\text{app}}^0 + m \frac{\partial V_{\text{app}}^0}{\partial m}, \quad (29)$$

where the concentration of the solute is expressed in molality ( $m$ ), and thus, the second term of Eq. (29) vanishes only at infinite dilution (when  $m = 0$ ). All reported experimental PMVs are to be understood, thus, as extrapolated values at infinite dilution.

PMVs are additive and can be estimated from group additivity schemes,<sup>[162–164]</sup> paralleling the behavior of QTAIM atomic properties expressed in Eq. (11). Further, the observed  $V^0$  can be considered as the resultant of two principal opposing contributions: A positive intrinsic steric bulk contribution due to the physical space occupied by each solute molecule, which displaces the surrounding solvent molecules ( $V_{\text{intr.}}^0$ ), and a generally negative electrostriction contribution ( $V_{\text{electr.}}^0$ ) due to the shrinking of the solvent shells surrounding the solute when both the solute and the solvent are polar ( $V_{\text{electr.}}^0$  can at times be positive if the solvent–solvent attraction is stronger than the solute–solvent attraction). Ignoring other contributions such as the temperature-dependent solute molecular vibration contribution (positive), we can write<sup>[163]</sup>

$$V^0 \approx V_{\text{intr.}}^0 + V_{\text{electr.}}^0, \quad (30)$$

which can be expressed in terms of QTAIM atomic contributions (after substituting the symbol for approximate equality by that of equality for simplicity):

$$V^0 = \sum_{\Omega} V_{\text{intr.}}^0(\Omega) + \sum_{\Omega} V_{\text{electr.}}^0(\Omega) = \sum_{\Omega} [V_{\text{intr.}}^0(\Omega) + V_{\text{electr.}}^0(\Omega)] = \sum_{\Omega} V^0(\Omega), \quad (31)$$

leading to the definition of an atomic contribution to the PMV as

$$V^0(\Omega) = V_{\text{intr.}}^0(\Omega) + V_{\text{electr.}}^0(\Omega), \quad (32)$$

where the atomic intrinsic volume is identified with the QTAIM atomic volume up to the  $\rho = 0.001$  a.u. isodensity envelope, namely, the van der Waals volume  $V(\text{vdW})_{\Omega} \equiv V_{\text{intr.}}^0(\Omega)$ , leading to the equivalent definition for a group (or a molecule)  $R$ :

$$V_{\text{intr.}}^0 = \sum_{\Omega} V_{\text{intr.}}^0(\Omega) \equiv V(\text{vdW})_R = \sum_{\Omega} V(\text{vdW})_{\Omega}, \quad (33)$$

where the sum runs over all the atoms in  $R$ .

The atomic electrostriction contribution in water is equated in this modeling to the magnitude of the atomic charge since a charge of either sign will pull water molecules toward the solute (with reversed orientation), ignoring the solvent–solute and electric charge sign-dependent difference in the magnitude of electrostriction. The justification of this assumption is empirical, that is, the ability of the resultant model(s) to fit experimental data with accuracy. For a molecule or a group of several atoms  $R$ , we identify the charge separation index (CSI),<sup>[165]</sup> namely the sum of the magnitudes of the atomic charges, as its electrostriction term:

$$V_{\text{electr.}}^0 = \sum_{\Omega} V_{\text{electr.}}^0(\Omega) \equiv \text{CSI} \equiv \sum_{\Omega} |q(\Omega)|, \quad (34)$$

where multiplication of the CSI by a constant of unit magnitude and of dimensions [volume/charge] is implied.

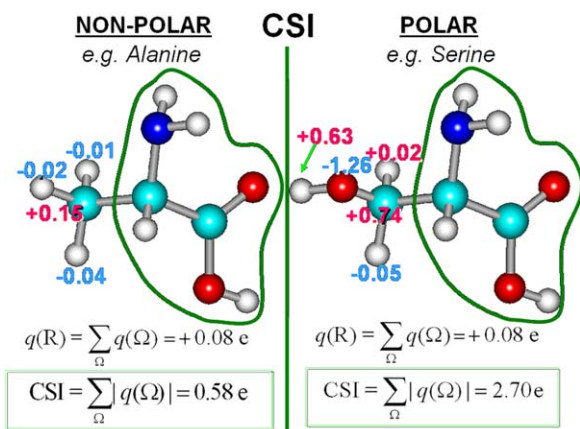
Figure 6 displays the atomic charges on the side chains of a nonpolar amino acid (alanine) and a polar amino amino acid (serine) of roughly similar molecular size. The total charge on the two side chains ( $R$ ) is  $+0.08$  indicating very little charge transfer to the  $-\text{C}_x\text{H}_z(\text{NH}_2)\text{COOH}$  group. The CSI clearly distinguishes the polar and nonpolar side chain being 0.58 and 2.70 a.u., respectively.

One can construct a QSAR model on the basis of the two contributions described above to model the experimental PMVs of the hydrogen-capped genetically encoded amino acids side chains<sup>[162]</sup> leading to<sup>[80]</sup>

$$V_{\text{AA}}^0 (\text{cm}^3 \text{mol}^{-1}) = 37.269 + 0.098 \times V(\text{vdW})_R (\text{a.u.}) - 0.889 \times \text{CSI}_R (\text{a.u.}), \quad (35)$$

$$[r^2 = 0.978, \text{RMSD} = 0.823, n = 20]$$

where the subscript AA refers to an entire amino acid and  $R$  refers to its side chain.



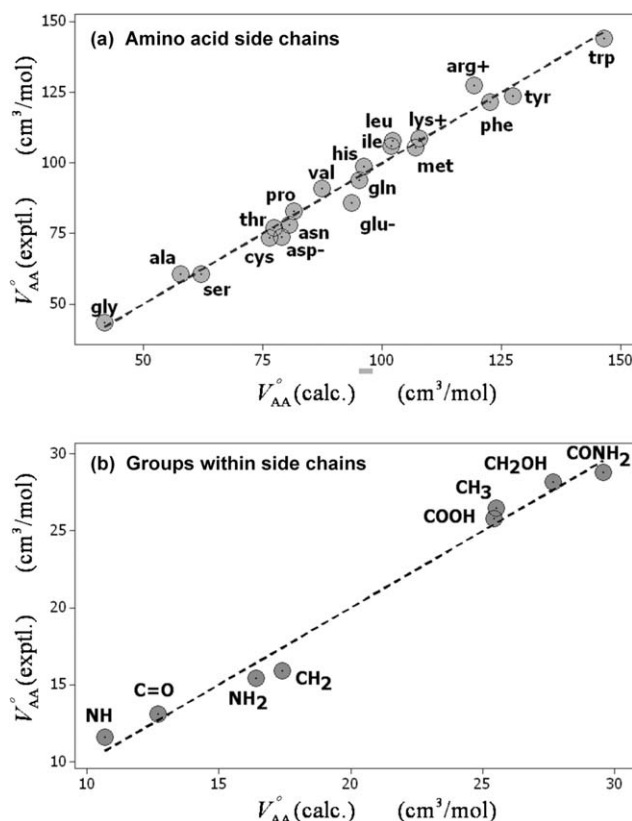
**Figure 6.** The CSI of the side chain of a nonpolar (alanine) and a polar (serine) amino acid. The numerical labels are the QAIM atomic charges on the side chain atoms in a.u. The two side chains are electrically neutral with sums of the atomic charges below 0.1 a.u. each. CSI defined in Eq. (34) is capable of distinguishing the nonpolar and polar amino acid side chains. The common group  $-\text{C}_\alpha\text{H}_2(\text{NH}_2)\text{COOH}$  is ignored since it is the same in all amino acids. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

The agreement between the experimental and calculated values according to this equation can be gleaned from the upper section of Table 4 and upper plot of Figure 7. The

**Table 4.** Experimental and calculated PMVs ( $V^0$ ) of the free amino acids in water (25°C) and some of the groups composing them and the terms used in the modeling.<sup>[a]</sup>

	$V^0$ (Exptl.)	$V^0$ (Calc)	$V(\text{vdW})$	CSI
Amino acid				
gly	43.3	41.9	47.25	0.009
ala	60.5	57.8	212.22	0.221
ser	60.6	62.0	277.93	2.707
cys	73.4	76.3	406.30	0.741
asp <sup>(-)</sup>	73.8	79.0	473.10	5.083
thr	76.9	77.3	435.30	2.805
asn	78.0	80.5	493.10	5.487
pro	82.8	81.4	461.70	1.070
glu <sup>(-)</sup>	85.9	93.6	625.23	5.350
val	90.8	87.3	519.70	0.737
gln	93.9	95.2	644.87	5.665
his	98.8	96.2	666.83	6.982
met	105.4	106.9	715.79	0.275
ile	105.8	101.9	670.83	1.010
leu	107.8	102.2	674.48	1.019
lys <sup>(+)</sup>	108.5	107.8	760.30	4.175
phe	121.5	122.5	879.93	0.698
tyr	123.6	127.3	947.65	2.826
arg <sup>(+)</sup>	127.3	119.1	926.27	9.679
trp	143.9	146.3	1146.63	3.238
Group				
-NH-	11.6	10.7	126.83	1.826
-C=O	13.1	12.7	167.78	3.119
-NH <sub>2</sub>	15.4	16.4	175.63	2.013
-CH <sub>2</sub> -	15.9	17.4	150.49	0.309
-COOH	25.8	25.5	306.26	5.077
-CH <sub>3</sub>	26.5	25.5	213.79	0.277
-CH <sub>2</sub> OH	28.2	27.7	277.93	2.707
-CONH <sub>2</sub>	28.8	29.6	344.49	5.363

[a] Experimental values taken from Ref. 162.  $V^0$  values are in  $\text{cm}^3 \text{mol}^{-1}$ , van der Waals volume and CSI are in a.u. The calculated  $V^0$  are obtained from Eqs. (35) and (36).



**Figure 7.** Experimental versus calculated PMVs (a) of the natural amino acids and (b) of functional groups composing them. (Redrawn after Ref. 80).

agreement between the calculated and experimental values is remarkable especially that only two parameters were used in the fitting of 20 PMVs (with a ratio of data points to parameters = 10:1). This agreement between the QSAR model and experiment extends to the PMVs of the chemical groups composing the amino acids.<sup>[80]</sup>

$$V_G^0 (\text{cm}^3 \text{mol}^{-1}) = -0.925 + 0.127 \times V(\text{vdW})_G (\text{a.u.}) - 2.456 \times \text{CSI}_G (\text{a.u.}),$$

$$[r^2 = 0.983, \text{RMSD} = 0.314, n = 8]$$
(36)

where  $V_G^0$  is the empirical additive group contributions to the PMV,<sup>[162]</sup> and the subscript G is a short for "group." The bottom part of Table 4 and bottom plot in Figure 7 illustrate the agreement between the empirical and calculated PMV of the groups composing the naturally occurring amino acids according to model (36).

The coefficients of the last two terms of the R.H.S.' of Eqs. (35) and (36) exhibit the expected signs: Positive for the intrinsic volume terms and negative for the electrostriction terms. There is, however, a subtle but important difference between these two equations: The former has a significant intercept while the latter has practically none. The intercept of about  $37.3 \text{ cm}^3 \text{mol}^{-1}$  in Eq. (35) represents the (predicted) PMV of the  $-\text{C}_\alpha\text{H}_2(\text{NH}_3^+)\text{COO}^-$  group which is common to all amino acids and which has not been directly modeled in this work. The (experimental) contribution of that group to the PMV can

be estimated using the additivity property of PMVs and the experimental values listed in Table 4:

$$\begin{aligned} V^0[\text{C}_x\text{H}_x(\text{NH}_3^+)\text{COO}^-] &\approx V^0[\text{H}_3\text{C}-\text{C}_x\text{H}_x(\text{NH}_3^+)\text{COO}^-] - V^0[-\text{CH}_3] \\ &= 60.5 - 26.5 = 34.0 \text{ cm}^3 \text{ mol}^{-1} \end{aligned} \quad (37)$$

The experimentally estimated value of  $34.0 \text{ cm}^3 \text{ mol}^{-1}$  is gratifyingly close to the intercept of Eq. (35). Further, the estimated PMV of the zwitter-ionic  $-\text{C}_x\text{H}_x(\text{NH}_3^+)\text{COO}^-$  group,  $37.3 \text{ cm}^3 \text{ mol}^{-1}$ , is significantly smaller than that of a hypothetical non-zwitter ionic  $-\text{C}_x\text{H}_x(\text{NH}_2)\text{COOH}$  group estimated to be  $52.7 \text{ cm}^3 \text{ mol}^{-1}$  using Eq. (36) and the group values listed in Table 4, and this is also expected due to electrostriction shrinkage accompanying the charge-separated zwitter-ionic form.

Another important class of biomolecules are the five nucleic acid bases: Adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U). Lee and Chalikian<sup>[166]</sup> determined accurately the PMVs of the five corresponding nucleosides as well as those of free A, C, T, and U bases at 18 and 55°C (the PMV of the free G base was not determined due to experimental impediments). In the following model, it is shown that the QSAR modeling using again an intrinsic volume term and an electrostriction term is sufficiently sensitive to capture not only the PMVs but also their small changes as a function of a temperature change of 37°.

The contribution of the sugar moiety to the PMV of the nucleosides can be evaluated using the additivity and transferability of groups contributions by assuming:<sup>[167]</sup>

$$V_{\text{sugar}}^0 \approx V_{\text{nucleoside}}^0 - V_{\text{base}}^0 \quad (38)$$

The entries in Table 5 demonstrate the simultaneous constancy (transferability) and additivity of the contribution of the sugar moiety to the PMV ( $80.4 \pm 1.0 \text{ cm}^3 \text{ mol}^{-1}$ ), Eq. (38), a value unaffected by the temperature change from 18 to 55°C within the combined experimental and statistical precision. The empirical PMV of guanine that has not been directly measured can be estimated from the difference between the PMV of guanosine minus the constant contribution of the sugar moiety, that is<sup>[167]</sup>

$$V_{\text{guanine}}^0 \approx V_{\text{guanosine}}^0 - V_{\text{sugar}}^0 = V_{\text{guanosine}}^0 - 80.4 \pm 1.0 \text{ cm}^3 \text{ mol}^{-1} \quad (39)$$

With a complete data set of PMV of all five nucleic acid bases, we can now proceed as before and model their PMV with an intrinsic and an electrostriction term to obtain<sup>[167]</sup>

$$\begin{aligned} V_{18\text{C}}^0 &= 0.0951 \times V(\text{vdW}) - 1.2530 \times \text{CSI}, \\ [r^2 &= 0.9798, \text{RMSD} = 0.5507, n = 5] \end{aligned} \quad (40)$$

and

$$\begin{aligned} V_{55\text{C}}^0 &= 0.1005 \times V(\text{vdW}) - 1.3093 \times \text{CSI}, \\ [r^2 &= 0.9666, \text{RMSD} = 0.7664, n = 5] \end{aligned} \quad (41)$$

The excellent agreement of the modeling at both temperatures can be seen from the listings in Table 5 and the plots of

**Table 5.** Experimental and calculated PMVs of nucleic acid bases, experimental PMVs of nucleosides, and estimates of the contribution of the ribose sugar to the PMV of the nucleosides.

Molecule	$V^0$ (18°C) (Exptl)	$V^0$ (18°C) (Calc)	$V^0$ (55°C) (Exptl)	$V^0$ (55°C) (Calc)
Experimental <sup>[a]</sup>				
Free bases				
Uracil	70.2 ± 0.4	69.1	74.6 ± 0.5	74.7
Cytosine	72.4 ± 0.4	74.0	75.8 ± 0.5	79.1
Thymine	86.3 ± 0.4	86.0	91.6 ± 0.5	90.3
Adenine	88.0 ± 0.4	87.5	93.5 ± 0.5	92.2
Guanine <sup>[b]</sup>	91.6 ± 1.0 <sup>[b]</sup>	91.8	97.5 ± 1.2 <sup>[b]</sup>	96.6
Nucleosides				
Uridine	150.7 ± 0.6		154.8 ± 0.7	
Cytidine	152.2 ± 0.6		156.4 ± 0.7	
Thymidine	166.4 ± 0.6		170.4 ± 0.7	
Adenosine	169.2 ± 0.6		175.3 ± 0.7	
Guanosine	172.0 ± 0.6		177.8 ± 0.7	
Differences <sup>[c]</sup>				
Sugar (U)	80.5 ± 1.0		80.2 ± 1.0	
Sugar (C)	79.8 ± 1.0		80.6 ± 1.0	
Sugar (T)	80.1 ± 1.0		78.8 ± 1.0	
Sugar (A)	81.2 ± 1.0		81.8 ± 1.0	
Sugar(average)	80.4 ± 1.0		80.4 ± 1.0	

[a] Experimental data, except for guanine, are obtained from Ref. 166.

[b] Calculated from Eq. (39). [c] Calculated from Eq. (38).

Figure 8. The increase in the average molecular kinetic energy and population shifts to higher vibrational levels increase the effective molecular volume at the higher temperature, which is reflected in the larger effective PMV at 55°C, an effect well-reproduced by the fitting.

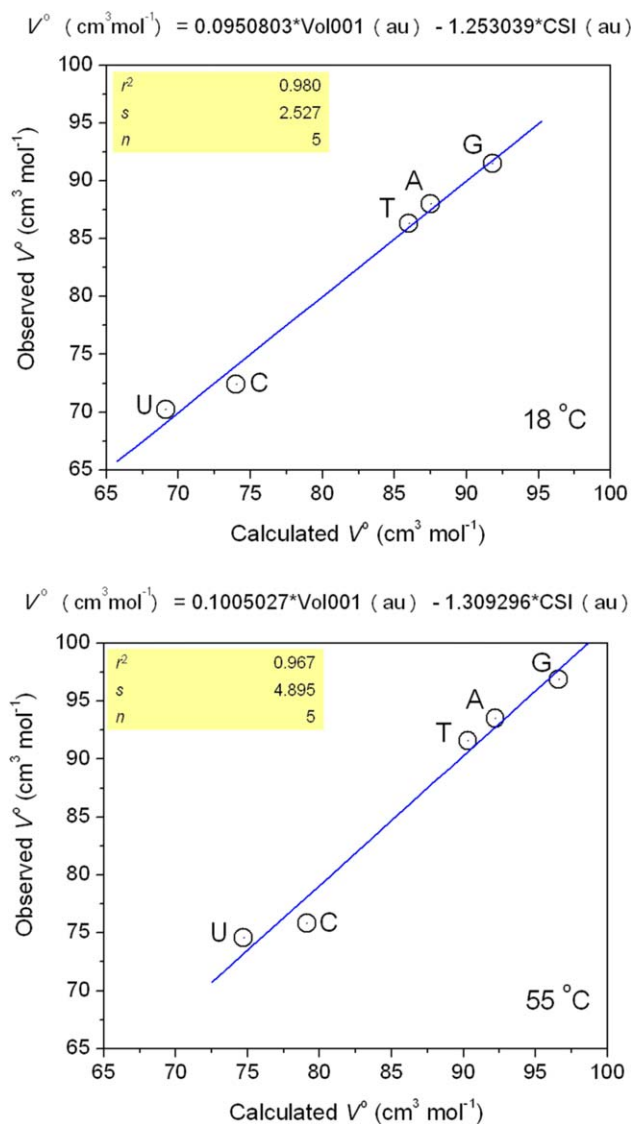
### Partitioning between immiscible liquids and log *P*

A solute in excess at equilibrium with two immiscible liquid phases sharing an interface will result in saturated solutions containing different concentrations of the solute in each of the two phases. In the absence of an excess of solute, the solute will be partitioned between the two phases in the same ratio if the activity coefficient remains approximately constant.<sup>[168]</sup> The ratio of the solute concentrations in two touching phases in equilibrium,  $C_1$  and  $C_2$ , respectively, namely the equilibrium constant  $K$  is also known as the distribution ratio, the distribution coefficient, or more commonly as the partition coefficient ( $P_{o/w}$  or simply  $P$ ). Often the two solvents include an organic hydrophobic/oily phase (often octanol) and an aqueous phase. Thus, the partition coefficient, defined as

$$K = K_{o/w} = \frac{C_o}{C_w} = \frac{\text{conc. in organic phase}}{\text{conc. in aqueous phase}} \equiv P_{o/w}, \quad (42)$$

measures the tendency of a compound to be preferentially distributed in the organic or in the aqueous phase. The concentrations (often molarities) rather than activities are used in the definition of  $P$  since at equilibrium the activities of the solute in both the organic and the aqueous phase are equal and in this case  $K$  would always be unity, which is not useful.

Due to the form of thermodynamic equations, it is more common to use log  $P$  in QSAR modeling rather than  $P$  itself.



**Figure 8.** Observed versus calculated aqueous PMVs at infinite dilution of the five nucleic acid bases adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) at 18 °C and 55 °C. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

The value of  $\log P$  is smaller than zero when the concentration of the solute is higher in the aqueous phase, that is, in the case of hydrophilic solutes, and conversely it is positive for hydrophobic solutes.

The partition coefficient plays a central role in pharmaceutical sciences and technology as it governs the absorption, distribution, and binding of drugs and determines a diversity of phenomena from the activity of narcotic drugs to the efficiency of solvent extraction of a compound from its solution in another solvent.<sup>[168]</sup> For these reasons, extensive compilations of  $\log P$  obtained from experimental measurement and/or estimated from group contributions have long been available.<sup>[2]</sup> The importance of  $\log P$  has been underscored, more recently, in the Lipinski's rule of five obeyed by the majority of orally active drugs. The rule states that orally active drugs must satisfy at least three of the following criteria:<sup>[169]</sup>

- A molecular mass less than 500 D.
- No more than five hydrogen bond donor groups.
- No more than 10 hydrogen bond acceptor groups.
- An octanol–water partition coefficient of less than 5 ( $\log P < 5$ ).

Drugs that significantly violate these rules, except compounds that are substrates for biological transporters, are poorly absorbed and have to be administered by other routes, for example by injection.<sup>[169]</sup> (The rule acquired its name apparently because all the numbers that appear in it are multiples of five).

Partitioning between two phases across an interface such as a biomembrane is common in the living system. For example, the serum of blood in contact with the neuronal membranes constitute the blood–brain-barrier that must be crossed by any drug intended to act inside the brain tissue. Another example is the necessary ability of a molecule to cross the mucosa of the small intestine before reaching the systemic circulation. This passage through the intestinal mucosa is frequently accomplished by passive diffusion through partitioning of the drug between the aqueous phase in the lumen of the intestine, the oily phase of the mucosa, and finally again an aqueous phase on the other side of the mucosa, that is, the blood serum.

Soluble serum proteins are often exploited as drug reservoirs for the slow and steady release of sparingly soluble drugs, the steady plasma concentration being maintained due to the partitioning with the protein reservoir. The equilibrium constant of binding of a molecules to a protein,  $K$ , can be predicted accurately from  $P$  by the equation:<sup>[2]</sup>

$$\log K = k_2 \log P + k_4, \quad (43)$$

where  $k_2$  and  $k_4$  are dimensionless fitting parameters, which have the values  $0.751 \pm 0.007$  and  $2.301 \pm 0.15$ , respectively, when the protein is bovine serum albumin ( $n = 42$  compounds), the standard error =  $0.159 \log K$  units, and an  $r^2 = 0.960$ .<sup>[11]</sup>

A pharmacological endpoint is often expressed at the L.H.S. of equations similar in form to Eq. (43) [Eq. (1)] with a R.H.S. consisting of  $f(P)$ . The linearity of Eq. (43) is deceptive since it is merely an approximation to a linear tail of a more general parabolic functional relationship between  $\log(1/C)$  and  $\log P$  when the range of values is enlarged. The more general relation has the form:<sup>[1–3,5]</sup>

$$\log \frac{1}{C} = -k_1 (\log P)^2 + k_2 \log P + k_3 \sigma + k_4, \quad (44)$$

where  $k_1$  and  $k_3$  are fitting parameters,  $\sigma$  is the Hammett substituent constant defined in Eq. (12), and where the second and fourth terms are the R.H.S. of Eq. (43). The third term,  $k_3 \sigma$ , is of importance only for molecules that possess a common skeleton with different substituents but loses its relevance when the molecular set used to construct the QSAR model do not share a common skeleton, in which case the equation simplifies to<sup>[5]</sup>



**Table 6.** Amino acid (AA) side chain (R) QTAIM descriptors (CSI and van der Waals' volume) against the second letter of the mRNA codons and its chemical class, and experimental and calculated standard Gibbs energies of transfer of the hydrogen capped side chains to water from the gas-phase, from cyclohexane, and from octane.<sup>[a]</sup>

AA	Descriptors		Modeled properties							
			mRNA 2nd letter	Base <sup>[b]</sup>	$\Delta G_{\text{gas} \rightarrow \text{w}}^0$		$\Delta G_{\text{cyc} \rightarrow \text{w}}^0$		$\Delta G_{\text{oct} \rightarrow \text{w}}^0$	
	CSI <sub>R</sub>	V(vdW) <sub>R</sub>			exptl.	calc.	exptl.	calc.	exptl.	calc.
arg <sup>+</sup>	9.679	926.27	G	I	-19.92	-19.28	-14.92	-12.65	-1.32	-0.58
his <sup>+</sup>	6.982	666.83	A	I	-10.27	-13.42	-4.66	-9.13	0.95	-0.42
gln	5.665	644.87	A	I	-9.38	-10.57	-5.54	-6.70	-0.07	-0.06
asn	5.487	493.10	A	I	-9.68	-10.18	-6.64	-7.39	-0.01	-0.42
glu <sup>-</sup>	5.350	625.23	A	I	-10.20	-9.88	-6.81	-6.22	-0.79	-0.01
asp <sup>-</sup>	5.083	473.10	A	I	-10.95	-9.30	-8.72	-6.73		-0.34
lys <sup>+</sup>	4.175	760.30	A	I	-9.52	-7.33	-5.55	-2.99	0.08	0.73
trp	3.238	1146.63	G	I	-5.88	-5.30	2.33	1.50	2.51	2.08
tyr	2.826	947.65	A	I	-6.11	-4.41	-0.14	0.95	1.63	1.67
thr	2.805	435.30	C	II	-4.88	-4.36	-2.57	-2.52	0.27	0.29
ser	2.707	277.93	G/C	I/II <sup>[c]</sup>	-5.06	-4.15	-3.40	-3.41	0.04	-0.11
pro	1.070	461.70	C	II		-0.59		1.06		0.91
leu	1.019	674.48	U	II	2.28	-0.48	4.92	2.62	1.76	1.51
ile	1.010	670.83	U	II	2.15	-0.46	4.92	2.61	2.04	1.50
cys	0.741	406.30	G	I	-1.24	0.12	1.28	1.33		0.87
val	0.737	519.70	U	II	1.99	0.13	4.04	2.11	1.18	1.18
phe	0.698	879.93	U	II	-0.76	0.21	2.98	4.66	2.09	2.17
met	0.275	715.79	U	II	-1.48	1.13	2.35	4.36	1.32	1.86
ala	0.221	212.22	C	II	1.94	1.25	1.81	1.02	0.52	0.51
gly	0.009	47.25	G	I	2.39	1.71	0.94	0.31	0.00	0.13

[a] CSI and van der Waals volumes are in a.u. and all energies are in kcal/mol. Experimental values were determined by Wolfenden et al.<sup>[172]</sup> [b] Chemical nature of the base: I = purine, II = pyrimidine. [c] Serine is the only amino acid with a degenerate middle letter of the triplet code.

$$\log \frac{1}{C} = -k_1(\log P)^2 + k_2 \log P + k_4. \quad (45)$$

The parabolic dependence of  $\log(1/C)$  on  $\log P$  in Eq. (45) expresses the requirement that a molecule must possess an optimal  $\log P$ , referred to as  $\log P^0$ ,<sup>[170]</sup> for maximal pharmacological activity—higher or lower  $\log P$  results in a reduction in activity. The optimal  $\log P$  ensures that the solubility in the oily phase is just enough to cross biological membrane barriers but not elevated enough to hinder the eventual release of the drug from the membrane to be redistributed on the other (aqueous) side where the site of action is located. Consistent with these remarks is the observation that a  $\log P^0 \approx 2.3$  seems to be a universal requirement for a molecule to be a general anesthetic,<sup>[170]</sup> which is the value obtained from the fitting of the activities of ethers for which  $k_1 = 0.22$ ,  $k_2 = 1.04$ , and  $k_4 = 2.16$ .<sup>[170]</sup>

The partition coefficient [Eq. (42)] may be expressed as a ratio of the equilibrium mole fractions (rather than molarities) in which case it is labeled with a prime ( $P'$ ). The reason for this modified definition is the appearance of  $P'$  in the thermodynamic expression of the standard free energy of transfer of a solute from solvent 1 to solvent 2:<sup>[2]</sup>

$$\Delta G_{1 \rightarrow 2}^0 = RT \ln P', \quad (46)$$

which embodies the experimental procedure of measuring the Gibbs energy of transfer from one solute to the other from partitioning experiments.

Conceptually, the most drastic partitioning between two phases is the partitioning of a solute between its gas-phase

and its aqueous solution phase. The Gibbs energy of transfer from the gas-phase to the aqueous medium measures the affinity of a given solute to water, a property crucial in determining proteins folding in water.<sup>[171]</sup> The molar free energy of transfer from the gas phase to the aqueous phase, termed the molar free energy of hydration or "hydration potential" ( $\Delta G_{\text{hydr.}}$ ), has been accurately determined by Wolfenden et al.<sup>[172]</sup> at 25°C after correction to pH 7 for each of the hydrogen-capped amino acid side chains ((R); since the amino acids themselves have very small vapor pressures due to their zwitter-ionization in water). These hydration potentials are accurately modeled using a single fitting parameter that measures the hydrophilicity/hydrophobicity, namely, the CSI<sup>[80]</sup> [defined in Eq. (34)]:

$$\Delta G_{\text{hydr.}} = 1.727 - 2.170 \times \text{CSI}_R. \quad (47)$$

$[r^2 = 0.923, \text{RMSD} = 0.381, n = 19]$

The experimental and calculated  $\Delta G_{\text{hydr.}}$  are listed along with CSI<sub>R</sub> in Table 6, which is sorted in order of decreasing CSI (hydrophilicity). The correlation of the experimental and calculated values obtained from Eq. (47) is displayed in the top of Figure 9.

It has long been recognized that the second position of the triplet genetic codon of an amino acid is the most important in determining the physical properties of that amino acid.<sup>[173]</sup> By sorting the amino acids on the basis of the hydration potentials of the side chains, Wolfenden et al.<sup>[172]</sup> noted that in the mRNA genetic code, hydrophilic amino acids have a purine base in the second position (A or G) while hydrophobic

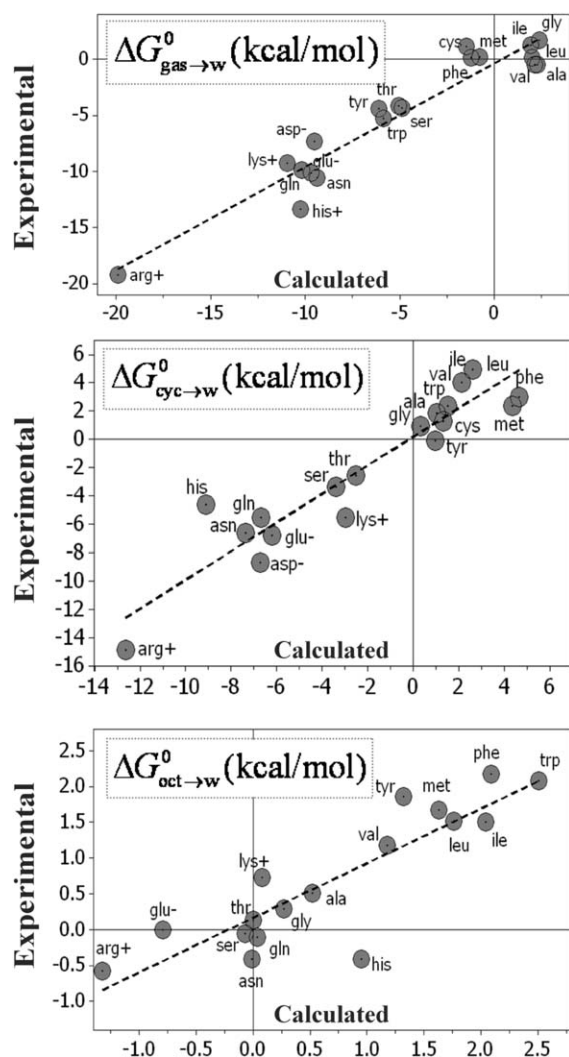


Figure 9. Experimental and calculated Gibbs energies of transfer of the hydrogen-capped amino acid side chains to water from gas-phase (top), cyclohexane (middle), octanol (bottom), in kcal mol<sup>-1</sup>.

amino acids tend to have a pyrimidine base (U or C) in the second position. Since the hydrophilicity/hydrophobicity of the amino acid side chains is highly correlated to the CSI, one can glean from Table 6, a correspondingly strong correlation between the chemical nature of the middle letter of the triplet genetic codon and the CSI of the side chain of the encoded amino acid. All amino acid side chains with  $9.68 > \text{CSI}_R \geq 2.83$  a.u., the hydrophilic amino acids according to CSI, possess a purine base in the second position without exceptions. Conversely, most hydrophobic amino acids with  $\text{CSI}_R \leq 2.81$  a.u. possess a pyrimidine base in the second position, with a few exceptions: (i) Serine, which is the only amino acid with a degenerate middle letter in its codons (middle letter can be either "G" or "C"), (ii) glycine, which is an exceptional amino acids in that it does not have a side chain and is the only amino acid with an achiral  $\alpha$ -carbon, and (iii) cysteine, which is the only amino acid that dimerizes through its side chain to form S—S bridges of cystine. To the author's knowledge, this is probably the most direct link between the genetic code and the electron density.

The experimental Gibbs energies of transfer of the 20 hydrogen-capped amino acid side chains from cyclohexane (cyc) and from octanol (oct) to water (w) were reported by Radzicka and Wolfenden.<sup>[174]</sup> These experimental  $\Delta G_{1 \rightarrow 2}^0$  can be modeled with an electrostriction term [ $\text{CSI}_R$ , Eq.(34)] and an intrinsic volume term [ $V(\text{vdW})_R$ , Eq.(33)].<sup>[15,80,167]</sup>

$$\Delta G_{\text{cyc} \rightarrow \text{w}}^0 = -1.963 \text{CSI}_R + 0.00685 V(\text{vdW})_R, \quad (48)$$

$[r^2=0.881, \text{RMSD}=0.415, n=19]$

and

$$\Delta G_{\text{oct} \rightarrow \text{w}}^0 = -0.321 \text{CSI}_R + 0.00272 V(\text{vdW})_R, \quad (49)$$

$[r^2=0.758, \text{RMSD}=0.125, n=17]$

where energies are in kcal mol<sup>-1</sup>, and CSI values and volumes are in a.u. Table 6 provides a listing of the experimental and calculated values corresponding to Eqs. (48) and (49) along with a listing of the amino acid descriptors  $\text{CSI}_R$  and  $V(\text{vdW})_R$  used in the modeling, while Figure 9 (middle, and bottom) displays the agreement between calculated and experimental values graphically. A number of other remarkable models were reported where  $\text{CSI}_R$  and  $V(\text{vdW})_R$  could be used in the accurate prediction of the change in the stability of proteins on mutation, that is, the change in Gibbs energy upon denaturation of the mutant compared to this change for the wild type.<sup>[80,167]</sup>

The versatility of the CSI index to model widely differing properties is quite notable. Recently, Sánchez-Floresy et al.<sup>[175]</sup> report accurate modeling of excitation energies in simple molecules using this single descriptor. For example, these workers show that the excitation energies for the low-lying singlet and triplet  $\pi \rightarrow \pi^*$  states of CO can be closely fitted to the following linear regression equation:<sup>[175]</sup>

$$\Delta E = -4.542 + 5.639 \text{CSI} \quad [r^2=0.988], \quad (50)$$

while transition energies between low-lying singlet and triplet  $\pi \rightarrow \pi^*$  states of benzene can be fitted to:

$$\Delta E = 2.730 + 7.115 \text{CSI} \quad [r^2=0.975], \quad (51)$$

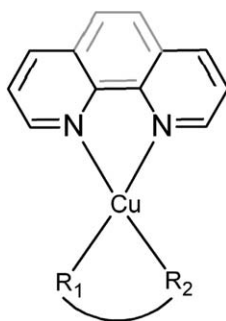
again using the ground-state CSI, which reinforces the arguments presented above in the opening of the section entitled "Bond Properties as Predictors of Spectroscopic Transitions and NMR Proton Chemical Shifts".

A general conclusion that can be drawn is that the QAIM ground-state CSI is capable of capturing a good deal of important molecular physics, in the QSAR sense, and is an important and valuable descriptor worthy of note and further exploration.

#### $\pi$ -Stacking of antineoplastic agents (Casiopenas<sup>®</sup>) and charge transfer from flanking DNA base pairs

Casiopenas<sup>®</sup> are a promising class of antineoplastic cytotoxic agents of uncertain mechanism of action.<sup>[176]</sup> These complexes

are denoted  $[\text{Cu}(\text{N}-\text{N})(\text{N}-\text{O})]\text{NO}_3$  and  $[\text{Cu}(\text{N}-\text{N})(\text{O}-\text{O})]\text{NO}_3$  where N–N symbolizes a substituted bipyridine or phenanthroline aromatic ring system, N–O an  $\alpha$ -aminoacide or peptide, and O–O acetylacetonate or salicylaldehyde. These compounds consist of tetradentate complexes of  $\text{Cu}^{2+}$  where the metal is chelated to a substituted bipyridine or phenanthroline ring from one side (top moiety in the structure below) and to an  $\alpha$ -aminoacide, a peptide, an acetylacetonate, or a salicylaldehyde moiety from the other side (bottom moiety). The general structure of these molecules can be represented by Scheme 4,



Scheme 4.

where the light gray part is variable, existing in the phenanthroline congeners but nonexistent in the bipyridine congeners, and where  $R_1$  and  $R_2$  are parts of a closed ring system where the two atoms bonded to the central  $\text{Cu}^{2+}$  may be both oxygen atoms or one can be an oxygen and the other a nitrogen.

Galindo-Murillo et al.<sup>[177]</sup> have recently suggested that these complexes can intercalate between DNA base pairs through their aromatic moiety and that the  $\pi$ -stacking interaction is driven by an electron depletion of the planar ligand (the substituted bipyridine or phenanthroline ring) due to the transfer of charge to the metal center which, in turn, drives charge transfer from the flanking DNA bases to the intercalating ligand. Using the integrated QTAIM charges summed over entire molecular fragments, these workers found a simple but strong statistical correlation between the complex stabilization energy of the adenine–Casiopieinas<sup>®</sup> complex and the net electron population transferred from adenine to the aromatic ligand:<sup>[177]</sup>

$$\Delta E(\text{kcal mol}^{-1}) = -34.542 - 254.83 \times \Delta N, \quad (52)$$

$[r^2 = 0.926, n = 21]$

which lends numerical support to the charge-transfer assisted  $\pi$ -stacking hypothesis advanced by these workers and, simultaneously, enhance the plausibility of their proposed mechanism of action initiated by stacking intercalation of Casiopieinas<sup>®</sup> between DNA base pairs. What is desirable for a future study is, perhaps, to correlate energies of stacking interaction with a direct measure of biological antineoplastic activity.

Before leaving this section, we note that there exists a number of similar empirical correlations of  $\pi$ -stacking interactions

modeled with QTAIM (bond) properties, we cite as examples those of Zhikol et al.,<sup>[178]</sup> Platts and co-workers,<sup>[179]</sup> and Parthasarathi and Subramanian.<sup>[180]</sup>

## Localization/Delocalization Matrices (LDM), Delocalization Matrices (DM), and Density-Weighted Connectivity Matrices (DWCM) in QSAR

### Definition of LDM ( $\zeta$ ), DM, and DWCM

Among the properties that can be associated with a bounded region of space is the number of electrons localized within that region. When this region is an atom in a molecule defined by QTAIM,  $\Omega$ , the average number of electrons localized within this atom, that is, not shared with any other atom in the molecule, is termed the “localization index” (LI, for short) and is given the symbol  $\Lambda(\Omega)$ . Since not all electrons are localized, in general  $\Lambda(\Omega) \leq N(\Omega)$ , the equality is approached at the closed-shell limit as in ionic bonding and is exact only for infinitely separated isolated atoms or ions. For an illustration, in the case of the ionic molecule  $\text{Li}^{+0.935}\text{F}^{-0.935}$  molecule (the superscripts are the atomic charges):  $[\Lambda(\text{F}) = 9.845] < [N(\text{F}) = 9.935]$  with a difference of  $N(\text{F}) - \Lambda(\text{F}) = 0.090$ . In contrast, for the covalent  $\text{F}_2$ , the corresponding values are:  $[\Lambda(\text{F}) = 8.549] < [N(\text{F}) = 9.000]$  with a significantly larger difference of  $N(\text{F}) - \Lambda(\text{F}) = 0.451$ . (QCISD/6-311++G(3df,2pd) level of theory).

The difference between  $N(\Omega)$  and  $\Lambda(\Omega)$  accounts for the number of electrons belonging to  $\Omega$  but shared with other atoms in the molecules. In the case of  $\text{Li}^{+0.935}\text{F}^{-0.935}$ ,  $0.090 e^-$  are contributed by F and an equal number is shared from the Li with a total of  $0.180 e^-$  shared between the two basins [denoted by  $\delta(\text{Li}, \text{F})$  and which is termed the “delocalization index” (DI), in QTAIM parlance]. In this manner, all electrons in the LiF molecule are accounted for:  $[\Lambda(\text{Li}) = 1.975] + [\Lambda(\text{F}) = 9.845] + [\delta(\text{Li}, \text{F}) = 0.180] = 12.000 e^-$ .

Thus, QTAIM defines a DI that counts the number of electrons shared between any two atomic basins in a molecule and which is generally denoted by  $\delta(\Omega_1, \Omega_2)$ . The formal definition and methods of calculation of the LI and DI are discussed elsewhere.<sup>[118,181–187]</sup>

Two important points must be emphasized before moving to examples of applications. The first is that, as mentioned above, LI and DI account for the whereabouts of all electrons in a molecule composed of  $n$  atoms  $[\Omega_i, i = 1, 2, \dots, n]$ , their general relation to an atomic electron population being:

$$N(\Omega_i) = \Lambda(\Omega_i) + \frac{1}{2} \sum_{j \neq i}^n \delta(\Omega_i, \Omega_j). \quad (53)$$

Given expression (53), the total molecular electron population can be expressed as the sum of two subpopulations:

$$N = \sum_{i=1}^n N(\Omega_i) = \sum_{i=1}^n \Lambda(\Omega_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \delta(\Omega_i, \Omega_j) = N_{\text{loc}} + N_{\text{deloc}}, \quad (54)$$

where

$$N_{\text{loc}} \equiv \sum_{i=1}^n \Lambda(\Omega_i), \quad (55)$$

and

$$N_{\text{deloc}} \equiv \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n \delta(\Omega_i, \Omega_j). \quad (56)$$

The second point, which directly follows from Eq. (53), is that *the DI measures the number of electrons (and not the number of pairs of electrons)*, an erroneous characterization that has propagated in many publications (including some by the present author). A source of the confusion may be that when two atoms  $\Omega_i$  and  $\Omega_j$  are bonded the number of electrons shared between them,  $\delta(\Omega_i, \Omega_j)$ , is numerically close to the number of shared pairs in the Lewis bonded structure and hence can be misconstrued as a bond order. One can assert, though, that  $\delta$  is generally *proportional* to the classical bond order when two atoms share a bond path. A simple example helps settle this confusion. For the  $\text{H}_2$  molecule,  $\delta(\text{H}, \text{H}') \approx 1.0$  and  $\Lambda(\text{H}) = \Lambda(\text{H}') \approx 0.5$ , which means that  $N_{\text{loc}} = 2 \times 0.5$ , leaving only 1 e to be delocalized (shared).

It is also worth reminding the reader that  $\delta(\Omega_i, \Omega_j)$  is not only defined between bonded atoms but also between any two atoms in a molecule, no matter how distant and regardless of the presence or absence of a bond path linking their nuclei. Obviously the DI cannot be related to any bond order if the atoms do not share a bond path, that is, when they are not bonded in the first place.

Ferro-Costas, Vila, and Mosquera (F-CVM)<sup>[188]</sup> have recently demonstrated that the anomeric effect in halogen-substituted methanols cannot be explained by hyperconjugation arguments based on the behavior of atomic populations and the QTAIM localization/delocalization indices. These authors have presented lower-triangular matrix-like tabulations of the delocalization indices and have used differences between these matrices in their argumentation.<sup>[188]</sup>

Given that the  $\delta$  indices were found to be excellent predictors of experimental quantities such as NMR JJ coupling constants between protons<sup>[185]</sup> and fluorine atoms,<sup>[187]</sup> it is tempting to construct matrices similar to those of F-CVM and use them as tools in QSAR-type studies. Initial explorations of these ideas are presented here for the first time, meanwhile more extensive studies in this direction are being undertaken in the author's laboratory.

Disregarding the nature of the matrix elements, the form of the matrices used by (F-CVM)<sup>[188]</sup> is identical to the oft-used graph-theoretical connectivity (adjacency) matrices. Connectivity matrices have been brought from the field of graph theory, a branch of pure mathematics, to mathematical chemistry by

scientists such as Balaban, Gutman, Hosoya, Nicolici, Randić, Trinajstić, and several others.<sup>[6,7,51,189–193]</sup> In a typical graph-theoretic connectivity matrix, one places 1 or 0 depending on whether two atoms are bonded or not in the hydrogen-suppressed graph. Atom-connectivity matrices have been proposed that do not suppress the hydrogen atoms and that include as off-diagonal entries properties that depend on atomic pairs such as interatomic distances, vibrational force constants, or the bond dissociation energies while atomic symbols are introduced along the diagonal elements.<sup>[51]</sup>

A natural extension of the graph-theoretical approach is to merge it with the topological analysis of the electron density. The idea does not appear to be new, already in 1981 Dmitriev, in a 155 page book on chemical graph theory, devotes pp. 110–115 to a section on Bader's "Topology of Molecular Charge Distributions," but makes no connections between it and the main theme of the book.<sup>[193]</sup> In 1994, Balasubramanian has suggested the "[i]ntegration of graph theory and quantum chemistry"<sup>[194]</sup> in a paper that concludes with a review of the basic concepts of the topological analysis of the electron density according to the theory now known as QTAIM. In his paper, Balasubramanian has emphasized (i) the remarkable correspondence between the graph-theoretical graph of a molecule and the set of bond paths constituting its QTAIM molecular graph<sup>[17]</sup> and (ii) the role of the Laplacian of the electron density,  $\nabla^2 \rho(\mathbf{r})$ , in determining the sites of nucleophilic and electrophilic attack in a molecule. The present author is unaware of other proposals to merge or integrate QTAIM and the graph-theoretical approach and presents below some thoughts on how this might possibly be achieved in the future.

Balasubramanian's integration is intuitively implemented through the construction and the use of "electron density weighted adjacency matrices (EDWAM)" (quoting the words and the idea of Massa (L. Massa, *Personal communication*, 2014)). An EDWAM places the value of the electron density at the BCP,  $\rho_{\text{BCP}}$  as the off-diagonal matrix element whenever a bond path exists between two atoms instead of the "1" of the graph-theoretic connectivity matrices. There is no hydrogen suppression in the proposed EDWAM. Figure 10 displays a EDWAM for chloroacetic acid. The respective research groups of Massa and of the present author are jointly exploring the use of EDWAMs in QSAR-type studies at the time of writing.

We now introduce a new matrix representation of molecules that encapsulates one- and two-electron information simultaneously and which will be referred to as "localization–delocalization matrix (LDM)." An early version of the LDM was introduced in 2001 in the documentation manual of the program AIMDELOC<sup>[195]</sup> and bears some similarity to the matrices used here and those used recently by F-CVM.<sup>[188]</sup>

The diagonal elements of LDMs proposed here are the atomic localization indices  $\Lambda(\Omega)$ , as originally proposed in 2001,<sup>[195]</sup> but the off-diagonal elements are the delocalization indices divided by 2, that is,  $\delta(\Omega_i, \Omega_j)/2$ , a feature distinguishing the LDMs in this work from the 2001 proposal and from that of F-CVM. Further, we preserve all the elements of these symmetric matrices (not just the lower or upper triangular part). Thus, we define the LDM that we denote by  $\zeta$  as

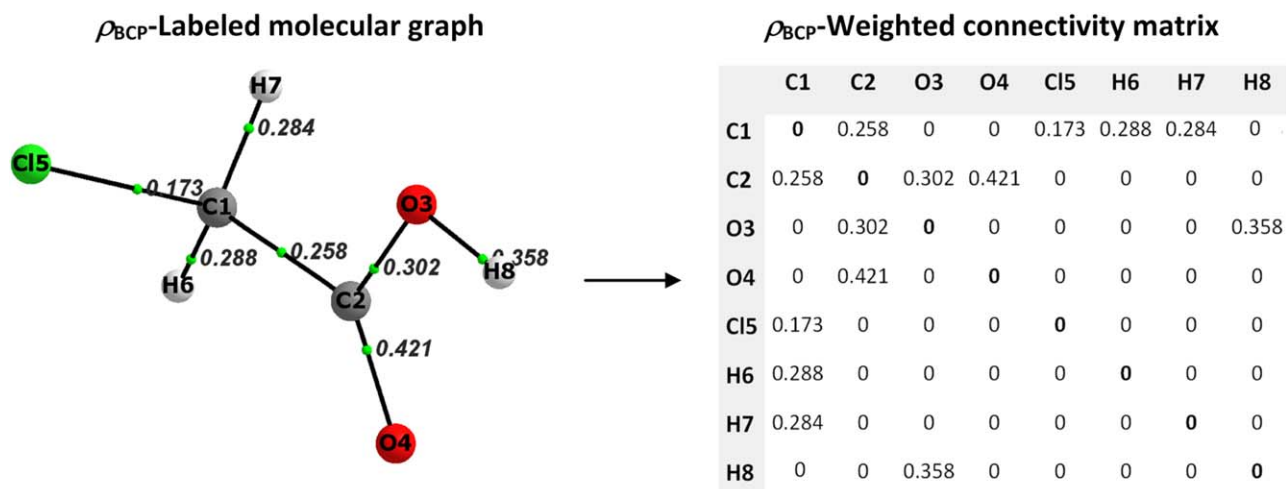


Figure 10. Molecular graph of chloroacetic acid showing the bond paths each labeled with the electron density at the BCP (the small sphere) in a.u. and the corresponding electron density weighted connectivity matrix (EDWCM). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

$$\zeta \equiv \begin{bmatrix} \Lambda(\Omega_1) & \delta(\Omega_1, \Omega_2)/2 & \cdots & \delta(\Omega_1, \Omega_n)/2 \\ \delta(\Omega_2, \Omega_1)/2 & \Lambda(\Omega_2) & \cdots & \delta(\Omega_2, \Omega_n)/2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta(\Omega_n, \Omega_1)/2 & \delta(\Omega_n, \Omega_2)/2 & \cdots & \Lambda(\Omega_n) \end{bmatrix}_{n \times n} \left. \begin{array}{l} \sum_{\text{row}} \\ = N(\Omega_1) \\ = N(\Omega_2) \\ \vdots \\ = N(\Omega_n) \end{array} \right\} \sum_{i=1}^n N(\Omega_i) = N. \quad (57)$$

$$\underbrace{\sum_{\text{column}}}_{= N(\Omega_1)} \quad \underbrace{\quad}_{= N(\Omega_2)} \quad \underbrace{\quad}_{= N(\Omega_n)} \quad \text{tr}(\zeta) = N_{\text{loc}}$$

$$\sum_{i=1}^n N(\Omega_i) = N$$

In this manner, and according to Eq. (53), the sum of any row or column of  $\zeta$  equals to the given atomic electron population  $N(\Omega)$ , while the sum of the column sums or row sums yields the total number of electrons in the molecule [Eq. (54)]. Further, the trace of  $\zeta$  is the total number of localized electrons ( $N_{\text{loc}}$ ) in the molecule [Eq. (55)], and the number of delocalized electrons is, then, given by difference:

$$N_{\text{deloc}} = N - \text{tr}(\zeta) = N - N_{\text{loc}}. \quad (58)$$

As known,  $\delta(\Omega_i, \Omega_j)$  decays exponentially with internuclear separation,<sup>[196]</sup> which implies that the  $\zeta$  matrix, by including all the delocalization indices in a molecule, does implicitly contain interatomic distance matrix information. The  $\zeta$  matrix is, thus, rich in coded information about not only the electron density distribution but also about the pair density and even contains some geometric distance information. For these reasons, this matrix can be expected to constitute the raw material for a new class of QSAR correlations.

All matrix representations of molecules have common and well-known drawbacks, which we briefly discuss below with the particular reference to LDMs.

**Problem 1: Nonuniqueness due to ambiguity of atom labelling** (i.e., a family of matrices related by permutations all code for the same structure). The matrices are not unique since they are labelling-dependent and an interchange of a pair of atomic

labels will result in the permutation of a pair of rows and the simultaneous permutation of a pair of columns. There are  $n!$  ways to label  $n$  atoms and correspondingly, there are as many different matrices describing the same system. This problem has prompted the graph theoretic community to search for “matrix invariants,” that is, properties associated with these matrices that are labelling-independent. Common invariants include the characteristic polynomial and its roots (the eigenvalues), the trace of the matrix, the determinant of the matrix and so forth. The labelling problem is of course independent of the nature of the matrix elements and hence is also relevant to the LDMs proposed here. We will hence follow the lead of the chemical graph-theoretic community and extract invariants from the LDMs at the cost of complicating the physical interpretation of the final (invariant) descriptors, which are, nevertheless, derived from well-characterized measures of electron localization and delocalization compactly collected in the LDMs. Since an LDM is by construction real and symmetric, then it is diagonalizable to a similar matrix  $\mathbf{D}$  (the “diagonalized localization–delocalization matrix,” or DLDM) through a similarity transformation:

$$\mathbf{P}^{-1} \zeta \mathbf{P} = \mathbf{D}. \quad (59)$$

The matrix  $\mathbf{D}$  is indispensable when the atomic numbering scheme cannot be made consistent across a molecular series

that lacks, for example, a common skeleton. In these instances, the original  $\zeta$  matrices of different molecules cannot be compared directly. At times, however, the  $\zeta$  matrices themselves can be compared directly when a consistent labelling scheme can be imposed on all molecules (e.g., molecules with the same skeleton as in the example of substituted acetic acids discussed in the section entitled “The LDM  $\zeta$  and the  $pK_a$  of Halogenated Acetic Acids” below). Matrices comparison can be accomplished, for example, by calculating the distance between them (e.g., the Frobenius distance, see below).

**Problem 2: Unequal matrix dimensions for differently sized molecules.** Unless the number of atoms in the compared molecules is the same, the matrix dimensions of  $\zeta$  will differ from one molecule to another rendering the direct comparison such as calculating the Frobenius distance impossible. A solution to this problem has been proposed by White and Wilson<sup>[197]</sup> (WW) in the context of image statistical pattern recognition where different sized (sub)graphs are treated on equal footing. The

approach of WW consists in determining first the matrix representative of the largest graph ( $n \times n$ ) and enlarge all other smaller matrices to  $n \times n$  by filling all new empty matrix elements with zeros after placing the smaller matrix in the upper left block of the enlarged matrix, that is, “padding” all matrices other than the largest with zeros to the same common  $n \times n$  size.<sup>[197]</sup>

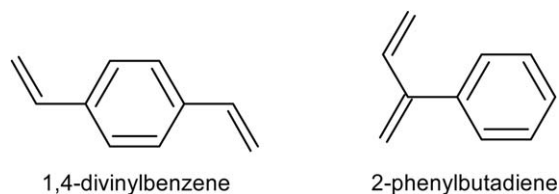
The WW approach is applicable to LDMs since a zero diagonal element means that there are zero electrons localized within a given atomic basin while the zeros in the other (off-diagonal) matrix elements of the row and column that correspond to that atom imply that there are zero electrons shared between that basin and all other atoms in the molecule. This is what computational chemists refer to as a “ghost atom”. Thus, if one wishes to compare the first four saturated aliphatic hydrocarbons, the matrix size will be that of butane (14  $\times$  14), and the LDMs of methane, ethane, and propane will be enlarged to that size by “zeros padding”. At the DFT-B3LYP/6-311G(d,p) level of theory, and given the atomic numbering schemes in the *Supporting Information*, to two decimal places, an LDM of butane is:

$$\zeta_{\text{butane}} = \begin{bmatrix} 3.87 & 0.50 & 0.49 & 0.47 & 0.47 & 0.02 & 0.02 & 0.02 & 0.01 & 0.00 & 0.00 & 0.02 & 0.02 & 0.02 \\ 0.50 & 3.95 & 0.02 & 0.02 & 0.02 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.48 & 0.48 & 0.48 \\ 0.49 & 0.02 & 3.87 & 0.02 & 0.02 & 0.50 & 0.47 & 0.47 & 0.02 & 0.02 & 0.02 & 0.01 & 0.00 & 0.00 \\ 0.47 & 0.02 & 0.02 & 0.46 & 0.02 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 \\ 0.47 & 0.02 & 0.02 & 0.02 & 0.46 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.01 & 0.00 \\ 0.02 & 0.01 & 0.50 & 0.00 & 0.00 & 3.95 & 0.02 & 0.02 & 0.48 & 0.48 & 0.48 & 0.00 & 0.00 & 0.00 \\ 0.02 & 0.00 & 0.47 & 0.01 & 0.00 & 0.02 & 0.46 & 0.02 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.02 & 0.00 & 0.47 & 0.00 & 0.01 & 0.02 & 0.02 & 0.46 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 \\ 0.01 & 0.00 & 0.02 & 0.00 & 0.00 & 0.48 & 0.00 & 0.00 & 0.45 & 0.02 & 0.02 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.02 & 0.00 & 0.00 & 0.48 & 0.01 & 0.00 & 0.02 & 0.46 & 0.02 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.02 & 0.00 & 0.00 & 0.48 & 0.00 & 0.01 & 0.02 & 0.02 & 0.46 & 0.00 & 0.00 & 0.00 \\ 0.02 & 0.48 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.45 & 0.02 & 0.02 \\ 0.02 & 0.48 & 0.00 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.46 & 0.02 \\ 0.02 & 0.48 & 0.00 & 0.01 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.02 & 0.02 & 0.46 \end{bmatrix}, \quad (60)$$

while a corresponding (ghost atom-padded) LDM of ethane is:

$$\zeta_{\text{ethane}} = \begin{bmatrix} 3.94 & 0.48 & 0.51 & 0.48 & 0.48 & 0.02 & 0.02 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.48 & 0.46 & 0.02 & 0.02 & 0.02 & 0.01 & 0.00 & 0.00 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.51 & 0.02 & 3.94 & 0.02 & 0.02 & 0.48 & 0.48 & 0.48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.48 & 0.02 & 0.02 & 0.46 & 0.02 & 0.00 & 0.01 & 0.00 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.48 & 0.02 & 0.02 & 0.02 & 0.46 & 0.00 & 0.00 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.02 & 0.01 & 0.48 & 0.00 & 0.00 & 0.46 & 0.02 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.02 & 0.00 & 0.48 & 0.01 & 0.00 & 0.02 & 0.46 & 0.02 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.02 & 0.00 & 0.48 & 0.00 & 0.01 & 0.02 & 0.02 & 0.46 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (61)$$

**Problem 3:** The mapping of the set of molecular graphs to the set of adjacency matrix representative can be surjective not injective (nonbijective), i.e., one and the same matrix invariant can code for more than one distinct graph. This problem is illustrated by the example recognized by Živković (Quoted through Ref. 192 due to unavailability of the original reference: T. Živković, *Report on Summer School*, Repino, Leningrad, 1973) and whereby two chemically distinct molecules (Scheme 5)



Scheme 5.

coded by their respective adjacency matrices yield one and the same characteristic polynomial ( $x^{10} - 10x^8 + 33x^6 - 44x^4 + 24x^2 - 4$ ). Naturally, this problem will never occur when the molecules are coded by their respective LDMs or electron density-weighted adjacency matrices with matrix elements that are very different than simple ones and zeros.

**Problem 4: Inability to distinguish optical isomers (enantiomeric ambiguity).** While the off-diagonal entries in LDMs are geometric distance-sensitive, since the DIs decay with distance, there are no matrix representatives that can capture the relative arrangement of groups in space that determines handedness. As mentioned previously in relation to QTMS, this is not a limitation as long as the experimental biological/pharmacological responses are induced using the correct optical isomer(s). For modeling of physicochemical properties, this issue is seldom relevant except in cases such as chiral chromatography where again the appropriate optical isomers must be used to generate the experimental data.

**Problem 5: Conformational flexibility.** The presence of multiple thermally accessible minima on conformational potential energy surfaces complicates any method that aims at extracting meaningful descriptors except when the descriptors are insensitive to the interatomic distance matrix (such as graph-theoretical indices). In principle, some form of conformational averaging has to be performed to mimic Nature, but in practice this may prove difficult especially when hundreds of molecules are under consideration. This problem is not specific to LDMs and will not be addressed here. We now explore examples illustrating the possible uses of  $\zeta$  in QSAR.

#### The LDM $\zeta$ and transferability in aliphatic alkanes (methane, ethane, propane, and butane)

The LDM matrices for the first four members of the aliphatic alkanes homologous series have been obtained at the B3LYP/6-311G\*\* (the complete set of the  $\zeta$  matrices and their eigenvalues are available in the *Supporting Information*). We note in passing that while the meaning of the LI and DI is strictly defined within Hartree–Fock theory, the indices obtained from

DFT calculations are often used as they yield numerical values of LIs and DIs that are close to the Hartree–Fock values.

First, we investigate one way by which the similarity or dissimilarity of these four molecules can be quantified using their DLDM representations. It is possible to define a “distance” between two matrices (of the same dimensions), even though such a definition is not unique. The definition used in this work is the Frobenius distance, that is, the “Frobenius norm” of the difference matrix:

$$d(\mathbf{A}, \mathbf{B}) \equiv \|\mathbf{A} - \mathbf{B}\| \equiv \sqrt{\sum_{i,j} |a_{ij} - b_{ij}|^2}. \quad (62)$$

There are  $n(n-1)/2$  different distances  $d(\mathbf{A}, \mathbf{B})$  between  $n$  molecules that may be arranged into a lower or upper triangular intermolecular distance matrix. The Frobenius distance matrix between the DLDMs ( $\mathbf{D}$ ) representing the first four members of the saturated alkane series is given in Table 7.

An examination of Table 7 reveals that the “contribution of a methylene group” to the distance of one member compared to the next in the homologous series follows expectations in the sense that the first addition of a  $\text{CH}_2$  to  $\text{CH}_4$  to obtain  $\text{C}_2\text{H}_6$  results in a somewhat anomalous change compared to the subsequent additions of  $\text{CH}_2$  to the growing alkane chain, and where this change appears to quickly reach a constant value (as emphasized by Bader in relation, for example, to the contribution of a methylene group to the heats of formation of aliphatic alkanes).<sup>[17]</sup> Thus, to two decimals,  $d(\text{CH}_4, \text{C}_2\text{H}_6) = 3.29$ ,  $d(\text{C}_2\text{H}_6, \text{C}_3\text{H}_8) = 3.04$ , and  $d(\text{C}_3\text{H}_8, \text{C}_4\text{H}_{10}) = 2.92$ , fast converging to an apparent asymptotic value around 3.

Since this is a relatively uncharted territory, it may be of interest to explore the geometry of this 14-D comparison space. One criterion worthy of consideration is the satisfaction of the triangle inequality as emphasized by Muskulus.<sup>[198]</sup> The inequality applied to intermolecular metrics of three molecules  $A$ ,  $B$ , and  $C$  is:

$$d(A, B) \leq d(A, C) + d(C, B). \quad (63)$$

A violation of this inequality indicates that the geometric representation is inadequate as a proximity measure of the studied objects<sup>[198]</sup> and hence would invalidate the use of  $d(X, Y)$  as a similarity or dissimilarity criterion of molecules  $X$  and  $Y$ . Figure 11 displays the triangular distance and angular relationships between the three possible triads of molecules in this group of four aliphatic alkanes. The figure demonstrates that the inequality is clearly satisfied given the intermolecular distance matrix in Table 7. Further, the angular relationships

Table 7. Frobenius distance matrix between the diagonalized LDMs (DLDMs) of the four first members of the aliphatic alkane series (DFT-B3LYP/6-311G( $d,p$ ) level of theory).

	Methane	Ethane	Propane
Ethane	3.2941	0	
Propane	4.8450	3.0398	0
Butane	6.0535	4.5281	2.9155

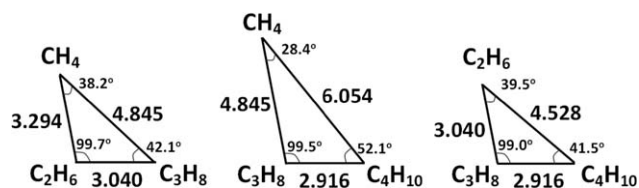


Figure 11. Triangle inequalities of the first four members of the aliphatic alkanes showing the Frobenius distance relations between their diagonalized LDMs.

displayed in the figure exhibit a curious regularity: The angles of the triangles formed by three consecutive members of the homologous series are almost identical (compare the first and third triangles). The almost constancy of the obtuse angle  $\sim 99^\circ$ – $100^\circ$  in all triangles cannot escape notice either. The significance of these observations is not evident to the author at the time of writing.

Another LDM invariant, other than the DLDM representation discussed above, is the trace  $\text{tr}(\zeta) = N_{\text{loc}}$  [Eq. (58)]. Since in most molecules, the majority of electrons are localized within their respective atomic basins, for a given homologous series  $N_{\text{loc}}$  is roughly proportional to the total number of electrons  $N$  and is thus an extensive variable that grows with size. In other words,

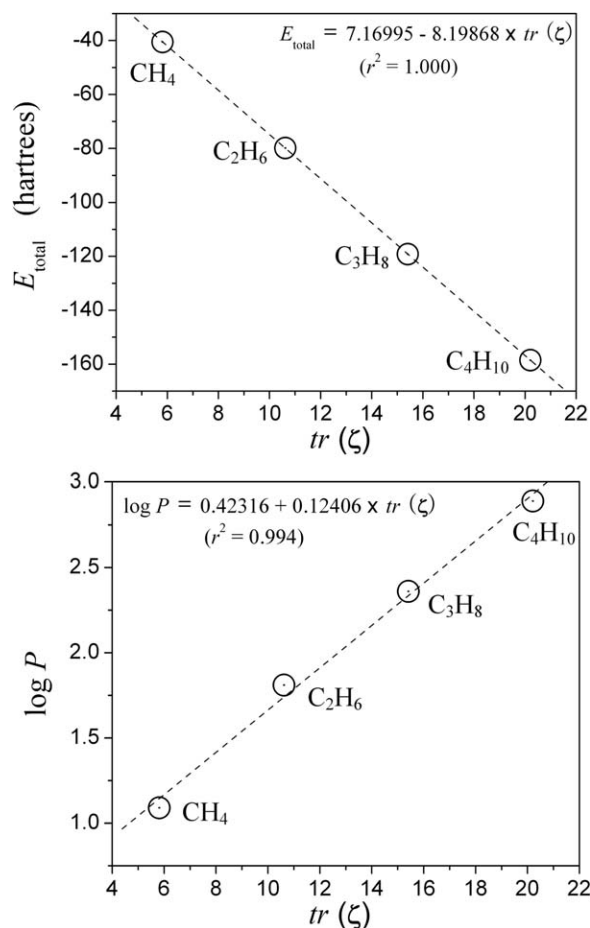


Figure 12. (Top) The correlation between the total molecular energy in a.u. and the trace of the LDMs of the four first members of the aliphatic alkane series. (Bottom) the correlation between experimental  $\log P$  and the trace (experimental values taken from Ref. 199).

Table 8. Frobenius distance matrix between the LDMs of the substituted acetic acid series (DFT-B3LYP/6-311G(d,p) level of theory).

$\text{CH}_i\text{X}_{3-i}$	$\text{CH}_3$	$\text{CH}_2\text{F}$	$\text{CHF}_2$	$\text{CF}_3$	$\text{CH}_2\text{Cl}$	$\text{CHCl}_2$
$\text{CH}_2\text{F}$	4.52	0				
$\text{CHF}_2$	7.78	4.76	0			
$\text{CF}_3$	11.16	7.89	5.05	0		
$\text{CH}_2\text{Cl}$	9.29	7.47	8.67	10.65	0	
$\text{CHCl}_2$	14.05	11.91	10.47	11.40	9.17	0
$\text{CCl}_3$	19.56	15.83	14.03	12.72	13.87	9.17

$\text{tr}(\zeta)$  can be expected to capture molecular properties that depend on steric bulk and/or lipid solubility. Figure 12 displays the correlation of  $\text{tr}(\zeta)$  of the first four aliphatic hydrocarbons with a typical size-extensive property, the calculated total energy  $E$ , yielding a correlation with an  $r^2 = 1.000$  (top), and with a lipid solubility, measured by  $\log P$ , with an  $r^2 = 0.994$  (bottom). These preliminary results call for further investigation with many more data points to provide definitive conclusions though.

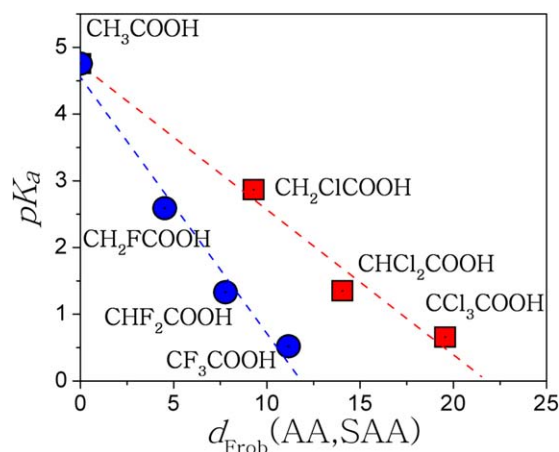
### The LDM $\zeta$ and the $pK_a$ of halogenated acetic acids

The LDMs of a set of six homohalogenated acetic acid ( $\text{CH}_2\text{FCOOH}$ ,  $\text{CHF}_2\text{COOH}$ ,  $\text{CF}_3\text{COOH}$ ,  $\text{CH}_2\text{ClCOOH}$ ,  $\text{CHCl}_2\text{COOH}$ , and  $\text{CCl}_3\text{COOH}$ ) are compared with that of the parent unsubstituted compound, acetic acid. (LDM  $\zeta$  matrices, their eigenvalues, and the atomic numbering scheme for all members of this molecular set are available in the *Supporting Information*). In this example, a consistent atomic numbering scheme can be adopted for all seven molecules, all of which have a common skeleton and the same number of atoms, obviating the need to diagonalize their LDMs or pad any matrices with ghost atoms, thus here the distances are calculated directly between the LDMs.

The Frobenius intermolecular distance matrix is given in Table 8. In this table, the first column, for example, lists the distance between every substituted acid from the parent unsubstituted acetic acid. A plot of these distances against the corresponding experimental  $pK_a$ 's is displayed in Figure 13. The figure shows that the experimental  $pK_a$ 's of all congeners substituted with one and the same halogen (either F or Cl) are almost perfectly linearly correlated with the Frobenius distances of their respective LDMs from that of unsubstituted acetic acid ( $r^2 = 0.98$  for both regression lines). The slopes of the two lines are different, however, which signals that the  $pK_a$  is not captured in its entirety by the LDM intermolecular Frobenius distance matrix.

The LDM is generally "biased" somewhat by its diagonal elements  $\Lambda(\Omega_i)$ , which typically have larger magnitudes than the off-diagonal  $\frac{1}{2}\delta(\Omega_i, \Omega_j)$ . In a homologous series, both  $N_{\text{loc}}$  [Eq. (55)] and  $N_{\text{deloc}}$  [Eq. (56)] increase roughly in proportion to the total number of electrons  $N$  [Eq. (54)]. The situation is different for a series of congeners molecules  $\text{R-X}$  differing only in the identity of a substituent, say with  $\text{X} = \text{F}, \text{Cl}, \text{Br}, \text{or I}$ . In this case, the traces of the LDMs will be significantly different since these substituents will contribute a large fraction of their atomic numbers of electrons to the count yet the delocalization indices in



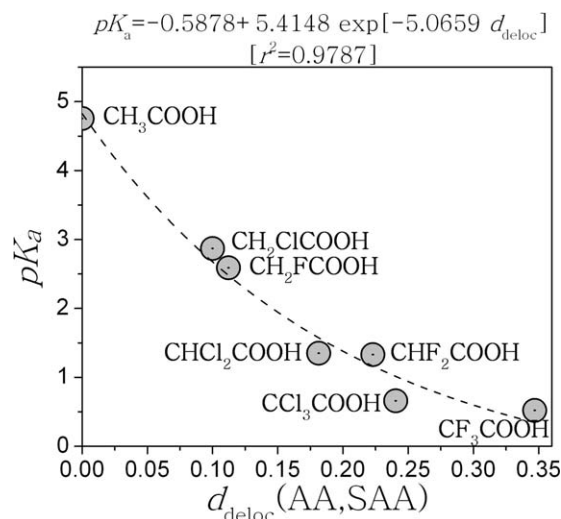


**Figure 13.** Correlations between the experimental  $pK_a$  values of fluorine- and chlorine-substituted acetic acids (F-SAA, Cl-SAA) against the Frobenius distance of their LDMs from that of unsubstituted acetic acid (AA). The two linear correlation can be fitted to  $pK_a(\text{F-SAA}) = 4.5511 - 0.3840 d(\text{AA, F-SAA})$  ( $r^2 = 0.9793$ ,  $\text{STD} = 0.3256$ ,  $n = 4$ ); and  $pK_a(\text{Cl-SAA}) = 4.73954 - 0.2173 d(\text{AA, Cl-SAA})$  ( $r^2 = 0.9834$ ,  $\text{STD} = 0.2872$ ,  $n = 4$ ). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

these congeners will, at most, be slightly perturbed by the substitution. For these four congeners  $\text{R-X}$ , then, the diagonal elements are the sole carrier of size-extensive information in their respective LDMs. For a property primarily driven by electronic structure, such LDMs, biased by steric bulk, may not work just as we have seen in the case of QTMS (see the section entitled “Quantum Topological Molecular Similarity (QTMS)” above). The inability to fit the two sets of congeners of substituted acetic acids ( $X_n = \text{F, Cl}$ ) on one line using the distances between their LDMs proves the point (Fig. 13).

We thus, define a size-independent electronic fingerprinting matrix representation of a molecule as the diagonal-suppressed-LDM or simply the “delocalization matrix” (DM) in which all diagonal elements are suppressed to zeros (no explicit localization information). The similarity can then be measured as the Frobenius distance between DMs when the modeled properties are primarily governed by electronic structure and not by steric bulk.

The experimental  $pK_a$ 's are plotted in Figure 14 against the distances between the DMs of the substituted and unsubstituted acetic acids, denoted as  $d_{\text{deloc}}(\text{AA, SAA})$ , and are listed in Table 9. This size-independent electronic fingerprinting distance correlates much better with the experimental  $pK_a$ 's than the size dependent LDMs-based distances bringing both sets of congeners ( $X_n = \text{F, Cl}$ ) together on a single exponential regression line ( $r^2 = 0.979$ ). A similar conclusion is reached within the QTMS framework that models  $pK_a$ 's accurately on its own without injecting extraneous size-dependent information in the model, since  $pK_a$  is inherently governed by electronics rather than by steric bulk. It may be concluded that *the full LDMs appear to capture properties that depend simultaneously on the electronic structure and size (steric bulk) while the DMs are better tuned to capture the properties primarily dependent on electronic structure.*



**Figure 14.** Correlations between the experimental  $pK_a$  values of fluorine- and chlorine-substituted acetic acids (F-SAA, Cl-SAA) against the Frobenius distance of their localization-suppressed-LDMs (DMs) from that of unsubstituted acetic acid (AA). The data can be closely fit to an exponential model  $pK_a(\text{SAA}) \approx -0.588 + 5.415 \times \exp[-5.066 \times d_{\text{deloc}}(\text{AA, SAA})]$  ( $r^2 = 0.979$ ,  $n = 7$ ).

We return to Table 8 for some final observations. The distances between  $\text{CH}_i\text{F}_{3-i}\text{-COOH}$  and  $\text{CH}_{i-1}\text{F}_{2-i}\text{-COOH}$  ( $i = 3, 2, 1$ ), that is, two homologous members differing by one fluorine atom, are 4.52, 4.76, and 5.05 for  $i = 3, 2$ , and 1, respectively. A fluorine atom's contribution to the distance is between  $\sim 4.5$  and 5.1 and it changes by roughly a constant amount of  $\sim 0.25$  with  $i$ . The corresponding distances between  $\text{CH}_i\text{Cl}_{3-i}\text{-COOH}$  and  $\text{CH}_{i-1}\text{Cl}_{2-i}\text{-COOH}$  in the case of chlorine substitutions are 9.29, 9.17, and 9.17, exhibiting an almost constant contribution to the distance for every addition of a chlorine atom. The table exhibits other regularities that we skip for brevity.

## Molecular Electrostatic Potential (MESP) and Field (MESF)

Bonaccorsi, Scrocco, and Tomasi<sup>[56]</sup> suggested in 1970 the use of the MESP ( $V(\mathbf{r})$ ) to locate the molecular regions most prone to electrophilic attack in a series of six small three-membered heterocyclic rings and cyclopropane. In this article, the authors calculate, plot, and analyze the topography of the MESP, for the first time, to extract more information from the calculated

**Table 9.** Frobenius distances between the diagonal-suppressed LDMs (DMs) of substituted acetic acids from the unsubstituted parent compound sorted in order of decreasing  $pK_a$  (DFT-B3LYP/6-311G( $d,p$ ) level of theory).

$\text{CH}_i\text{X}_{3-i}$	$pK_a^{[a]}$	$d_{\text{deloc}}(\text{CH}_3-, \text{CH}_i\text{X}_{3-i}-)$
$\text{CH}_3-$	4.756	0.0000
$\text{CH}_2\text{Cl}-$	2.87	0.1000
$\text{CH}_2\text{F}-$	2.59	0.1122
$\text{CHCl}_2-$	1.35	0.1814
$\text{CHF}_2-$	1.33	0.2229
$\text{CCl}_3-$	0.66	0.2404
$\text{CF}_3-$	0.52	0.3470

[a] Experimental values obtained from Ref. 200.

wavefunctions than merely atomic charges. The use of the MESP in determining the low energy (preferred) relative orientations of reactants, of guests with respect to their hosts, and of substrates or inhibitors with respect to their active sites has expanded into a wide area of computational and theoretical chemistry.<sup>[57,58,60,62,63,201]</sup> MESP can also be calculated directly from X-ray crystallographic scattering data.<sup>[25–27,61,67]</sup> It is impossible to review this field of research here, instead, a few examples are selected for illustration as they relate to QSAR.

The total charge density expressed in Eq. (4), that is, the discrete charge distribution of the nuclei in addition to the continuous diffuse electron density, give rise to an electrostatic potential surrounding the molecule. In principle, the MESP can be calculated from<sup>[56]</sup>

$$V(\mathbf{r}) = \sum_{\substack{\alpha=1 \\ \mathbf{R}_\alpha \neq \mathbf{r}}}^M \frac{Z_\alpha}{|\mathbf{R}_\alpha - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}', \quad (64)$$

in a.u., where  $Z_\alpha$  is the charge of  $\alpha$ th nucleus located at  $\mathbf{R}_\alpha$ , and where the  $\alpha$ th term in the first sum must be eliminated when  $\mathbf{R}_\alpha = \mathbf{r}$  (to evaluate the MESP at the position of the  $\alpha$ th nucleus). The dimensions of  $V$  are [energy]  $\times$  [charge<sup>-1</sup>], but often the displays of the MESP feature contour lines that are labeled in units of energy where the multiplication by a test unit charge (unity in a.u.) is implied if not stated explicitly.

The first HK theorem tells us that  $\rho(\mathbf{r})$  fixes the external potential, that is, the positions and charges of the nuclei, that is, the first term in the R.H.S. of Eq. (64). Then  $\rho(\mathbf{r})$ , which appears as the integrand of the second term, fixes the R.H.S. of Eq. (64) completely and hence there exists a bijective mapping between  $\rho(\mathbf{r})$  and  $V(\mathbf{r})$ .

Before discussing a few applications, we emphasize that  $V(\mathbf{r})$  has the same fundamental status as a basic descriptor of the ground state as  $\rho(\mathbf{r})$ . In 1962, Wilson has shown that the energy of a molecule can be expressed as a functional of  $\rho(\mathbf{r})$  and its change in response to a scaling of the nuclear charges by a continuous parameter  $\lambda$ .<sup>[202]</sup> Starting from this result, Politzer and Parr demonstrated that the total energy of a molecule  $E$  can be expressed as a sum of  $M$  terms of the form<sup>[203,204]</sup>

$$E = \sum_{\alpha=1}^M \frac{1}{2} \left\{ Z_\alpha V_\alpha^0 - \int_0^{Z_\alpha} \left[ Z'_\alpha \left( \frac{\partial V_\alpha^0}{\partial Z'_\alpha} \right) - V_\alpha^0 \right] dZ'_\alpha \right\}, \quad (65)$$

where  $V_\alpha^0$  is the MESP at the position of the  $\alpha$ th nucleus. The result in Eq. (65) is exact and demonstrates the fundamental role of  $V(\mathbf{r})$  even if it is not implementable in practice without approximations.

While the electrostatic potential of an isolated atom is always positive everywhere in space,<sup>[205]</sup> clearly in the case of molecules, according to Eq. (64), the MESP at  $\mathbf{r}$  can have either sign that diagnose regions that are primarily dominated by the bare nuclear potential, that is, the first term on the R.H.S. of Eq. (64), or dominated by the electron density (the second term). While it is quite possible and common for neutral mole-

cules to exhibit local minima (other than at the nuclear positions), it is impossible for any neutral molecule in the ground or excited states to exhibit true local maxima in the MESP, as demonstrated by Pathak and Gadre<sup>[206]</sup> on the basis of classical electrostatic arguments (only saddle points and directional maxima are allowed).

Minima in the MESP ( $V_{\min}$ ) pinpoint the sites favoring electrophilic attacks, while the magnitudes of corresponding  $V_{\min}$ 's in related molecules often correlate with reactivity toward electrophiles. The converse correlation with reactivity toward nucleophiles is nonexistent due to the absence of local maxima in the MESP.

In 1982, Weiner, Langridge, Blaney, Schaefer, and Kollman (WLBSK) proposed to color-code molecular 3-D surfaces with the values of the MESP as a function of position on these surfaces.<sup>[207]</sup> In this manner, the structural 3-D information conveyed by the surface of the molecule is augmented with electrostatic information provided by the color coding of the MESP on that surface, creating a "four-information-dimensional image," to quote Kahn, Pau, and Hehre (KPH).<sup>[208]</sup>

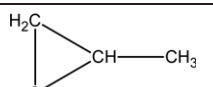
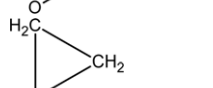
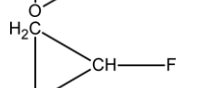

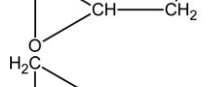
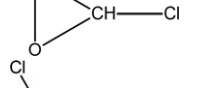
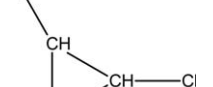
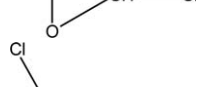
In the original proposal by WLBSK, the molecular surface was taken as the Connolly surface<sup>[207]</sup> (the solvent exclusion cavity in the bulk solvent). Nowadays, molecular isoelectron density envelopes are often selected for the color coding by the values of the MESP, as first suggested by KPH.<sup>[208]</sup> While true 3-D maxima in the MESP are not possible, 2-D local maxima on a given molecular surface, such as an isodensity isosurface, are possible.

Politzer has pioneered the application of MESP to solve chemical, environmental, and drug design problems. It is fitting then to briefly discuss one of his classical early works that provide a fast and accurate screening of the carcinogenic potential of epoxides.<sup>[201,209]</sup> Table 10 lists a series of epoxides sorted in decreasing magnitude of the  $V_{\min}$  near the oxygen atom. The table suggests a threshold of around  $-30$  kcal mol<sup>-1</sup> above which a compound is highly suspected of carcinogenic activity. This hypothesis has been confirmed by further extensive studies by Politzer and Murray<sup>[201]</sup> of over 60 epoxides, several of them calculated on the request of the *US Environmental Protection Agency*. The first step in DNA damage by epoxides is believed to involve protonation of the oxygen atom, which in turn promotes ring opening, which is consistent with Politzer's finding that highly negative  $V_{\min}$  near the oxygen atom promotes carcinogenic activity.

The second more recent example relates the highest positive MESP on the solvent-accessible surface (SAS) to the hydrogen-bonding-donor ability of the molecule.<sup>[210]</sup> First, the experimental quantification of molecular hydrogen-donor (or acceptor) ability deserves a brief introduction.

Abraham et al.<sup>[211]</sup> express what they terms "solute property (SP)" as a sum of four variables gauged against saturated alkanes for which all properties are taken as zero. The four terms are: (1)  $E$ : The excess molar refraction (in units of cm<sup>3</sup> mol<sup>-1</sup>/100), which models the excess dispersion interaction capability over saturated alkanes that  $\pi$ - and  $n$ -electrons impart; (2)  $S$ : solute polarity/polarizability, a term that captures the strength of solute-solvent interaction; (3)  $A$ : solute overall (sum) of

**Table 10.** Minimum MESP near the oxygen atom of some epoxides sorted in order of decreasing magnitudes (calculated at the HF/STO-5G level of theory).<sup>[a]</sup>

Molecule	$V_{\min}(\text{kcal mol}^{-1})$	Carcinogenic activity
	-53.4	Active
	-51.3	Active
	-47.7	Active
	-43.1	Active
	-38.1	Active
	-23.1	Inactive
	-17.1	Inactive
	-9.2	Inactive

[a] After Ref. 201.

H-bond acidity, the term subject of the modeling in the reviewed example below; (4) *B*: solute overall H-bond basicity; and (5)  $V_M$ : the McGowan characteristic molar volume (in units of  $\text{cm}^3 \text{mol}^{-1}/100$ ), which is a fragment contribution approximate molecular volume—not to be confused with the symbol for MESP. The SP is then taken as the dependent variable of a multivariate regression model of the form:<sup>[211]</sup>

$$SP = c + eE + sS + aA + bB + vV_M, \quad (66)$$

where the small letters characterize the solvent effect on the solute and the capital letters the effect of the solute on the solvent. As a concrete example, consider the regression equation:<sup>[211]</sup>

$$\log P_{O/w} = 0.088 + 0.562E - 1.054S - 0.032A - 3.460B + 3.814V, \\ [r^2 = 0.995, \text{STD} = 0.116, F = 23161.6, n = 613]$$

(67)

where STD is standard deviation, obtained for the logarithm of the octanol–water partition coefficient using a set of 613 com-

pounds. In this set,  $a \approx 0$  implies that octanol has the same basic hydrogen bonding power as water (essentially an oxygen atom acting as a hydrogen acceptor). In the meantime, water is far more acidic (better proton donor) than octanol, which is reflected in a significantly negative *b*.

Ghahourian and Dearden<sup>[210]</sup> model the experimentally determined *A* values for a set of 63 solute molecules using the MESP's most positive value on the SAS,  $V^+$ . Intuitively, the hydrogen atoms with the higher magnitude of  $V^+$  can be expected to be more acidic. These authors present 17 statistical models constructed from a variety of combinations of Gas-teiger charges,  $V^+$ , and  $E_{\text{LUMO}}$  (and molecular sets that include or exclude two outliers) to predict the Abraham hydrogen bonding donor *A* parameter. The best two models including the two outliers are<sup>[210]</sup>

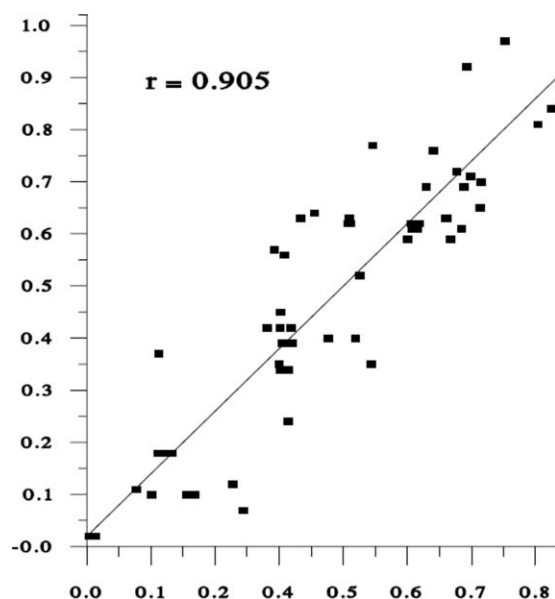
$$A = -0.490 + 0.0338 \times V^+, \quad (68) \\ [r^2 = 0.735, \text{STD} = 0.127, F = 166, n = 62]$$

and

$$A = -0.301 + 0.0290 \times V^+ - 0.0504 \times E_{\text{LUMO}}, \quad (69) \\ [r^2 = 0.819, \text{STD} = 0.106, F = 133, n = 62]$$

which are both slightly improved to  $r^2$  of 0.783 and 0.858, respectively, on exclusion of outliers ( $n = 57$  in both latter cases). The performance of Eq. (69) in predicting experimental values is displayed graphically in Figure 15, which is excellent in the light of the smallness of the parameters-to-data ratio of only 1:31.

The next example considers experimentally determined MESP maps. Today, advances in crystallography and the availability of multipolar refinement parameters databases<sup>[87,96–109]</sup> are such that atomic resolution protein structures are within



**Figure 15.** Correlations between the experimentally determined Abraham acidic overall hydrogen bonding capacity *A* and those estimated from the regression model of Eq. (69). (Reproduced from Ref. 210 with permission © 2004 Elsevier Masson SAS).

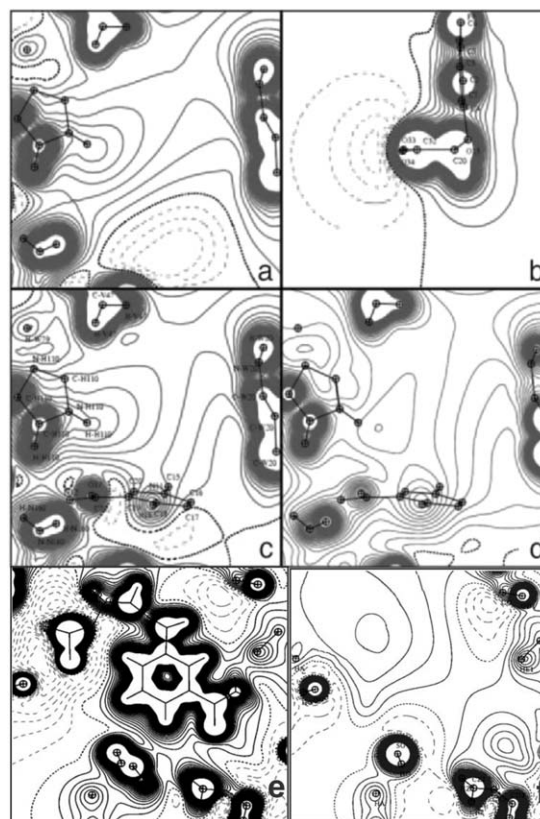
reach, which enables a better grasp of the electrostatic complementarity between guests and their protein hosts and the role played by the coenzymes in fine-tuning the active site electrostatics.<sup>[106,212–214]</sup> High resolution X-ray structures of enzymes complexed with their substrates, inhibitors, and/or coenzymes (which could be cosubstrates such as nicotinamide adenine dinucleotide phosphate cation (NADP<sup>+</sup>)) have been used to generate accurate maps of the MESP in the active site. These maps are particularly revealing if they are generated for the active site with and without its host(s).

The group in Nancy (Lecomte, Gillot, Jelsh, and co-workers) and the group in Strasbourg (Podjarny et al.) report experimental and calculated MESP maps of the inactive form (apoenzyme) of human aldol reductase (hALR2) (E); the active holoenzyme (apoenzyme complexed with its cosubstrate NADP<sup>+</sup>) but without the substrate or inhibitors (EC); an inhibitor of aldol reductase (IDD594) alone (I); and the enzyme–cosubstrate–inhibitor (ECI) complex. All four MESP maps are calculated from the experimental structure of the ECI complex by selective removal of the unwanted molecules.<sup>[106,212–214]</sup> The study provides a striking visualization of the electrostatic lock-and-key complementarity of the MESP of the inhibitor and that of the active site. The cosubstrate is shown to accentuate the inhibitor–enzyme complementarity (Fig. 16).

Another application of the MESP in drug-design relates to bioisosterism. Bioisosteres are pharmacologically interchangeable groups used to optimize the pharmacodynamics and the pharmacokinetics [absorption, distribution, metabolism, excretion, and toxicity (ADMET)] of a drug.<sup>[215,216]</sup> The bioisosterism of chemically unrelated pairs such as ethanoic (acetic) acid anion (EA) and 5-methyltertrazole (5MT) is puzzling.<sup>[217,218]</sup> The puzzle is resolved by examining the geometric disposition of the  $V_{\min}$  surrounding the two bioisosteres, a disposition that is not much altered by changing the capping groups.<sup>[219]</sup> The MESP exhibits four local minima around each bioisostere in a quasi-identical geometric and superimposable planar arrangement as can be seen from the distance matrices between the centroids of these minimas (Fig. 17). This similarity of the 3-D topography of the MESP is what the receptor site would “see,” and presumably has complementary electrophilic groups properly oriented in space (Fig. 17).<sup>[219]</sup>

It is noted in passing that the average group electron densities of these two chemically (very) different bioisosteres, EA and 5MT, is almost identical. Table 11 lists the atomic populations, volumes, and average electron densities defined as the electron population divided by the volume. The table lists the respective atomic, group, and molecular values of both isosteres. It is remarkable that the average density of the  $-\text{COO}^-$  group is 0.0659 a.u. and that of the  $-\text{CN}_4^-$  group is 0.0660, differing by only 0.0001 a.u. ( $\sim 0.15\%$ ) despite of being composed of different number and types of atoms and bonds, while the methyl groups in the two molecules have respective densities of 0.0386 a.u. (EA) and 0.0395 a.u. (5MT), that is, differing by an order of magnitude more than the bioisosteric groups in absolute terms and by  $\sim 2.3\%$  in relative terms.

Finally, it is sometime suggested to use the electric field  $\mathbf{E}(\mathbf{r})$  as opposed to  $V(\mathbf{r})$ , since the former provides a more direct

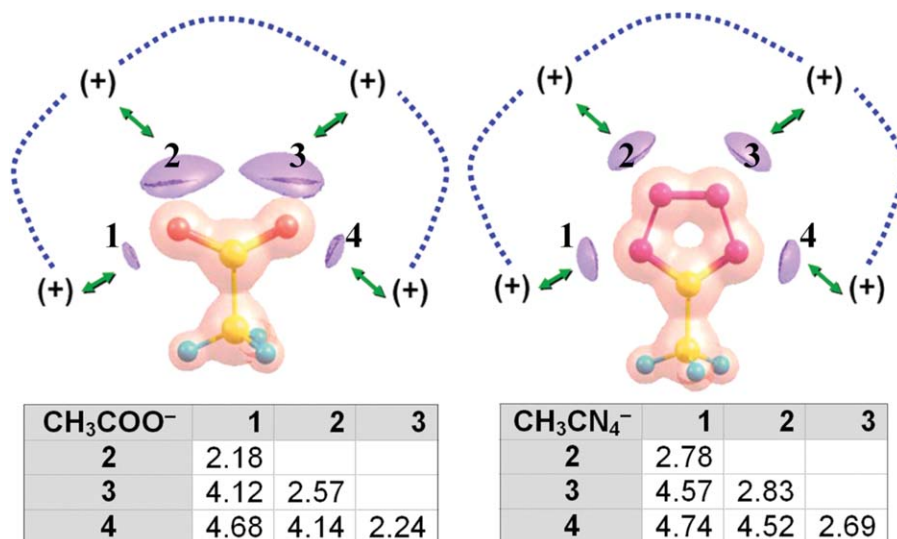


**Figure 16.** (a–d) MESP in the plane perpendicular to the inhibitor carboxylate group. (a) The active site in the apoenzyme (E) from DFT calculations at the experimental geometry of the complex. (b) The inhibitor alone (I) from DFT calculations at the experimental geometry. (c) The holoenzyme (apoenzyme + coenzyme NADP<sup>+</sup>, or EC) from DFT calculations at the experimental geometry. (d) The experimental MESP map corresponding to the calculated one in (c) obtained using a multipolar parameters database. (e) MESP of the holoenzyme computed with transferred multipolar parameters in the active site in the plane of the nicotinamide ring of the cofactor NADP<sup>+</sup>. (f) The MESP in the same region and in the same orientation as in (e) but after removing the contribution of the cofactor NADP<sup>+</sup>. Contour levels are  $\pm 0.05$  e/Å. Solid lines for positive contours; dashed ones for negative; and dotted line for zero level. (a–d) Reproduced from Ref. 106 with permission © 2003 National Academy of Science of the USA, and e–f from Ref. 214 with permission © 2004 Wiley Periodicals).

display of the direction of approach of an infinitesimal positive charge to the undisturbed molecular charge distribution. From elementary electrostatics:

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}). \quad (70)$$

The representation of  $\mathbf{E}$  instead of  $V$  is more convenient in cases when, for example, we are interested in the relative orientations of motion-restricted host molecules in the electric field generated by the restricting guest. A display of experimentally determined electric field lines in the active site of fatty acid binding protein, without its host (oleic acid), (B. Guillot, A. Podjarny, *Private communication* 2013), is provided in Figure 18. The field lines appear as bundles emanating from electropositive atoms and end at electro-negative atom sinks. The apparent crossing of field lines in this image is due to the projection of 3-D curved field lines



**Figure 17.** The two biososteres ethanoic acid anion and 5-methyltetrazole anion and a model of the region of a receptor complementary to the  $-\text{COO}^-$  and  $-\text{CN}_4^-$  regions. The figure displays the isopotential surfaces of the MESP as transparent envelopes. Pink transparent envelopes are those for  $\text{MESP} > 0$  and violet transparent envelopes indicate  $\text{MESP} < 0$ . The displayed isosurfaces  $\pm 0.250$  au for  $\text{CH}_3-\text{COO}^-$  and  $\pm 0.275$  au for  $\text{CH}_3-\text{CN}_4^-$ . The bottom part of the figure is the distance matrix (in Å) relating the four minima in the MESP of each biosostere.

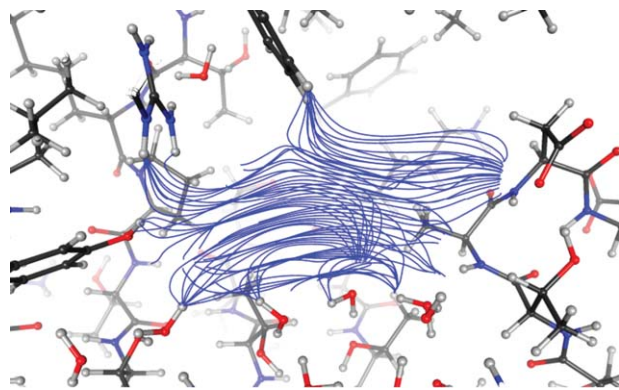
on the plane of the figure (field lines, of course, never cross). The intensity of the electric field in this enzyme active site is reflected in the density of the field lines and averages to a magnitude of the order of  $10^9$  V m<sup>-1</sup> near the centroid of the cavity (B. Guillot, A. Podjarny, *Private communication* 2013). These electric field strengths, typically encountered in enzyme active sites, can alter the rates of chemical reactions significantly and can have detectable IR vibrational Stark shifts.<sup>[220–232]</sup>

Through the use of calibration curves, vibrational Stark shifts exhibited by small host molecules or chromophoric groups in molecules trapped into enzyme active sites can be used to probe the strengths and directions of the local electric fields. Such studies have been carried out for a number of host–protein systems: CO bound to the Fe of the porphyrin of myoglobin,<sup>[225]</sup> and  $\text{C}\equiv\text{N}$  (nitrile)-containing substrates complexed at the human aldose reductase enzyme (hALR2) active site.<sup>[226,227]</sup> are two notable examples. The changes in the local fields

**Table 11.** Atomic, group, and molecular electron populations, volumes, and average electron densities and the numbering schemes of the anionic biososteres EA (acetic acid) and 5-methyltetrazole.<sup>[a,b]</sup>

Ethanoic acid anion (EA)				5-Methyltetrazole anion (5MT)			
Atom ( $\Omega$ )	$N(\Omega)$	$\text{Vol}(\Omega)$	$\langle \rho \rangle$	Atom ( $\Omega$ )	$N(\Omega)$	$\text{Vol}(\Omega)$	$\langle \rho \rangle$
C1	6.0102	73.06	0.0823	C1	5.9629	70.79	0.0842
H2	1.0397	54.30	0.0192	H2	1.0174	52.54	0.0194
H3	1.0392	54.67	0.0190	H3	1.0255	53.18	0.0193
H4	1.0397	54.81	0.0190	H4	1.0114	52.01	0.0195
$\Sigma-\text{CH}_3$	<b>9.1288</b>	<b>236.83</b>	<b>0.0386</b>	$\Sigma-\text{CH}_3$	<b>9.0171</b>	<b>228.52</b>	<b>0.0395</b>
C5	4.3728	38.76	0.1128	N5	7.7152	124.55	0.0619
O6	9.2513	153.98	0.0601	N6	7.2050	113.09	0.0637
O7	9.2466	154.10	0.0600	N7	7.7174	124.85	0.0618
$\Sigma$ biosostere	<b>22.8707</b>	<b>346.84</b>	<b>0.0659</b>	N9	7.2067	113.18	0.0637
$\Sigma$ molecule	<b>31.9995</b>	<b>583.67</b>	<b>0.0548</b>	C8	5.1390	54.71	0.0939
				$\Sigma$ biosostere	<b>34.9832</b>	<b>530.38</b>	<b>0.0660</b>
				$\Sigma$ molecule	<b>44.0003</b>	<b>758.90</b>	<b>0.0580</b>

[a] Rows list the properties of individual atoms except the row " $\Sigma-\text{CH}_3$ " which lists the properties of the capping methyl group, the row " $\Sigma$  biosostere" which lists the properties of the entire biososteric groups, and the row " $\Sigma$  molecule" which lists the molecular properties. [b] Data taken from Ref. 219.



**Figure 18.** Experimental electric field lines in the active site of oleic acid in fatty acid binding protein. The intensity of the field in the enzyme active site is of the order of  $10^9 \text{ V m}^{-1}$ . (Obtained from a private communication and used with the permission of the copyright holders, © 2013 B. Guillot and A. Podjarny).

accompanying point mutations in the active site have been monitored through the observation of the corresponding vibrational Stark shifts of the nitrile group IR fingerprint by Boxer et al.<sup>[226,227]</sup> (See also Refs. 228–231).

## Conclusions

This review focuses on the intersection of two areas of research: One is QSAR-type research and the other is the study of observable molecular fields derived from either theoretical calculations or experiment. QSAR-type studies have, in general, a dominant utilitarian flavor in the sense that it places the highest import on the speedy and accurate prediction of activity with perhaps a secondary focus on insight as long as the QSAR model yields robust statistics. The purpose here is not to criticize this type of QSAR studies since they are of vital importance to cure disease and improve the lives of humans. Rather, what is being argued is that by choosing descriptors directly derived from fundamental fields such as the electron density or the electrostatic potential, then by necessity those fields must be the source of all predictions. It follows that density- and/or MESP-derived descriptors can yield robust QSAR relationships in addition to insight.

A density-based descriptor as simple as the CSI has been shown to have a particularly wide reach of applicability. This one descriptor, which measures the local degree of charge separation at an atomic resolution, can be used in the predictive modeling of molecular properties that include PMVs, partition coefficients ( $\log P$ ), the change in the protein stability on mutation, the structure of the genetic code, and excitation energies of simple molecules. The latter prediction follows directly from the HK theorem as discussed at the start of the section entitled “*Bond Properties as Predictors of Spectroscopic Transitions and NMR Proton Chemical Shifts*” since the ground state electron density encodes information about excited states as well.

We have also seen how the concept of QTAIM atomic energy,  $E(\Omega)$ , and that of atomic electric charge,  $q(H)$ , converge, through QSAR modeling, to two formulas of identical

form despite of being reached via totally different considerations:

$$pK_a(\text{exptl.}) = -150.486 - 0.1725 \times E(H) (\text{kJ mol}^{-1}), \quad (26)$$

$$pK_a(\text{exptl.}) = 69.5 (\pm 5.2) - 112.5 (\pm 8.9) \times q(H) (\text{a.u.}), \quad (27)$$

where the first, due to Adam,<sup>[158]</sup> is derived from statistical mechanical arguments and approximations followed by a fitting to experimental  $pK_a$  values, and the second is obtained by Hollingsworth et al.<sup>[160]</sup> from a direct fitting to experiment on grounds of chemical intuition. The connection between these two QSAR equations can be rationalized as follows: The more an acid is acidic the more positively charged is its acidic (dissociating) hydrogen, and hence by Eq. (27) it has a lower  $pK_a$ . In the meantime, the more positive the hydrogen, the less is its electron population, and the lower the electron population the lower the stabilization due to the nuclear–electron Coulombic attraction, and hence that leads to a higher atomic energy (less negative), which leads to a lower  $pK_a$  as well by Eq. (26).

Popelier’s QTMS approach has been briefly reviewed in the light of being a molecular structure fingerprinting tools, invaluable in QTAIM-based QSAR. We have seen how the inability of QTMS to capture size and hydrophobicity/hydrophilicity information can actually be a strength rather than a weakness since it can be used as a diagnostic to indicate if a given property is primarily driven by bulk and/or by lipid/aqueous solubility or is primarily a reflection of electronic structure. Popelier and coworkers have complemented QTMS models, when necessary, with bulk and solubility descriptors extraneous to QTAIM, in analogy with QSAR strategies based on Hammett constants. This is an example of the complementarity and synergism between QTAIM and other more traditional approaches in QSAR.

The QTMS approach has the flexibility of being applicable to any desired portion of the molecular framework. One can construct a QTMS QSAR using any desired isolated subgraph of the molecule such as the subgraph representing the active site fragment. The ability to isolate and focus on a molecular active site is of particular value when such a site is common to a series of molecules that otherwise differ significantly in their molecular skeletons, and which are hence not amenable to direct comparison.

QTMS is based on bond properties, that is, a sampling of the electron density at the BCP. There are other approaches that share this focus on BCP properties in the construction of predictive tools but without constructing the BCP abstract mathematical space followed by the use of Eq. (13). An example is the work of BAA. These workers use BCP properties to predict NMR and UV spectroscopic transitions,<sup>[145–147]</sup> while others use bond properties in the empirical prediction of  $\pi$ -stacking interaction energies.<sup>[178–180]</sup> Cortés-Guzmán and coworkers<sup>[177]</sup> use the net group charge transfer as a predictor of the strength of  $\pi$ -stacking between anticancer Casiopeinas<sup>®</sup> and adenine to support their hypothesis for the antineoplastic activity of this class of drugs through DNA intercalation.

There are a number of parallels between features of QTMS and the newly proposed LDMs and EDWAMs. First, both LDMs

and EDWAMs can be defined for an arbitrary molecular fragment subgraph, in which case the sum of sums in Eq. (57) for an LDM would yield the electronic population of the fragment. Second, the LDMs have a built-in measure of molecular size for congeners with the same number of atoms and a common framework (see the section entitled “The LDM  $\zeta$  and the  $pK_a$  of Halogenated Acetic Acids”), namely, the sum of the diagonal elements  $\text{tr}(\zeta)$ . The LDMs, hence, may not be ideal to model properties such as  $pK_a$  that strongly depend on electronic structure but only weakly on molecular size. In these cases, the diagonal suppressed LDMs (the DMs) perform far better being primarily an electronic fingerprinting tool, as the example of halogenated acetic acids demonstrates.

The reasoning leading to LDMs can be extended to any property that can be partitioned into atomic “self” and two-atom “interaction” terms, to use the Oviedo Group’s IQA terminology.<sup>[119–121]</sup> Within the IQA approach, the energy of a molecule is rigorously decomposed as:

$$E = \sum_{i=1}^n E_{\text{self}}(\Omega_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n E_{\text{int}}(\Omega_i, \Omega_j), \quad (71)$$

which is identical in form to the R.H.S. of the second equality of Eq. (54) and hence can be conveniently organized in a matrix format similar to the LDM:

$$\eta \equiv \begin{bmatrix} E_{\text{self}}(\Omega_1) & E_{\text{int}}(\Omega_1, \Omega_2)/2 & \cdots & E_{\text{int}}(\Omega_1, \Omega_n)/2 \\ E_{\text{int}}(\Omega_2, \Omega_1)/2 & E_{\text{self}}(\Omega_2) & \cdots & E_{\text{int}}(\Omega_2, \Omega_n)/2 \\ \vdots & \vdots & \ddots & \vdots \\ E_{\text{int}}(\Omega_n, \Omega_1)/2 & E_{\text{int}}(\Omega_n, \Omega_2)/2 & \cdots & E_{\text{self}}(\Omega_n) \end{bmatrix}, \quad (72)$$

where the sum of any row or column is the so-called “additive” energy of atom  $\Omega_i$ ,  $E_{\text{add}}(\Omega_i)$ , since the total molecular energy is equal to the sum of all  $E_{\text{add}}(\Omega_i)$ . The same QSAR manipulations expressed in this article for the LDMs ( $\zeta$ ) can be applied to  $\eta$ .

Closely related to the total charge density [Eq. (4)] is the MESP  $V(\mathbf{r})$  and its associated electrostatic field  $\mathbf{E}(\mathbf{r})$ . The MESP and field have the same experimentally-accessible and fundamental qualities of  $\rho$  itself [Eq. (65)]. Modern ultrahigh resolution X-ray crystallography is capable of delivering accurate experimental MESP and electric fields even within the confines of an enzyme active site. Probing the electric field inside an active site is of paramount importance in understanding enzyme catalysis as has been emphasized in the recent literature.<sup>[220–232]</sup>

Lecomte, Guillot, Jelsh, Podjarny and coworkers<sup>[106,212–214]</sup> (and B. Guillot, A. Podjarny, *Private communication* 2013) map the electric fields in the active site of enzymes with and without their hosts (whether cofactors, substrates, or inhibitors, and combinations of those). This strategy yields key information about the electrostatic characteristics necessary for optimal inhibition which can guide drug-design by “reverse engineering” in a striking visualization and use of Hermann Emil Fischer’s “lock-and-key” hypothesis.

## Acknowledgments

The author is indebted to the late Professor Richard F. W. Bader (1931–2012) for numerous discussions over the years. The author thanks Dr. Todd A. Keith and Professor Fernando Cortés-Guzmán for making their manuscript available prior to publication;<sup>[175]</sup> Dr. Benoit Guillot and Dr. Alberto D. Podjarny for sharing the electric field map displayed in Figure 18 prior to publication (B. Guillot, A. Podjarny, *Private communication* 2013); Professor Lou Massa, Dr. Nenad Trinajstić (NT), Dr. Sonja Nikolić (SN) for discussions regarding graph-theoretical matrices, and NT and SN for their hospitality in Zagreb and for offering the author a copy of Refs. 51 and 189; and Professor Paul L. A. Popelier for helpful comments on the manuscript regarding QTMS.

## Note added in Proof

Very recently, a paper has appeared<sup>[233]</sup> that demonstrates strong correlations between the number of hydrogen-hydrogen close contacts (hydrogen-hydrogen weak bonding interactions)<sup>[234–236]</sup> in alkane dimers and boiling points. These weak interactions also appear to have a secondary role in the relative stability of branched alkanes with respect to their less branched isomers.<sup>[233]</sup> This provides another example of a direct link between an observable property of the topology of the electron density and a measurable physical property of the corresponding compound. A review that examines the role of these weak interactions in the stability of molecules and crystals is being coauthored by the present writer and will be submitted to publication soon.

**Keywords:** electron density · QSAR · physicochemical properties of molecules · QTAIM · quantum theory of atoms in molecules · biophysical chemistry · quantitative structure-activity relationships

How to cite this article: Matta, C. F., *J. Comput. Chem.* **2014**, *35*, 1165–1198. DOI: 10.1002/jcc.23608



Additional Supporting Information may be found in the online version of this article.

- [1] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, **1995**.
- [2] A. Leo, C. Hansch, D. Elkins, *Chem. Rev.* **1971**, *71*, 525.
- [3] C. Hansch, *Acc. Chem. Res.* **1969**, *2*, 232.
- [4] C. Hansch, J. E. Quinlan, G. L. Lawrence, *J. Org. Chem.* **1968**, *33*, 347.
- [5] M. S. Tute, In *Advances in Drug Research*, Vol. 6; N. J. Harper, A. B. Simmonds, Eds.; Academic Press: London, **1971**; p. 1.
- [6] L. H. Hall, L. B. Kier, *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: Boston, **1976**.
- [7] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, 2nd ed., Vols. I and II; Wiley-VCH: Weinheim, **2009**.
- [8] P. P. Mager, *Multidimensional Pharmacology: Design of Safer Drugs*; Academic Press: London, **1984**.
- [9] R. Carbó-Dorca, D. Robert, L. Amat, X. Gironé, E. Besalú, *Molecular Quantum Similarity in QSAR and Drug Design*; Springer: Berlin, **2000**.
- [10] W. G. Richards, *Quantum Pharmacology*, 2nd ed.; Butterworth & Co.: London, **1983**.
- [11] F. Helmer, K. Kiehs, C. Hansch, *Biochemistry* **1968**, *7*, 2858.

- [12] D. A. Winkler, *Brief. Bioinformatics* **2002**, *3*, 73.
- [13] A. V. Gubskaya, In *Quantum Biochemistry: Electronic Structure and Biological Activity*, Vol. 2; C. F. Matta, Ed.; Wiley-VCH: Weinheim, **2010**; p. 693.
- [14] R. V. Bensasson, S. Sowlati-Hashjin, V. Zoote, D. Dauzonne, C. F. Matta, *Int. Rev. Phys. Chem.* **2013**, *32*, 393.
- [15] C. F. Matta, A. A. Arabi, *Future Med. Chem.* **2011**, *3*, 969.
- [16] D. Chopra, *J. Phys. Chem. A* **2012**, *116*, 9791.
- [17] R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, **1990**.
- [18] P. L. A. Popelier, *Atoms in Molecules: An Introduction*; Prentice Hall: London, **2000**.
- [19] C. F. Matta, R. J. Boyd, Eds., *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*; Wiley-VCH: Weinheim, **2007**.
- [20] R. F. W. Bader, *Chem. Rev.* **1991**, *91*, 893.
- [21] R. F. W. Bader, *J. Phys. Chem. A* **2007**, *111*, 7966.
- [22] R. F. W. Bader, *Monatsh Chem* **2005**, *136*, 819.
- [23] R. F. W. Bader, *Can. J. Chem.* **1998**, *76*, 973.
- [24] R. F. W. Bader, P. L. A. Popelier, T. A. Keith, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 620.
- [25] V. G. Tsirelson, R. P. Ozerov, *Electron Density and Bonding in Crystals: Principles, Theory and X-ray Diffraction Experiments in Solid State Physics and Chemistry*; Institute of Physics Publishing: New York, **1996**.
- [26] P. Coppens, *X-ray Charge Densities and Chemical Bonding*; Oxford University Press: New York, **1997**.
- [27] T. S. Koritsansky, P. Coppens, *Chem. Rev.* **2001**, *101*, 1583.
- [28] C. Gatti, P. Macchi, Eds., *Modern Charge Density Analysis*; Springer: Germany, **2012**.
- [29] P. Macchi, *Cryst. Rev.* **2013**, *19*, 58.
- [30] A. A. Rykounov, V. G. Tsirelson, *J. Mol. Struct. (Theochem)* **2009**, *906*, 11.
- [31] M. Jablonski, M. Palusiak, *J. Phys. Chem. A* **2010**, *114*, 12498.
- [32] C. F. Matta, *J. Comput. Chem.* **2010**, *31*, 1297.
- [33] V. R. Coluci, R. Vendrame, R. S. Braga, D. S. Galvão, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1479.
- [34] O. Llorens, J. J. Perez, H. O. Villar, *Int. J. Quantum Chem.* **2002**, *88*, 107.
- [35] F. T. Hatch, M. G. Kinze, M. E. Colvin, *Environ. Mol. Mutagen.* **2001**, *38*, 268.
- [36] L. F. Motta, W. P. Almeida, *Int. J. Drug Discov.* **2011**, *3*, 100.
- [37] M. Shahlaei, *Chem. Rev.* **2013**, *113*, 8093.
- [38] R. Kiralja, M. M. C. Ferreira, *J. Chemom.* **2010**, *24*, 681.
- [39] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, 864.
- [40] W. A. Kato, *Commun. Pure Appl. Math.* **1957**, *10*, 151.
- [41] I. Riess, W. Münch, *Theor. Chim. Acta.* **1981**, *58*, 295.
- [42] R. F. W. Bader, P. Becker, *Chem. Phys. Lett.* **1988**, *148*, 452.
- [43] P. G. Mezey, *Mol. Phys.* **1999**, *96*, 169.
- [44] R. F. W. Bader, P. F. Zou, *Chem. Phys. Lett.* **1992**, *191*, 54.
- [45] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 3rd ed.; Oxford University Press: Oxford, **1958**.
- [46] R. F. W. Bader, C. F. Matta, *J. Phys. Chem. A* **2004**, *108*, 8385.
- [47] L. P. Hammett, *Physical Organic Chemistry: Reaction Rates, Equilibria, and Mechanisms*, 2nd ed.; McGraw-Hill Book Company: New York, **1970**.
- [48] T. K. Attwood, D. J. Parry-Smith, *Introduction to Bioinformatics*; Prentice Hall: London, **1999**.
- [49] C. Gibas, P. Jambeck, *Developing Bioinformatics Computer Skills*; O'Reilly: Cambridge, **2001**.
- [50] D. W. Mount, *Bioinformatics: Sequence and Genome Analysis*, 2nd ed.; Cold Spring Harbor Laboratory Press: New York, **2004**.
- [51] D. Janezic, A. Milicevic, S. Nikolic, N. Trinajstic, *Graph Theoretical Matrices in Chemistry (Mathematical Chemistry Monographs, Vol. 3)*; University of Kragujevac: Kragujevac, **2007**.
- [52] J.-P. Doucet, J. Weber, *Computer-Aided Molecular Design: Theory and Applications*; Academic Press: London, **1996**.
- [53] R. Carbó, L. Leyda, M. Arnau, *Int. J. Quantum Chem.* **1980**, *17*, 1185.
- [54] R. Carbó-Dorca, P. G. E. Mezey, *Advances in Molecular Similarity, Vol. 1*; Jai Press: London, **1996**.
- [55] R. Carbó-Dorca, P. G. E. Mezey, *Advances in Molecular Similarity, Vol. 2*; Jai Press: London, **1998**.
- [56] R. Bonaccorsi, E. Scrocco, J. Tomasi, *J. Chem. Phys.* **1970**, *52*, 5270.
- [57] C. Petrongolo, J. Tomasi, *Int. J. Quantum Chem.: Quantum Biol. Symp.* No. 2 **1975**, 181.
- [58] R. Bonaccorsi, E. Scrocco, J. Tomasi, *J. Am. Chem. Soc.* **1976**, *98*, 4049.
- [59] J. Tomasi, In *Chemical Applications of Atomic and Molecular Electrostatic Potentials. Reactivity, Structure, Scattering; Energetics of Organic, Inorganic, and Biological Systems*; P. Politzer, D. G. Truhlar, Eds.; Plenum Press: New York, **1981**, p. 257.
- [60] J. Tomasi, C. Cappelli, B. Mennucci, R. Cammi, In *Quantum Biochemistry: Electronic Structure and Biological Activity, Vol. 1*; C. F. Matta, Ed.; Wiley-VCH: Weinheim, **2010**; p. 131.
- [61] P. Politzer, D. G. Truhlar, Eds. *Chemical Applications of Atomic and Molecular Electrostatic Potentials. Reactivity, Structure, Scattering, Energetics of Organic, Inorganic, and Biological Systems*; Plenum Press: New York, **1981**.
- [62] J. S. Murray, P. Politzer, In *Encyclopedia of Computational Chemistry*; P. v. R. Schleyer, Ed.; Wiley: Chichester, **1998**; p. 912.
- [63] P. Politzer, J. S. Murray, In *Reviews in Computational Chemistry, Vol 2*; K. B. Lipkowitz, D. B. Boyd, Eds.; Wiley: Hoboken, NJ, **1991**; p. 273.
- [64] S. R. Gadre, S. A. Kulkarni, C. H. Suresh, I. H. Shrivastava, *Chem. Phys. Lett.* **1995**, *239*, 273.
- [65] C. H. Suresh, S. R. Gadre, *J. Am. Chem. Soc.* **1998**, *120*, 7049.
- [66] S. R. Gadre, In *Computational Chemistry: Reviews of Current Trends*; J. Leszczynski, Ed.; World Scientific: Singapore, **1999**; p. 1.
- [67] S. R. Gadre, R. N. Shirsat, *Electrostatics of Atoms and Molecules*; Universities Press: Hyderabad, **2000**.
- [68] D. Roy, P. Balanarayan, S. R. Gadre, *J. Chem. Phys.* **2008**, *129*, 174103.
- [69] J. Cioslowski, A. Nanayakkara, *J. Am. Chem. Soc.* **1993**, *115*, 11213.
- [70] P. Constans, L. Amat, R. Carbó-Dorca, *J. Comput. Chem.* **1997**, *18*, 826.
- [71] A. J. McMahon, P. M. King, *J. Comput. Chem.* **1997**, *18*, 151.
- [72] N. Cohen, S. W. Benson, *Chem. Rev.* **1993**, *93*, 2419.
- [73] N. Cohen, S. W. Benson, In *The Chemistry of Alkanes and Cycloalkanes*; S. Patai, Z. Rappoport, Eds.; Wiley: New York, **1992**; p. 215.
- [74] S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, R. Walsh, *Chem. Rev.* **1969**, *69*, 279.
- [75] C. Chang, R. F. W. Bader, *J. Phys. Chem.* **1992**, *96*, 1654.
- [76] R. F. W. Bader, D. Bayles, *J. Phys. Chem. A* **2000**, *104*, 5579.
- [77] R. F. W. Bader, F. J. Martín, *Can. J. Chem.* **1998**, *76*, 284.
- [78] F. J. Martín, *Theoretical Synthesis of Macromolecules from Transferable Functional Groups*. Ph.D. Thesis. McMaster University: Hamilton, 2001.
- [79] R. F. W. Bader, C. F. Matta, F. J. Martín, In *Medicinal Quantum Chemistry*; F. Alber, P. Carloni, Eds.; Wiley-VCH: Weinheim, **2003**; p. 201.
- [80] C. F. Matta, R. F. W. Bader, *Proteins: Struct. Funct. Genet.* **2003**, *52*, 360.
- [81] C. F. Matta, R. F. W. Bader, *J. Phys. Chem. A* **2006**, *110*, 6365.
- [82] R. F. W. Bader, C. F. Matta, *Found. Chem.* **2013**, *15*, 253.
- [83] R. F. W. Bader, *Theor. Chem. Acc.* **2001**, *105*, 276.
- [84] L. Delle Site, *Theor. Chem. Acc.* **2002**, *107*, 378.
- [85] L. Delle Site, *IMA J. App. Math.* **2002**, *67*, 411.
- [86] C. F. Matta, *J. Phys. Chem. A* **2001**, *105*, 11088.
- [87] S. Scheins, M. Messerschmidt, P. Luger, *Acta Cryst. B* **2005**, *61*, 443.
- [88] L. Huang, L. Massa, J. Karle, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12690.
- [89] L. Huang, L. Massa, J. Karle, *Int. J. Quantum Chem.* **2005**, *103*, 808.
- [90] L. Huang, L. Massa, J. Karle, *Biochemistry* **2005**, *44*, 16747.
- [91] L. Huang, L. Massa, J. Karle, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1233.
- [92] L. Huang, L. Massa, J. Karle, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1731.
- [93] L. Huang, L. Massa, J. Karle, In *Quantum Biochemistry: Electronic Structure and Biological Activity, Vol 1*; C. F. Matta, Ed.; Wiley-VCH: Weinheim, **2010**; p. 3.
- [94] L. Huang, H. Bohorquez, C. F. Matta, L. Massa, *Int. J. Quantum Chem.* **2011**, *111*, 4150.
- [95] L. Huang, L. Massa, C. F. Matta, *Carbon* **2014**, accepted, in the press.
- [96] T. Koritsansky, A. Volkov, P. Coppens, *Acta Cryst. A* **2002**, *58*, 464.
- [97] A. Volkov, X. Li, T. Koritsansky, P. Coppens, *J. Phys. Chem. A* **2004**, *108*, 4283.
- [98] P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, *J. Chem. Theor. Comput.* **2007**, *3*, 232.
- [99] A. Volkov, M. Messerschmidt, P. Coppens, *Acta Cryst. D* **2007**, *63*, 160.
- [100] V. Pichon-Pesme, C. Lecomte, R. Wiest, M. Bénard, *J. Am. Chem. Soc.* **1992**, *114*, 2713.
- [101] V. Pichon-Pesme, C. Lecomte, H. Lachekar, *J. Phys. Chem.* **1995**, *99*, 6242.



- [102] C. Jelsch, V. Pichon-Pesme, C. Lecomte, A. Aubry, *Acta Cryst.* **1998**, *D54*, 1306.
- [103] D. Housset, F. Benabicha, V. Pichon-Pesme, C. Jelsch, A. Maierhofer, S. David, J. C. Fontecilla-Camps, C. Lecomte, *Acta Cryst.* **2000**, *D56*, 151.
- [104] M. V. Fernández-Serra, J. Junquera, C. Jelsch, C. Lecomte, E. Artacho, *Solid State Commun.* **2000**, *116*, 395.
- [105] B. Guillot, L. Viry, R. Guillot, C. Lecomte, C. J. Jelsch, *J. App. Cryst.* **2001**, *34*, 214.
- [106] N. Muzet, B. Guillot, C. Jelsch, E. Howard, C. Lecomte, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8742.
- [107] B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman, *Acta Cryst. D* **2006**, *62*, 1325.
- [108] P. Luger, B. Dittrich, In *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, C. F. Matta, R. J. Boyd, Eds.; Wiley-VCH: Weinheim, **2007**; p. 317.
- [109] P. Luger, *Org. Biomolec. Chem.* **2007**, *5*, 2529.
- [110] F. Cortés-Guzmán, R. F. W. Bader, *Chem. Phys. Lett.* **2003**, *379*, 183.
- [111] R. F. W. Bader, K. M. Gough, K. E. Laidig, T. A. Keith, *Mol. Phys.* **1992**, *75*, 1167.
- [112] R. F. W. Bader, T. A. Keith, *J. Chem. Phys.* **1993**, *99*, 3683.
- [113] F. Cortés-Guzmán, R. F. W. Bader, *J. Phys. Org. Chem.* **2004**, *17*, 95.
- [114] C. M. Breneman, L. W. Weber, In *NATO ASI Series: The Application of Charge Density Research to Chemistry and Drug Design*, G. A. Jeffrey, J. F. E. Piniella, Eds.; Plenum Press: New York, **1991**; p. 357.
- [115] C. M. Breneman, T. R. Thompson, M. Rhem, M. Dung, *Comput. & Chem.* **1995**, *19*, 161.
- [116] C. M. Breneman, M. Rhem, *J. Comput. Chem.* **1997**, *18*, 182.
- [117] N. Sukumar, C. M. Breneman, In *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, C. F. Matta, R. J. Boyd, Eds.; Wiley-VCH: Weinheim, **2007**; p. 473.
- [118] X. Fradera, M. A. Austen, R. F. W. Bader, *J. Phys. Chem. A* **1999**, *103*, 304.
- [119] M. A. Blanco, A. Martín-Pendás, E. Francisco, *J. Chem. Theor. Comput.* **2005**, *1*, 1096.
- [120] E. Francisco, A. Martín-Pendás, M. A. Blanco, *J. Chem. Theory Comput.* **2006**, *2*, 90.
- [121] A. Martín-Pendás, E. Francisco, M. A. Blanco, *J. Phys. Chem. A* **2006**, *110*, 12864.
- [122] P. L. A. Popelier, *Mol. Phys.* **1996**, *87*, 1196.
- [123] P. L. A. Popelier, *J. Phys. Chem. A* **1999**, *103*, 2883.
- [124] S. E. O'Brien, P. L. A. Popelier, *Can. J. Chem.* **1999**, *77*, 28.
- [125] S. E. O'Brien, P. L. A. Popelier, In *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2000)*, International Center for Numerical Methods in Engineering, Barcelona, **2000**; p. 1.
- [126] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764.
- [127] U. A. Chaudry, P. L. A. Popelier, *J. Phys. Chem. A* **2003**, *107*, 4578.
- [128] P. L. A. Popelier, P. J. Smith, U. A. Chaudry, *J. Comput-Aided Mol. Des.* **2004**, *18*, 709.
- [129] U. A. Chaudry, P. L. A. Popelier, *J. Org. Chem.* **2004**, *69*, 233.
- [130] P. J. Smith, P. L. A. Popelier, *Org. Biomol. Chem.* **2005**, *3*, 3399.
- [131] N. Singh, R. J. Loader, P. J. O'Malley, P. L. A. Popelier, *J. Phys. Chem. B* **2006**, *110*, 6498.
- [132] P. L. A. Popelier, P. J. Smith, *Eur. J. Med. Chem.* **2006**, *41*, 862.
- [133] R. J. Loader, N. Singh, P. J. O'Malley, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1249.
- [134] K. Roy, P. L. A. Popelier, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2604.
- [135] I. Bytheway, M. G. Darley, P. L. A. Popelier, *ChemMedChem.* **2008**, *3*, 445.
- [136] K. Roy, P. L. A. Popelier, *QSAR Comb. Sci.* **2008**, *27*, 1006.
- [137] A. P. Harding, D. C. Wedge, P. L. A. Popelier, *J. Chem. Inf. Mod.* **2009**, *49*, 1914.
- [138] M. Devereux, P. L. A. Popelier, I. M. McLay, *J. Comput. Chem.* **2009**, *30*, 1300.
- [139] M. Devereux, P. L. A. Popelier, I. M. McLay, *Phys. Chem. Chem. Phys.* **2009**, *11*, 1595.
- [140] G. I. Hawe, I. Alkorta, P. L. A. Popelier, *J. Chem. Inf. Model.* **2010**, *50*, 87.
- [141] S. Kar, A. P. Harding, K. Roy, P. L. A. Popelier, *SAR QSAR Env. Res.* **2010**, *21*, 149.
- [142] P. L. A. Popelier, In *Quantum Biochemistry: Electronic Structure and Biological Activity*, Vol 2, C. F. Matta, Ed.; Wiley-VCH: Weinheim, **2010**; p. 669.
- [143] M. Z. Griffiths, P. L. A. Popelier, *J. Chem. Inf. Model.* **2013**, *53*, 1714.
- [144] A. J. Green, P. L. A. Popelier, *J. Chem. Inf. Model.* **2014**, *54*, 553.
- [145] B. Buttingsrud, B. K. Alsberg, P.-O. Åstrand, *J. Mol. Struct. (THEOCHEM)* **2007**, *810*, 15.
- [146] B. Buttingsrud, B. K. Alsberg, P.-O. Åstrand, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2226.
- [147] B. Buttingsrud, B. K. Alsberg, P.-O. Åstrand, *J. Comput. Chem.* **2007**, *28*, 15.
- [148] A. Williams, *Free Energy Relationships in Organic and Bio-Organic Chemistry*. Royal Society of Chemistry (RSC): Cambridge, UK, **2003**.
- [149] C. Xue, P. L. A. Popelier, *J. Phys. Chem. B* **2009**, *113*, 3245.
- [150] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
- [151] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe, *Chem. Rev.* **1994**, *94*, 1873.
- [152] P. L. A. Popelier, U. A. Chaudry, P. J. Smith, *J. Chem. Soc., Perkin Trans. 2* **2002**, 1231.
- [153] H. Kubinyi, In *The Encyclopedia of Computational Chemistry*, Vol. 1; P. Schleyer, Ed.; Wiley: Chichester **1998**, p. 448.
- [154] C. L. Waller, J. D. McKinney, *J. Med. Chem.* **1992**, *35*, 3660.
- [155] S. E. O'Brien, P. L. A. Popelier, *J. Chem. Soc., Perkin Trans. 2* **2002**, 478.
- [156] C. D. Selassie, R. P. Verma, S. Kapur, A. J. Shusterman, C. Hansch, *J. Chem. Soc., Perkin Trans. 2* **2002**, 1112.
- [157] R. Gaudoin, K. Burke, *Phys. Rev. Lett.* **2004**, *93*, 173001.
- [158] K. R. Adam, *J. Phys. Chem. A* **2002**, *106*, 11963.
- [159] C. F. Matta, A. A. Arabi, T. A. Keith, *J. Phys. Chem. A* **2007**, *111*, 8864.
- [160] C. A. Hollingsworth, P. G. Seybold, C. M. Hadad, *Int. J. Quantum Chem.* **2002**, *90*, 1396.
- [161] T. V. Chalikian, *J. Phys. Chem. B* **2008**, *112*, 911.
- [162] H.-J. Hinz, Ed. *Thermodynamic Data for Biochemistry and Biotechnology*; Springer-Verlag: Berlin, **1986**.
- [163] F. J. Millero, A. L. Surodo, C. Shin, *J. Phys. Chem.* **1978**, *82*, 784.
- [164] T. H. Lilley, In *Chemistry and Biochemistry of the Amino Acids*, G. C. Barrett, Ed.; Chapman and Hall: London, **1985**; p. 591.
- [165] E. R. Collantes, W. J. I. Dunn, *J. Med. Chem.* **1995**, *38*, 2705.
- [166] A. Lee, T. V. Chalikian, *Biophys. Chem.* **2001**, *92*, 209.
- [167] C. F. Matta, In *Quantum Biochemistry: Electronic Structure and Biological Activity*, Vol 1, C. F. Matta, Ed.; Wiley-VCH: Weinheim, **2010**; p. 423.
- [168] A. Martin, J. Swarbrick, A. Cammarata, *Physical Pharmacy: Physical Chemical Principles in the Pharmaceutical Sciences*, 3rd ed.; Lea & Febiger: Philadelphia, **1983**.
- [169] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. Feeney, *J. Adv. Drug Delivery Rev.* **1997**, *23*, 3.
- [170] G. L. Patrick, *An Introduction to Medicinal Chemistry*; Oxford University Press: Oxford, **2005**.
- [171] T. E. Creighton, *Proteins: Structures and Molecular Principles*; W. H. Freeman and Co.: New York, **1983**.
- [172] R. Wolfenden, L. Andersson, P. M. Cullis, C. C. B. Southgate, *Biochemistry* **1981**, *20*, 849.
- [173] C. Alff-Steinberger, *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 584.
- [174] A. Radzicka, R. Wolfenden, *Biochemistry* **1988**, *27*, 1664.
- [175] E. I. Sánchez-Floresy, R. Chávez-Calviloz, T. A. Keith, G. Cuevas, T. Rocha-Rinzay, F. Cortés-Guzmán, *J. Comput. Chem.* **2014**, *35*, 820.
- [176] M. J. Clarke, F. Zhu, D. R. Frasca, *Chem. Rev.* **1999**, *99*, 2511.
- [177] R. Galindo-Murillo, J. Hernandez-Lima, M. Gonzalez-Rendon, F. Cortés-Guzmán, L. Ruiz-Azuarac, R. Moreno-Esparza, *Phys. Chem. Chem. Phys.* **2011**, *13*, 14510.
- [178] O. A. Zhikol, O. Shishkin, K. A. Lyssenko, J. Leszczynski, *J. Chem. Phys.* **2005**, *122*, 144104.
- [179] M. P. Waller, A. Robertazzi, J. A. Platts, D. E. Hibbs, P. A. Williams, *J. Comput. Chem.* **2006**, *27*, 491.
- [180] R. Parthasarathi, V. Subramanian, *Struct. Chem.* **2005**, *16*, 243.
- [181] R. F. W. Bader, A. Streitwieser, A. Neuhaus, K. E. Laidig, P. Speers, *J. Am. Chem. Soc.* **1996**, *118*, 4959.
- [182] G. Merino, A. Vela, T. Heine, *Chem. Rev.* **2005**, *105*, 3812.
- [183] E. Matito, M. Duran, M. Solà, *J. Chem. Edu.* **2006**, *83*, 1243.
- [184] J. Poater, M. Solà, M. Duran, X. Fradera, *Theor. Chem. Acc.* **2002**, *107*, 362.
- [185] C. F. Matta, J. Hernández-Trujillo, R. F. W. Bader, *J. Phys. Chem. A* **2002**, *106*, 7369.

- [186] Y.-G. Wang, C. F. Matta, N. H. Werstuijk, *J. Comput. Chem.* **2003**, *24*, 1720.
- [187] N. Castillo, C. F. Matta, R. J. Boyd, *J. Chem. Inf. Mod.* **2005**, *45*, 354.
- [188] D. Ferro-Costas, A. Vila, R. A. Mosquera, *J. Phys. Chem. A* **2013**, *117*, 1641.
- [189] D. Bonchev, D. H. Rouvray, Eds. *Complexity in Chemistry: Introduction and Fundamentals*; Taylor and Francis: London, **2003**.
- [190] M. D. Diudea, I. Gutman, J. Lorentz, *Molecular Topology*; Nova Science Publishers: Hauppauge, **1999**.
- [191] D. Bonchev, D. H. Rouvray, *Chemical Graph Theory: Introduction and Fundamentals*; OPA: Amsterdam, **1991**.
- [192] M. Randić, In *Concepts and Applications of Molecular Similarity*; M. A. Johnson, G. M. Maggiora, Eds.; Wiley: New York, **1990**, p. 77.
- [193] I. S. Dmitriev, *Molecules without Chemical Bonds (English Translation)*; Mir Publishers: Moscow, **1981**.
- [194] K. Balasubramanian, *SAR & QSAR Environ. Res.* **1994**, *2*, 59.
- [195] C. F. Matta, AIMDELOC: Program to calculate AIM localization and delocalization indices (Quantum Chemistry Program Exchange QCPE 0802), Indiana University, Indiana, **2001**.
- [196] I. Cukrowski, C. F. Matta, *Chem. Phys. Lett.* **2010**, *499*, 66.
- [197] D. White, R. C. Wilson, Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), University of South Florida and IEEE Computer Society, Tampa, Florida, **2008**, p. 1.
- [198] M. Muskulus, *Distance-Based Analysis of Dynamical Systems and Time Series by Optimal Transport*, PhD Thesis. Leiden University: Leiden, The Netherlands, **2010**.
- [199] C. Hansch, A. Albert Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, **1979**.
- [200] D. R. Lide, *CRC Handbook of Chemistry and Physics*, 88th ed.; CRC Press, **2007-2008**.
- [201] J. S. Murray, P. Politzer, In *Quantum Medicinal Chemistry*; P. Carloni, F. Alber, Eds.; Wiley-VCH: Weinheim, **2003**; p. 233.
- [202] E. B. Wilson, *J. Chem. Phys.* **1962**, *36*, 2232.
- [203] P. Politzer, R. G. Parr, *J. Chem. Phys.* **1974**, *61*, 4258.
- [204] P. Politzer, J. S. Murray, *Theor. Chem. Acc.* **2002**, *108*, 134.
- [205] H. Weinstein, P. Politzer, S. Srebrenik, *Theor. Chim. Acta* **1975**, *38*, 159.
- [206] R. K. Pathak, S. R. Gadre, *J. Chem. Phys.* **1990**, *93*, 1770.
- [207] P. K. Weiner, R. Langridge, J. M. Blaney, R. Schaefer, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3754.
- [208] S. D. Kahn, C. F. Pau, W. Hehre, *J. Int. J. Quantum Chem. (Suppl. Proc. Int. Symp. Quantum Chem. Solid-State Theor. Comput. Methods)* **1988**, *34* (Issue Supplement 22), 575.
- [209] P. Politzer, *Toxicol. Lett.* **1988**, *43*, 257.
- [210] T. Ghafourian, J. C. Dearden, *Il Farmaco* **2004**, *59*, 473.
- [211] M. H. Abraham, A. Ibrahim, A. M. Zissimos, Y. H. Zhao, J. Comer, D. P. Reynolds, *Drug Discov. Today* **2002**, *7*, 1056.
- [212] B. Fournier, E.-E. Bendeif, B. Guillot, A. Podjarny, C. Lecomte, C. Jelsch, *J. Am. Chem. Soc.* **2009**, *131*, 10929.
- [213] D. Liebschner, M. Elias, S. Moniot, B. Fournier, K. Scott, C. Jelsch, B. Guillot, C. Lecomte, E. Chabriere, *J. Am. Chem. Soc.* **2009**, *131*, 7879.
- [214] C. Lecomte, B. Guillot, C. Jelsch, A. Podjarny, *Int. J. Quantum Chem.* **2005**, *101*, 624.
- [215] G. A. Patani, E. J. LaVoie, *Chem. Rev.* **1996**, *96*, 3147.
- [216] D. A. Williams, T. L. Lemke, *Foye's Principles of Medicinal Chemistry*. Lippincott Williams & Wilkins, **2001**.
- [217] C. Biot, H. Bauer, R. H. Schirmer, E. Davioud-Charvet, *J. Med. Chem.* **2004**, *47*, 5972.
- [218] R. Herr, *J. Bioorg. Med. Chem.* **2002**, *10*, 3379.
- [219] C. F. Matta, A. A. Arabi, D. F. Weaver, *Eur. J. Med. Chem.* **2010**, *45*, 1868.
- [220] S. Franzen, R. F. Goldstein, S. G. Boxer, *J. Phys. Chem.* **1990**, *94*, 5135.
- [221] A. Gopher, Y. Blatt, M. Schonfeld, M. Y. Okamura, G. Feher, M. Montal, *Biophys. J.* **1985**, *48*, 311.
- [222] Z. D. Popovic, G. J. Kovacs, P. S. Vincett, G. Alegria, P. L. Dutton, *Chem. Phys.* **1986**, *110*, 227.
- [223] P. M. Moroney, T. A. Scholes, P. C. Hinkle, *Biochemistry* **1984**, *23*, 4991.
- [224] A. A. Arabi, C. F. Matta, *Phys. Chem. Chem. Phys. (PCCP)* **2011**, *13*, 13738.
- [225] H. Lehle, J. M. Kriegl, K. Neienhaus, P. Deng, S. Fengler, G. U. Nienhaus, *Biophys. J.* **2005**, *88*, 1978.
- [226] I. T. Suydam, C. D. Snow, V. S. Pande, S. G. Boxer, *Science* **2006**, *313*, 200 (erratum 29 September 2006).
- [227] L. Xu, A. E. Cohen, S. G. Boxer, *Biochemistry* **2011**, *50*, 8311.
- [228] J.-H. Choi, K.-I. Oh, H. Lee, C. Lee, M. Cho, *J. Chem. Phys.* **2008**, *128*, 134506.
- [229] X. Wang, X. He, J. Z. H. Zhang, *J. Phys. Chem. A* **2013**, *117*, 6015.
- [230] D. J. Sandberg, A. N. Rudnitskaya, J. A. Gascón, *J. Chem. Theory Comput.* **2012**, *8*, 2817.
- [231] S. Sowlati-Hashjin, C. F. Matta, *J. Chem. Phys.* **2013**, *139*, 144101.
- [232] S. Shaik, S. P. de Visser, D. Kumar, *J. Am. Chem. Soc.* **2004**, *126*, 11746.
- [233] N. K. V. Monteiro, C. L. Firme, *J. Phys. Chem. A* **2014**, *118*, 1730.
- [234] C. F. Matta, *Applications of the Quantum Theory of Atoms in Molecules to Chemical and Biochemical Problems*; Ph.D. Thesis; McMaster University: Hamilton, Canada, **2002**.
- [235] C. F. Matta, J. Hernández-Trujillo, T. H. Tang, R. F. W. Bader, *Chem. Eur. J.* **2003**, *9*, 1940.
- [236] J. Hernández-Trujillo, C. F. Matta, *Struct. Chem.* **2007**, *18*, 849.

Received: 23 December 2013  
Revised: 16 March 2014  
Accepted: 21 March 2014  
Published online on 29 April 2014