For reprint orders, please contact reprints@future-science.com

Localization-delocalization matrices and electron density-weighted adjacency matrices: new electronic fingerprinting tools for medicinal computational chemistry

“...extracting molecular descriptors from measurable 3D scalar fields such as the electron density $[\rho(r)]$ and the electrostatic potential $[V(r)]$, both fundamentally and uniquely related to the quantum state of the system, can yield in addition to robust QSAR models invaluable insight into the mode of action of drugs and their interaction with their receptors.”

Keywords: electronic fingerprinting • medicinal chemistry • quantitative-structure–activity relationship modeling

Quantitative-structure-activity-relationships (QSAR) are essential in the search and prediction of the activities of compounds with tailored pharmacological activities [1–4]. The structural descriptors incorporated into a QSAR model may be experimental, calculated (quantum chemically), connectivity descriptors, etc. A common assumption in QSAR is:

$$\text{biological activity} = f(x) = \frac{1}{\tilde{e} \cdot \tilde{g}} \cdot \left(\text{constant} + a_{1,1}x_1^1 + a_{1,2}x_1^2 + \dots + a_{1,m}x_1^m + \dots + a_{2,1}x_2^1 + a_{2,2}x_2^2 + \dots + a_{2,m}x_2^m + \dots + a_{n,1}x_n^1 + a_{n,2}x_n^2 + \dots + a_{n,m}x_n^m \right)$$

Equation 1

where the biological activity is expressed as the reciprocal of the concentration that elicit a given biological, pharmacological or toxicological endpoint such as ED_{50} (effective dose 50%), IC_{50} (inhibitory concentration 50%) or LD_{50} (lethal dose 50%), and where x_i is the i^{th} predictor (from a set of n predictors) raised to the powers from 1 to m and weighted by its coefficients a_{ij} obtained from a multivariate statistical fitting [4,5]. Usually, most of the coefficients a_{ij} can be zero, leading to much simpler looking QSAR models than the general form in Equation 1, as in the example

below. Often, other functions of x_i are used in the QSAR model such as the logarithmic or exponential function instead of a simple power expansion as in Equation 1.

As a simple illustration of the operation of Equation 1 is the following example of a QSAR model to predict the inhibitory effect of 3-nitroflavones (3NF) based on a single predictor, x (the energy of the highest occupied molecular orbital, E_{HOMO}) raised to the power 1 (see also [6]). Invasive tumor growth and metastasis are sustained by angiogenesis (the formation of new blood vessels [7,8]), which led Folkman to suggest anticancer therapies by the use of antiangiogenic agents [9,10]. Lichtenberg *et al.* have shown that the antiangiogenicity of six 3NF is correlated with the energies of the highest occupied molecular orbital, E_{HOMO} , the lower the oxidation potential the greater the inhibitory effects on several biomarkers of angiogenesis [11,12]. These authors report the following correlation between the E_{HOMO} and the IC_{50} of six 3NF for inhibition of proliferation of microvascular endothelial HMEC-1 cells after 72 h of incubation:

$$IC_{50} (\mu\text{M}) = -1487 - 163.6 E_{\text{HOMO}} \\ [n = 6, r = 0.93].$$

Equation 2

Chérif F. Matta

Department of Chemistry & Physics,
Mount Saint Vincent University, Halifax,
Nova Scotia, B3M2J6 Canada
and

Department of Chemistry, Dalhousie
University, Halifax, Nova Scotia,
B3H4J3 Canada
and

Department of Chemistry, Saint Mary's
University, Halifax, Nova Scotia, B3H3C3
Canada

Tel.: +1 902 457 6142

Fax: +1 902 457 6134

cherif.matta@msvu.ca

FUTURE
SCIENCE

part of

fsg

between the E_{HOMO} and the IC_{50} for inhibition of endothelial migration:

$$\text{IC}_{50} (\mu\text{M}) = -3319 + 357 E_{\text{HOMO}} \\ [n = 6, r = 0.95],$$

Equation 3

and between the E_{HOMO} and the relative% of inhibition of activity of metalloprotease MMP-2 expressed in angiogenesis observed at a concentration of 50 μM of the six 3NFs:

$$\% \text{inhibition} = 619 + 59.1 E_{\text{HOMO}} \\ [n = 6, r = 0.96].$$

Equation 4

As seen from the above and many more examples [6], and generally, the statistical correlations, invaluable as they may be in designing new more active and less toxic compounds, do not necessarily lead to a clear cause–effect relation between the predictor(s) and the predicted activity. What E_{HOMO} has to do with these three different facets of biological activity of 3NFs is unclear. In general, what counts first and foremost is the predictive strength of the QSAR

model whether or not it sheds light on the mechanism of action.

It has recently been emphasized that extracting molecular descriptors from measurable 3D scalar fields such as the electron density $[\rho(\mathbf{r})]$ and the electrostatic potential $[V(\mathbf{r})]$, both fundamentally and uniquely related to the quantum state of the system, can yield in addition to robust QSAR models invaluable insight into the mode of action of drugs and their interactions with their receptors [13–17].

The electron density exhibits a characteristic topography that engenders a natural partitioning of the molecular 3D space into nonoverlapping atomic basins according to Bader's quantum theory of atoms in molecules [14–15,18–20]. This partitioning leads to the possibility of defining bond paths, a definition that brings concepts such as chemical bonding and structural stability into coincidence within real 3D space. Figure 1 depicts the electron density in a chosen plane of the acetic acid molecule, a small molecule with a broad spectrum of biological and pharmacological activities [21]. The figure depicts the electron isodensity contours and their associated gradient vector field lines which point in the direction of steepest ascent in the density field. Each group of gradient vector field lines terminate at one and only one nucleus effectively associating that nucleus with a given region of electron density called atomic basin, which together with the nucleus constitute an atom in a molecule defined as a real bounded 3D object [22]. These atomic basins (Ω), depicted in Figure 1 and demarcated from one another by the thick partitioning lines, give rise to an entire quantum theory of subsystems, namely, quantum theory of atoms in molecules (for an introduction see Ref. [14]). The double integral of the Fermi hole density over a single atomic basin or over two basins yields, respectively, the number of electrons localized within a basin or shared/delocalized between two given basins, or the so-called localization and delocalization indices (LIs and DIs) (briefly discussed in Ref. [14]). These indices, when arranged in matrix format, can be treated by the tools of chemical graph theory [3,23–28] to extract graph invariants for use into QSAR modeling ([13,29]).

The LIs and DIs/2 can be organized as a matrix termed a localization-delocalization matrix (LDM, or the ζ -matrix), which has been shown promising as an electronic fingerprinting tool [13]. The LDM matrices of different molecules can be compared and the distances between these matrices defined and used as a measure of their dissimilarity. The intermolecular dissimilarity distances provide predictive statistical models to predict $\log P$'s, pK_a 's and UV λ_{max} 's with r^2 in excess of 0.97 outperforming even Hammett constants modeling ([13,29]).

Figure 1. Contour map of the electron density $\rho(\mathbf{r})$ in the H–C–C(O)–OH plane of acetic acid. The outermost contour has the isodensity value of 0.001 au followed by 2×10^n , 4×10^n , and 8×10^n atomic units with n starting at -3 and increasing in steps of unity. The lines connecting the nuclei are the bond paths, and the lines delimiting atoms are the intersection of the interatomic surfaces with the plane of the figure and which define the atomic basins of the atoms in the molecule. The intersection of a bond path with an interatomic surface occurs at the bond critical point, BCP, where $\rho(\mathbf{r}) = 0$ indicated by the small red dot. The faint lines terminating at the nuclei are the gradient vector field lines associated with the electron density and which partition the space into separate atoms in the molecule.

Let's briefly discuss an illustrative example. The experimental pK_a 's of a series of seven halogenated acetic acids can be modeled with the distances of their diagonal-suppressed matrices leading to the model:

$$pK_a = -0.588 + 5.415 \exp(-5.066 d_{\text{deloc}}),$$

Equation 5

with $r^2 = 0.979$ [13]. The model fits the experimental data extremely well then. Now, what insight can we gain into the relationship of structure (as captured by the LDM)-to-activity (pK_a)? Let us examine the LDMs of the two acetic acids with extremes of activity (pK_a 's) of the set of seven acetic acids: Unsubstituted acetic acid (AA) which has a $pK_a = 4.76$ and trifluoroacetic acid (TFAA), the pK_a of which is 0.52.

Figure 2. The localization-delocalization matrix by example. (A) The general form of the localization-delocalization matrix (LDM, or ζ -matrix) where the diagonal elements are the localization indexes (Λ) and the off-diagonals 1/2 of the delocalization index δ between atom Ω_1 and atom Ω_2 . The sum of the i^{th} row or column equal the number of electrons associated with the atom owning this row or column $N(\Omega_i)$, and the sum of all these sums is just N the total number of electrons in the molecule. (See Ref. [13] for details). **(B)** The atom numbering scheme of acetic acid [$X = H$ **(C)** and $X = F$ **(D)**]. **(C)** The LDM of unsubstituted acetic acid (AA). **(D)** The LDM of trifluoroacetic acid (TFAA).

With the numbering scheme for acetic acid in Figure 2B (X = H, F or Cl), an LDM for AA is given in Figure 2C and a corresponding LDM for TFAA is given in Figure 2D. In Figure 2C & D, a column sum (or row sum, not shown) is the atomic electron populations $N(\Omega)$ of atom Ω which when subtracted from the atomic number Z_Ω yield the atomic charges $q(\Omega)$ expressed in atomic units [$q(\Omega) = Z_\Omega - N(\Omega)$]. It can be easily verified that the sum of all atomic electron populations (sum of column or row sums) is 32, the total number of electrons in AA from Figure 2C, and 56, the total number of electrons in TFAA from Figure 2D.

Let us focus on the last column (or last row) of the acidic hydrogen (H8) of AA (Figure 2C), since this hydrogen is responsible for the acidity of the molecule in aqueous medium through its dissociation as a proton. First, the sum of all DI/2 and the LI for this atom yield only 0.42 electrons associated with the basin of this atom, clearly the only significantly positively charged hydrogen (protonic) atom in the molecule with a net positive charge of +0.58. Second, electrons in the basin of H8 are significantly shared only with the attached hydroxyl oxygen atom (O3) as can be seen from the entry at the intersection of their respective rows and columns. The total number of electrons shared between O3 and H8 is the DI and is equal to $2 \times 0.321 = 0.642$.

TFAA is far more acidic than AA with an acid dissociation constant K_a approximately 1.7×10^4 larger than that of AA. A comparison of the LDM of TFAA with that of AA in Figure 2D & C, respectively, reveals that all matrix elements are different. We note particularly: each of the three fluorine atoms carry a substantial negative charge close to approximately -0.6 instead of the quasi-neutral methyl hydrogen atoms in AA; as a result of this flow of electronic charge to the terminal F_3C- group the total charge of the carboxylic group in TFAA [$q(\text{TFAA-COOH}) = q(\text{C2}) + q(\text{O4}) + q(\text{O3}) + q(\text{H8}) = +0.067$] is 0.215 electronic charge less negative than this group in AA where $q(\text{AA-COOH}) = -0.148$. The COOH group in TFAA is, thus, far more capable of accommodating electronic charge than the already negatively charged carboxylic group of the unsubstituted compound explaining its far greater readiness to donate its H8 as a proton and accommodate the resulting excess electron on its carboxylate group; the acidic hydrogen, H8, is more positive (acidic) in TFAC [$q(\text{TFAA-H8}) = +0.601$] than in AA [$q(\text{AA-H8})$

$= +0.580$] by 0.021 electronic charge; and finally, the electrons in the basin of H8 are less shared with O3 ($2 \times 0.309 = 0.618$) in TFAA than in AA where the DI amounts to 0.642 indicating a weaker more readily severed bond consistent with its considerably higher acidity.

From this example, LDM appears to be promising in QSAR-type research with a similar robustness as traditional or *ad hoc* descriptors but with valuable added insight into the observed behavior ([13,29]).

An alternative and simpler electronic fingerprinting is the bond critical point matrix termed the electron density weighted connectivity matrix (EDWCM) [MASSA L, PERS. COMM. (2014)]. This matrix consists of organizing in matrix format the values of $\rho(\mathbf{r})$ at the bond critical points (the small red spheres in Figure 1). The matrix, thus, has entries only at the intersection of columns and rows of atoms that are bonded one to another and the value of the density at the bond critical point which reflects the strength of the given chemical bond. All other entries in an EDWCM are exactly zero, unlike the LDM where all matrix elements have a value (even if small) since electrons are delocalized between any two atoms in a molecule whether or not they are chemically bonded. The EDWCM may be less rich in electronic information than the LDMs but it is: accessible from both theory and X-ray diffraction experiments; and cheaper to obtain since $\rho(\mathbf{r})$ is only sampled at the positions of the critical points (no space integrations involved).

These new and exciting developments have not yet been directly applied to pharmacological properties, but the first steps in the modeling of physicochemical properties are promising. Extensively, explorations are necessary and are currently being undertaken to establish the utility of these novel electronic fingerprinting tools in medicinal chemistry, their domain of applicability and their limitations.

Financial & competing interests disclosure

Funding for this work was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI) and Mount Saint Vincent University. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

- Hansch C, Leo A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society, WA, USA (1995).
- Hansch C. A quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* 2, 232–239 (1969).
- Todeschini R, Consonni V. *Molecular Descriptors for Chemoinformatics (2nd Edition)* (Vols. I and II). Wiley, NJ, USA (2009).
- Mager PP. *Multidimensional Pharmacochemistry: Design of Safer Drugs*. Academic Press, Inc., London, UK (1984).
- Richards WG. *Quantum Pharmacology (2nd Edition)*. Butterworth & Co., Ltd., London, UK (1983).
- Bensasson RV, Sowlati-Hashjin S, Zoote V, Dauzonne D, Matta CF. Physicochemical properties of exogenous molecules correlated with their biological efficacy as protectors against carcinogenesis and inflammation. *Int. Rev. Phys. Chem.* 32, 393–434 (2013).
- Folkman J. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* 285, 1182–1186 (1971).
- Folkman J, Klagsbrun M. Angiogenic factors. *Science* 235, 442–447 (1987).
- Folkman J. Fighting cancer by attacking its blood supply. *Sci. Amer.* 275, 150–154 (1996).
- Folkman J. Role of angiogenesis in tumor growth and metastasis. *Semin. Oncol.* 29, 15–18 (2002).
- Lichtenberg M. *Diplomarbeit. Investigation of anti-angiogenic effects displayed by 3-nitroflavone and 3-nitroflavone 8-acetic acid derivatives*. Fakultät für Biowissenschaften der Ruprecht-Karls-Universität, Heidelberg, Germany (2005).
- Lichtenberg M, Dauzonne D, Zoote V, Gerhäuser C, Bensasson RV. Quantitative structure activity relationships of 3-nitroflavones acting as potential inhibitors of angiogenesis, target of cancer inhibition. *Chem. Biol. Interact.* 161, 217 (2006).
- Matta CF. Modeling biophysical and biological properties from the characteristics of the molecular electron density, electron localization and delocalization matrices, and the electrostatic potential. *J. Comput. Chem.* 35(16), 1165–1198 (2014).
- Matta CF, Arabi AA. Electron-density descriptors as predictors in quantitative structure activity/property relationships and drug design. *Future Med. Chem.* 3(8), 969–994 (2011).
- Matta CF, Massa L. Subsystem quantum mechanics and *in-silico* medicinal and biological chemistry. *Future Med. Chem.* 3(16), 1971–1974 (2011).
- Gubskaya AV. Quantum-chemical descriptors in QSAR/QSOR modeling: achievements, perspectives and trends. In: *Quantum Biochemistry: Electronic Structure and Biological Activity*. Matta CF (Ed.) Wiley-VCH, Weinheim, Germany, 693–721 (2010).
- Matta CF, Arabi AA, Weaver DF. The bioisosteric similarity of the tetrazole and carboxylate anions: clues from the topologies of the electrostatic potential and of the electron density. *Eur. J. Med. Chem.* 45(5), 1868–1872 (2010).
- Bader RFW. *Atoms in Molecules: A Quantum Theory*. Oxford University Press, Oxford, UK (1990).
- Popelier PLA. *Atoms in Molecules: An Introduction*. Prentice Hall, London, UK (2000).
- Matta, CF, Boyd, RJ (Eds). *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*. Wiley-VCH, Weinheim, Germany (2007).
- Johnston CS, Gaas CA. Vinegar: medicinal uses and antiglycemic effect. *Medscape J. Med.* 8(2), 61–79 (2006).
- Bader RFW, Matta CF. Atoms in molecules as non-overlapping, bounded, space-filling open quantum systems. *Found. Chem.* 15, 253–276 (2013).
- Balaban AT. Applications of graph theory in chemistry. *J. Chem. Inf. Comput. Sci.* 25(3), 334–343 (1985).
- Balaban AT. *Chemical Applications of Graph Theory*. Academic Press, NY, USA (1976).
- Diudea MD, Gutman I, Lorentz J. *Molecular Topology*. Nova Science Publishers, Inc., NY, USA (1999).
- Hall LH, Kier LB. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, MA, USA (1976).
- Dmitriev IS. *Molecules without Chemical Bonds (English Translation)*. Mir Publishers, Moscow, Russia (1981).
- Janežić D, Milicević A, Nikolić S, Trinajstić N. *Graph Theoretical Matrices in Chemistry, Mathematical Chemistry Monographs* (Vol. 3). University of Kragujevac, Kragujevac, Serbia (2007).
- Sumar I, Ayers PW, Matta CF. Electron localization and delocalization matrices in the prediction of pKa's and UV-wavelengths of maximum absorbance of p-benzoic acids and the definition of super-atoms in molecules. *Chem. Phys. Lett.* 612, 190–197 (2014).