

Chemical Applications of Electron Localization-Delocalization Matrices (LDMs)
with an emphasis on predicting molecular properties

By
Ismat Sumar

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
for the Degree of Master of Science in Applied Science

September 12, 2016

©Ismat Sumar, 2016

Approved: Dr. Chérif Matta
Co-Supervisor

Approved: Dr. Paul Ayers
Co-Supervisor

Approved: Dr. Cory Pye
Committee Member

Approved: Dr. Zhongmin Dong
Committee Member

Approved: Dr. Erin Johnson
Examiner

Date: September 12, 2016

Chemical Applications of Electron Localization-Delocalization Matrices (LDMs)
with an emphasis on predicting molecular properties

by Ismat Sumar

Abstract

A matrix is constructed where the vertices (atoms) are connected by edges (bonds) resulting in a square matrix that is symmetrical. The localization index (unshared electrons) occupies the long diagonal where the delocalization index (shared electrons between two different atoms divided by 2) represent the off-diagonal elements. Such a matrix is called a localization-delocalization matrix or LDM. These matrices have shown promise as a novel Quantitative Structure Activity Relationship (QSAR) method via the Frobenius Distance, a method to compare matrices of similar sizes that returns a Euclidean distance. Some notable results that will be expanded upon are that for a series of 14 para-substituted benzoic acids for pKa prediction ($r^2 = 0.986$), and a series of 13 polycyclic benzenoid hydrocarbons (PBH) separated by inner and outer rings ($r^2 = 0.97$). A program (AIMLDM) was developed in Python 3.4.1 to construct these matrices and perform the required calculations.

September 12, 2016

Acknowledgements

First and foremost I thank my two supervisors and friends Dr. Chérif F. Matta and Dr. Paul W. Ayers for giving me the opportunity to work with them. They provided guidance when I was first trying to come up with the AIMLDM program. Dr. Matta made a script for the first iteration of the program, and Dr. Ayers helped me to understand how to move the whole program to Python instead of going back and forth between Shell Scripts, Python, and Excel. They filled the gaps in my knowledge whether it was taking a class with Dr. Matta or visiting Dr. Ayers at McMaster. They were very patient with me when I was not ready to defend on our initial date, and were very supportive throughout the Masters. Most importantly I thank them for the friendship they extended towards me which made this Masters a very enjoyable experience.

I thank my committee members Dr. Cory Pye and Dr. Zhongmin Dong for their helpful discussions and suggestions during our meetings; it was Dr. Pye who first introduced us to the idea of pruning matrices. I give a special thanks to Dr. Ronald Cook who showed us new ways in which to use the LDMs and for his invaluable insight for helping test and further develop AIMLDM.

I thank Dr. Erin Johnson for her patience and agreeing to be my external examiner. I thank Saint Mary's University for accepting me and allowing me to TA, Mount Saint Vincent University for allowing me to use their facilities, and McMaster University.

Table of Contents

Abstract	i
Acknowledgements	ii
Table of Contents	iii
List of Figures	vi
List of Tables	ix
1 Introduction to LDMs	1
1.1 Introduction	2
1.2 Definition of the LDM	3
1.3 Application of LDMs	9
1.4 Limitations of LDMs	10
1.4.1 Ambiguity of Atomic Labelling.	10
1.4.2 Matrices of Different Size	12
1.4.3 Other Limitations of LDMs	17
1.5 EDWCM	18
1.6 Acknowledgements	20
1.7 References	20
2 AIMLDM	24
2.1 The AIMLDM Programme	24
2.2 Numerical Illustrative Testing	28
2.3 QSAR Studies with LDMs	34
2.3.1 LDM-Eigenvalues as Predictors in QSAR	35
2.4 Conclusion	40

2.5	Acknowledgements	42
2.6	References	43
3	Predicting pK_a and λ_{max}	47
3.1	Modelling of pK_a	49
3.2	Modelling of λ_{max}	52
3.3	Conclusion	57
3.4	Computational Methods	58
3.5	Acknowledgements	58
3.6	References	59
4	Aromaticity from LDMs	62
4.1	Rings-in-molecules (RIMs)	63
4.2	The molecular set	66
4.3	Computational details	67
4.4	Aromaticity measures and Eigenvalues	68
4.4.1	Definitions of the measures of aromaticity considered in this work	69
4.4.2	Correlations of the Frobenius distances from benzene with es- tablished measures of aromaticity	71
4.4.3	Correlations of aromaticity with the eigenvalues of the RIM-LDM	74
4.5	Conclusion	76
4.6	Acknowledgements	77
4.7	References	78
5	Eigenvalues and Important Atoms in a Molecule	87
5.1	A look at the Eigenvalues	87

5.2 Important Atoms	89
Appendices	98
A Operating Instructions for AIMLDM	99
B Benzoic Acid Structures with their LDMs	106
C Frobenius Distance on a per-atom basis	120

List of Figures

1.1	Connectivity Matrices	3
1.2	Contours of Electron Density	6
1.3	Perturbed Group	14
2.1	A flowchart explaining the logic of operation of the AIMLDM programme.	26
2.2	Isoelectronic Series and corresponding energies	31
2.3	Isoelectronic Series Bond Lengths	32
2.4	Isoelectronic Series vs Various properties	33
2.5	Comparing Basis Sets vs LDMs	34
2.6	Shepard plot of the data in Table 2.3	40
2.7	2-d projection of 6-d eigenvalues	41
3.1	Numbering Scheme for matrices	48
3.2	Frobenius distances of partial (DMs) vs pK_a	51
3.3	λ_{max} vs Frobenius distances of COOH subgraph	56
4.1	Phenanthrene and its atom and ring labelling scheme.	64
4.2	Molecular set supplying the “rings-in-molecules (RIMs)” for this study.	67
4.3	Correlations between Frobenius distance (d_{Frob}) RIMs and common aromaticity measures	72
4.4	Two-dimensional MDS projection of the dissimilarity matrix in Table 4.3	76
4.5	LDM Eigenvalue distances vs aromaticity measures	77

List of Tables

2.1	Frobenius Distance for isoelectronic series	29
2.2	“Scrambled” LDM for acetic acid and resulting eigenvalues	36
2.3	Eigenvalues of LDMs from a series of carboxylic acids	37
3.1	Frobenius distances comparing the subgraph(s) to pK_a and λ_{max}	50
3.2	Frobenius distances from the DMs vs pK_a	52
3.3	Frobenius distances from the LDMs vs λ_{max}	56
4.1	RIMs Frobenius distance vs Popular Aromaticity measures	68
4.2	Aromatic ranking agreement of various aromaticity indices with the Frobenius distance dissimilarity to benzene.	73
4.3	Pairwise vector angles (in degrees ($^\circ$)) matrix for the ring in molecules to three decimals	75
5.1	Comparing Eigenvalues of $BACOCH_3$ with its atom electron popula- tions $N(\Omega_i)$	88
5.2	Comparing Eigenvalues of Benzene (carbon atoms only) with its atom electron population $N(\Omega_i)$ (values are arranged from smallest to largest)	89
5.3	Ranking of Frobenius distance of individual atoms with respect to pK_a (LI only)	91
5.4	Ranking of Frobenius distance of individual atoms with respect to pK_a (DI only)	92
5.5	Ranking of Frobenius distance of individual atoms with respect to λ_{max} (LI only)	93

5.6	Ranking of Frobenius distance of individual atoms with respect to λ_{max} (DI only)	94
5.7	Ranking of Frobenius distance of individual atoms with respect to HOMA (LI only)	95
5.8	Ranking of Frobenius distance of individual atoms with respect to HOMA (LI only, cyclohexane removed)	95
5.9	Ranking of Frobenius distance of individual atoms with respect to HOMA (DI only, cyclohexane removed)	96
C.1	Frobenius Distance between individual atoms for only the LI to be compared directly with pK_a , the most acidic molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)	120
C.2	Frobenius Distance between individual atoms for only the DI to be compared directly with pK_a , the most acidic molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)	121
C.3	Frobenius Distance between individual atoms for only the LI to be compared directly with λ_{max} , the lowest λ_{max} molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)	122
C.4	Frobenius Distance between individual atoms for only the DI to be compared directly with λ_{max} , the lowest λ_{max} molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)	122
C.5	Frobenius Distance between individual atoms for only the LI to be compared directly with the aromaticity measure HOMA, the highest HOMA value molecule's atoms are taken as the reference atoms. (Aro- matic Series)	123

C.6 Frobenius Distance between individual atoms for only the DI to be compared directly with the aromaticity measure HOMA, the highest HOMA value molecule's atoms are taken as the reference atoms. (Aromatic Series)	123
--	-----

Localization-Delocalization and Electron Density-Weighted Connectivity Matrices: A Bridge between the Quantum Theory of Atoms in Molecules and Chemical Graph Theory¹

“The development of chemistry has both led to, and been made possible by, the evolution of certain primary concepts. These concepts, without which there would be neither correlation nor prediction of the observations of descriptive chemistry, are: (1) the existence of atoms of functional groupings of atoms in molecules as evidenced by characteristic sets of properties; (2) the concept of bonding; and (3) the associated concepts of molecular structure and molecular shape. These concepts logically (but not historically) are consequences of fundamental topological properties of the charge distribution (electronic and nuclear) in a molecular system. In terms of the Born-Oppenheimer approximation the electronic distribution $\rho(r)$ is the scalar field defined in the real three-dimensional space with Euclidean metric. The universal topological properties of $\rho(r)$ are characterized by its gradient field $\rho\nabla(r)$.”

I.S. Dmitriev (1981)

¹Based on the Chapter: Applications of Topological Methods in Molecular Chemistry Volume 22 of the series Challenges and Advances in Computational Chemistry and Physics pp 53–88

1.1 Introduction

A molecule can be abstracted as a network of points (vertices) connected by lines (edges) and hence constituting a graph. Molecular graphs formed from a set of edges each consisting of what chemists normally call a “chemical bond” can be – but generally are not – *complete*. (A “complete graph” is one in which every vertex is connected by an edge, a trivial example being the graph of a diatomic molecule). A graph based on any pair-wise property such as inter-nuclear distance, nuclear-nuclear repulsion, or a count of electrons delocalized between any two pairs of atoms in the molecule is a complete graph.

Molecular graphs, complete or incomplete, can be conveniently represented by connectivity matrices as can be seen in the examples in Fig. 1.1 and in documents [1-9]. A complete graph where connectedness is indicated by 1 and disjointedness by 0 will have a non-zero entry for every non-diagonal element of the matrix while an incomplete graph has finite entries only for connected vertices and zero elsewhere in the matrix (Fig. 1.1).

A matrix representative of a complete graph with n vertices whereby connectivity is assigned “1” as in Fig. 1.1(a) is thus filled with ones except along the diagonal and hence has $n\frac{(n-1)}{2}$ edges, the number of non-diagonal elements of its matrix representative. In practice, a complete graph such as the delocalization matrix (DM), described below, may have zero (negligible) entries other than along the diagonal when the delocalization index between a given pair of atoms in a molecule has a magnitude below the precision to which the numerical entries are reported.

Within Richard F. W. Bader’s Quantum Theory of Atoms in Molecules (QTAIM) [10-12] a molecular graph is defined as the set of connected bond paths found in the

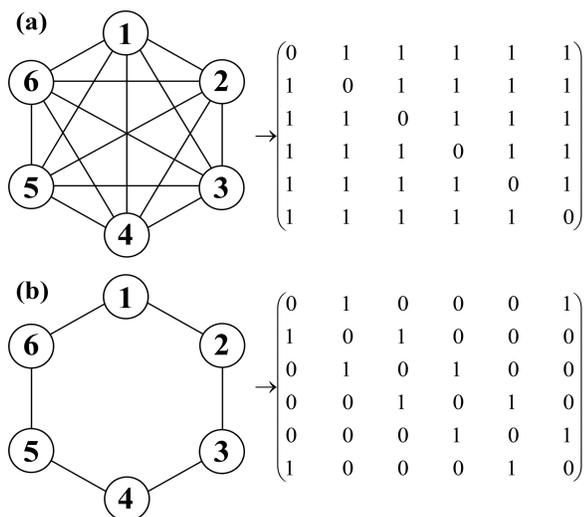


Figure 1.1: (a) An example of a complete graph with 6 vertices (K_6) with $6 - \frac{5}{2} = 15$ edges along with its matrix representative according to the numbering scheme. (b) An example of an incomplete graph with the same number of vertices and numbering scheme as in (a) along with its matrix representative.

molecular electron density. The molecular graph, so defined, is generally incomplete in the graph-theoretic sense since generally not every atom is sharing a bond path with every other atom in the molecule (except in diatomics and possibly a few other exceptions). The same theory, QTAIM, also defines delocalization indices (DIs), *vide infra*, that define a “complete graph” since there is a non-directed DI between every pair of atoms in the molecule whether sharing a bond path or not. As already mentioned, while in principle a DI graph is complete, in numerical practice it may not be so.

1.2 Definition of the LDM

Dmitriev, in his introductory book on Chemical Graph Theory (CGT), discusses the relation between molecular topology, graph theory, and what is known today

as QTAIM. The author outlines the topological underpinnings of QTAIM in the differential topology and topography of the electron density $\rho(r)$ culminating with the Poincaré -Hopf relationship relating the numbers and types of different critical points (CPs) in the electron density scalar field (points where the gradient of the electron density vanishes, that is, $\nabla\rho_{CP} = 0$).

QTAIM locates the various critical points in the density and uses each bond critical point (BCP) as a starting point for the search of the inter-atomic surfaces of zero-flux in the gradient vector field of the electron density separated and shared by pairs of bonded atoms. A BCP is also used in tracing its associated bond path which is a unique line of maximal electron density that links the nuclei of two bonded atoms [13-15] and which characterizes the nature and strength of chemical bonding [16].

The bond path is always found to be accompanied by a shadow graph, the virial path, first discovered by Keith *et al.* [17]. The virial path is a line of maximally-negative potential energy density in three-dimensional space that links the same pair of atoms that share a bond path and an interatomic surface of zero-flux. There is no mathematical proof that *requires* the presence of a virial path as a doppelgänger of every bond path that links two chemically bonded atoms, however, there is no known computational violation of this observation to date to the authors' best knowledge. The presence of the virial path links the concept of chemical bonding directly with the concept of energetic stability as amply discussed in literature on QTAIM.

The partitioning of the space into separate non-overlapping atomic basins, exhausting all three-dimensional space, entails the definition of "atomic properties" that add up to yield the corresponding molecular counterparts. Such atomic properties are obtained by integrating each corresponding property density over the bounded region of real space occupied by the atomic basin.

Fig. 1.2 shows the intersection of the atomic basins with the H-C-C(O)-OH plane in ethanoic (acetic) acid. The figure displays isodensity contours of the electron density, a representative set of gradient vector field lines traced by the gradient of the electron density, the intersections of interatomic surfaces (IASs) with the plane of the figure, the set of bond paths that are coplanar with the plane of the figure, and the bond critical points each of which lies simultaneously on the IAS and the associated bond path. For atoms exposed on the molecular surface (and hence that extend to infinity), the atomic basins are usually delimited by the intersections of their IASs with the outer isodensity contour of $\rho_{vdW} = 0.001$ atomic unit (a.u.), the van der Waals envelope (1 a.u. of electron density = 1 electron per cubic bohr).

As explained above, numerical integration (using readily available robust software such as Keith's AIMAll [18]) yields atomic quantum mechanical averages of properties such as atomic electron populations ($N(\Omega)$), number of electrons localized within the basin ($\Lambda(\Omega)$), and number of electrons delocalized (shared) between one atomic basin and every other basin in the molecule ($\delta(\Omega, \Omega')$).

The amount of electron delocalization (shared) between atomic basins Ω_i and Ω_j can be measured by the delocalization index (DI), $\delta(\Omega_i, \Omega_j)$. For a closed-shell molecule, the DI at the Hartree-Fock level of theory is defined [19]:

$$\delta(\Omega_i, \Omega_j) = 2|F^\alpha(\Omega_i, \Omega_j)| + 2|F^\beta(\Omega_i, \Omega_j)| \quad (1.1)$$

where

$$F^\sigma(\Omega_i, \Omega_j) = - \sum_k^{\text{occ}} \sum_l^{\text{occ}} \int_{\Omega_i} \int_{\Omega_j} \varphi_k^*(r_1) \varphi_l(r_1) \varphi_l^*(r_2) \varphi_k(r_2) dr_1 dr_2 \quad (1.2)$$

$$= - \sum_k^{\text{occ}} \sum_l^{\text{occ}} S_{kl}(\Omega_i) S_{lk}(\Omega_j) \quad (1.3)$$

is the Fermi correlation, and where $S_{kl}(\Omega_i) = S_{lk}(\Omega_i)$ is the overlap integral of two spin orbitals φ_k and φ_l within Ω_i , and where σ refers to spin (α or β). For single determinantal methods, the first-order density matrix—printed in standard electronic structure software—is sufficient to determine all properties. For post-Hartree-Fock methods, the Müller approximation is used by AIMAll, the software used to obtain the LIs and DIs, to obtain an approximate second-order density matrix from the first-order density matrix.

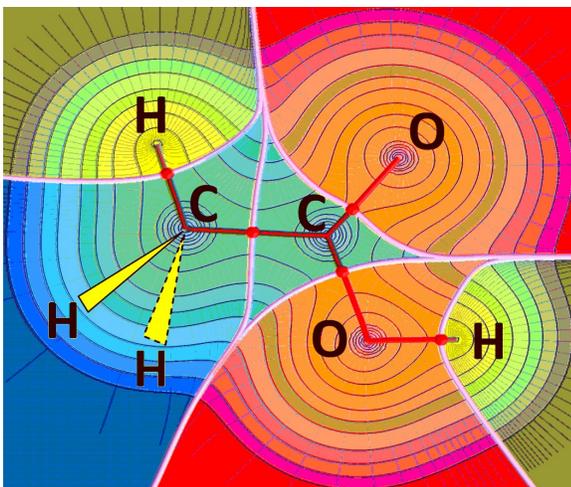


Figure 1.2: Contours of the electron density in the molecular plane of ethanoic (acetic) acid. The contours from outside inwards have the values (in atomic units (a.u.)): 0.001 a.u. and then 2×10^n , 4×10^n , and 8×10^n , n starting at -3 and increasing in steps of unity. Nuclei are linked by bond paths and atomic basins are separated by the intersections of the interatomic surfaces with the molecular plane, every atomic basin being distinguished by an element-specific dominant colour. Each BCP appears at the intersection of the associated bond path and interatomic surface and is depicted as a *small red dot*.

If $i = j$ in Eqs. 1.2-1.3, $(S_{kl}(\Omega_i)S_{lk}(\Omega_j) \rightarrow [S_{kl}(\Omega_i)]^2)$, then both integrals are over the same atomic basin giving the total Fermi correlation for the electrons contained within that basin. At the limit of total localization this double integral approaches $-N^\sigma(\Omega_i)$, the negative of the σ -spin population of Ω_i . This limit is reached only when atoms are infinitely separated since in any molecule electrons in a given atomic basin always exchange with electrons in every other atomic basin to some extent and $|F^\alpha(\Omega_i, \Omega_i)| \leq N^\alpha(\Omega_i)$. This localization index (LI) is thus defined [19]:

$$\Lambda(\Omega_i, \Omega_i) = |F^\alpha(\Omega_i, \Omega_i)| + |F^\beta(\Omega_i, \Omega_i)| \quad (1.4)$$

In a molecule, the electron population of an atom is always shared to some extent with other basins, *i.e.*, there always exists a degree of electron sharing or delocalization.

Since electrons can either be localized within a basin or shared with other basins in the molecule, then the LI of an atom plus half of the sum of its $(n - 1)$ DIs shared with the remaining atoms in the molecule (where n is the number of atoms in the molecule), must necessarily equal its electron population $N(\Omega_i)$ [19]:

$$N(\Omega_i) = \Lambda(\Omega_i) + \frac{1}{2} \sum_{j \neq i}^n \delta(\Omega_i, \Omega_j) = \int_{\Omega_i} \rho(r) dr \quad (1.5)$$

The population $N(\Omega_i)$ obtained via the bookkeeping of electrons' whereabouts embodied in the first equality of Eq. (1.5) or through the integration of the electron density over Ω_i (second equality of Eq. (1.5)) determines the atomic charge which, given the atomic number Z_{Ω_i} , is defined (in a.u.):

$$q(\Omega_i) = Z_{\Omega_i} - N(\Omega_i) \quad (1.6)$$

Since the total molecular electron population N is the sum of the atomic populations then it is expressible as the sum of two (sub-)populations: The molecular average number of localized electrons (N_{loc}) plus the molecular average of delocalized electrons (N_{deloc}) [20]:

$$N = \sum_{i=1}^n N(\Omega_i) = \sum_{i=1}^n \Lambda(\Omega_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j \neq 1}^n \delta(\Omega_i, \Omega_j) = N_{loc} + N_{deloc} \quad (1.7)$$

where

$$N_{loc} \equiv \sum_{i=1}^n \Lambda(\Omega_i) \quad (1.8)$$

and

$$N_{deloc} \equiv \frac{1}{2} \sum_{i=1}^n \sum_{j \neq 1}^n \delta(\Omega_i, \Omega_j) = N - tr(\zeta) = N - N_{loc} \quad (1.9)$$

Further, the full set of molecular LIs and DIs can be organized in a localization-delocalization matrix (LDM, or ζ -matrix) [20-24]:

$$\zeta \equiv \left[\begin{array}{cccc} \Lambda(\Omega_1) & \frac{\delta(\Omega_1, \Omega_2)}{2} & \dots & \frac{\delta(\Omega_1, \Omega_n)}{2} \\ \frac{\delta(\Omega_2, \Omega_1)}{2} & \Lambda(\Omega_2) & \dots & \frac{\delta(\Omega_2, \Omega_n)}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta(\Omega_n, \Omega_1)}{2} & \frac{\delta(\Omega_n, \Omega_2)}{2} & \dots & \Lambda(\Omega_n) \end{array} \right]_{n \times n} \left. \begin{array}{l} \sum_{row} = N(\Omega_1) \\ = N(\Omega_2) \\ \vdots \\ = N(\Omega_n) \end{array} \right\} \sum_{i=1}^n N(\Omega_i) = N \quad (1.10)$$

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

In the LDM, the sum of the matrix elements in any row or corresponding column equals the atomic population $N(\Omega_i)$ (by the first equality of Eq. 1.5) and hence the sum of the column sums or row sums equals the total molecular electron population. The trace of the LDM is the localized electron population (N_{loc}) of the molecule (Eq. 1.8), and the delocalized electron population can be obtained by (Eq. 1.9).

The LDM is a representation of a complete molecular graph where all atoms (vertices) are interconnected by non-directional DI links (edges), and where the diagonals are non-zero giving the number of electrons localized in a given atomic basin. This last point distinguishes the LDM graph from a typical “complete graph” of the type shown in Fig. 1.1(a) in that vertices are connected back to themselves through their respective LIs.

1.3 The LDM as a Molecular Fingerprinting and Similarity Assessment Tool

The distances between the localization-delocalization matrices (LDMs) of different molecules can be used as a measure of their dissimilarity. The greater or smaller the “distance” between two LDMs the lesser or more similar are the molecules they represent.

The inter-molecular distance between two molecules A and B, each represented by an $n \times n$ LDM, is defined as the Frobenius norm of the difference matrix, that is:

$$d(A, B) \equiv \|A - B\| \equiv \sqrt{\sum_{i,j} |\alpha_{ij} - \beta_{ij}|^2} \quad (1.11)$$

where α_{ij} and β_{ij} are two corresponding elements in the matrices \mathbf{A} and \mathbf{B} that

represent each molecule in the pair.

After the electronic structure calculation yields a wavefunction file, AIMAll/AIM-Studio program [18] is used to calculate the localization and delocalization indices. A Python program (AIMLDM), developed by Sumar *et al.* [25], extracts the localization and delocalization indices from AIMAll’s output and manipulates it to extract the matrix invariants as well as the Frobenius distances.

1.4 Limitations of LDMs, and possible Solutions

LDMs share well-known limitations with all matrix representatives of molecular graphs when used as a tool for comparing different molecules. These limitations are briefly outlined along with possible solutions.

1.4.1 Ambiguity of Atomic Labelling.

Any matrix representation of the molecular graph, complete or incomplete, is labelling-dependent since there exists $n!$ ways to label the n -atoms composing a given molecule. Unless all compared molecules have very similar graphs and can be given consistent atomic labelling, e.g. benzoic acids substituted, say, at the *para*-position by monoatomic substituents such as halogens, one must rely on “matrix invariants”.

Labelling-independent invariants extracted from a matrix representation of a molecular graph include, for example, the characteristic polynomial, the eigenvalues, the trace, and the determinant. LDMs, by being real and symmetric, are diagonalizable by a similarity transformation:

$$P^{-1}\zeta P = D \tag{1.12}$$

where \mathbf{D} is the diagonalized LDM. The eigenvalues can then be organized as a vector sorted in a consistent order of, say, increasing value.

For example, an LDM- ζ of methane is:

$$\zeta_{CH_4} = \begin{array}{c} C1 \\ H2 \\ H3 \\ H4 \\ H5 \\ \sum \end{array} \begin{array}{ccccc} C1 & H2 & H3 & H4 & H5 \\ \left(\begin{array}{ccccc} 4.040 & 0.492 & 0.492 & 0.492 & 0.492 \\ 0.492 & 0.444 & 0.021 & 0.021 & 0.021 \\ 0.492 & 0.021 & 0.444 & 0.021 & 0.021 \\ 0.492 & 0.021 & 0.021 & 0.444 & 0.021 \\ 0.492 & 0.021 & 0.021 & 0.021 & 0.444 \end{array} \right)_{5 \times 5} \\ \sum \end{array} \begin{array}{c} \sum \\ 6.007 \\ 0.998 \\ 0.998 \\ 0.998 \\ 0.998 \\ 10.000 \end{array} \quad (1.13)$$

of which the total number of localized electrons is given by its trace, $tr(\zeta_{CH_4}) = 5.815$, while its determinant $det(\zeta_{CH_4}) \approx 0.082$, and the corresponding \mathbf{D} written either as a matrix or a column vector is:

$$D_{CH_4} = \begin{array}{c} \left(\begin{array}{ccccc} 0.251 & 0 & 0 & 0 & 0 \\ 0 & 0.423 & 0 & 0 & 0 \\ 0 & 0 & 0.423 & 0 & 0 \\ 0 & 0 & 0 & 0.423 & 0 \\ 0 & 0 & 0 & 0 & 4.295 \end{array} \right) \equiv \begin{array}{c} \left[\begin{array}{c} 0.251 \\ 0.423 \\ 0.423 \\ 0.423 \\ 4.295 \end{array} \right]_{5 \times 1} \\ \sum = 5.815 \end{array} \end{array} \quad (1.14)$$

where the sum of the elements of \mathbf{D} represent the total number of localized electrons since the trace of a matrix is invariant upon diagonalization. The Frobenius distance can be calculated using \mathbf{D} without regard to the arbitrariness of the labelling scheme.

$$\mathbf{D}_{C_2H_6} = \begin{bmatrix} 0.284 \\ 0.323 \\ 0.430 \\ 0.430 \\ 0.440 \\ 0.440 \\ 3.638 \\ 4.632 \end{bmatrix}_{8 \times 1} \quad (1.16)$$

In order to compute the Frobenius distance between ethane and methane, we enlarge the matrix representative of methane with ghost atoms to:

$$\zeta_{CH_4} = \begin{matrix} & C1 & H2 & H3 & H4 & H5 & & & & \\ \begin{matrix} C1 \\ H2 \\ H3 \\ H4 \\ H5 \\ \\ \\ \\ \end{matrix} & \left(\begin{array}{ccccccccc} 4.040 & 0.492 & 0.492 & 0.492 & 0.492 & 0 & 0 & 0 \\ 0.492 & 0.444 & 0.021 & 0.021 & 0.021 & 0 & 0 & 0 \\ 0.492 & 0.021 & 0.444 & 0.021 & 0.021 & 0 & 0 & 0 \\ 0.492 & 0.021 & 0.021 & 0.444 & 0.021 & 0 & 0 & 0 \\ 0.492 & 0.021 & 0.021 & 0.021 & 0.444 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) & \rightarrow \mathbf{D}_{CH_4} = \begin{bmatrix} 0.000 \\ 0.000 \\ 0.000 \\ 0.251 \\ 0.423 \\ 0.423 \\ 0.423 \\ 4.259 \end{bmatrix}_{8 \times 1} \end{matrix} \quad (1.17)$$

which, in its \mathbf{D} -form, can now be compared with the corresponding vector in Eq. 1.16 for ethane (yielding a methane-ethane Frobenius distance (for the diagonalized

LDMs) of *ca.* 3.294).

While the padding with zeroes appears ideal for homologous series such as the aliphatic hydrocarbons, other approaches may be more adequate when there exists a “common skeleton” with a substituent at a particular location that perturbs the active group of interest. These substituents may or may not have the same number of atoms, but are all attached to the same atom of the common skeleton. An example is provided by the substituted benzoic acid series.

Fig. 1.3 represents the series of *para*-substituted benzoic acids, whereby we can consider the carboxylic group as the active center responsible for “activity”, here the pK_a . In this case, the active center is being perturbed through a common skeleton (the aromatic ring) which transmits the perturbation of a substituent S of variable size and nature (in this example, S is at position 15 attached to C8 in Fig. 1.3).

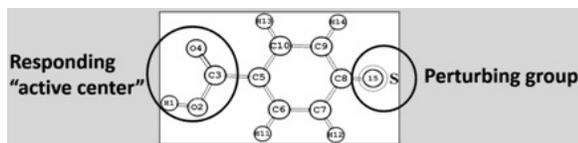


Figure 1.3: *p*-Benzoic viewed as an active center $-(\text{COOH})$ perturbed by a distant substituent (S) attached at carbon C8.

In the example of the substituted benzoic acids, all matrices are equalized in size by condensing all the atoms of S into a “super-atom”, that is a collection of nuclei and their associated atomic basins that are taken as one self-contained group. The idea of a super-atom implements the concept of pruning the branches introduced by Pye and Poirier [27, 28].

The number of localized electrons within the super-atom S is the sum of the localized electrons in each composing atoms plus the number of electrons delocalized within the group (that is between the constituent atoms). Thus, we define the localization

index of the super atom [21]:

$$\Lambda(\Omega_{super}) = \sum_{i=1}^{n_{super}} \Lambda(\Omega_i) + \sum_{\substack{i=1 \\ i \neq j \\ i \in \Omega_{super}}}^{n_{super}} \delta(\Omega_i, \Omega_j) \quad (1.18)$$

It is non-coincidental that Eq. 1.18 bears a striking resemblance to Eq. 1.7 since at the limit where the super-atom is enlarged to consist of the full molecule is a case by which the number of electrons localized within the bounds of the full super-atom (which includes N_{loc} and N_{deloc}) is none else than N , the total number of electrons in the molecule.

On the other hand, the number of electrons shared between the super-atom S and an atom k outside of S is given by the sum of the delocalization indices of every atom within S to that atom [21]:

$$\delta(\Omega_{super}, \Omega_k) = \sum_{\substack{i=1 \\ i \in \Omega_{super}}} \delta(\Omega_i, \Omega_k) \quad (1.19)$$

leading to off-diagonal entries of $\frac{1}{2}\delta(\Omega_{super}, \Omega_k)$ between the super-atom and the k^{th} atom in the molecule.

As an example, and following the numbering scheme in Fig. 1.3, an LDM of *p*-nitrobenzoic acid is a 17×17 matrix:

	H1	O2	C3	C4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	O16	O17
H1	0.07	0.31	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
O2	0.31	8.10	0.44	0.15	0.04	0.02	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
C3	0.01	0.44	2.83	0.66	0.48	0.02	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C4	0.01	0.15	0.66	8.20	0.05	0.01	0.00	0.01	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00
C5	0.00	0.04	0.48	0.05	3.91	0.67	0.04	0.04	0.04	0.67	0.02	0.00	0.02	0.00	0.00	0.00	0.00
C6	0.00	0.02	0.02	0.01	0.67	3.94	0.70	0.04	0.05	0.03	0.47	0.02	0.00	0.00	0.01	0.00	0.00
C7	0.00	0.00	0.00	0.00	0.04	0.70	3.93	0.67	0.03	0.05	0.02	0.46	0.00	0.00	0.03	0.02	0.01
C8	0.00	0.00	0.00	0.01	0.04	0.04	0.67	3.77	0.66	0.04	0.00	0.02	0.00	0.02	0.42	0.05	0.05
C9	0.00	0.00	0.00	0.00	0.04	0.05	0.03	0.66	3.93	0.70	0.00	0.00	0.02	0.46	0.03	0.01	0.02
C10	0.00	0.01	0.03	0.02	0.67	0.03	0.05	0.04	0.70	3.94	0.00	0.00	0.47	0.02	0.01	0.00	0.00
H11	0.00	0.02	0.00	0.00	0.02	0.47	0.02	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00
H12	0.00	0.00	0.00	0.00	0.00	0.02	0.46	0.02	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.02	0.00
H13	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.02	0.47	0.00	0.00	0.37	0.00	0.00	0.00	0.00
H14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.46	0.02	0.00	0.00	0.00	0.36	0.00	0.00	0.02
N15	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.42	0.03	0.01	0.00	0.00	0.00	0.00	4.44	0.83	0.83
O16	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.01	0.00	0.00	0.02	0.00	0.00	0.83	7.30	0.21
O17	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.05	0.02	0.00	0.00	0.00	0.00	0.02	0.83	0.21	7.30

(1.20)

in which the matrix elements belonging to the atoms composing the super-atom are in *italicized-bold* font for easy distinction. This matrix reduces to a 15×15 matrix upon treating the $-NO_2$ group as a super-atom, which, with columns and rows sums explicitly shown, is:

	H1	O2	C3	C4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NO_215	Σ
H1	0.07	0.31	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41
O2	0.31	8.10	0.44	0.15	0.04	0.02	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	9.09
C3	0.01	0.44	2.83	0.66	0.48	0.02	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	4.49
C4	0.01	0.15	0.66	8.20	0.05	0.01	0.00	0.01	0.00	0.02	0.00	0.00	0.02	0.00	0.00	9.13
C5	0.00	0.04	0.48	0.05	3.91	0.67	0.04	0.04	0.04	0.67	0.02	0.00	0.02	0.00	0.01	6.00
C6	0.00	0.02	0.02	0.01	0.67	3.94	0.70	0.04	0.05	0.03	0.47	0.02	0.00	0.00	0.01	5.99
C7	0.00	0.00	0.00	0.00	0.04	0.70	3.93	0.67	0.03	0.05	0.02	0.46	0.00	0.00	0.07	5.98
C8	0.00	0.00	0.00	0.01	0.04	0.04	0.67	3.77	0.66	0.04	0.00	0.02	0.00	0.02	0.52	5.80
C9	0.00	0.00	0.00	0.00	0.04	0.05	0.03	0.66	3.93	0.70	0.00	0.00	0.02	0.46	0.07	5.98
C10	0.00	0.01	0.03	0.02	0.67	0.03	0.05	0.04	0.70	3.94	0.00	0.00	0.47	0.02	0.01	5.98
H11	0.00	0.02	0.00	0.00	0.02	0.47	0.02	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.93
H12	0.00	0.00	0.00	0.00	0.00	0.02	0.46	0.02	0.00	0.00	0.00	0.36	0.00	0.00	0.02	0.91
H13	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.02	0.47	0.00	0.00	0.37	0.00	0.00	0.92
H14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.46	0.02	0.00	0.00	0.00	0.36	0.02	0.91
NO_215	0.00	0.00	0.00	0.00	0.01	0.01	0.07	0.52	0.07	0.01	0.00	0.02	0.00	0.02	22.76	23.50
Σ	0.41	9.09	4.49	9.13	6.00	5.99	5.98	5.80	5.98	5.98	0.93	0.91	0.92	0.91	23.50	86.00

(1.21)

where $N(NO_2) = 23.50e^-$ indicating a net electron withdrawal of $0.50e^-$ from the common skeleton.

The super-atom is useful when there exists a “common skeleton” for a family of molecules. Another way to look at this family is to instead use a truncated matrix. A truncated matrix may be more suitable since the variance of the super-atom can be high in a family of molecules, thus one might want to omit the super-atom. We will look more into the idea of truncated matrices in Chapter 3.

1.4.3 Other Limitations of LDMs

As discussed in Ref.[20], some matrix invariants within the context of chemical graph theory may occasionally be identical despite being derived from different molecular graphs. A known example is that of the characteristic polynomial of 1,4-divinylbenzene and that of 2-phenylbutadiene which are identical ($x^{10} - 10x^8 + 33x^6 - 44x^4 + 24x^2 - 4$). This problem is extremely unlikely when the molecules are coded not by topological connectivity matrices consisting of ones and zeroes but rather by their respective LDMs (or electron density-weighted adjacency matrices, discussed below) since these matrices cannot contain elements that are all of identical magnitudes.

Another common limitation of all known connectivity graphs - complete or incomplete - of their matrix surrogates is their inherent insensitivity to optical isomerism. This limitation is circumvented if the experimental dataset includes the active isomers and rejects the inactive ones from the set.

Finally, and as any other method for use in empirical modeling of experimental data, conformational averaging has to be performed whenever there exists more than one thermally-accessible rotamer that compete significantly for the molecular population as governed by the Boltzmann-distribution at the temperature of interest.

1.5 The Electron Density-Weighted Connectivity Matrix (EDWCM)

The LDM requires for its determination a quantum chemical calculation since the calculation of the LIs and DIs requires the electron density and the electron pair density contained in the second-order density matrix which is inaccessible from experiment (or an appropriate approximation of the second-order density matrix given the first-order density matrix if the latter is not printed by the electronic structure calculation software [29]).

The usage of matrix representatives of molecules is not restricted to LDMs and can be extended to quantities directly derivable from both theory and experiment such as the matrix of Coulombic nuclear-nuclear repulsion, the distance matrix, or the matrices of bond critical point (BCP) properties such as the electron density-weighted adjacency matrix (EDWAM) [22-24,30].

The chemical graph theoretic hydrogen-suppressed connectivity matrix of ethane is:

$$\begin{array}{cc} & \begin{array}{cc} \text{C1} & \text{C2} \end{array} \\ \begin{array}{c} \text{C1} \\ \text{C2} \end{array} & \left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right) \end{array} \quad (1.22)$$

with a determinant of -1 , a vector $\mathbf{D} = (1, -1)$, and the characteristic polynomial:

$$\lambda^2 - 1 \quad (1.23)$$

The unique features and properties of this molecule are captured with a higher fi-

delity and specificity if (a) the “ones” in the above matrix are multiplied (weighted) by the value of the electron density (in a.u.) at the bond critical point (BCP) for the corresponding bonds, and (b) if all atoms are kept including hydrogen atoms to yield an EDWAM representative of this molecule. The idea of EDWAM was first communicated to one of us (CM) by Professor Lou Massa in the form of a private communication [30].

An EDWAM representation of ethane is:

$$\begin{array}{c}
 \begin{array}{cccccccc}
 & \text{C1} & \text{H2} & \text{C3} & \text{H4} & \text{H5} & \text{H6} & \text{H7} & \text{H8} \\
 \text{C1} & \left(\begin{array}{cccccccc}
 0 & 0.273 & 0.238 & 0.273 & 0.273 & 0 & 0 & 0 \\
 0.273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0.238 & 0 & 0 & 0 & 0 & 0.273 & 0.273 & 0.273 \\
 0.273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0.273 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0.273 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0.273 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0.273 & 0 & 0 & 0 & 0 & 0 & 0
 \end{array} \right) & & & & & & & & \\
 \text{H2} & & & & & & & & \\
 \text{C3} & & & & & & & & \\
 \text{H4} & & & & & & & & \\
 \text{H5} & & & & & & & & \\
 \text{H6} & & & & & & & & \\
 \text{H7} & & & & & & & & \\
 \text{H8} & & & & & & & &
 \end{array}
 \end{array} \quad (1.24)$$

which yields a determinant of zero, and $\mathbf{D} = (-0.607, -0.369, 0.000, 0.000, 0.000, 0.000, 0.369, 0.607)$, and a characteristic polynomial:

$$1.000\lambda^8 - 0.504\lambda^6 + 0.050\lambda^4 \quad (1.25)$$

The molecular graph is generally incomplete since not every pair of atoms share a

bond path. The EDWAM has the advantage of being accessible from experiment and relatively inexpensive to calculate theoretically since it does not involve any numerical integration over atomic basins post-electronic structure calculation. Because of these practical advantages, the EDWAM may be well-suited for quantitative structure activity relationship (QSAR) studies that involve large molecular sets typical of the *in silico* phase of drug design for example (this is the last time the EDWAM will be discussed in this thesis).

The same limitations and solutions that are discussed for LDMs in Section 1.4 apply equally to the EDWAM.

1.6 Acknowledgements

The authors thank Dr. Todd A. Keith, Professor Lou Massa, Dr. Nenad Trinajstić, Dr. Sonja Nikolić, and Mr. Matthew J. Timm for helpful discussions. Financial support of this work was provided by the *Natural Sciences and Engineering Research Council of Canada* (NSERC), *Canada Foundation for Innovation* (CFI), *Saint Mary's University*, *McMaster University*, and *Mount Saint Vincent University*.

1.7 References

- [1] I.S. Dmitriev; *Molecules without Chemical Bonds (English Translation)*; Mir Publishers: Moscow, 1981.
- [2] A.T. Balaban; *Chemical Application of Graph Theory*; Academic Press: New York, 1976.
- [3] A.T. Balaban; Applications of graph theory in chemistry. *J. Chem. Inf. Comput.*

Sci. **1985**, *25*, 334-343.

- [4] L.H. Hall, L.B. Kier; *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: Boston, 1976.
- [5] Bonchev, D. Rouvray, D.H.; *Chemical Graph Theory: Introduction and Fundamentals*; OPA: Amsterdam, 1991.
- [6] K. Balasubramanian; Integration of graph theory and quantum chemistry for structure-activity relationship. *SAR & QSAR Environ Res.* **1994**, *2*, 59-77.
- [7] M.D. Diudea, I. Gutman, J. Lorentz; *Molecular Topology*; Nova Science Publishers, Inc.: Hauppauge NY, 1999.
- [8] D. Janezic, A. Milicevic, S. Nikolic, N. Trinajstic; *Graph Theoretical Matrices in Chemistry (Mathematical Chemistry Monographs, Vol. 3)*; University of Kragujevac: Kragujevac, 2007.
- [9] R. Todeschini, V. Consonni; *Molecular Descriptors for Chemoinformatics (Second Edition; Vols. I and II)*; Wiley-VCH Weinheim: Weinheim, 2009.
- [10] R.F.W. Bader; *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- [11] P.L.A. Popelier; *Atoms in Molecules: An Introduction*; Prentice Hall: London, 2000.
- [12] 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.
- [13] R.F.W. Bader; A bond path: A universal indicator of bonded interactions. *J. Phys. Chem. A* **1998**, *102*, 7314-7323.
- [14] R.F.W. Bader; Bond paths are not chemical bonds. *J. Phys. Chem. A* **2009**, *113*, 10391-10396.
- [15] A. Martin Pendas, E. Francisco, M.A. Blanco, C. Gatti; Bond paths as

- privileged exchange channels. *Chem. Eur. J.*, **2007**, *13*, 9362-9371.
- [16] G.R. Runtz, R.F.W. Bader, R.R. Messer; Definition of bond paths and bond directions in terms of the molecular charge distribution. *Can. J. Chem.* **1977**, *55*, 3040-3045.
- [17] T.A. Keith, R.F.W. Bader, Y. Aray; Structural homeomorphism between the electron density and the virial field. *Int. J. Quantum Chem.* **1996**, *57*, 183-198.
- [18] T.A. Keith; AIMAll/AIMStudio, <http://aim.tkgristmill.com/>**2015**.
- [19] X. Fradera, M.A. Austen, R.F.W. Bader; The Lewis model and beyond. *J. Phys. Chem. A* **1999**, *103*, 304-314.
- [20] C.F. Matta; 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.* **2014**, *35*, 1165-1198.
- [21] I. Sumar, P.W. Ayers, C.F. Matta; Electron localization and delocalization matrices in the prediction of pK_a 's and UV-wavelengths of maximum absorbance of *p*-benzoic acids and the definition of super-atoms in molecules. *Chem. Phys. Lett.* **2014**, *612*, 190-197.
- [22] M.J. Timm, C.F. Matta, L. Massa, L. Huang; The localization-delocalization matrix and the electron density-weighted connectivity matrix of a finite graphene flake reconstructed from kernel fragments. *J. Phys. Chem. A* **2014**, *118*, 11304-11316.
- [23] C.F. Matta; Localization-delocalization matrices and electron density-weighted adjacency matrices: New electronic fingerprinting tools for medicinal computational chemistry. *Future Med. Chem.* **2014**, *6*, 1475-1479.
- [24] B. Dittrich, C.F. Matta; Contributions of charge-density research to medicinal

- chemistry. *Int. U. Cryst J, (IUCrJ)* **2014**, *1*, 457-469.
- [25] I. Sumar, R. Cook, P.W. Ayers, C.F. Matta; AIMLDM: A program to generate and analyze electron localization-delocalization matrices (LDMS). *Comp. Theor. Chem.* **2015**, *1070*, 55-67.
- [26] D. White, R.C. Wilson; Parts-based generative models for graphs. *19th Intl. Conf. on Pattern Recognition (ICPR 2008)* **2008**, 1-4.
- [27] C.C. Pye, R.A. Poirier; Graphical approach for defining natural internal coordinates. *J. Comput. Chem.* **1998**, *19*, 504-511.
- [28] C.C. Pye; *Applications of Optimization to Quantum Chemistry, PhD Thesis*; Memorial University of Newfoundland: Saint John's (NF), Canada, **1997**.
- [29] A.M.K. Müller; Explicit approximate relation between reduced two- and one-particle density matrices. *Phys. Lett. A* **1984**, *105*, 446-452.
- [30] L. Massa; *personal communication* **2014**.

AIMLDM: A Program to Generate and Analyze Electron Localization-Delocalization Matrices (LDMs)¹

We report a programme called AIMLDM [1] written in Python 3.4.1 that extracts the localization and delocalization indices from the output of QTAIM numerical integration analysis software AIMAll/AIMStudio [2] (the .sum file), creates the LDMs, condenses atomic groups into super-atoms (pruning), extracts matrix invariants such as the LDMs' eigenvalues, and calculates the molecule-to-molecule Frobenius distance matrices. In addition, AIMLDM can also print the diagonal suppressed LDM (or DM) and the off diagonal suppressed LDM (or LM) (both of them which are pruned). [The software has recently been updated to now also include characteristic polynomial calculations]

2.1 The AIMLDM Programme

AIMLDM [1] extracts the elements (the LIs and DIs) of the LDM matrix from the relevant sections of the AIMAll output of every molecule in the molecular set (the .sum file, which lists the LIs first, separately from the DIs). The LI values are extracted from a given molecule's .sum file and placed along the diagonal of the LDM followed by half of the DI values which are entered as the off-diagonal elements. AIMLDM lists the atoms, their LIs and their DIs using the same numbering scheme as in the .sum file which originates from the numbering sequence of the original wavefunction

¹Based on the Paper: Computational and Theoretical Chemistry 1070 (2015) 55–67

obtained from the electronic structure computational software such as Gaussian or GAMESS.

Following Pye and Poirier's lead [3,4], the program prunes all matrices to match the matrix size of the smallest matrix in the set (discussed in the previous chapter). All atoms in a given substituent to a common skeleton are condensed to a super-atom by implementing Eqs. 1.18,1.19 [5]. Precaution is taken in atom labelling so that corresponding atoms and/or super-atoms in the entire molecular set receive the same numerical labels (refer to AIMLDM: Operating Instructions in **Appendix A**).

Once pruning has been achieved, a similarity distance matrix obtained from the Frobenius distances between LDMs (or the diagonalized LDMs) via Eq. 1.11 is constructed. While distances between matrices are not uniquely defined, the Frobenius distance has the appeal of being effectively an Euclidean distance in the $\{\lambda_i, \frac{1}{2}\delta_{ij}, i \neq j; i, j = 1 \dots m\}$ m^2 -dimensional space, and that it also has been shown to satisfy the triangle inequality [6].

Fig. 2.1 presents a flowchart describing the logical pathways of the programme AIMLDM. The user inputs the location of the .sum files of the molecular set (all in one directory) and also the preferred location for the output. The program then creates a variable for every .sum file and places all the text in a given .sum file in that variable. Keywords are then used to locate the start of the relevant sections listing the LIs and the DIs and to locate the end of each of these two sections. The text between the beginning and end keywords is stored while the remainder of the text of the .sum file is discarded then LIs and DIs are stored as separate arrays.

In order to organize what LI/DI array is assigned to which file, a dictionary is created so that each file is associated with its own LI/DI array. The LI and DI arrays are now combined to form a LI/DI array that corresponds to each file stored in a

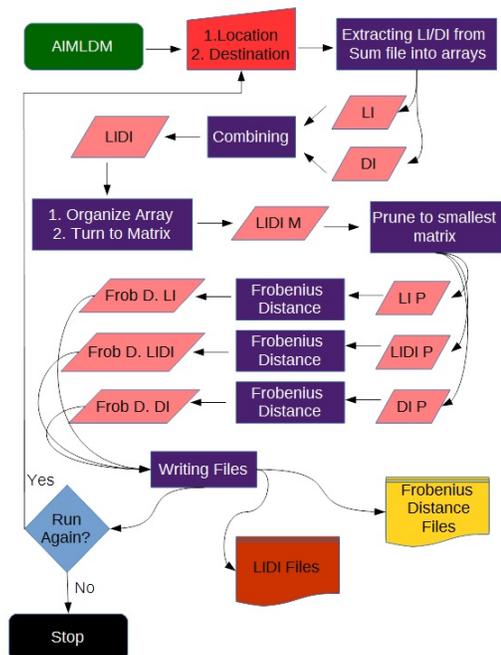


Figure 2.1: A flowchart explaining the logic of operation of the AIMLDM programme.

dictionary (the list of molecules in the set). The elements of the array are then cast into a symmetric matrix after dividing every DI by 2 to satisfy Eq. 1.5 so that the row or column sums are the atomic electron population of the atom labelling the column or row. Now the LDMs are ready for matrix operation and extraction of matrix invariants (eigenvalues) as well as the pruning steps as described above.

At this stage, the program searches through every LI array (since the length of the LI array gives the length of each row, and the square of the length of the row is the total number of elements in a given LDM) and recognizes the smallest molecule in the set. Pruning is then applied to every matrix to match the smallest matrix size as already described (Eqs. 1.18,1.19) [5].

Three matrices are produced by the programme in addition to the full LDM Eq. 1.10 after pruning has been complete (even if all the molecules are the same size a

pruning directory will still be created, these files will be identical of course to the full non-pruned LDM). These matrices are (a) the full-pruned LDM, (b) the off-diagonal suppressed matrix (pruned), that is, the delocalization matrix (DM) where zeroes are entered along the diagonal instead of the LIs, and (c) the localization matrix (pruned) (LM) that only has entries along the diagonals (LIs) but zeroes elsewhere. The reasons for creating the DMs and LMs is that they have been found to be more useful in QSAR than the full LDM in certain cases [6]. Three corresponding Frobenius distance matrices are then calculated, one for each matrix subtype. In addition to all these matrices, the eigenvalues for each matrix (pruned and not pruned) even for the Frobenius distance matrices are produced.

The program now has completed its calculations and prints the output files. Each .sum file spawns nine files (one lists the LIs/DIs, another casts the LIs and DIs/2 in the LDM format, three for each of the (pruned) LDM, LM, DM, and four for the eigenvalues for each matrix). In addition to these nine files (per .sum file), three Frobenius distance files are created that list the distance matrix between all molecules based on the Frobenius distance and their LDM representations using the pruned LDM, LM, and DM where every molecule is taken as a reference in a cyclical manner to exhaust all molecules in the set. Lastly three Frobenius distance eigenvalue files are created for each of the pruned LDM, LM, and DM.

The summary of the AIMLDM programme operations can be captured in the following few points:

1. Start.
2. Manual input of .sum files and desired output destination file directory.
3. Extract LIs/DIs of each molecule and store into separate arrays.

4. Combine LIs and DIs of each molecule into a single array (generation of molecular LDMs).
5. Organization of the combined LIs/DIs arrays into conventional matrix format.
6. Prune all matrices to that of the smallest molecule in the set.
7. Create pruned LDM, LM, and DM, as well as their eigenvalue matrices.
8. Compute Frobenius distance matrices from pruned matrices, as well as their eigenvalue matrices.
9. Write the output files (nine per molecule).
10. Write six Frobenius distance files (LDM, LM, DM, and eigenvalues for each one) comparing all molecules in the set.
11. Option to perform matrix operations on another set of files.
12. Stop.

Sample input and output files of the programme can be found in **Appendix A**. Successive improved and expanded versions will be available from the authors in the future.

2.2 Numerical Illustrative Testing

Several (but not all) of the properties of isoelectronic series are known to change gradually across a given ordered series [7-10]. First we test whether the gradual change in some of these properties are reflected in the Frobenius distance from the

Table 2.1: Frobenius molecule-molecule distance matrix using the full LDM

Molecule	ArO_4	ClO_4^-	SO_4^{2-}	PO_4^{3-}	SiO_4^{4-}
ArO_4	0	2.1845	3.9641	4.6977	5.0721
ClO_4^-	2.1845	0	1.8028	2.5680	2.9678
SO_4^{2-}	3.9641	1.8028	0	0.7888	1.2095
PO_4^{3-}	4.6977	2.5680	0.7888	0	0.4326
SiO_4^{4-}	5.0721	2.9678	1.2095	0.4326	0

last (or first) member of a series. For this purpose, we use the $N = 50e^-$ series (SiO_4^{4-} , PO_4^{3-} , SO_4^{2-} , ClO_4^- , and ArO_4) that was recently examined [7-10].

The Gaussian 09 [11] software was used to (a) optimize the geometry (followed by a frequency calculation that ensured all real frequencies) and (b) to generate “wavefunctions” at the optimized geometry for every one of the five molecules in the set. These calculations were conducted at a level of theory defined by the second order Møller-Plesset perturbation theory (MP2) in conjunction with a Pople 6-311+G(d,p) basis set. Handy and Schaefer’s Z-vector correction procedure [12] was then applied to the SCF density matrix to generate an effective correlated “relaxed” or “gradient” density matrix. These effective correlated wavefunctions were then subjected to numerical integration to calculate the LIs and DIs using AIMAll/AIMStudio [2]. Finally AIMLDM was used to extract the relevant information from the output of AIMAll and generate the LDMs representing the five molecules of the set and calculate the distance matrix shown in Table 2.1.

The distance matrix in Table 2.1 clearly shows that there is a gradual but non-monotonic change in the dissimilarity distance going down a given column whereby the molecule listed along the diagonal is taken as a reference (zero-distance from itself). For example the difference between the distances listed as two consecutive entries in column 1 (taking ArO_4 as reference) are: $d(SO_4^{2-}, ClO_4^-) - d(ArO_4, ClO_4^-)$

$= 3.9641 - 2.1845 = 1.7796$; the differences between subsequent entries (in order going down the same column, or across the first row) are 1.7796, 0.7336, and 0.3744 that is not following a discernable pattern except that the difference between two members of the series gets smaller the higher the atomic number of the central atom.

Table 2.1 also reveals an important property of the space being studied. The triangle inequality is obeyed with a central angle close to 180° . For example, from the distance matrix one reads: $d(\text{SiO}_4^{4-}, \text{ArO}_4) = 5.0721$, which is very close to the distance obtained by the distances sums, say, $d(\text{SiO}_4^{4-}, \text{SO}_4^{2-}) + d(\text{SO}_4^{2-}, \text{ArO}_4) = 1.2095 + 3.9641 = 5.1736$, or $d(\text{SiO}_4^{4-}, \text{PO}_4^{3-}) + d(\text{PO}_4^{3-}, \text{SO}_4^{2-}) + d(\text{SO}_4^{2-}, \text{ClO}_4^-) + d(\text{ClO}_4^-, \text{ArO}_4) = 0.4326 + 0.7888 + 1.8028 + 2.1845 = 5.2087$, etc. The discrepancy is possibly due to the slight departure from the Euclidean geometry of the mathematical dissimilarity distance space under study.

Plots of the values listed in the first row or column of the molecule-molecule distance matrix based on their LDMs in Table 2.1 and the corresponding row/column of the distance matrix based on the DMs (not shown) against the total energy ($E_{total} = E_{el} + E_{nn}$) and against the nuclear-nuclear repulsions energy (E_{nn}) are displayed in Fig. 2.2.

Both energies exhibit a roughly linear correlation with the DM-based distance from ArO_4 . The correlation becomes non-linear when the distances from ArO_4 are obtained from the LDMs (Table 2.1) as can be seen on the plots to the right of Fig. 2.2. This shows that *global molecular energetic properties* in this isoelectronic series are highly correlated with inter-molecule distances from a chosen reference, ArO_4 in this case. It is perhaps remarkable that E_{nn} , a classical Coulombic energy term that only depends on the charge and position of the nuclei that determine the “external potential”, is strongly correlated with a similarity measure based on electron

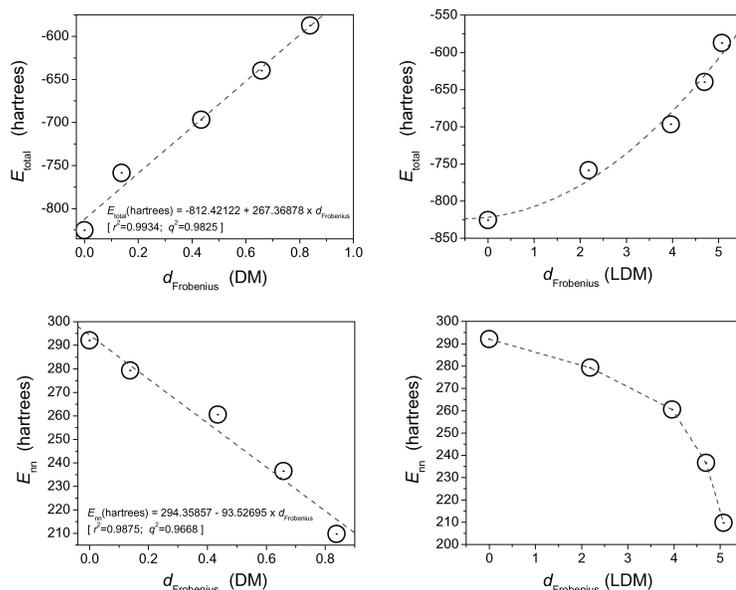


Figure 2.2: *Top:* Total energy (electronic + nuclear-nuclear) of the $N = 50e^-$ isoelectronic series SiO_4^{4-} (last data point on the far right), PO_4^{3-} , SO_4^{2-} , ClO_4^- , and ArO_4 (first data point on the far left taken as reference) against the Frobenius molecule-molecule distance using the diagonal suppressed LDM (or delocalization matrix, DM) (*left*) and from the full LDM (*right*). *Bottom:* Nuclear-nuclear repulsion energy for the same series of molecules against the Frobenius distance obtained using the DMs (*left*) and the LDMs (*right*).

localization/delocalization.

Fig. 2.3 displays two often reported experimentally-determinable *molecular properties* as functions of the similarity distance from ArO_4 , that is, the bond length (B.L.) in Å and the isotropic polarizability $\langle \alpha \rangle = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$ in a.u. obtained from the quantum chemical calculations. Both properties exhibit a non-linear dependence on both the DM- and on the LDM-based distances from the reference molecule without any obvious outliers.

Fig. 2.4 shows correlation with three *local properties*: (*Top*) The electron density at the nucleus of the central (non-oxygen) atom (p_n); (*Middle*) The maximum electrostatic potential (V) on the outer molecular Van der Waals isodensity surface ($p = 0.001a.u.$) associated with the central atom; (*Bottom*) the electron-nuclear at-

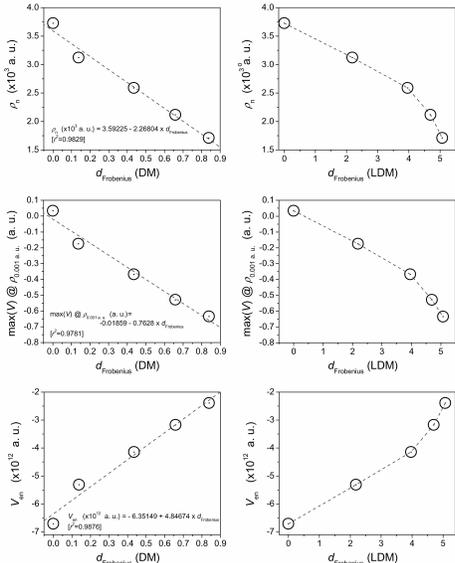


Figure 2.3: *Top:* Bond lengths (equal to bond path lengths to at least 4 decimals) in Å of the $N = 50e^-$ isoelectronic series SiO_4^{4-} (last data point on the far right), PO_4^{3-} , SO_4^{2-} , ClO_4^- , and ArO_4 (first data point on the far left taken as reference) against the Frobenius molecule-molecule distance using the diagonal suppressed LDM (or delocalization matrix, DM) (*left*) and from the full LDM (*right*). *Bottom:* Isotropic (average) polarizability $\frac{(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})}{3}$ for the same series of molecules against the Frobenius distance obtained using the DMs (*left*) and the LDMs (*right*)

traction contribution to the virial field at the central nucleus. All three plotted against the DM- and LDM-based Frobenius inter-molecular distances from ArO_4 . The three local properties appear to be roughly linearly correlated with the distance but when the full LDM is taken as a basis for comparison the correlations are non-linear, but strong nevertheless.

We next investigate the correlation of LDM distances with the total energy of small molecules calculated with different basis sets with the E_{total} from Hartree-Fock (HF) electronic structure calculation (SCF level). The basis sets used are STO-3-6G, 3-21G, 3-21+G, SVP, 6-31G, 6-31G(d), 6-31+G(d), 6-31+G(d,p), 6-311++G(2d,p), TZVP, UGBS, cc-pVDZ, cc-pVTZ, and cc-pVQZ. The quality of the basis set is reflected in the energy since HF is variational and generally results in a smaller LDM-Frobenius distance from the best value (calculated with the basis set that delivers the lowest energy). As an illustration, these calculations are performed on four small molecules of differing polarity (CH_4 , CH_3OH , H_2O , and NH_3) with a number of commonly used standard basis sets.

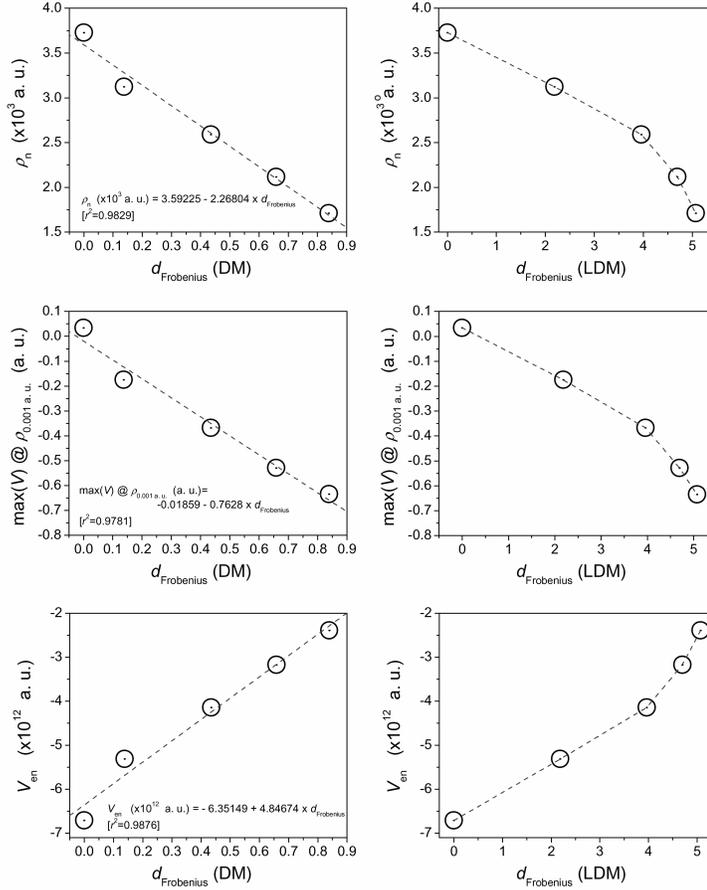


Figure 2.4: *Top:* Electron density at the nucleus of the central (non-oxygen) atom (p_n) in a.u. of the $N = 50e^-$ isoelectronic series SiO₄⁴⁻ (last data point on the far right), PO₄³⁻, SO₄²⁻, ClO₄⁻, and ArO₄ (first data point on the far left taken as reference) against the Frobenius molecule-molecule distance using the diagonal suppressed LDM (or delocalization matrix, DM) (*left*) and from the full LDM (*right*). *Middle:* Maximum electrostatic potential (V) in a.u. on the 0.001 a.u. isodensity surface against the Frobenius distance obtained using the DMs (*left*) and the LDMs (*right*). *Bottom:* Electron-nuclear attraction contribution to virial field at the central nucleus in a.u. against the Frobenius distance obtained using the DMs (*left*) and the LDMs (*right*).

For all four molecules, the lowest energy is obtained at the HF/cc-pvqz level of theory, which is used as the reference in the Frobenius distance calculations. Fig. 2.5 plots E_{total} as a function of the LDM-based Frobenius distance from the HF/cc-pvqz calculation with respect to each molecule. Each data-point on every one of the four plots is also labelled with the basis set that was used. The plots show that, generally,

2.3.1 LDM-Eigenvalues as Predictors in QSAR

Principal Component Analysis (PCA) [17] can be used to reduce the dimensions of LDMs and to extract QSAR descriptors from them. In this approach, an orthogonal transformation converts a matrix of (possibly correlated) variables into a set of linearly uncorrelated variables termed principal components which are less than or equal to the number of original variables. The first principal component has the largest variance and accounts for as much of the variability in the data as possible, and each succeeding component has maximal variance under the constraint that it is orthogonal to the preceding components (uncorrelated with it). Thus, principal components are orthogonal since they are the eigenvectors of the covariance matrix, which is a symmetric matrix. One can think of PCA as fitting an n -dimensional ellipsoid to the data, where each orthogonal axis of the ellipsoid represents a principal component. Small axes of the ellipse correspond to small variance along that axis. Omitting small axes (small principal components) from the LDM results in a commensurately small loss of information.

We have consistently observed a strong correlation between LDMs' eigenvalues and the number of electrons in atomic basins. Hydrogen atoms have the smallest electron populations and hence their contributions to the eigenvalues extracted from the LDM by the PCA transformation can be neglected. This is not dissimilar to the hydrogen-suppressed graphs pioneered by Kier and Hall in their "Atom Level Electrotopological State" [18].

As an initial exploration of the validity of extracting QSAR descriptors from LDMs using PCA transformations we investigate a series of carboxylic acids that extend the set in Refs.[5,6]. We first observe that as long as we retain the pair-wise values for

the LI (e.g. C1 to C1, O7 to O7, *etc.*) and the DI (C1 to O7, O7 to H4, *etc.*), then the ordering of the LDM does not affect the eigenvalues that are produced from the LDM (Table 2.2).

Table 2.2: “Scrambled” LDM for acetic acid and resulting eigenvalues

Table 2.2.A								
	C1	C2	H3	H4	H5	O6	O7	H8
C1	4.010	0.464	0.475	0.477	0.477	0.060	0.053	0.004
C2	0.464	2.793	0.017	0.021	0.021	0.654	0.417	0.005
H3	0.475	0.017	0.397	0.017	0.017	0.009	0.007	0.001
H4	0.477	0.021	0.017	0.419	0.018	0.010	0.006	0.001
H5	0.477	0.021	0.017	0.018	0.419	0.010	0.006	0.001
O6	0.060	0.654	0.009	0.010	0.010	8.276	0.143	0.008
O7	0.053	0.417	0.007	0.006	0.006	0.143	8.171	0.321
H8	0.004	0.005	0.001	0.001	0.001	0.008	0.321	0.075
	F1	F2	F3	F4	F5	F6		
<i>Eigen</i> ^(*)	3.314	1.816	0.968	0.762	0.559	0.541		
<i>Var.</i> % ^(*)	41.4	22.7	12.1	9.5	7.0	6.8		
<i>Cum.</i> % ^(*)	41.4	64.1	76.2	85.8	92.7	99.5		
Table 2.2.B								
	H3	O7	H4	H5	O6	C1	H8	C2
H3	0.397	0.007	0.017	0.017	0.009	0.475	0.001	0.017
O7	0.001	8.171	0.006	0.006	0.143	0.053	0.321	0.417
H4	0.017	0.006	0.419	0.018	0.010	0.477	0.001	0.021
H5	0.017	0.006	0.018	0.419	0.010	0.477	0.001	0.021
O6	0.009	0.143	0.010	0.010	8.276	0.060	0.008	0.654
C1	0.475	0.053	0.477	0.477	0.060	4.010	0.004	0.464
H8	0.001	0.321	0.001	0.001	0.008	0.004	0.0075	0.005
C2	0.017	0.417	0.021	0.021	0.654	0.464	0.005	2.793
	F1	F2	F3	F4	F5	F6		
<i>Eigen</i> ^(*)	3.314	1.816	0.968	0.762	0.559	0.541		
<i>Var.</i> % ^(*)	41.4	22.7	12.1	9.5	7.0	6.8		
<i>Cum.</i> % ^(*)	41.4	64.1	76.2	85.8	92.7	99.5		

(*)Eigen. = eigenvalues, Var.% = percent variability, Cum.% = cumulative percentage.

The largest six eigenvalues extracted using the PCA method generally account for more than 95% of the variance in the LDM as can be seen from Table 2.3. The

unaccounted-for variance (especially in the larger molecules) is principally due to the hydrogen atoms.

Table 2.3: Eigenvalues of the LDMs from a series of carboxylic acids from PCA (non-traditional names are used to highlight the functional groups attached to the C-COOH skeleton.)

Compounds	pK_a	F1	F2	F3	F4	F5	F6
2,2,2-trimethylacetic acid	5.03	3.5044	3.4968	2.6103	1.6493	1.1271	0.8440
2-methylacetic acid	4.88	3.2949	2.6510	1.6724	0.9231	0.7578	0.5431
2,2-dimethylacetic acid	4.84	3.6629	2.8154	2.0677	1.5495	0.9162	0.7466
2-ethylacetic acid	4.82	3.4123	2.6025	2.4643	1.5942	0.9186	0.7470
acetic acid	4.76	3.3139	1.8157	0.9681	0.7624	0.5594	0.5410
2,2-diethylacetic acid	4.71	3.6628	2.9875	2.5368	2.4024	1.9023	1.4590
2-phenylacetic acid	4.31	2.9070	2.3874	2.2992	2.1406	1.6469	1.4845
2-hydroxyacetic acid	3.83	2.7729	2.2391	1.7063	0.9008	0.7231	0.5576
2-methoxyacetic acid	3.57	3.5054	2.5375	1.6007	1.1461	0.8160	0.6961
2-mercaptoacetic acid	3.55	2.7687	2.2753	1.6337	0.8686	0.6791	0.5247
chloroacetic acid	2.87	2.8650	1.8055	1.3091	0.8645	0.5579	0.5375
fluoroacetic acid	2.59	2.8279	1.7929	1.3095	0.8673	0.5903	0.5444
2-cyanoacetic acid	2.45	2.8160	2.2821	1.6958	0.8868	0.7173	0.4891
glycine	2.37	3.0611	2.6926	1.7311	0.9443	0.7691	0.5314
N-methylglycine	2.35	3.4219	2.4983	2.0691	1.5747	0.8849	0.7399
N,N-dimethylglycine	2.04	3.5031	3.1997	2.3859	1.5197	1.0354	0.8589
difluoroacetic acid	1.34	2.4528	1.6669	1.3705	1.1398	0.8523	0.3994
dichloroacetic acid	1.26	2.4669	1.6702	1.3612	1.1442	0.8436	0.4083
trifluoroacetic acid	0.52	2.2115	1.4565	1.1411	1.1403	1.1120	0.7399
trichloroacetic acid	0.51	2.2149	1.4654	1.1462	1.1454	1.0845	0.7442

It would be instructive to compare pairs of molecules by mapping each in an n -dimensional abstract mathematical eigenvalue space (obtained from a PCA of the LDM) then determine if the respective locations of the molecules in this space coincide with chemical knowledge. It is not possible to readily visualize spatial relationships beyond three dimensions, and hence, even for the 6-dimensional space that corresponds to the PCs listed in Table 2.3 reduction of dimensionality is needed. This is achievable through a number of methods clamped together in what is known as multi-dimensional scaling (MDS) algorithms [19-25]. These algorithms aim at projecting the

complicated “distance matrix” between the compared object from the n -dimensional space to 2- or 3- dimensions under the constraint to minimize the changes on inter-object distances. To achieve this goal, MDS algorithms minimize a criterion termed the “*Kruskal Stress (S)*”, $0 \leq S \leq 1$, defined as [22]:

$$S = \sqrt{\frac{\sum_{i,j} [f(x_{ij}) - d_{ij}]^2}{\sum_{i,j} d_{ij}^2}} \quad (2.1)$$

where d_{ij} is the distance measured between points i and j and $f(x_{ij})$ is the transformation of the raw input data x_{ij} whereby when $f(x_{ij}) = x_{ij}$ the raw data is compared to the distances on the lower dimensional map directly (metric scaling) otherwise f is a (weakly monotonic) transformation used to minimize S . The closer the stress is to zero, the better the 2- or 3-dimensional representation of the n -dimensional space.

Rigorous statistical methods that evaluate the quality of a MDS representation are not available at the time of writing. A plot called a “*Shepard diagram*” is often used as a qualitative indicator of the quality of the lower-dimensional representation [23,24].

The Shepard diagram is essentially a scatter plot in which the abscissa represents the inter-objects distances in the full n -dimensional space while the ordinate represents the distance between every given pair of objects projected on the lower-dimensional space obtained from the MDS. Larger spread (scatter of data away from the line of best fit) is a diagnostic of an unreliable multidimensional scaling map. On the other hand, when all points lie on the same line, then the quality is perfect, but for any realistic example some scatter is expected, the smaller the scatter the more reliable is the MDS projection.

Fig. 2.6 displays a Shepard diagram using the data listed in Table 2.3 after MDS

treatment using the programme XLSTATTM [26]. The Shepard plot reveals that S is low and that the scatter-plot is linear. Fig. 2.7 displays a 2-dimensional projection of the 6-dimensional eigenvalue descriptors obtained from the LDMs of a series of substituted acetic acids. This mapping groups acids with electron-withdrawing substituents together (upper left quadrant) while those acids with electron donating substituents are grouped together and far from the first group (lower right quadrant), in line with chemical expectation.

For closely related series of carboxylic acids, such as halogenated acetic acids, the positions on the map is expected to be strongly correlated with physical properties. Such a correlation has indeed been reported between the pK_a 's of fluorinated and chlorinated acetic acids, that is, substituted acetic acids (SAA) where S = F, Cl, and the Frobenius distance of their DMs from that of unsubstituted acetic acid (AA) [6]:

$$pK_a(SAA) \approx -0.588 + 5.415e^{[-5.066d_{deloc}(AA,SAA)]} \quad (2.2)$$

$$(r^2 = 0.979, n = 7)$$

Now if we regress the distances (d) of the same set of six chlorine and fluorine substituted acetic acids from the unsubstituted reference molecule generated from the MDS projected map displayed in Fig. 2.7 we get:

$$pK_a(SAA) \approx 8.4075e^{[-0.644d]} \quad (2.3)$$

$$(r^2 = 0.996, n = 7)$$

The strength of the correlation in Eq. 2.1 indicates that the 2-dimensional projection

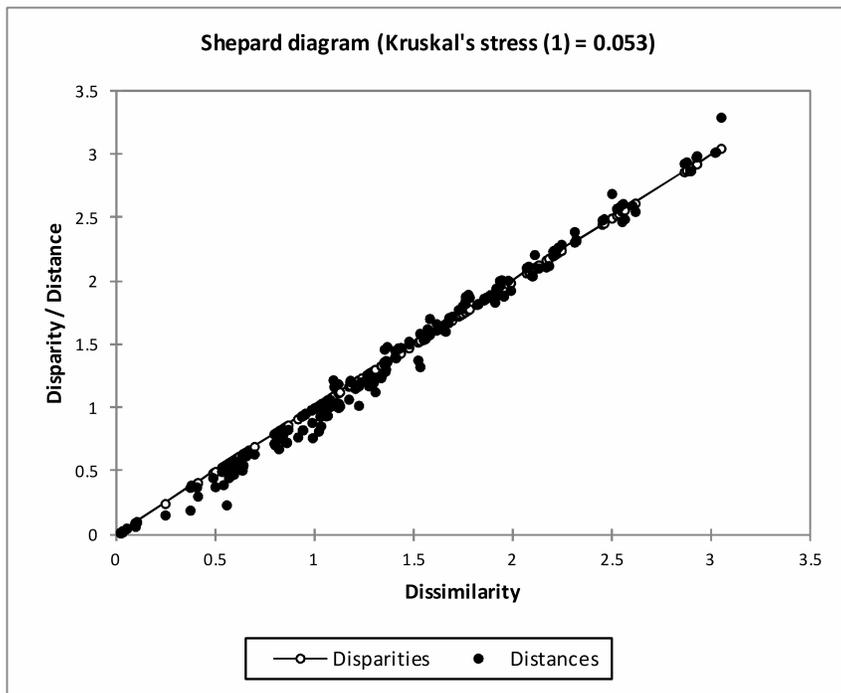


Figure 2.6: Shepard plot of the multidimensional scaling (MDS) transformed six dimensional data listed in Table 2.3

of the 6-dimensional eigenvalue descriptor set for these molecules retains most of the information contained in their LDM (or DM) representations.

2.4 Conclusion

LDMs and related matrices have been shown promising in QSAR-type studies. The size of the data sets in the past has been limited by the necessity of manual construction and manipulation of these matrices. The first release of a programme that automates the essential steps necessary for the LDM-based analysis is presented here and instructions on how to operate are in **Appendix A**. The AIMLDM programme's principal usage is to extract LDMs and related matrices from as many AIMAll output files as desired. In other words, what AIMLDM achieves is essentially extracting and

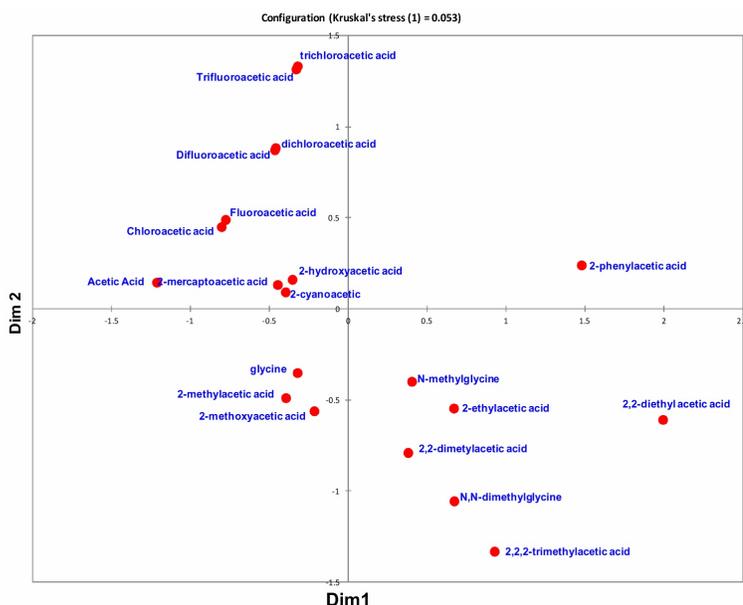


Figure 2.7: Two-dimensional projection of the 6-dimensional eigenvalue descriptors of the LDMs of a series of substituted acetic acids.

constructing the LDMs and DMs of large molecular datasets from the AIMAll output and, subjecting said matrices to basic manipulations. In this way, AIMLDM is a programme that operates at a different level than AIMAll, the latter being concerned with one molecule at a time while the former uses the output of AIMAll for each molecule in a set to create their matrix representatives for further processing. This is the main goal of AIMLDM. It is *not* the scope of AIMLDM to cover every aspect of the manipulation of the matrices it extracts from a set of AIMAll outputs.

The first release of AIMLDM is not claimed to be flawless and will naturally be improved in subsequent releases that will be made available in the future by the authors. Other programmes such as XLSTATTM [39] can be used to apply analyses such as those based on multidimensional scaling once the matrices for all the molecules are generated by AIMLDM.

Numerical examples suggest the programme's numerical stability since no unex-

pected outliers can be identified using the $N = 50e^-$ isoelectronic series (SiO_4^{4-} , PO_4^{3-} , SO_4^{2-} , ClO_4^- , and ArO_4) as a test set. Hartree-Fock calculations on four small molecules (CH_4 , CH_3OH , H_2O , and NH_3) with a variety of basis sets demonstrate that the LDM Frobenius distance from the most flexible basis set increases with the total energy as the basis set's quality is reduced. This result suggests that among potential uses of the LDM-analysis would be the comparison and assessment of the quality of basis sets and possibly also the testing of new density functional theory (DFT) functionals. However, the main area of anticipated use of the LDM analysis is in the domain of quantitative structure-activity relationship (QSAR) studies largely used in drug and materials design as the examples outlined in this paper and in the literature cited therein suggest.

It is also shown that the analysis of the eigenvalues of the LDMs using the principal component analysis constitute another promising approach to extract condensed or "pruned" descriptors from the LDMs. Compared to the full LDMs or DMs, intermolecular dissimilarity distances calculated using a combination of principal component analysis and multidimensional scaling yield a simple exponential model that accurately predicts $\text{p}K_a$'s.

2.5 Acknowledgements

The authors would like to acknowledge Matthew Timms (PhD Candidate, University of Toronto) for the input geometries of the isoelectronic series, and Dr. Todd Keith (SemiChem, Inc.) and Prof. Lou Massa (Hunter College, City University of New York) for several useful discussions. Financial support of this work was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada

Foundation for Innovation (CFI), Saint Mary's University, McMaster University, and Mount Saint Vincent University.

2.6 References

- [1] I. Sumar, Electron Localization-Delocalization Matrices (LDMs): Theory and Chemical Applications, M.Sc. Thesis (Saint Mary's University, Halifax, 2015).
- [2] T.A. Keith, AIMAll, [Http://Aim.Tkgristmill.Com/](http://Aim.Tkgristmill.Com/) (2011).
- [3] C.C. Pye and R.A. Poirier, Graphical approach for defining natural internal coordinates, *J. Comput. Chem.* 19 (1998) 504-511.
- [4] C.C. Pye, Applications of Optimization to Quantum Chemistry, PhD Thesis (Memorial University of Newfoundland, Saint John's (NF), Canada, 1997).
- [5] I. Sumar, P.W. Ayers and C.F. Matta, 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 (2014) 190-197.
- [6] C.F. Matta, 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 (2014) 11651198.
- [7] P. Pyykkö, Ab initio predictions for new chemical species, *Phys. Script.* 33 (1990) 52-53.
- [8] P. Pyykkö, Ab initio study of bonding trends among cyanamidophosphates ($[\text{PO}_n(\text{NCN})_{4-n}]^{3-}$) and related systems, *Chem. Eur. J.* 6 (2000) 2145-2151.
- [9] R. Lindh, W.P. Kraemer and M. Kämper, On the Thermodynamic Stability of

ArO₄, *J. Phys. Chem. A* 103 (1999) 8295-8302.

[10] M. Timm and C.F. Matta, Primary retention following nuclear recoil in β -decay: Proposed Synthesis of a metastable rare gas oxide (³⁸ArO₄) from (³⁸ClO₄) and the evolution of chemical bonding over the nuclear transmutation reaction path, *Appl. Rad. Isotopes* 94 (2014) 206-215.

[11] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski and D.J. Fox, *Gaussian 09*, Revision B.01 (Gaussian Inc., Wallingford CT, 2010).

[12] Handy N. C. and H.F. Schaefer III, On the evaluation of analytic energy derivatives for correlated wave-functions, *J. Chem. Phys.*, 81 (1984) 5031-33 81 (1984) 5031-5033.

[13] R. Cook, I. Sumar, P.W. Ayers and C.F. Matta, *Electron Localization-Delocalization Matrices (LDMs): Theory and Applications* (Springer International Publishing AG, Cham, Switzerland, 2016).

[14] C.F. Matta , I. Sumar, R. Cook and P.W. Ayers, *Localization-delocalization*

- and electron density-weighted connectivity matrices: A bridge between the quantum theory of atoms in molecules and chemical graph theory, In: Applications of Topological Methods in Molecular Chemistry (Challenges and Advances in Computational Chemistry and Physics Series); Chauvin, R.; Silvi, B.; Alikhani, E.; Lepetit, C. (Eds.) (Springer, 2015). electrostatic potential, *J. Comput. Chem.* 35 (2014) 1165-1198.
- [15] C.F. Matta, Localization-delocalization matrices and electron density-weighted adjacency matrices: New electronic fingerprinting tools for medicinal computational chemistry, *Future Med. Chem.* 6 (2014) 1475-1479.
- [16] B. Dittrich and C.F. Matta, Contributions of charge-density research to medicinal chemistry, *Int. U. Cryst. J. (IUCrJ)* 1 (2014) 457-469.
- [17] P.P. Mager, *Multidimensional Pharmacochimistry: Design of Safer Drugs* (Academic Press, Inc., London, 1984).
- [18] L.B. Kier , L.H. Hall and J.W. Frazer, An index of electrotopological state for atoms in molecules, *J. Math. Chem.* 7 (1991) 229-241.
- [19] I. Borg and P. Groenen, *Modern multidimensional scaling: theory and applications* (Springer, New York, 1997).
- [20] T.F. Cox and M.A.A. Cox, *Multidimensional Scaling* (Chapman & Hall, . London, 1994).
- [21] J.B. Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika* 29 (1964) 1-27.
- [22] J.B. Kruskal, Nonmetric multidimensional scaling: a numerical method, *Psychometrika* 29 (1964) 115-129.
- [23] R.N. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. II, *Psychometrika* 27 (1962) 219-246.

- [24] R.N. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. I., *Psychometrika* 27 (1962) 125-140.
- [25] W.S. Torgerson, Multidimensional scaling: I. Theory and method, *Psychometrika* 17 (1952) 401-419.
- [26] T. Fahmy and E. Jakobowicz, XLSTAT[®]-Pro (Version 2015.2.02, www.xlstat.com/) / Microsoft[®]Excel 2007 (Addinsoft, Inc., Brooklyn, New York, 2015).

Electron Localization and Delocalization Matrices in the Prediction of pK_a 's and UV-Wavelengths of Maximum Absorbance of *p*-Benzoic Acids¹

As has been shown throughout this thesis, the combination of chemical graph theory and the quantum theory of atoms in molecules (QTAIM) is very powerful in QSAR. In this chapter we further emphasize the importance of the localization and delocalization indices (LIs and DIs). This is demonstrated through the modelling of pK_a 's and λ_{max} 's of a series of *para*-substituted benzoic acids.

Distances between the LDM representations of a set of *para*-substituted benzoic acids (BAS) is shown to capture the ordering of their respective pK_a 's and UV- λ_{max} (BA refers to unsubstituted benzoic acid and S refers to the substituent). The studied molecular set consists of the following 14 members (labeled by S in order of increasing pK_a): NO₂, CN, COCH₃, CHO, Cl, F, H, NHCOCH₃, CH₃, OCH₃, OH, NH₂, and N(CH₃)₂ \approx NHCH₃; where the parent unsubstituted benzoic acid is the member with S = H.

We examine the series of 14 *para*-substituted benzoic acids (BAS) referred to above in which the common fragment (the benzene ring and the carboxylic group) are in one-to-one correspondence across the series but where S differs not only in atomic identities but also in the number of composing atoms and in the pattern by which they are bonded together. How can then these molecules be compared on equal footing?

The solution has been presented earlier in **Chapter 1**, we use the idea of the

¹Based on the Paper: Chemical Physics Letters 612 (2014) 190—197

super-atom (Eq. 1.18) so that each substituent ($S = \text{F}, \text{OH}, \text{N}(\text{CH}_3)_2$, etc.) will be treated as a single atom thus keeping all matrices of equal size.

Fig. 3.1 displays the numbering scheme of the *p*-substituted benzoic acids where the substituent S at position 15 can be a hydrogen atom (in the parent compound, benzoic acid), another atom such as F (in *p*-fluorobenzoic acid), or a (pruned) super-atom such as an OH group (in *p*-hydroxybenzoic acid).

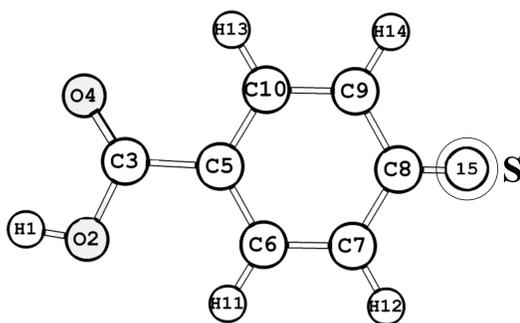


Figure 3.1: Atomic numbering scheme adopted for all the matrices in this work. Position 15 can be an atom of a super-atom as defined in the text.

We are now in a position to use the pruned LDMs in the modelling of two important properties of the studied set of *para*-substituted benzoic acids: (1) The pK_a and (2) the $\text{UV-}\lambda_{max}$. The LDMs of all studied molecules are available in **Appendix B** to 3-decimal precision in both their unpruned and pruned forms along with the corresponding atom numbering schemes.

We first note that the Frobenius distance Eq. 1.11 is a scalar distance between the matrix representatives of the studied molecules. Thus, as such, this distance contains no “direction” information, that is to say, two molecules can be equidistant from a third but flanking it on two sides, yet their Frobenius distances from that third would be identical. This is no impediment for accurate modelling of physical properties as long as the triangle inequality holds, as discussed in detail in Ref.[1],

and which has been verified in the present study. This insensitivity to the direction of the difference is not specific to LDMs but to all Euclidean distance measures of molecular (dis)similarity. A notable example of such Euclidean distance measures that has been demonstrated to be of wide versatility and predictability with respect to a wide range of properties is the so-called quantum molecular similarity approach (QTMS) of Popelier and coworkers [2-8].

Because of the direction insensitivity of dissimilarity measures the reference molecule (the origin of the distance measurement) must be one that exhibits an extreme value of the studied property, maximum or minimum within the molecular set. In this work, the molecule with the smallest value of the studied property is taken as the reference. Thus, in the case of pK_a , the reference molecule is the one with the lowest value (the most acidic molecule), namely *p*-nitrobenzoic acid, pK_a (BANO2)=3.44, while for the λ_{max} unsubstituted benzoic acid itself is the reference since it has the shortest wavelength of maximal absorption, λ_{max} (BA) = 230 nm.

3.1 Modelling of pK_a

It has been argued recently that the LDM can be biased by the diagonal elements that have magnitudes that are typically significantly larger than the off diagonal elements. Further, the diagonal elements (the LIs), scale much more rapidly with N , the total number of electrons in the molecule. The full LDM can, hence, sometimes fail to correlate with properties that are primarily electronic and independent of the core electrons such as pK_a 's, which is confirmed by our findings in the present study. Table 3.1 lists the squared correlation coefficients (r^2) obtained between inter-matrix

Frobenius distances and two experimental molecular properties, namely, pK_a and λ_{max} . The table gives r^2 values for correlation between these molecular properties and the distances from the respective reference molecules.

It can be seen from Table 3.1 that the pK_a is always best correlated with the matrix representative of the subgraph of the “active site”, that is, [COOH]. This is closely followed by the smaller subgraph [OH]. This observation indicates an almost equal capability of the LM, DM, or LDM to locate the active site “automatically” so to speak since the inclusion of more atoms (e.g. taking the matrix representatives of the full molecules) destroys the correlation. This automatic zooming on the active centre is not dissimilar to what has been achieved previously in the QTMS context [2]. The subgraph of the full carboxylic group performs slightly better in its correlation with pK_a than the [OH] subgraph which can be expected given that acidity is dependent on the ability of the entire group to accommodate a delocalized negative charge. The r^2 values for the full LM, DM, and LDM with the super-atom row/column omitted is not displayed in Table 3.1, those correlations are all still poor but are a slight improvement compared to the respective full LM, DM, and LDM.

Table 3.1: Pearson squared correlation coefficients (r^2) between calculated Frobenius distances and two molecular properties (pK_a , λ_{max}). The Frobenius distance is obtained from the localization matrices (LMs), delocalization matrices (DMs), and localization-delocalization matrices (LDMs) for the full molecule and for two subgraphs, namely, [COOH] and [OH].^(a)

Property	LM			DM			LDM		
	Full	COOH	OH	Full	COOH	OH	Full	COOH	OH
$pK_a^{(b)}$	0.027	0.981	0.972	0.349	0.986	0.966	0.159	0.970	0.973
$\lambda_{max}(nm)^{(c)}$	0.443	0.967	0.858	0.757	0.970	0.926	0.445	0.972	0.931

(a) Entries in bold typeset highlight particularly strong correlations ($r \geq 0.9$)

(b) The reference molecule is *p*-nitrobenzoic acid, $pK_a(\text{BA-NO}_2) = 3.44$

(c) The reference molecule is unsubstituted benzoic acid, $\lambda_{max}(\text{BA}) = 230 \text{ nm}$

Fig. 3.2 displays the correlations between the Frobenius distances of the DMs of the subgraphs [COOH] and [OH] from that of the reference molecule, *p*-

nitrobenzoic acid, given symbols of the form, $d_{matrixtype}^{[subgraph]}$ (reference, BA(S)), which are self-explanatory. The best model, using the DM of the [COOH] subgraph, yields the following linear fit:

$$pK_a = 3.456 + 72.990 \times d_{DM}^{COOH}(BANO_2, BAS) \quad (3.1)$$

$$[r^2 = 0.986, St.Err. = 0.0641, n = 14]$$

in which the number of parameters to data points is 1:14. The corresponding leave-one-out cross-validated linear regression coefficient is $q^2 = 0.982$, a value of only 0.4% lower than the crude r^2 which shows the absence of over-fitting and also that the model is predictive [9].

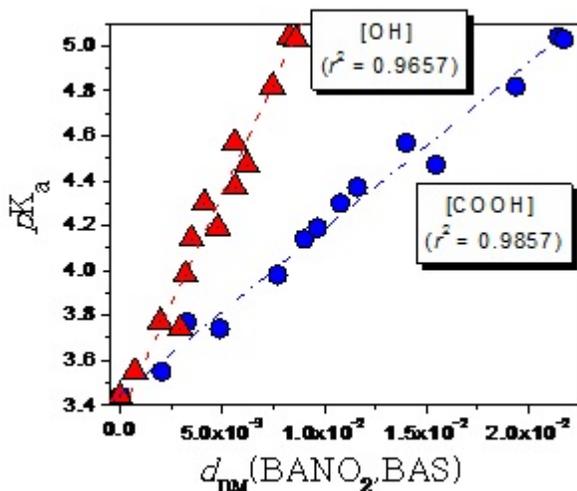


Figure 3.2: Plots of experimental pK_a 's against Frobenius distances between corresponding partial delocalization matrices (DMs) from the most acidic member of the group, BA- NO_2 (*p*-nitrobenzoic acid), which is taken as a reference. The upper plot is obtained from the partial matrices including all the atoms of the carboxylic group [COOH] while the lower plot includes only the acidic hydrogen atom and its bonded oxygen, viz., the [OH] group.

Table 3.2: Frobenius distances calculated from the DM representatives of the *para*-substituted benzoic acid derivatives from the most acidic member of the set (*p*-nitrobenzoic acid) and their corresponding experimental and calculated pK_a values.

-S	$d_{DM}^{COOH}(BANO_2, BAS) \times 10^3$	$pK_a(\text{exptl})$	Ref.	$pK_a(\text{calc})^a$	$pK_a(\text{calc})^b$
-NO ₂	0	3.44	27	3.46	3.46
-CN	2.03	3.55	1	3.60	3.62
-COCH ₃	4.88	3.74	18	3.81	3.82
-CHO	3.26	3.77	16	3.69	3.68
-Cl	7.71	3.98	1	4.02	4.02
-F	9.02	4.14	1	4.11	4.11
-H	9.66	4.19	1	4.16	4.16
-NHCOCH ₃	10.8	4.30	18	4.24	4.24
-CH ₃	11.6	4.37	1	4.30	4.30
-OCH ₃	15.5	4.47	1	4.59	4.60
-OH	14.0	4.57	1	4.48	4.47
-NH ₂	19.4	4.82	18	4.87	4.88
-N(CH ₃) ₂	21.7	5.03	18	5.04	5.05
-NHCH ₃	21.4	5.04	16	5.02	5.02

^a Calculated values were obtained from Eq. 3.1

^b Calculated values from a cross-validated leave-one-out regression model with $q^2 = 0.982$

3.2 Modelling of λ_{max}

The quantum mechanical calculation of electronic transition spectra normally requires a high level of configuration interaction. Empiricism, hence, may have a practical advantage in the prediction of such spectra. Despite that the first Hohenberg-Kohn theorem [10] has been proven for non-degenerate *ground states*, the ground state density $\rho(r)$ specifies the Hamiltonian operator $\hat{H}[\rho(r)]$ uniquely, and through the time independent many-particle Schrödinger equation, $\rho(r)$ also determines the excited states and their properties.

Thus, excited states and their properties, including their energies, are functionals of the ground-state density, even though a Hohenberg-Kohn theorem relating the excited state properties to the excited state density does not exist [11]. Since the ground state density is mapped to the energies of the ground and excited states, it is equally mapped to the differences between these energies and hence to the UV elec-

tronic transition energies and their associated wavelengths. This is why the modelling of λ_{max} given properties derived from the ground state density of wavefunctions are possible, as has recently been emphasized [1]. Buttingsrud, Alsberg, and Åstrend, for example, use optimized ground-state bond lengths and QTAIM bond critical point descriptors to accurately predict λ_{max} and excitation energies ΔE_{hv} of 191 substituted azobenzene dyes [12]. Here it is shown that the LDMs can also be used to model λ_{max} of substituted benzoic acids.

Protonated *para*-benzoic acids exhibit two UV-bands, one of high absorptivity termed the primary band centered around 230 nm and a secondary weaker band around 270 nm [13-15]. The first band, the one examined here, is attributed to an intramolecular charge transfer (CT) [15], and the second, to a shifted benzene band. The 230 nm band undergoes a bathochromic shift upon substitution of the aromatic ring with a substituent S, irrespective of the electron donating or withdrawing nature of S [13]. Electron withdrawing groups do not alter the wavelength of the secondary band unless these substituents are themselves chromophores such as $-\text{NO}_2$ and $-\text{NHCOCH}_3$ [14] (due to their significant π -character). Benzoic acids substituted by these two chromophoric substituents were excluded from the statistical correlation due to this interference. By only shifting the primary band to longer wavelengths without affecting the secondary band, non-chromophoric electron withdrawing groups can hence lead to the overlap of the secondary and the primary bands in some cases. On the other hand, electron donating groups, in addition to their bathochromic shift of the primary band also increase the wavelength and the intensity of the secondary band [13]. In this work, the (shifted) first band is the subject of the modelling with both the Hammett σ -constants, as a standard reference, and with the Frobenius inter-matrix distances to elucidate their predictive performance. Table 3.1 shows that the 230 nm

λ_{max} value of unsubstituted BA is best predicted by the LDM of [COOH], followed closely by the [OH] LDM. The inclusion of additional atoms considerably reduces the r^2 value. The best correlation of λ_{max} is obtained with the LDM-distance of the [COOH] ($r^2 = 0.972$) closely followed by its DM-distance ($r^2 = 0.970$). Using the full molecule for the LDM distance calculation yields an r^2 value of 0.445, using instead the full molecule minus the S substituent yields an r^2 value of 0.851. An LDM distance using the benzene ring only also yields an r^2 value of 0.851.

The correlation between $d_{LDM}^{[COOH]}(BA, BAS)$ and the eight available experimental λ_{max} values (seven substituted benzoic acids, in addition to the parent compound, after excluding $-\text{NO}_2$ and $-\text{NHCOCH}_3$), is displayed in Fig. 3.3 (a) which is the best model with $r^2 = 0.972$. The statistical fitting yields the following regression equation:

$$\lambda_{max}(nm) = 223.50 + 3.4171 \times 10^3 \times d_{LDM}^{[COOH]}(BA, BAS) \quad (3.2)$$

$$[r^2 = 0.973, St.Err. = 5.74, n = 8]$$

in which the number of parameters to data points is 1:8. The corresponding leave-one-out cross-validated linear regression coefficient $q^2 = 0.944$, again indicating little over-fitting and strong predictivity [9].

Table 3.3 is sorted in order of increasing experimental λ_{max} values from the shortest wavelength of 230 nm (BA to the longest in the set of 315 nm (p -(CH_3)₂N-BA) and the corresponding Frobenius distances from BA ($d_{LDM}^{[COOH]}(BA, BAS)$). The experimental λ_{max} values and those calculated from the model expressed in Eq. 3.2 agree to within a mean absolute deviation (MAD) of about 4.0 nm and a root mean square deviation (RMSD) of 4.9 nm. The equivalent cross-validated values are MAD = 8.9 nm and RMSD = 11.1 nm. Table 3.3 also lists the Hammett σ_{para} -constants obtained from

the monograph by Hansch and Leo [19] and the corresponding values calculated from the following regression model:

$$\lambda_{max}(nm) = 238.55 - 75.46176 \times \sigma_{para} \quad (3.3)$$

$$[r^2 = 0.859, St.Err. = 12.86, n = 8]$$

a correlation which is also displayed graphically in Figure 3.3 (b). The cross-validated q^2 corresponding to Figure 3.3 (b) is only 0.652, significantly lower, indicating poor predictivity of the model based on the Hammett constants. The correlation of λ_{max} with the Hammett constants features a significant outlier: S=Cl. As mentioned above, *the bathochromic shift is independent of the direction of electronic charge flow to or from the substituent*, yet Hammett constants by construction account for such directional charge flow and which is reflected into the sign of the σ -constants ($0 < \sigma$ for electron withdrawing groups and $0 > \sigma$ for electron donating groups). Since Cl is the only member listed in Table 3.3 that is electron withdrawing and which was *not* excluded from the statistical fittings, it clearly reduces the strength of the statistical correlation based on Hammett constants. If this outlier is removed, however, the following fitted equation results:

$$\lambda_{max}(nm) = 225.99 - 96.8434 \times \sigma_{para} \quad (3.4)$$

$$[r^2 = 0.970, St.Err. = 4.18, n = 7]$$

which has a linear correlation coefficient that is significant higher than Eq. 3.3

(and a q^2 of 0.940) yet still outperformed by the model based on the LDM especially given that the latter incorporates the S=Cl atom and also as a result has 8 data points as opposed to 7.

Table 3.3: Frobenius distances calculated from the LDM representatives of nine *para*-substituted benzoic acid derivatives from the member with the shortest λ_{max} of the set (unsubstituted benzoic acid) and their corresponding experimental and calculated λ_{max} values.

-S	$d_{LDM}^{[OH]}(BA, BAS) \times 10^3$	σ_{para}^a	$\lambda_{max}(exptl)$	Ref.	$\lambda_{max}(g.c.)^a$	$\lambda_{max}(calc)^b$	$\lambda_{max}(calc)^d$
-H	0	0.00	230	14	230	223	239
-CH ₃	5.37737	-0.17	240	14	240	241	251
-Cl	7.47910	0.23	242	14	240	248	221
-OH	8.54205	-0.37	254	14	255	252	266
-OCH ₃	11.91509	-0.27	256	14	255	263	259
-NO ₂	25.65405	0.78	262 ^e	14	NA ^e	e	e
-NHCOCH ₃	3.86811	0.00	269 ^e	20	275	e	e
-NH ₂	19.45621	-0.66	288 ^b	13	288	289	288
-NHCH ₃	23.94033	-0.84	303 ^b	13	303	304	302
-N(CH ₃) ₂	25.12666	-0.83	315 ^b	13	315	308	301
$-r^2$	0	0	0	0	0.995	0.973	0.859
$-q^2^f$	0	0	0	0	0.992	0.944	0.652

a The empirical Hammett σ_{para} constants are obtained from Ref.18

b Calculated from group contributions (g.c.)

c Calculated from Eq. 3.2

d Calculated from Eq. 3.3

e The -NHCOCH₃ and -NO₂ groups are $\pi - \pi^*$ chromophores that contribute bands with λ_{max} that overlap with that of benzoic acid and hence were excluded from the modelling (see text).

f Leave-one-out cross validated squared linear regression coefficient.

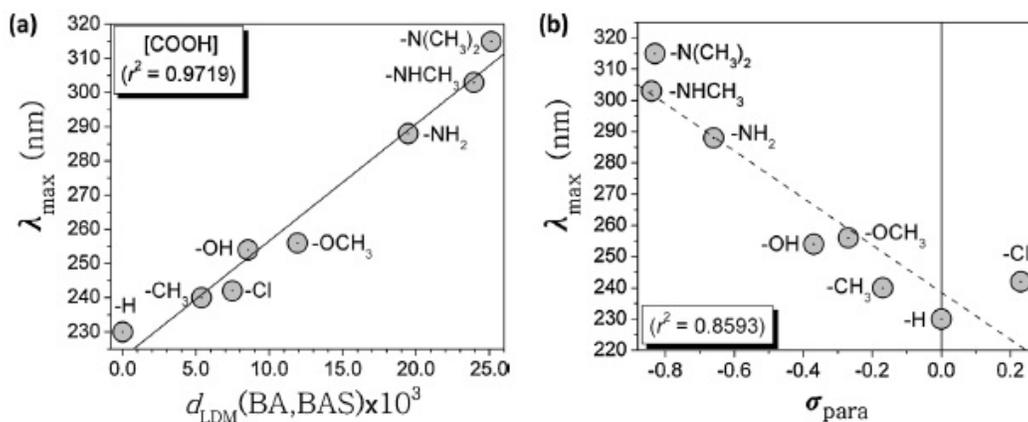


Figure 3.3: (a) Plots of experimental λ_{max} values against Frobenius distances between corresponding partial localization-delocalization matrices (LDMs) of the [COOH] subgraph taking unsubstituted benzoic acid taken as the reference ($\lambda_{max} = 230nm$). (b) Plot of experimental λ_{max} values against the Hammett σ_{para} substituent constants.

3.3 Conclusion

These are promising results that call for further verification with more compounds and test cases. It is remarkable, yet not uncommon, that the pK_a which is the negative logarithm of the equilibrium acidity constant ($-\log K_a$) that depends on both the acid and its conjugate base in aqueous medium, can be predicted from an examination of the properties of the undissociated acid in the gas-phase.

The entry for the pK_a in several successive issues of the *CRC Handbook of Chemistry and Physics* [16,17] for *p*-dimethylaminobenzoic acid (*p*-DMABA) is erroneously entered as 6.03, a value which when incorporated into our initial modelling constituted a significant outlier. This value is inconsistent with a similar molecule, namely, *p*-methylaminobenzoic acid (*p*-MABA), which cannot be expected to have a considerably different pK_a and which has an entry of 5.04 in the *CRC Handbook* [16]. Further search of literature for this pK_a confirmed our suspicion and Ref.[18] gives a value of 5.03 for *p*-DMABA which we included in Table 3.2 rather than the much higher 6.03 of the *CRC Handbook*. Moreover, the authoritative monograph by Hansch and Leo [19] gives a Hammett σ -constant of -0.66 in the case of *p*-aminobenzoic acid (*p*-ABA), which translates into a $pK_a(p\text{-ABA}) = pK_a(\text{BA}) - \sigma = 4.19 + 0.66 = 4.85$ (consistent with the tabulated value in Table 3.2 obtained from Ref.[18]) and which cannot also be expected to be that different from the pK_a value of *p*-DMABA. On the other hand, the tabulated σ value [20] for $-N(CH_3)_2$ is -0.83 which yields a pK_a of 5.02 which is close to the reported directly-determined value in the literature [18]. We undertake this opportunity to correct the record especially since the erroneous value of 6.03 has propagated in numerous other references and websites.

The modelling based on the LDM is also shown capable of empirical prediction

of the substituents effects on the UV absorption well, better than the Hammett constants. The failure of the latter has recently been noted by Smith et al. and has been attributed to their roots in the ground state equilibrium constants or bond dissociation energies, while UV transitions reflect energy gaps between the ground and excited states [14].

3.4 Computational Methods

The level of theory used in this work is density functional theory [21,22] (DFT), with the hybrid B3LYP functional[23,24] along with the 6-311++G(d,p) basis set, denoted by B3LYP 6-311++G(d,p). Geometries were optimized and the final wavefunction/s/densities obtained at the same level of theory, followed by (harmonic) vibrational frequency analysis to ensure the absence of any imaginary frequencies. All electronic structure calculations and harmonic frequencies were calculated using the Gaussian 09 software[25]. The subsequent QTAIM analysis was performed using the AIMAll/AIMStudio package[26].

3.5 Acknowledgements

The authors are indebted to Professor Cory C. Pye (*Saint Mary's University*) for helpful discussion regarding the pruning of peripheral groups. Financial support for this work was provided by the *Natural Sciences and Engineering Research Council of Canada* (NSERC), *Canada Foundation for Innovation* (CFI), *Saint Mary's University*, *McMaster University* and *Mount Saint Vincent University*.

3.6 References

- [1] C.F. Matta, *J. Comput. Chem.* 35 (2014) 1165
- [2] Popelier, P. L. A. Quantum Molecular Similarity. 1. BCP space. *J. Phys. Chem. A* 1999, 103, 2883-2890.
- [3] O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. Part 2: the relation between properties in BCP space and bond length. *Can. J. Chem.* 1999, 77, 28-36.
- [4] O'Brien, S. E.; Popelier, P. L. A. Quantum molecular similarity. 3. QTMS descriptors. *J. Chem. Inf. Comput. Sci.* 2001, 41, 764-775.
- [5] O'Brien, S. E.; Popelier, P. L. A. Quantum topological molecular similarity. Part 4. A QSAR study of cell growth inhibitory properties of substituted (E)-1-phenylbut-1-en-3-ones. *J. Chem. Soc., Perkin Trans. 2* 2002, 478-483.
- [6] Popelier, P. L. A.; Chaudry, U. A.; Smith, P. J. Quantum topological molecular similarity. Part 5. Further development with an application to the toxicity of polychlorinated dibenzo-p-dioxins (PCDDs). *J. Chem. Soc., Perkin Trans. 2* 2002, 1231-1237.
- [7] Harding, A. P.; Wedge, D. C.; Popelier P. L. A. pKa Prediction from "quantum chemical topology" descriptors. *J. Chem. Inf. Mod.* 2009, 49, 1914-1924.
- [8] P.L.A. Popelier, in: C.F. Matta (Ed.), *Quantum Biochemistry: Electronic Structure and Biological Activity*, Wiley-VCH, Weinheim, 2010, p. 669.
- [9] R.R. Picard, R.D. Cook, *J. Am. Stat. Assoc.* 79 (1984) 575.
- [10] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* 1964, 136, 864-871.
- [11] Gaudoin, R.; Burke, K. Lack of Hohenberg-Kohn theorem for excited states.

- Phys. Rev. Lett. 2004, 93, 173001(1-4).
- [12] Buttingsrud, B.; Alsberg, B. K.; strand, P.-O. Quantitative prediction of the absorption maxima of azobenzene dyes from bond lengths and critical points in the electron density. *Phys. Chem. Chem. Phys.* 2007, 9, 2226-2233.
- [13] Pavia, D. L.; Lampman, G. M.; Kriz, G. S.; Vyvyan, J. R. *Introduction to Spectroscopy* (4th Edition); Brooks/Cole Cengage Learning: Belmont, CA, USA, 2009.
- [14] H.-B. Guo, F. He, B. Gu, L. Liang, J.C. Smith, *J. Phys. Chem. A* 116 (2012) 11870.
- [15] Kamath, B. V.; Mehta, J. D.; Bafna, S. L. Ultraviolet absorption spectra: Some substituted benzoic acids. *J. Appl. Chem. Biotechnol.* 1975, 25, 743-751.
- [16] Lide, D. R. *CRC Handbook of Chemistry and Physics* 88th Edition; CRC Press: 2007-2008.
- [17] Lide, D. R. *CRC Handbook of Chemistry and Physics* 87th Edition; CRC Press: 2006.
- [18] Jover, J. ; Bosque, R.; Sales, J. QSPR prediction of pKa for benzoic acids in different solvents. *QSAR Combin. Sci.* 2008, 27, 563-581.
- [19] Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
- [20] C. Hansch, A. Leo, *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, 1995.
- [20] Lide, D. R.; Milne, G. W. A. *Handbook of Data on Organic Compounds* (3rd Edition), Vols. 1- 7; CRC Press: Boca Raton, FL, USA, 1994.
- [21] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.

- [22] Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, (Second Edition); Wiley-VCH: New York, 2001.
- [23] Becke, A. Density-functional thermochemistry .3. The role of exact exchange. *J. Chem. Phys.* 1993, 98, 5648-5652.
- [24] Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* 1988, 37, 785-789.
- [25] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J. A.; Peralta, J. E. ; Ogliaro, F.; Bearpark, M.; Heyd, J. J. ; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., *Gaussian 09*, Revision B.01 (Gaussian Inc., Wallingford CT, 2010).
- [26] Keith, T. A. AIMAll. <http://aim.tkgristmill.com/> 2011.
- [27] PubChem, The National Center for Biotechnology Information, NIH, <http://pubchem.ncbi.nlm.nih.gov/>, 2014.

Aromaticity of Rings-in-Molecules (RIMs) from Electron Localization-Delocalization Matrices (LDMs)¹

There has been a resurgence of intense interest in quantifying or even defining the concept of aromaticity especially with the advent of measures of electron delocalization in the 1990s [1-10]. Aromaticity is loosely defined as the tendency of an aromatic ring to react by substitution rather than addition. The various definitions tend to fall into groups that are primarily structural [11-13], reactive [14-17], energetic and thermochemical [17-20], magnetic [1,3,7,21-25], electronic [5-7,9,10,17,26-52], those based on the topological properties of the electron density and/or of the electrostatic potential [6,25,29,52-55], and chemical graph theoretical (CGT) [30,56-61].

Schleyer *et al.* define aromaticity as “*a manifestation of electron delocalization in closed circuits, either in two or three dimensions*” [4]. Several authors have already developed measures of aromaticity that are based on the characteristics of electron delocalization as quantified, for example, by the delocalization indices [45,62,63] of the Quantum Theory of Atoms in Molecules (QTAIM) [64-66]. DIs, whether two-centered [45,62,63] or multi-centered [9, 31, 32], are ideally suited for the study of aromaticity since they can quantify the delocalization of the electronic charge in closed two- or three-dimensional rings manifested in the observed structural, magnetic, and energetic characteristics of aromaticity [26].

LDMs and their closely related delocalization only matrices are used in this study to measure the similarity distance of a ring from benzene and investigate the cor-

¹Based on the Paper: Phys. Scr. 91 (2016) 013001 (13pp)

relation of this distance with well-established structural, electronic, and magnetic aromaticity measures. In other words, we report similarity distances of rings-in-molecules (RIMs) to benzene followed by a statistical comparison to some commonly used/popular aromaticity indices.

No known single criterion can encapsulate or measure aromaticity which is inherently multifaceted and multidimensional. Because of that, aromaticity measures not infrequently disagree in ranking the aromaticities of RIMs [67].

In this work, the similarity of a six-membered carbon ring in a molecule to the carbon ring in benzene, as quantified by LDMs distance (Eq. 1.8), is correlated with independent known measures of aromaticity. Next we investigate the correlation of the eigenvalues, invariants that are independent of comparisons with a reference such as benzene, with aromaticity indices.

4.1 Rings-in-molecules (RIMs)

The rings considered in this study are all six-membered carbon rings that occur in polycyclic benzenoid hydrocarbons. As the number of hydrogen atoms attached to a ring in a molecule depends on the immediate neighbourhood, only the carbon skeleton of a given ring-in-molecule (RIM) is considered. Carbon atoms that belong to more than one ring simultaneously are included in each of the rings being considered. For example, phenanthrene (Fig. 4.1), is split into three separate ring-LDMs (labeled A-C).

Following the labelling in Fig. 4.1, the three RIM-LDMs of phenanthrene are written (at the HF/6-31G(d) level, to three decimals):

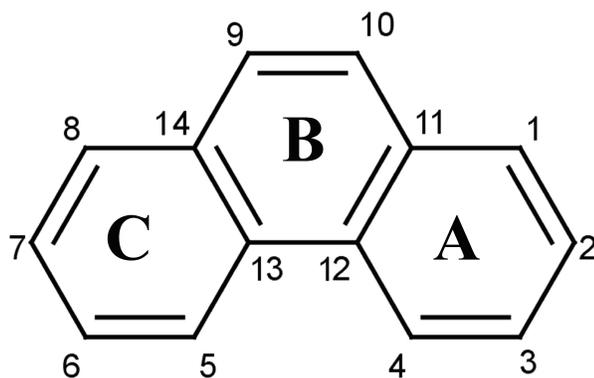


Figure 4.1: Phenanthrene and its atom and ring labelling scheme.

$$LDM_A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 12 & 11 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 12 \\ 11 \end{matrix} & \begin{pmatrix} 3.957 & 0.746 & 0.036 & 0.048 & 0.031 & 0.637 \\ 0.746 & 3.953 & 0.653 & 0.036 & 0.038 & 0.034 \\ 0.036 & 0.653 & 3.950 & 0.744 & 0.034 & 0.037 \\ 0.048 & 0.036 & 0.744 & 3.951 & 0.640 & 0.031 \\ 0.031 & 0.038 & 0.034 & 0.640 & 3.900 & 0.654 \\ 0.637 & 0.034 & 0.037 & 0.031 & 0.654 & 3.891 \end{pmatrix} \end{matrix} \quad (4.1)$$

$$LDM_B = \begin{matrix} & \begin{matrix} 9 & 10 & 11 & 12 & 13 & 14 \end{matrix} \\ \begin{matrix} 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \end{matrix} & \begin{pmatrix} 3.958 & 0.822 & 0.033 & 0.024 & 0.027 & 0.573 \\ 0.822 & 3.958 & 0.573 & 0.027 & 0.024 & 0.033 \\ 0.033 & 0.573 & 3.891 & 0.654 & 0.026 & 0.017 \\ 0.024 & 0.027 & 0.654 & 3.900 & 0.559 & 0.026 \\ 0.027 & 0.024 & 0.026 & 0.559 & 3.900 & 0.654 \\ 0.573 & 0.033 & 0.017 & 0.026 & 0.654 & 3.891 \end{pmatrix} \end{matrix} \quad (4.2)$$

$$LDM_C = \begin{matrix} & \begin{matrix} 5 & 6 & 7 & 8 & 14 & 13 \end{matrix} \\ \begin{matrix} 5 \\ 6 \\ 7 \\ 8 \\ 14 \\ 13 \end{matrix} & \begin{pmatrix} 3.951 & 0.744 & 0.036 & 0.048 & 0.031 & 0.640 \\ 0.744 & 3.950 & 0.653 & 0.036 & 0.037 & 0.034 \\ 0.036 & 0.653 & 3.953 & 0.746 & 0.034 & 0.038 \\ 0.048 & 0.036 & 0.746 & 3.957 & 0.637 & 0.031 \\ 0.031 & 0.037 & 0.034 & 0.637 & 3.891 & 0.654 \\ 0.640 & 0.034 & 0.038 & 0.031 & 0.654 & 3.900 \end{pmatrix} \end{matrix} \quad (4.3)$$

In contrast with the full molecular LDM, the sum of the matrix elements of these partial LDMs will generally not yield an integer number of electrons since some electrons will always be shared with the hydrogen atoms, the substituents, or the other fused rings.

The Frobenius distance of the LDM representative of a RIM to the LDM repre-

sentative of the carbon circuit of benzene is invariable to labelling as long as the ring atoms are labelled in the same order as benzene. For example, if we choose to construct the RIM-LDM matrix by listing one of the *ortho*-carbon atoms as the second atom (C2) immediately following any arbitrary choice (and the only arbitrary choice) for the *ipso*-carbon atom (C1), the *meta*-carbon atom attached to C2 as the third (C3), the *para*- as the fourth (C4), the second *meta*- as the fifth (C5), and the second *ortho*- as the sixth (C6), then the Frobenius distance from benzene is insensitive to the arbitrary choice of C1 as long as we follow the same numbering algorithm for both the RIM and for benzene.

4.2 The molecular set

The chemical structures of the molecular set used in this study are depicted in Fig. 4.2. The set includes the reference molecule (benzene), three linear cata-condensed polycyclic aromatic benzenoid hydrocarbons (PABH) (naphthalene, anthracene, and naphthacene), two zigzag cata-condensed PABHs (phenanthrene, chrysene), a branched cata-condensed PABH (triphenylene), and cyclohexane in the most stable (chair) conformations as an extreme reference for a non-aromatic ring.

There are in total 8 molecules and 13 symmetry-distinct rings. We introduce the following symbols for the 13 different rings where (I) and (O) symbolizes the inner- or outer-ring respectively. Benzene = Ben, naphthalene = N, anthracene = A, naphthacene = Nc, phenanthrene = P, chrysene = Ch, triphenylene = T, and cyclohexane = Cyc. Thus the symbol P(O) signifies the outer ring in phenanthrene. The complete set of symbols for every ring listed in Table 4.1.

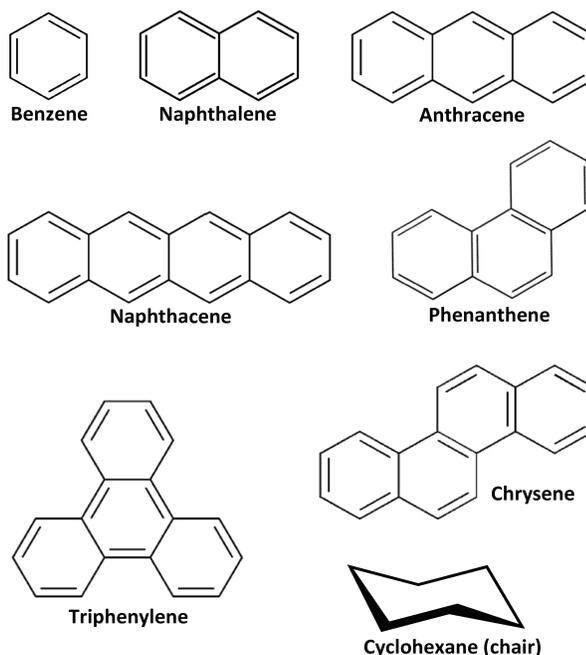


Figure 4.2: Molecular set supplying the “rings-in-molecules (RIMs)” for this study.

4.3 Computational details

Quantum chemical calculations were performed at the Hartree-Fock (HF) level using a 6-31G(d) basis set, the same level of theory used in previous studies [42], with which the current results are being compared. Geometries were first optimized then the wavefunctions obtained at this level of theory which is denoted in standard notation as HF/6-31G(d). All electronic structure calculations were performed using the Gaussian software [68]. The resulting wavefunctions were then subjected to QTAIM integrations using AIMAll/AIMStudio program [69] to calculate the LIs and DIs. The program AIMLDM [70] was then applied to the AIMAll sum files to extract the LDM for the entire molecular set followed by the extraction of matrix invariants and Frobenius distances. In total we have 13 symmetry-unique different RIMs being analyzed in this work.

Table 4.1: Aromatic rings in the molecules displayed in Fig. 4.2 sorted in order of increasing dissimilarity to benzene as measured by the Frobenius distance and four corresponding common indices of aromaticity.

Molecule	Ring	Code ^(a)	d_{FROB}	HOMA ^(b)	PDI ^(b)	FLU ^(b)	NICS(0) ^(b)
Benzene		Ben	0.000	1.00	0.105	0.00	-11.5
Triphenylene	Outer	T(O)	0.163	0.930	0.086	0.003	-10.6
Phenanthrene	Outer	P(O)	0.199	0.902	0.082	0.005	-11.4
Chrysene	Outer	Ch(O)	0.230	0.859	0.079	0.008	-11.1
Anthracene	Inner	A(I)	0.242	0.884	0.070	0.007	-14.2
Naphthalene		N	0.282	0.779	0.073	0.012	-10.9
Naphthacene	Inner	Nc(I)	0.294	0.774	0.063	0.011	-13.8
Chrysene	Inner	Ch(I)	0.357	0.553	0.052	0.019	-8.2
Anthracene	Outer	A(O)	0.386	0.517	0.059	0.024	-8.70
Phenanthrene	Inner	P(I)	0.403	0.402	0.053	0.025	-6.80
Triphenylene	Inner	T(I)	0.431	0.067	0.025	0.027	-2.60
Naphthacene	Outer	Nc(O)	0.442	0.325	0.051	0.031	-6.70
Cyclohexane	Chair	Cyc	0.741	-4.34	0.007	0.091	-2.10
$r^{2(c)}$				0.978	0.917	0.858	0.608
adjusted- $r^{2(c)}$				0.973	0.909	0.845	0.572
Order of polyn.				2	1	1	1

(a) Unique short-hand code notation to identify the 13 symmetry-unique rings subject of this work.

(b)...Data obtained from Ref:65.

(c) The statistical model is a polynomial of the form: Aromaticity index = $a_0 + a_1 \times d_{Frob} + a_2 \times d_{Frob}^2$. The model yields the following fitting constants: HOMA: $a_0 = 0.6821$, $a_1 = 5.3303$, $a_2 = -16.2087$; PDI: $a_0 = 0.107$, $a_1 = -0.140$; FLU: $a_0 = -0.0193$, $a_1 = 0.1232$; NICS(0): $a_0 = -14.48$, $a_1 = 16.70$.

4.4 Aromaticity measures and Eigenvalues

We first investigate the statistical correlations between the Frobenius distances of the RIMs in the molecular set in Fig. 4.2 and some of the well-established aromaticity criteria, namely, the harmonic oscillator model of aromaticity HOMA (*structural*) [12,13], the nucleus independent chemical shift (NICS) (*magnetic*) [1], the aromatic fluctuation index (FLU) [42,46], and the para delocalization index (PDI) (*electron delocalization*) [40].

4.4.1 Definitions of the measures of aromaticity considered in this work

The structural index we consider in this study is the popular Krygowski HOMA index which is defined as [12,13]:

$$HOMA = 1 - \frac{\alpha}{m} \sum_{i=1}^m (R_{opt} - R_i)^2 \quad (4.4)$$

where m is the number of bonds in the ring ($m = 6$ for all rings considered in this study), α is a parameter which equals 257.7 for carbon-carbon bonds that yields 0 (non-aromatic ring) \leq HOMA \leq 1 (benzene, where all bond lengths are equal in lengths $R_i = R_{opt} = 1.388\text{\AA}$).

The NICS index, extensively studied by Schleyer and coworkers, is the chemical shift at the ring center and has a negative value for aromatic systems and a positive value for anti-aromatic systems. This quantity is called NICS(0) to indicate that it is evaluated in the ring plane [3], and is the sole NICS that is considered in this work, hence we will drop the (0) designation from now on. The more negative the value of NICS indicates a more aromatic system. There are however odd results as some rings (e.g. central ring in anthracene) can give values for NICS that are more negative than benzene itself [42]. Such artefacts prompted the workers in this domain to introduce modifications into the NICS e.g. by measuring above the center of the ring by a given distance perpendicular to the ring plane [7]. However, NICS evaluated at the center of the ring appears to remain the most used and is the one considered in the comparisons described below.

Important electron aromaticity indices, developed and extensively studied by Solá

and coworkers, include the aromatic fluctuation index (FLU) [42,46] and the para delocalization index (PDI) [40]. The first aromaticity index, the FLU, measures the fluctuation of the DI among neighbouring atoms within a ring. Just as the structural HOMA index, a lack of fluctuation indicates a higher aromaticity as long as the value of the DI is close to that of the prototype aromatic molecule, benzene. The FLU index is, thus, an excellent electronic counterpart to the HOMA as it captures the cyclic delocalization of electrons in a given RIM. The index has been defined as [42]:

$$FLU = \frac{1}{m} \sum_{\Omega-\Omega'}^{RIM} \left\{ \left[\frac{V(\Omega')}{V(\Omega)} \right]^\alpha \left[\frac{\delta(\Omega, \Omega') - \delta(\Omega, \Omega')_{ref}}{\delta(\Omega, \Omega')_{ref}} \right] \right\}^2 \quad (4.5)$$

where the summation runs over all atoms sharing a bond path (bonded/neighbouring atoms) in the ring, m = the number of atoms forming the ring ($m = 6$ for all the 13 rings considered in the present work), $\delta(\Omega, \Omega')_{ref} = 1.4$ (the value obtained at the HF/6-31G(d) level for benzene), and $V(\Omega)$ is defined as:

$$V(\Omega) = \sum_{\Omega' \neq \Omega} \delta(\Omega, \Omega') \quad (4.6)$$

and termed the “*global delocalization*” (or valency) of Ω (which equals to twice of the sum of the row or column of the off-diagonal elements of the LDM labeled Ω), and $\alpha = \pm 1$ to ensure $[V(\Omega')/V(\Omega)]^\alpha \geq 1$.

The second electronic aromaticity index we consider here is known as the PDI [40]. This index is the average of the DI between *para*-atoms in a ring and hence is limited to 6-membered rings (6-MRs), which is not a limitation in this study since all the 13

rings we consider are 6-MRs. Thus this index can be written compactly as:

$$PDI(\Omega) = \frac{1}{3} \sum_{i=1}^3 \delta(\Omega_i, \Omega_{i+3}) \quad (4.7)$$

4.4.2 Correlations of the Frobenius distances from benzene with established measures of aromaticity

Three of the four studied aromaticity indices (HOMA, PDI, and FLU) are strongly correlated statistically with the Frobenius distance from benzene ($r^2 > 0.86$). NICS is not as strongly correlated and exhibits more scatter along the trend line in addition to some apparent outliers. The trends of these correlations are essentially linear in the cases of FLU and PDI and non-linear in the case of HOMA and NICS, all of which are displayed in Fig. 4.3 and the values upon which the figure is based on appear in Table 4.1. Despite falling on the general trend lines and its inclusion in the statistical analysis, the results of which appear at the bottom of Table 4.1, cyclohexane has been excluded from Fig. 4.3.

The strongest correlation of the Frobenius distance is with HOMA (r^2 -adjusted = 0.97) and is clearly nonlinear. The PDI, that measures the average QTAIM para-DIs within a 6-MR, is the second most strongly (and linearly) correlated to the Frobenius distance with an r^2 -adjusted of 0.91. The next in strength of correlation is the FLU which measures the fluctuation in the DI within a ring (r^2 -adjusted = 0.85, linear).

The NICS, which has a generally more negative value for the more aromatic ring, is *not* maximally negative for benzene (the inner ring of anthracene has this title) has a generally increasing trend with distance from benzene but the correlation is not as strong as the other indices as can be seen from Fig. 4.3.

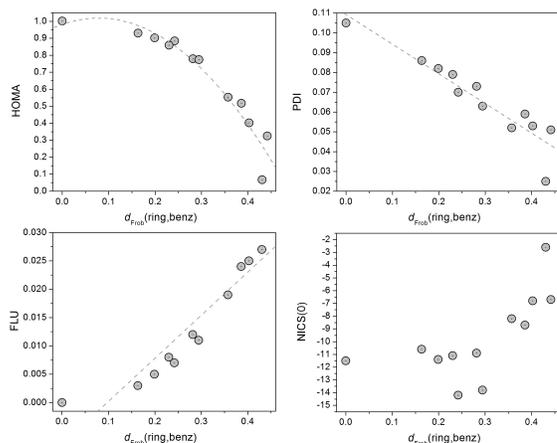


Figure 4.3: Correlations between the Frobenius distance from benzene of rings in the molecules listed in Table 4.1 and depicted in Fig. 4.2 and four common aromaticity indices: HOMA, PDI, FLU, and NICS(0). The Frobenius distance clearly captures more than one aspect of aromaticity as measured by these widely differing criteria (see text for further discussion)

Table 4.1 has been sorted in order of increasing distance (increasing dissimilarity) from benzene. The lower members in this table are thus the least similar to benzene. Cyclohexane appears at the very bottom of Table 4.1 as expected and is the least aromatic according to all criteria that are listed in the table. In summary, Table 4.1 and Fig. 4.3 show that these different aromaticity measures generally, *but not always*, are well correlated among themselves and with the Frobenius distance from benzene.

The qualitative ranking of aromaticity by various methods can be gleaned from Table 4.2. This table lists the RIMs starting by the most aromatic at the top then lists the various numbers of disagreements with the other studied methods. Unsurprisingly, HOMA, PDI, and FLU all list benzene – also the reference for the Frobenius distance calculation – as the most aromatic ring in the set. Surprisingly, however, and in disagreement with all other methods, NICS predicts that the inner ring of anthracene is more aromatic than benzene, and so is naphthalene’s inner ring as well.

All four methods rank the inner ring of anthracene as more aromatic than the outer

Table 4.2: Aromatic ranking agreement of various aromaticity indices with the Frobenius distance dissimilarity to benzene.

d_{Frob}	HOMA	PDI	FLU	NICS(0)
Benzene	Benzene	Benzene	Benzene	Anthracene(I)
Triphenylene(O)	Triphenylene(O)	Triphenylene(O)	Triphenylene(O)	Naphthacene(I)
Phenanthrene(O)	Phenanthrene(O)	Phenanthrene(O)	Phenanthrene	Benzene
Chrysene(O)	Anthracene(I)	Chrysene(O)	Anthracene(I)	Phenanthrene(O)
Anthracene(I)	Chrysene(O)	Naphthalene	Chrysene(O)	Chrysene(O)
Naphthalene	Naphthalene	Anthracene(I)	Naphthacene(I)	Naphthalene
Naphthacene(I)	Naphthacene(I)	Naphthacene(I)	Naphthalene	Triphenylene(O)
Chrysene(I)	Chrysene(I)	Anthracene(O)	Chrysene(I)	Anthracene(O)
Anthracene(O)	Anthracene(O)	Phenanthrene(I)	Anthracene(O)	Chrysene(I)
Phenanthrene(I)	Phenanthrene(I)	Chrysene(I)	Phenanthrene(I)	Phenanthrene(I)
Triphenylene(I)	Naphthacene(O)	Naphthacene(O)	Triphenylene(I)	Naphthacene(O)
Naphthacene(O)	Triphenylene(I)	Triphenylene(I)	Naphthacene(O)	Triphenylene(I)
Cyclohexane	Cyclohexane	Cyclohexane	Cyclohexane	Cyclohexane
No. disag. d_{Frob}	4	7	4	10
No. disag. HOMA		6	3	7
No. disag. PDI			9	9
No. disag. FLU				10

ring, the Frobenius distance criterion appears to place it at a reasonable relative ranking whereby the outer ring is four ranks below the inner ring (5 ranks below according to both HOMA and FLU, 7 ranks below according to NICS, and only 2 ranks below according to PDI). The reverse situation is observed for phenanthrene where all methods rank the outer ring as more aromatic and where the Frobenius rankings appear as a good compromise. The ranking ordering of the d_{Frob} is closest to the HOMA and furthest from NICS. Only FLU is closer to the ranking of HOMA than the Frobenius distance with three disagreements, but the disagreements between d_{Frob} are slight and consist of the interchange of two neighbouring-ranking pairs: Anthracene (inner) and chrysene (outer), and naphthacene (outer) and triphenylene (inner).

4.4.3 Correlations of aromaticity with the eigenvalues of the RIM-LDM

One of the earliest introductions to empirical “rules of thumb” that we are exposed to in our chemical education is “*like dissolves like*”. This and several similar empirical rules of thumb has been made rigorous in the form of Hansen’s Solubility Parameters [71]. The “*like dissolves like*” rule is fundamentally based on the concept of chemical similarity. Chemical (or molecular) similarity has its basis in the observation that similar compounds have similar properties. Chemical/molecular similarity is one of the most important concepts in the field of cheminformatics where it plays an important role in predicting the properties of compounds, selecting sets of chemical compounds with predefined sets of properties and screening large structure databases to find “hits”, that is, possible new active drugs.

What we would like to examine here is whether similarities of one matrix invariant (the eigenvalues) of the RIMs’ LDMs parallel established aromaticity measures. For each of the “ring in molecule” there are six carbon atoms represented by a 6×6 LDM. Therefore there will be six eigenvalues for each RIM. The six eigenvalues extracted from the LDMs can be thought of as the rings vector location in six dimensional space.

The similarities of the RIM can be assessed through a pairwise similarity matrix generated by comparing the vector angle of the vectors from the LDMs eigenvalues of each of the ring in molecules. The angle between two vectors is given as usual by:

$$\alpha = \arccos \left(\frac{\vec{v}_1 \cdot \vec{v}_2}{v_1 v_2} \right) \quad (4.8)$$

where the vectors represent the position of the RIM in the 6-dimensional eigenvalue space.

Table 4.3 lists the pairwise vector angles for the 13 studied RIMs. One way to appreciate the similarity of these ring in molecules would be to map the molecules in n -dimensional abstract mathematical space and use the distance between the rings as a measure of aromaticity when compared to benzene. It is difficult to visualize relationships beyond three dimensions and, consequently, dimensionality reduction is necessary if we are to visualize similarity distance between sets of rings. This dimensionality reduction is achieved through the so-called multidimensional scaling (MDS) methods (discussed previously) [72-78].

Table 4.3: Pairwise vector angles (in degrees ($^{\circ}$)) matrix for the ring in molecules to three decimals*.

RIM	Ben	A(O)	A(I)	P(O)	P(I)	N	Nc(O)	Nc(I)	Ch(O)	Ch(I)	T(O)	T(I)	Cyc
Ben	0.00	0.94	0.94	0.44	1.14	0.64	1.12	1.07	0.51	0.95	0.33	1.76	3.93
A(O)	0.94	0.00	0.39	0.52	0.50	0.31	0.20	0.40	0.45	0.42	0.62	1.41	3.49
A(I)	0.94	0.39	0.00	0.52	0.29	0.39	0.50	0.15	0.47	0.21	0.61	1.11	3.20
P(O)	0.44	0.52	0.52	0.00	0.74	0.20	0.71	0.65	0.07	0.55	0.11	1.48	3.64
P(I)	1.14	0.50	0.29	0.74	0.00	0.60	0.54	0.21	0.68	0.20	0.83	0.92	3.01
N	0.64	0.31	0.39	0.20	0.60	0.00	0.51	0.49	0.14	0.43	0.32	1.42	3.56
Nc(O)	1.12	0.20	0.50	0.71	0.54	0.51	0.00	0.46	0.64	0.52	0.82	1.44	3.46
Nc(I)	1.07	0.40	0.15	0.65	0.21	0.49	0.46	0.00	0.59	0.23	0.75	1.07	3.12
Ch(O)	0.51	0.45	0.47	0.07	0.68	0.14	0.64	0.59	0.00	0.50	0.18	1.45	3.61
Ch(I)	0.95	0.42	0.21	0.55	0.20	0.43	0.52	0.23	0.50	0.00	0.64	1.02	3.16
T(O)	0.33	0.62	0.61	0.11	0.83	0.32	0.82	0.75	0.18	0.64	0.00	1.53	3.70
T(I)	1.76	1.41	1.11	1.48	0.92	1.42	1.44	1.07	1.45	1.02	1.53	0.00	2.23
Cyc	3.93	3.49	3.20	3.64	3.01	3.56	3.46	3.12	3.61	3.16	3.70	2.23	0.00

* The symbols for the RIMs are: Benzene = Ben, Naphthalene = N, Anthracene = A, Naphthalene = Nc, Phenanthrene = P, Chrysene = Ch, Triphenylene = T, and Cyclohexane = Cyc; (I) = Inner ring and (O) = Outer ring

The mapping of the vector angle dissimilarities of the ring in molecules to a 2-dimensional space is, displayed in Fig. 4.4. The plot is in line with our chemical intuition: Generally, the outer RIMs are more similar to benzene than the inner ring RIMs and cyclohexane is by far the most dissimilar to benzene, as expected on the basis of aromaticity. In terms of similarity/dissimilarity one may expect a correlation between the x- and y-coordinates of the RIMs in Fig. 4.4 with the various aromaticity

measures. To that end we computed the Euclidean distances of the RIMs in this figure and regressed them against the aromaticity measures.

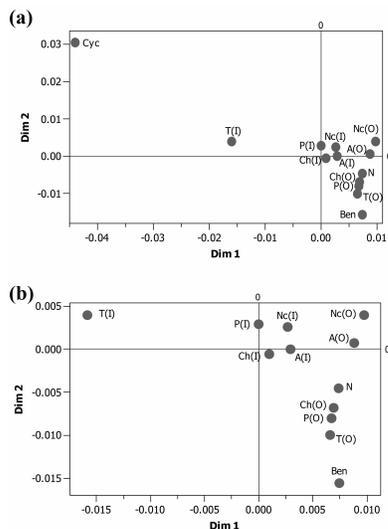


Figure 4.4: Two-dimensional MDS projection of the dissimilarity matrix in Table 4.3 (Kruskal stress (S) - 0.014). The symbols for the RIMs are: Benzene = Ben, Naphthalene = N, Anthracene = A, Naphthalene = Nc, Phenanthrene = P, Chrysene = Ch, Triphenylene = T, and Cyclohexane = Cyc; (I) = Inner ring and (O) = Outer ring. (a) The dataset including the extreme value of cyclohexane at the far upper left, (b) excluding cyclohexane to zoom on the 12 aromatic RIMs better showing their spread.

Fig. 4.5 shows the relationship between the Euclidean distance from benzene regressed against the aromaticity measures HOMA and PDI. The other aromaticity measures NICS except FLU showed any significant correlation to the Euclidean distance from benzene. From these results we see that (at least for HOMA and PDI) that the dissimilarities of the LDMS for the RIMs represented by the pairwise vector angles of the eigenvalues of the LDMS have a very strong correlation with the aromaticity parameters.

4.5 Conclusion

The aromaticity of a RIM is a property associated with cyclical electron delocalization around closed rings of atoms and which is generally recognized with ease by practising chemists yet, to this date, it remains a working concept lacking a unique or unambiguous definition. Thus it is no surprise that the DM was able to provide a

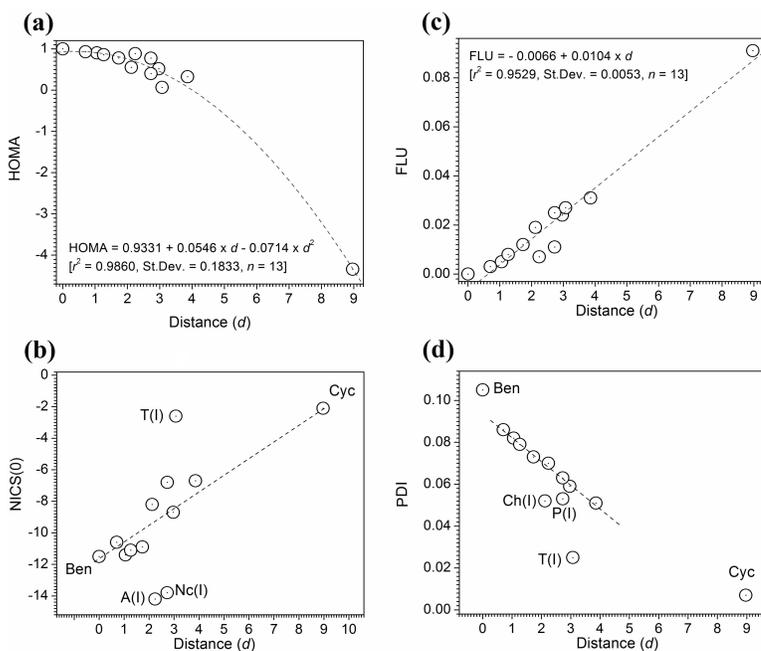


Figure 4.5: Distances obtained from the eigenvalues vectors versus the four aromaticity measures. Regression results are given only for well-behaved full datasets (a,c) while outliers are indicated otherwise (b, d). The symbols for the RIMs are: Benzene = Ben, Naphthalene = N, Anthracene = A, Naphthacene = Nc, Phenanthrene = P, Chrysene = Ch, Triphenylene = T, and Cyclohexane = Cyc; (I) = Inner ring and (O) = Outer ring.

good correlation since it is fully focused on the DI of the atoms in the RIM. There is no shortage of aromaticity indices, each designed to capture one particular aspect of this phenomenon, be it energetic, electronic, magnetic, structural, or reactive. These different aromaticity indices showed high correlations and the ability to map it onto a two-dimensional space is an effective tool for easy visualization.

4.6 Acknowledgements

Financial support of this work was provided by the *Natural Sciences and Engineering Research Council of Canada (NSERC)*. *Canada Foundation for Innovation (CFI)*, *Saint Mary's University*, *McMaster University*, and *Mount Saint Vincent University*

4.7 References

- [1]. Schleyer P.v-R.; Maerker, C.; Dransfeld, A.; Jiao, H.; Hommes, N. J. R. v. E. Nucleus-Independent chemical shifts: A simple and efficient aromaticity probe. *J. Am. Chem. Soc.* 1996, 118, 6317-6318.
- [2]. Krygowski, T. M.; Cyranski, M. K.; Czarnocki, Z.; Hfeling, G; Katritzky. Aromaticity: a Theoretical Concept of Immense Practical Importance. *Tetrahedron* 2000 56 1783-1796.
- [3]. Gomes, J. A. N. F.; Mallion, R. B. Aromaticity and ring currents. *Chem. Rev.* 2001, 101, 1349-1383.
- [4]. Chen, Z.; Wannere, C. S.; Corminboeuf, C.; Puchta, R.; Schleyer, P. v. R. Nucleus-independent chemical shifts (NICS) as an aromaticity criterion. *Chem. Rev.* 2005, 105, 3842-3888.
- [5]. Merino, G.; Vela, A.; Heine, T. Description of electron delocalization via the analysis of molecular fields. *Chem. Rev.* 2005, 105, 3812-3841.
- [6]. Mandado, M.; Gonzalez Moa, M. J.; Mosquera, R. A. Aromaticity: Exploring Basic Chemical Concepts with the Quantum Theory of Atoms in Molecules; Nova Science Publishers, Inc.: New York, 2008.
- [7]. Sol, M.; Feixas, F.; Jimnez-Halla, J. O. C.; Matito, E.; Poater, J. A critical assessment of the performance of magnetic and electronic indices of aromaticity. *Symmetry* 2010, 2, 1156-1179.
- [8]. Chattaraj, P. K. Ed. Aromaticity and Metal Clusters; CRC Press: New York, 2011.
- [9]. Feixas, F.; Matito, E.; Poater, J.; Sol, M. Quantifying aromaticity with electron delocalisation measures. *Chem. Soc. Rev.* 2015, 4, 6434-6451.

- [10]. Fernández, I.; Frenking, G.; Merino, G. Aromaticity of metallabenzenes and related compounds. *Chem. Soc. Rev.* 2015, 44, 6452-6463.
- [11]. Krygowski, T. M.; Szatyłowicz, H.; Stasyuk, O. A.; Dominikowska, J.; Palusiak, M. Aromaticity from the viewpoint of molecular geometry: Application to planar systems. *Chem. Rev.* 2014, 114, 6383-6422.
- [12]. Krygowski, T. M.; Cyranski, M. K. Structural aspect of aromaticity. *Chem. Rev.* 2001, 101, 1385-1419.
- [13]. Kruszewski, J.; Krygowski, T. M. Definition of aromaticity basing on the harmonic oscillator model. *Tetrahedron Lett.* 1972, 3839-3842.
- [14]. Hu, X.; Li, H.; Wang, C. The reactivity of all-metal aromatic complexes: A theoretical investigation on the methane activation reaction. *J. Phys. Chem. B* 2006, 110, 14046-14049.
- [15]. Li, S.; Jiang, Y. Bond lengths, reactivities, and aromaticities of benzenoid hydrocarbons based on the valence bond calculations. *J. Am. Chem. Soc.* 1995, 117, 8401-8406.
- [16]. Sainsbury, M. *Aromatic Chemistry*; Oxford Science Publications: Oxford, 1994.
- [17]. Badger, G. M. *Aromatic Character and Aromaticity*; Cambridge University Press: Cambridge, 1969.
- [18]. Cyranski, M. K. Energetic aspects of cyclic π -electron delocalization: Evaluation of the methods of estimating aromatic stabilization energies. *Chem. Rev.* 2005, 105, 3773-3811.
- [19]. Sivaramakrishnan, R.; Tranter, R. S.; Brezinsky, K. Ring conserved isodesmic reactions: a new method for estimating the heats of formation of aromatics and PAHs. *J. Phys. Chem. A* 2005, 109, 1621-1628.
- [20]. Slayden, S. W.; Liebman, J. F. The energetics of aromatic hydrocarbons: an

- experimental thermochemical perspective. *Chem. Rev.* 2001, 101, 1541-1566.
- [21]. Mitchell, R. H. Measuring aromaticity by NMR. *Chem. Rev.* 2001, 101, 1301-1315.
- [22]. Keith, T. A.; Bader, R. F. W. Use of electron charge and current distributions in the determination of atomic contributions to magnetic properties. *Int. J. Quantum Chem.* 1996, 60, 373-379.
- [23]. Keith, T. A.; Bader, R. F. W. Topological analysis of magnetically induced molecular current distributions. *J. Chem. Phys.* 1993, 99, 3669-3682.
- [24]. Schleyer P.v-R.; Manoharan, M.; Wang, Z.-X.; Kiran, B.; Jiao, H.; Puchta, R.; Hommes, N. J. R. v. E. Dissected nucleus-independent chemical shift analysis of π -aromaticity and antiaromaticity. *Org. Lett.* 2001, 3, 2465-2468.
- [25]. Foroutan-Nejad, C.; Badri, Z.; Shahbazian, S.; Rashidi-Ranjbar, P. The Laplacian of electron density versus NICS_{zz} scan: Measuring magnetic aromaticity among molecules with different atom types. *J. Phys. Chem. A* 2011, 115, 12708-12714.
- [26]. Mandado, M.; Gonzalez-Moa, M. J.; Mosquera, R. A. QTAIM N-center indices as descriptors of aromaticity in mono and poly heterocycles. *J. Comput. Chem.* 2007, 28, 127-136.
- [27]. Clar, E. *The Aromatic Sextet*; John Wiley and Sons Ltd.: London, 1972.
- [28]. Goldstein, M. J.; Hoffmann, R. Symmetry, topology, and aromaticity. *J. Am. Chem. Soc.* 1971, 93, 6193-6204.
- [29]. Palusiak, M.; Krygowski, T. M. Application of AIM parameters at ring critical points for estimation of π -electron delocalization in six-membered aromatic and quasi-aromatic rings. *Chem. Eur. J.* 2007, 13, 7996-8006.
- [30]. Mandado, M.; Gonzalez-Moa, M. J.; Mosquera, R. A. Chemical graph theory

- and n-center electron delocalization indices: A study on polycyclic aromatic hydrocarbons. *J. Comput. Chem.* 2007, 28, 1625-33.
- [31] Ponec R and Mayer I 1997 Investigation of some properties of multicenter bond indices *J. Phys. Chem. A* 101 173841.
- [32]. Bultinck, P.; Ponec, R.; Van Damme, S. Multicenter bond indices as a new measure of aromaticity in polycyclic aromatic hydrocarbons. *J. Phys. Org. Chem.* 2005, 18, 706-718.
- [33]. Bultinck, P.; Rafat, M.; Ponec, R.; Van Gheluwe, B.; Carb-Dorca, R.; Popelier, P. L. A. Electron delocalization and aromaticity in linear polyacenes: atoms in molecules multicenter delocalization index. *J. Phys. Chem. A* 2006, 110, 7642-7648.
- [34]. Bultinck, P.; Fias, S.; Ponec, R. Local aromaticity in polycyclic aromatic hydrocarbons: electron delocalization versus magnetic indices. *Chem. Eur. J.* 2006, 12, 8813-8818.
- [35]. Bultinck, P. Critical analysis of the local aromaticity concept in polyaromatic hydrocarbons. *Faraday Disc.* 2007, 135, 347-365 .
- [36]. Bultinck, P.; Ponec, R.; Carb-Dorca, R. Aromaticity in linear polyacenes: Generalized population analysis and molecular quantum similarity approach. *J. Comput. Chem.* 2007, 28, 152-160.
- [37]. Fias, S.; Fowler, P. W.; Delgado, J. L.; Hahn, U. ; Bultinck, P. Correlation of delocalization indices and current - density maps in polycyclic aromatic hydrocarbons. *Chem. Eur. J.* 2008, 14, 3093-3099.
- [38]. Fradera, X.; Poater, J.; Simon, S.; Duran, M.; Sol, M. Electron-pairing analysis from localization and delocalization indices in the framework of the atoms-in-molecules theory. *Theor. Chem. Acc.* 2002, 108, 214-224.
- [39]. Poater, J.; Sol, M.; Duran, M.; Fradera, X. The calculation of electron

- localization and delocalization indices at the Hartree-Fock, density functional and post-Hartree-Fock levels of theory. *Theor. Chem. Acc.* 2002, 107, 362-371.
- [40]. Poater, J.; Fradera, X.; Duran, M.; Sol, M. The delocalization index as an electronic aromaticity criterion: application to a series of planar polycyclic aromatic hydrocarbons. *Chem. Eur. J.* 2003, 9, 400-406.
- [41]. Poater, J.; Fradera, X.; Duran, M.; Sol, M. An insight into local aromaticities of polycyclic aromatic hydrocarbons and fullerenes. *Chem. Eur. J.* 2003, 9, 1113-1122.
- [42]. Matito, E.; Duran, M.; Sol, M. The aromatic fluctuation index (FLU): A new aromaticity index based on electron delocalization. *J. Chem. Phys.* 2005, 122, 014109.
- [43]. Poater, J.; Duran, M.; Sol, M.; Silvi, B. Theoretical evaluation of electron delocalization in aromatic molecules by means of atoms in molecules (AIM) and electron localization function (ELF) topological approaches. *Chem. Rev.* 2005, 105, 3911-3947.
- [44]. Portella, G.; Poater, J.; Bofill, J. M.; Alemany, P.; Sol, M. Local aromaticity of [n]acenes, [n]phenacenes, and [n]helicenes ($n = 1-9$). *J. Org. Chem.* 2005, 70, 2509-2521.
- [45]. Matito, E.; Duran, M.; Sol, M. A novel exploration of the Hartree-Fock homolytic bond dissociation problem in the hydrogen molecule by means of electron localization measures. *J. Chem. Edu.* 2006, 83, 1243-1248.
- [46]. Matito E, Poater J and Sol M 2007 In: *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design* ed C F Matta and R J Boyd (Weinheim: Wiley) chapter 15
- [47]. Matito, E.; Feixas, F.; Sol, M. Electron delocalization and aromaticity measures within the Hückel molecular orbital method. *J. Mol. Struct.*

(THEOCHEM) 2007, 811, 3-11.

[48]. Feixas, F.; Matito, E.; Poater, J.; Sol, M. On the performance of some aromaticity indices: A critical assessment using a test set. *J. Comput. Chem.* 2008, 29, 1543-1554.

[49]. Feixas, F.; Matito, E.; Sol, M.; Poater, J. Analysis of Hückels $[4n + 2]$ rule through electronic delocalization measures. *J. Phys. Chem. A* 2008, 112, 13231-13238.

[50]. Feixas, F.; Matito, E.; Sol, M.; Poater, J. Patterns of π -electron delocalization in aromatic and antiaromatic organic compounds in the light of Hückels $4n + 2$ rule. *Phys. Chem. Chem. Phys. (PCCP)* 2010, 12, 7126-7137.

[51]. Feixas Gerons, F. Analysis of Chemical Bonding and Aromaticity from Electronic Delocalization Descriptors, PhD Thesis; University of Girona: Girona (Spain), 2010.

[52]. Matta, C. F.; Hernández-Trujillo, J. Bonding in polycyclic aromatic hydrocarbons in terms of the the electron density and of electron delocalization. *J. Phys. Chem. A* 2003, 107, 7496-7504 (Correction: *J. Phys. Chem A*, 2005, 109, 10798).

[53]. Howard, S. T.; Krygowski, T. M. Benzenoid hydrocarbon aromaticity in terms of charge density descriptors. *Can. J. Chem.* 1997, 75, 1174-1181.

[54]. Suresh, C. H.; Gadre, S. R. Clar's aromatic sextet theory revisited via molecular electrostatic potential topography. *J. Org. Chem.* 1999, 64, 2505-2512.

[55]. Cyranski, M. K.; Stepień, B. T.; Krygowski, T. M. Global and local aromaticities of linear and angular polyacenes. *Tetrahedron* 2000, 56, 9663-9667.

[56]. Balaban, A. T. Applications of graph theory in chemistry. *J. Chem. Inf. Comput. Sci.* 1985, 25, 334-343.

- [57]. Randić, M. Aromaticity of polycyclic conjugated hydrocarbons. *Chem. Rev.* 2003, 103, 3449-3606.
- [58]. Gutman, I.; Stankovi, S. Testing the Y-rule in Clar theory. *Polycyc. Arom. Comp.* 2007, 27, 425-436.
- [59]. Balaban, A. T.; urevi, J.; Gutman, I.; Jeremi, S.; Radenkovi, S. Correlations between local aromaticity indices of bipartite conjugated hydrocarbons. *J. Phys. Chem. A* 2010, 114, 5870-5877.
- [60]. Aihara, J. Simple topological theory of aromaticity for annulenes and radialenes. *Bull. Chem. Soc. Jpn.* 1980, 53, 1751-1752.
- [61]. Aihara, J. Aromaticity and superaromaticity in cyclopolyacenes. *J. Chem. Soc., Perkin Trans. 2* 1994, 971-974.
- [62]. Fradera, X.; Austen, M. A.; Bader, R. F. W. The Lewis model and beyond. *J. Phys. Chem. A* 1999, 103, 304-314 .
- [63]. Matta, C. F.; Hernandez-Trujillo, J.; Bader, R. F. W. Proton spin-spin coupling and electron delocalisation. *J. Phys. Chem. A* 2002, 106, 7369-7375.
- [64]. Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- [65] Popelier P L A 2000 *Atoms in Molecules: An Introduction* (London: Prentice-Hall)
- [66] Matta C F and Boyd R J (ed) 2007 *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design* (Weinheim: Wiley)
- [67]. Krygowski, T. M.; Ciesielski, A.; Bird, C. W.; Kotschy, A. Aromatic character of the benzene ring present in various topological environments in benzenoid hydrocarbons. Nonequivalence of indices of aromaticity. *J. Chem. Inf. Comput. Sci.* 1995, 35, 203-210

- [68]. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakaajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr, J. A.; Peralta, J. E. ; Ogliaro, F.; Bearpark, M.; Heyd, J. J. ; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision B.01 (Gaussian Inc., Wallingford CT, 2010).
- [69]. Keith, T. A. AIMAll. <http://aim.tkgristmill.com/> 2011.
- [70]. Sumar, I. ; Cook, R.; Ayers, P. W.; Matta, C. F. AIMLDM: A Program to Generate and Analyze Electron Localization-Delocalization Matrices (LDMS). *Comput. Theor. Chem.* 2015, In press.
- [71]. Hansen, C. M. Hansen solubility parameters: A user's handbook; CRC Press and Taylor & Francis Group: Boca Raton, Florida, 2007.
- [72]. Borg, I.; Groenen, P. Modern multidimensional scaling: theory and applications; Springer: New York, 1997.
- [73]. Cox, T. F.; Cox, M. A. A. Multidimensional Scaling; Chapman & Hall: London, 1994.
- [74]. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a

nonmetric hypothesis. *Psychometrika* 1964, 29, 1-27.

[75]. Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method.

Psychometrika 1964, 29, 115-129.

[76]. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* 1962, 27, 219-246.

[77]. Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* 1962, 27, 125-140.

[78]. Torgerson, W. S. Multidimensional scaling: I. Theory and method.

Psychometrika 1952, 17, 401-419.

Eigenvalues and Important Atoms in a Molecule

5.1 A look at the Eigenvalues

In **Chapter 1** it was shown that for every $n \times n$ LDM there exists $n!$ ways to label this matrix. A way to circumvent this labelling problem was to diagonalize the LDMs, where the eigenvalues were arranged from smallest to largest along the long diagonal. One can then proceed as normal and take the Frobenius distance between the diagonalized LDMs.

In **Chapter 4** the eigenvalues were used in a different way. The eigenvalues were sorted in a vector from smallest to largest for each molecule and the vector angle could then be taken between each molecule as a measure of similarity. These similarity measures were correlated very strongly with the aromaticity measure PDI.

We have not currently done any serious studies into the physical significance of the eigenvalues, but the aim of this section is to discuss the trends we have observed for the benzoic acid series, and the aromatic series in the context of the eigenvalues.

In the benzoic acid series a very strong correlation between the eigenvalues and the total electron population of the atom $N(\Omega_i)$ was observed. And the size of $N(\Omega_i)$ seemed to dictate the size of the eigenvalues. Table 5.1 displays the molecule BACOCCH_3 (BA = Benzoic Acid, COCH_3 = substituent in the para-position) with its $N(\Omega_i)$ and its eigenvalues as well as $\text{BACOCCH}_3\text{-P}$ (where -P stands for pruned and where COCH_3 = super-atom) with its $N(\Omega_i)$ and its eigenvalues (both $N(\Omega_i)$ and the eigenvalues were arranged from smallest to largest).

Notice that in Table 5.1 that the size of $N(\Omega_i)$ has an effect on the size of the eigenvalues (most notable for the super-atom COCH_3). The number of hydrogen

Table 5.1: Comparing eigenvalues of both BACOCH₃ and BACOCH₃_P (P indicates pruning and COCH₃ is the super-atom) largest eigenvalue and largest atom electron population $N(\Omega_i)$ are in bold. The r^2 value between the eigenvalues and $N(\Omega_i)$ is > 0.98 (both $N(\Omega_i)$ and the eigenvalues were arranged from smallest to largest)

Atom	BACOCH ₃ ($N(\Omega_i)$)	Eigenvalues	Atom	BACOCH ₃ _P($N(\Omega_i)$)	Eigenvalues
H1	0.412	0.061	H1	0.412	0.061
H14	0.925	0.269	H14	0.925	0.308
H13	0.929	0.308	H13	0.929	0.323
H11	0.938	0.323	H11	0.938	0.328
H20	0.955	0.328	H12	0.971	0.351
H12	0.971	0.351	C3	4.500	2.449
H18	0.978	0.396	C10	5.996	2.787
H19	0.978	0.403	C9	5.998	3.325
C3	4.500	2.435	C6	5.999	3.395
C15	5.027	2.630	C5	6.008	4.557
C10	5.996	2.963	C7	6.016	4.626
C9	5.998	3.325	C8	6.033	5.461
C6	5.999	3.483	O2	9.091	7.999
C5	6.008	4.311	O4	9.138	8.429
C17	6.016	4.626	COCH ₃ 15	23.045	22.272
C7	6.016	4.644			
C8	6.033	5.485			
O2	9.091	7.999			
O16	9.092	8.245			
O4	9.138	8.429			

atoms in BACOCH₃ and BACOCH₃_P is different (they are actually the same but some of the hydrogen atoms are contained within the super-atom COCH₃), this is reflected in the number of eigenvalues that are < 1 . In fact a general trend for each molecule was observed, depending on the size and number of the eigenvalues one could determine the number of hydrogen, carbon, or oxygen atoms in the molecule. Eigenvalues are not being assigned to an atom, just stating a trend that was consistent.

Based on the results of Table 5.1 it was presumed that the eigenvalues were correlated with the atom's electron population $N(\Omega_i)$.

When looking at the eigenvalues for the aromatic series this trend was not observed. For the molecule Benzene all the carbon atoms are equivalent, thus if the

eigenvalues were correlated with $N(\Omega_i)$ then all the eigenvalues should be the same but they are not. Table 5.2 displays Benzene with its $N(\Omega_i)$ and its eigenvalues both arranged from smallest to largest.

Table 5.2: Comparing Eigenvalues of Benzene (carbon atoms only) with its atom electron population $N(\Omega_i)$ (values are arranged from smallest to largest)

Atom	Benzene($N(\Omega_i)$)	Eigenvalues
C1	5.479	2.579
C2	5.479	3.271
C3	5.479	3.271
C4	5.479	4.569
C5	5.479	4.569
C6	5.479	5.479

There is clearly no correlation between $N(\Omega_i)$ and the eigenvalues based on Table 5.2.

What we can say definitively is that the sum of the eigenvalues is equivalent to the trace of the matrix which is the total localization index of the molecule. What we can deduce from these examples is that the largest eigenvalue reflects the size of the largest atom in the molecule, as is indicated by the BACOCH_3 example and by the Benzene example (and has been observed consistently for both the benzoic acid and aromatic series).

5.2 Important Atoms in a Molecule

When comparing $\text{p}K_a$ against the Frobenius distance for the benzoic acid series in **Chapter 3**, we were able to identify the group of atoms primarily responsible for $\text{p}K_a$ i.e. the COOH group. This was no surprise since it was obvious that this group of atoms is primarily responsible for $\text{p}K_a$. What if someone with very limited chemical

knowledge was to perform the exact same study, it would be hard for them to find the group of atoms responsible for pK_a .

We come to an important question, is it possible to automatically locate the group of atoms responsible for the property of interest? That is the question this section aims to discuss.

We will look at both the benzoic acid and aromatic series and the properties pK_a , λ_{max} , and the aromaticity measure HOMA.

For the benzoic acid series strong correlations were observed for both pK_a and λ_{max} . This was done by truncating the LDMs such that the COOH group was zoomed in on (the OH group was also looked at but here the focus is only on the COOH group). One way to automate this would be to look at every possible truncated matrix one can create from an LDM, take the Frobenius distance between all corresponding truncated matrices, correlate them with the measure pK_a and determine which group of truncated matrices had the best correlation with the property. But this can be rather costly for big LDMs, and for a big set of molecules.

A different approach would be to look at the atoms individually, and see how they correlate with the property of interest. The idea is that the atoms that make up the “active site” should contribute to the studied property more so than atoms that do not make up the active site.

This is the approach we took:

1. Take the LM/DM only (this is the matrix with only the LI/DI values)
2. Take the Frobenius distance between the reference atom and the atom from the other molecule (i.e. Frobenius distance between H1 on BANO₂ and H1 on BA)
3. Do step 2 for all atoms except for the super-atom

4. Correlate the Frobenius distance with the pK_a for H1, O2, O3, C4, etc.
5. Rank the r^2 values for each atom
6. Check if the active site is recovered by the highest ranking r^2 values

Frobenius distance tables for the atoms and the studied properties can be found in **Appendix C**. Table 5.3 shows the ranking of the Frobenius distance of individual atoms based on their r^2 values for the LI only with respect to pK_a . From this table it is clear that the “active site” is recovered (refer to **Appendix B** for benzoic acid molecules and their atomic labelling scheme). The first four atoms are the COOH group, however the atom O2 has a much smaller r^2 value than the other 3 members of the COOH group.

Table 5.3: Ranking of Frobenius distance of individual atoms with respect to pK_a the first four atoms are indeed the active site (COOH) (LI only)

atom	r^2
O4	0.979
C3	0.965
H1	0.964
O2	0.857
H11	0.803
C5	0.798
C9	0.774
C7	0.708
C8	0.649
H13	0.580
H12	0.563
H14	0.434
C10	0.204
C6	0.134

Table 5.4 shows the ranking of Frobenius distance of individual atoms based on their r^2 values for the DI only with respect to pK_a . In this table the first 3 atoms

belong to the COOH group. This time however O4 has a low ranking of 0.708. The COOH group was not recovered in this instance. The rankings of C5 and C8 have risen from Table 5.3 (C5 had a previous ranking of 0.8, and C8 had a previous ranking of 0.649). One possible reason for this is because C5 and C8 are bridges between the COOH and super-atom respectively, so their DI is expected to be significant.

Table 5.4: Ranking of Frobenius distance of individual atoms with respect to pK_a the first three atoms make up the active site (COOH) (DI only)

molecule	r^2
O2	0.979
C3	0.965
H1	0.964
C5	0.857
C8	0.803
H11	0.798
C7	0.774
O4	0.708
C6	0.649
H12	0.580
C9	0.563
C10	0.434
H14	0.204
H13	0.134

Atom O4 consistently has a higher LI value than O2 and has a higher total electron population, conversely O2 consistently has a higher DI value than O4. This can help explain the reason as to why O4 is ranked higher than O2 in Table 5.3 and why O2 is ranked higher than O4 in Table 5.4.

Both Tables 5.3 and 5.4 show very high correlations for a majority of the members of the COOH group, they also highlight the importance of the atoms C5 and C8 based on the climb in ranking observed from Table 5.3 to Table 5.4.

This ranking seems to highlight the “important atoms” in a molecule. The COOH

group had very high rankings (except for O4 in Table 5.4) and atoms C5 and C8 climbed the rankings in Table 5.4. Atom H11 is ranked high in both but it is close to the COOH group and its ranking is relatively unchanged.

The environment of atoms C5 and C8 is very different than the environments of the other carbon atoms in the Benzene ring, similarly the environment of H1 is very different than all of the other hydrogen atoms on the Benzene ring. The oxygen atoms are both in different environments; one in a double bond and the other in a single bond. Based on the pK_a and the LI/DI, the atoms that are in different environments seem to be the ones that receive a high r^2 value.

We now look at the benzoic acid series and the property λ_{max} . Using the same methodology as described above, the results are in Table 5.5 for the ranking of Frobenius distance of individual atoms for LI only with respect to λ_{max} , and in Table 5.6 for the DI only.

Table 5.5: Ranking of Frobenius distance of individual atoms with respect to λ_{max} (LI only)

molecule	r^2
O2	0.876
H1	0.808
C3	0.802
C8	0.798
C6	0.724
O4	0.723
C10	0.692
H13	0.561
H14	0.422
C7	0.254
H12	0.217
C5	0.178
C9	0.147
H11	0.123

Table 5.5 shows very low r^2 values when only looking at the LI. The highest r^2

value is 0.876 for O2, H1 and C3 have r^2 values of 0.808 and 0.802 respectively, but these are fairly low.

Table 5.6: Ranking of Frobenius distance of individual atoms with respect to λ_{max} (DI only)

molecule	r^2
C5	0.942
C8	0.918
C3	0.881
O2	0.874
C7	0.815
H1	0.811
O4	0.753
C9	0.637
C10	0.619
H14	0.606
H13	0.606
C6	0.509
H11	0.508
H12	0.169

Looking at Table 5.6 however C5, C8, C3, and O2 have fairly high r^2 values of 0.942, 0.918, 0.881, and 0.874 respectively. Perhaps a reason as to why C5 and C8 are ranked so high is because Benzene itself has a primary absorption band at 184 nm [1]. The atoms C5 and C8 are bridges to the COOH and substituents respectively, and are thus important since addition of a substituent to the Benzene ring shifts the primary band. Upon attaching a COOH group to C5 the primary band shifts to 230 nm [1], a significant change from Benzene's primary band. The substituents attached to C8 also have an impact on the primary band (see Table 3.3 for details). C7 is ranked high as well possibly because it is bonded to C8 and the way the substituents are oriented they are close in proximity to C7.

Tables 5.5 and 5.6 highlight the atoms C5, C8, C3, and O2. This gives us a clue as to where we should look for atoms that contribute to the property λ_{max} . The atoms

C5 and C8 are important because they are the bridges to the Benzene ring. Atoms C3 and O2 are important because they are part of the COOH group.

It is perhaps not surprising that Table 5.5 shows weaker correlations since it focuses only on the LI which is not significant compared to the DI for λ_{max} .

Now moving on to the aromatic series and proceeding just like before we will compare the Frobenius distance of the individual atoms to the aromaticity index HOMA. We should expect that the DI not the LI correlates well with HOMA. Table 5.7 shows the ranking of Frobenius distance of the individual atoms compared with the HOMA index (LI only).

Table 5.7: Ranking of Frobenius distance of individual atoms with respect to HOMA (LI only)

molecule	R ²
C3	0.980
C6	0.977
C4	0.935
C5	0.932
C1	0.886
C2	0.878

Table 5.7 has the outlier cyclohexane, if that is removed the r^2 values differ significantly.

Table 5.8: Ranking of Frobenius distance of individual atoms with respect to HOMA (LI only, cyclohexane removed)

molecule	R ²
C3	0.696
C1	0.538
C6	0.421
C4	0.393
C2	0.126
C5	0.107

Table 5.8 shows a poor correlation between the HOMA index and the Frobenius

distance of the individual atoms for only the LI, as expected.

Table 5.9: Ranking of Frobenius distance of individual atoms with respect to HOMA (DI only, cyclohexane removed)

molecule	R^2
C2	0.885
C4	0.871
C3	0.869
C1	0.860
C5	0.854
C6	0.838

Table 5.9 shows that all 6 carbon atoms have an r^2 value > 0.8 . This is what is expected since each carbon atom should contribute to the aromaticity of the ring. It is not expected that each carbon atom should have the same correlation since the rings are all different (some atoms are part of more than one ring which can affect its DI and is quite possibly the reason why some of the correlations are not as strong i.e. C6).

What is important about these findings is that their might be potential to locate the important atoms in a family of molecules based on two things:

1. The property being measured (in these few cases pK_a , λ_{max} , and aromaticity index HOMA).
2. The studied index (either LI, DI, and/or LDI).

It is important to determine which index might be more suitable for the studied measure as was clearly seen in the last example.

While currently this method is not automated (it will be at some point in the future) it can tell the user based off of LI, DI, and/or LI/DI in conjunction with the property being measured what atoms are important to both the index and the studied

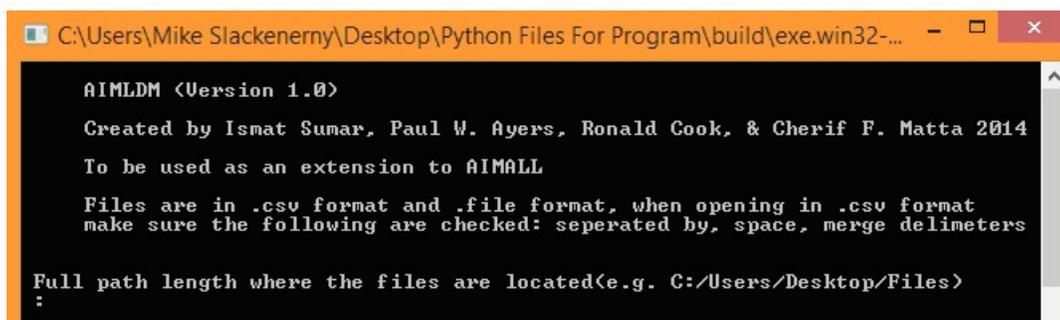
property. This could potentially be of great value, especially for big molecules with thousands of atoms, as this method can act as a quick scan to highlight region(s) of interest.

[1] Pavia, D. L.; Lampman, G. M.; Kriz, G. S.; Vyvyan, J. R. Introduction to Spectroscopy (4th Edition); Brooks/Cole Cengage Learning: Belmont, CA, USA, 2009.

Appendices

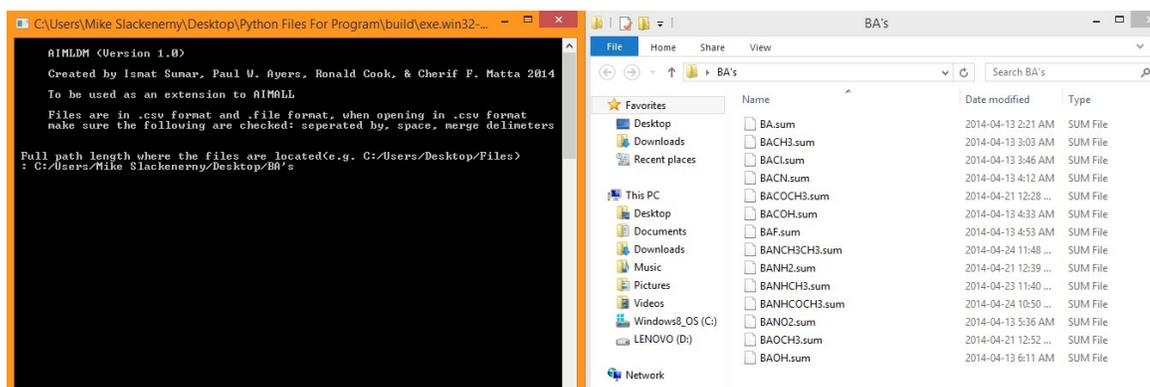
Operating Instructions for AIMLDM

Download the provided AIMLDM zipped file. Once downloaded, unzip the provided archive file. The executable programme file is called AIMLDM.exe while the other accompanying files in the folder are required for AIMLDM.exe to run. Run the file AIMLDM.exe (there is no installation required), the screen should be similar to the following:



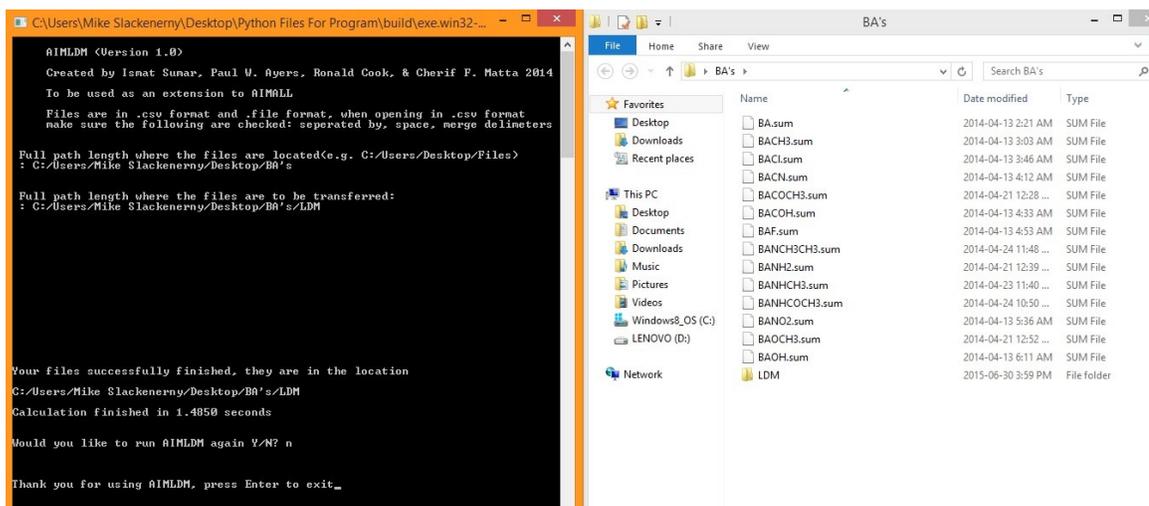
```
C:\Users\Mike Slackenem\\Desktop\Python Files For Program\build\exe.win32-...  
AIMLDM (Version 1.0)  
Created by Ismat Sumar, Paul W. Ayers, Ronald Cook, & Cherif F. Matta 2014  
To be used as an extension to AIMALL  
Files are in .csv format and .file format, when opening in .csv format  
make sure the following are checked: seperated by, space, merge delimiters  
Full path length where the files are located(e.g. C:/Users/Desktop/Files)  
:
```

Place the AIMAll output “.sum” files of all the molecules in the molecular set being studied in a chosen directory/folder and give the full path of that directory to AIMLDM, then press return.



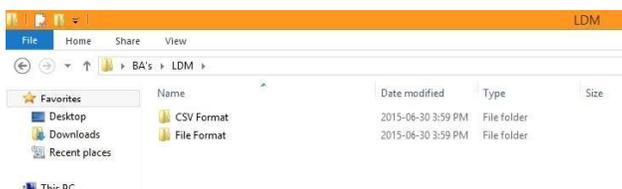
You will be prompted to enter the desired location of the output files (give the full path where you want the output to be produced). If you give an address to a directory that does not exist then AIMLDM will create it.

If everything is successful AIMLDM will tell you where your files are located and how long the calculation took, along with the option to run the program again.

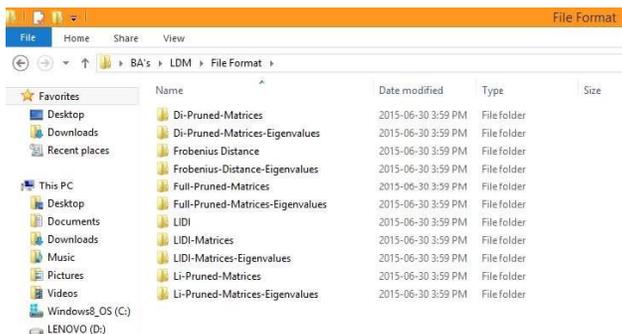


In the folder LDM there exists two directories one with the “.file” format which can be opened by a simple text editor (we recommend *notepad++* which can be downloaded for free from www.ninite.com under “Developer Tools”). The other with the “.csv” format for *Excel* type programs.

The following images show the layout of the folder LDM:



Inside each folder are the following folders.



Here is an example of a “.file” for benzoic acid (BA) and for *p*-methylbenzoic acid (BACH₃) using *notepad++*. Note that even though it says BA is pruned it is not because in this set BA was the smallest matrix. It is just in the folder *Full-Pruned-Matrices*, and thus has the associated word “*pruned*”.

	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	H15
6 H1	0.0743087	0.3179415	0.0060006	0.0101162	0.0044065	0.0006136	0.0000404	0.0000628	0.0001619	0.0006527	0.0004374	0.0000090	0.0000420	0.0000528	0.0000056
7 O2	0.3179415	8.0966135	0.4354779	0.1487853	0.0427766	0.0189898	0.0017091	0.0024805	0.0016532	0.0074471	0.0155243	0.0000726	0.0008484	0.0003063	0.0001810
8 C3	0.0060006	0.4354779	2.8525975	0.6519313	0.4852089	0.0260445	0.0045634	0.0054771	0.0047660	0.0281626	0.0025596	0.0007201	0.0031834	0.0007189	0.0002613
9 O4	0.0101162	0.1487853	0.6519313	8.2114501	0.0533585	0.0153630	0.0026417	0.0069294	0.0020595	0.0238550	0.0010183	0.0004241	0.0155993	0.0000757	0.0004281
10 C5	0.0044065	0.0427766	0.4852089	0.0533585	3.9126359	0.6701446	0.0356872	0.0470488	0.0358067	0.6672723	0.0213639	0.0049233	0.0211531	0.0049567	0.0024531
11 C6	0.0006136	0.0189898	0.0260445	0.0153630	0.6701446	3.9490550	0.6985277	0.0370649	0.0486776	0.0334410	0.4698397	0.0244865	0.0045055	0.0025321	0.0048825
12 C7	0.0000404	0.0017091	0.0045634	0.0026417	0.0356872	0.6985277	3.9547412	0.6919981	0.0358067	0.0486776	0.0358067	0.0023687	0.0048974	0.0221580	0.0240403
13 C8	0.0000628	0.0024805	0.0054771	0.0069294	0.0470488	0.0370649	0.6919981	3.9560175	0.6898329	0.0372669	0.0045197	0.0241733	0.0045217	0.0240670	0.4804069
14 C9	0.0001619	0.0016532	0.0047660	0.0020595	0.0358067	0.0486776	0.0486776	0.0358067	0.6898329	3.9544999	0.7011149	0.0023687	0.0048974	0.0221580	0.4804511
15 C10	0.0006527	0.0074471	0.0281626	0.0238550	0.6672723	0.0334410	0.0490284	0.0372669	0.7011149	3.9478288	0.0045599	0.0025766	0.4686546	0.0245031	0.0048704
16 H11	0.0004374	0.0155243	0.0025596	0.0010183	0.0213639	0.0213639	0.0223751	0.0045197	0.0023687	0.0045599	0.3925855	0.0025766	0.4686546	0.0001582	0.0008126
17 H12	0.0000090	0.0000726	0.0007201	0.0004241	0.0049233	0.0244865	0.4805857	0.0241733	0.0048974	0.0025766	0.0029582	0.4239047	0.0001573	0.0009073	0.0029736
18 H13	0.0000420	0.0008484	0.0031834	0.0155993	0.0211531	0.0045055	0.0023278	0.0045217	0.0221580	0.4686546	0.0009397	0.0001573	0.3872300	0.0029009	0.0008252
19 H14	0.0000528	0.0003063	0.0007189	0.0000757	0.0049567	0.0223251	0.0048771	0.0240670	0.4804511	0.0245031	0.0001582	0.0009073	0.0029009	0.4230364	0.0029547
20 H15	0.0000056	0.0001810	0.0002613	0.0004281	0.0024531	0.0048825	0.0240403	0.4804069	0.0239214	0.0048704	0.0008126	0.0029736	0.0008252	0.0029547	0.4238566

Following is the pruned matrix for BACH₃. Note how the super atom is labelled as CHHH which is in fact the CH₃ group in place of the regular H.

	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	CHHH15
6 H1	0.0746464	0.3185401	0.0060389	0.0102760	0.0043897	0.0006104	0.0000396	0.0000602	0.0001613	0.0006551	0.0004332	0.0000084	0.0000415	0.0000519	0.0000142
7 O2	0.3185401	8.0972663	0.4350389	0.1481794	0.0430096	0.0188621	0.0016602	0.0024107	0.0017624	0.0073935	0.0152818	0.0000660	0.0008368	0.0003001	0.0004850
8 C3	0.0060389	0.4350389	2.8560166	0.6508641	0.4883591	0.0264695	0.0047864	0.0054706	0.0049589	0.0280453	0.0025710	0.0007117	0.0031402	0.0007147	0.0010367
9 O4	0.0102760	0.1481794	0.6508641	8.2149881	0.0535899	0.0156915	0.0028492	0.0067727	0.0020154	0.0233642	0.0010359	0.0004078	0.0153614	0.0000720	0.0012336
10 C5	0.0043897	0.0430096	0.4883591	0.0535899	3.9129268	0.6714137	0.0354133	0.0443860	0.0354594	0.6617055	0.0213721	0.0048772	0.0209510	0.0049487	0.0067911
11 C6	0.0006104	0.0188621	0.0264695	0.0156915	0.6714137	3.9481200	0.6980409	0.0358078	0.0477029	0.0361000	0.4700531	0.0244604	0.0045014	0.0025252	0.0071299
12 C7	0.0000396	0.0016602	0.0047864	0.0028492	0.0354133	0.6980409	3.9574222	0.6790971	0.0354401	0.0468170	0.0225097	0.4775911	0.0222442	0.0047626	0.0462718
13 C8	0.0000602	0.0024107	0.0054706	0.0067727	0.0443860	0.0358078	0.6790971	3.8886286	0.6693425	0.0360415	0.0044620	0.0233323	0.0045066	0.0230398	0.5729535
14 C9	0.0001613	0.0017624	0.0049589	0.0020154	0.0354594	0.0477029	0.0354401	0.6693425	3.9595710	0.7089672	0.0023495	0.0047850	0.0225706	0.4782536	0.0435445
15 C10	0.0006551	0.0073935	0.0280453	0.0233642	0.6617055	0.0336100	0.0468170	0.0360415	0.7089672	3.9464743	0.0045999	0.0025490	0.4689292	0.0248532	0.0074285
16 H11	0.0004332	0.0152818	0.0025710	0.0010359	0.0213721	0.0213721	0.0225097	0.0044620	0.0023495	0.0045999	0.3940439	0.0030519	0.0009649	0.0001561	0.0009502
17 H12	0.0000084	0.0000660	0.0007117	0.0004078	0.0048772	0.0244604	0.4775911	0.0233323	0.0047850	0.0025490	0.0030519	0.4271421	0.0001558	0.0010321	0.0123670
18 H13	0.0000415	0.0008368	0.0031402	0.0153614	0.0209510	0.0045014	0.0022442	0.0045066	0.0225706	0.4689292	0.0009649	0.0001558	0.3889340	0.0030305	0.0009699
19 H14	0.0000519	0.0003001	0.0007147	0.0000720	0.0049487	0.0225252	0.0047626	0.0230398	0.4782536	0.0248532	0.0001561	0.0010321	0.0030305	0.4278028	0.0094598
20 CHHH15	0.0000142	0.0004850	0.0010367	0.0012336	0.0067911	0.0071299	0.0462718	0.5729535	0.0435445	0.0074285	0.0009502	0.0123670	0.0009699	0.0094598	8.2320562

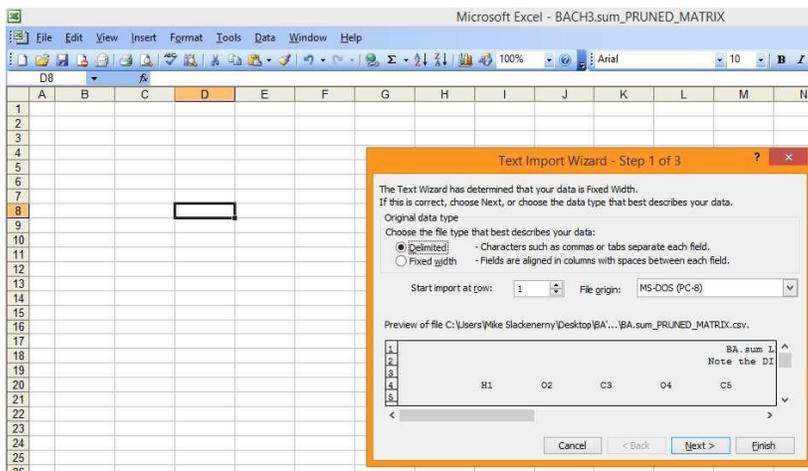
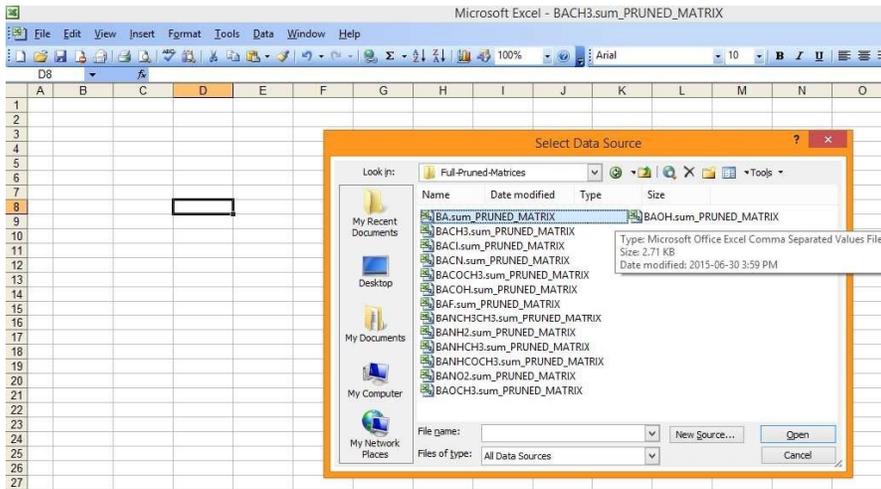
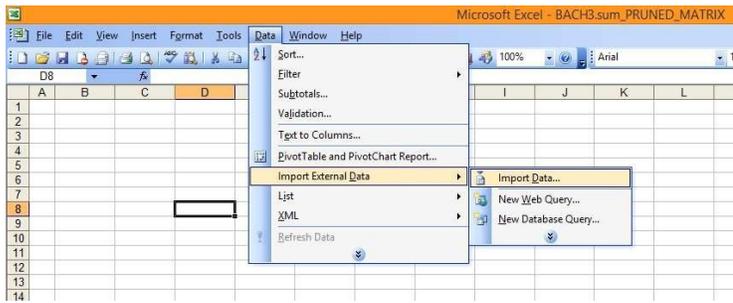
The next few images explain the steps to follow if you are using the “.csv” file formats and *Excel* does not ask for data delimiter, since the files are not “true” *Excel* files it will probably look distorted.

In *Excel* first choose *data* → *import external data* → *import data*

Then choose the file you wish to import:

Next select “*Delimited*”:

Select “*Space*” as your delimiter:



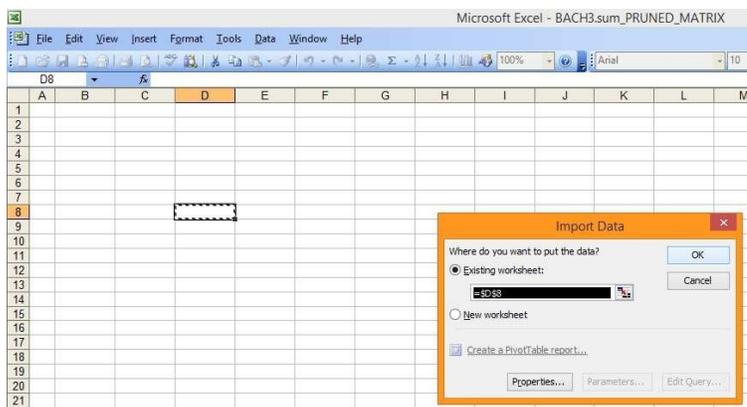
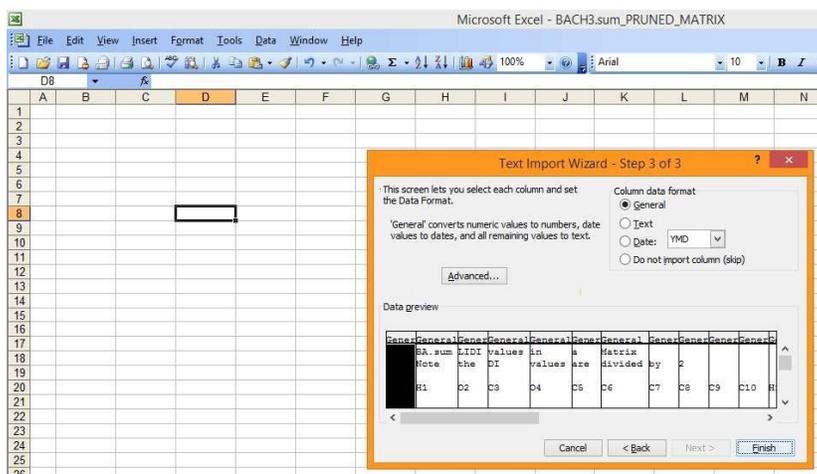
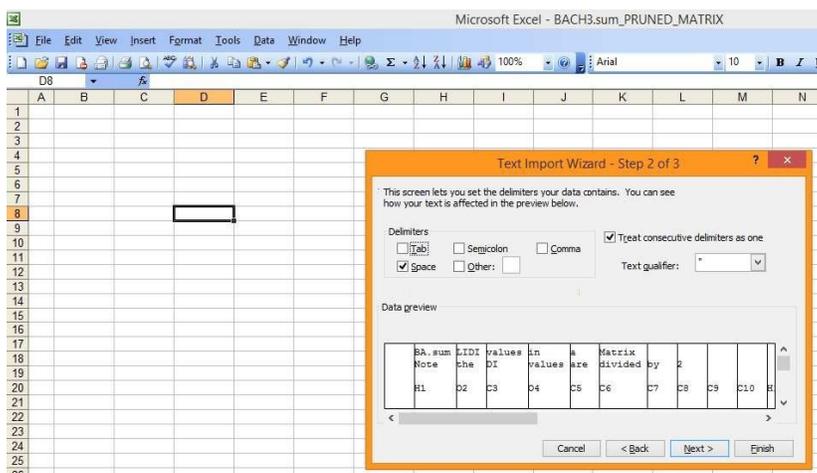
Select “*Finish*”:

Finally select “*OK*”:

These steps will produce the same matrices in *Excel* format:

Including the pruned matrix as well:

Elementary *single* group pruning is decided in the following steps (not much flex-

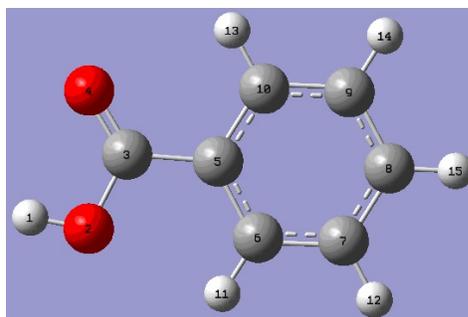


ibility in this version of the programme): Below is a ball-and-stick labelled diagram of benzoic acid (BA) (15 atoms) which is the smallest molecule in the substituted

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	BACH3 sum	LIDI	values	in	a	Matrix									
2	Note	the	DI	values	are	divided	by	2							
4	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	H15

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	BACH3 sum	LIDI	values	in	a	Matrix									
2	Note	the	DI	values	are	divided	by	2							
4	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	CHHH15

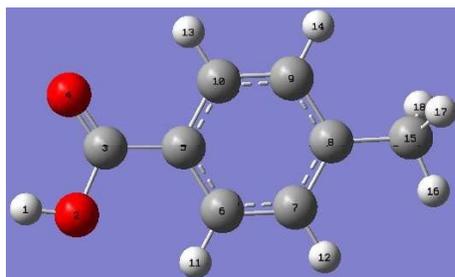
benzoic acid set. Every atom beyond atom #15 will thus be pruned into one single super-atom.



The following is a ball-and-stick labelled diagram of BACH₃. Then by defaults atoms 15-18 will be pruned into a super-atom as can be seen from the LDM presented above (the group CHHH15 represents the atoms 15-18 in this image below):

Subsequent versions of AIMLDM may be posted for download from the site:

<http://www.cmatta.ca/software>

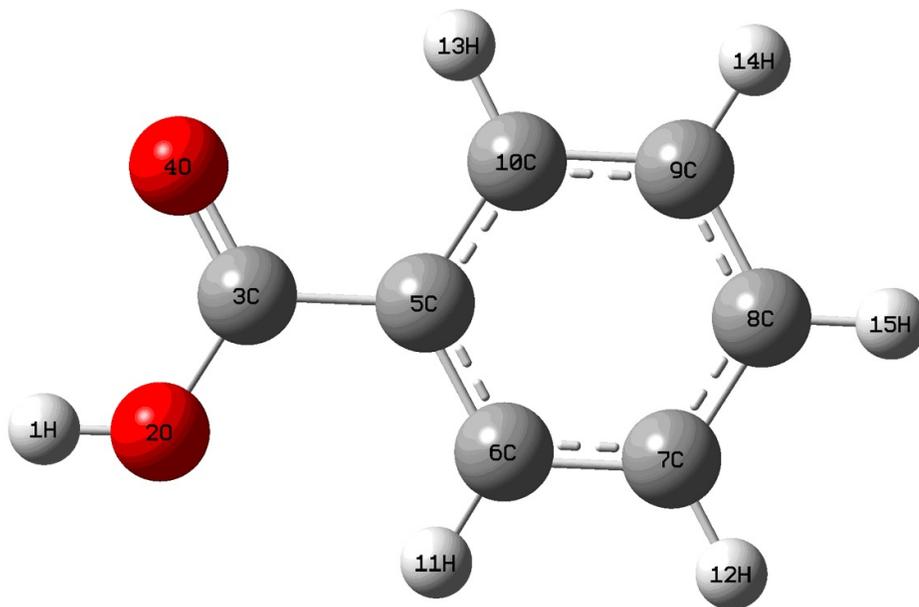


References

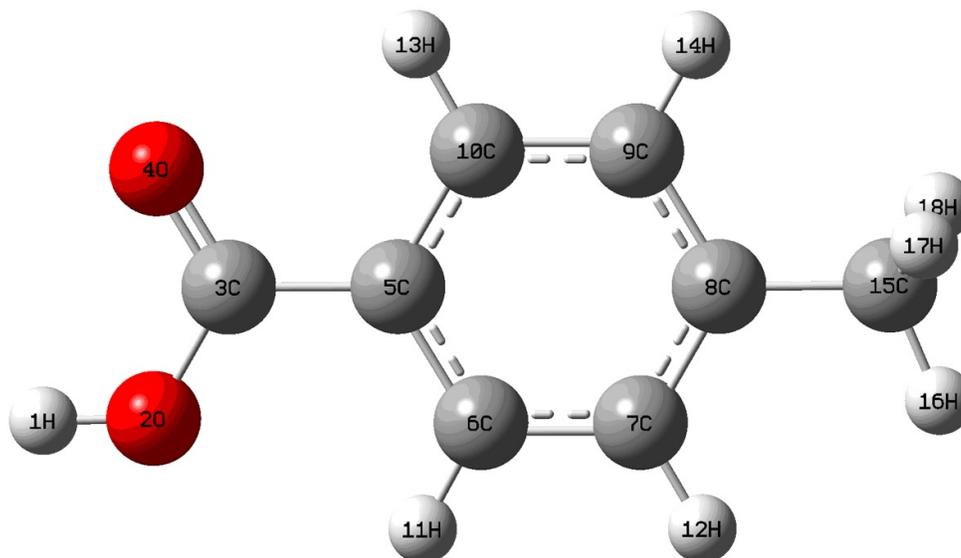
- [1] I. Sumar, R. Cook, P.W. Ayers and C.F. Matta, AIMLDM: A Program to Generate and Analyze Electron Localization-Delocalization Matrices (LDMs), *Comput. Theor. Chem.* 1070 (2015) 55-67.
- [2] R. Cook, I. Sumar, P.W. Ayers and C.F. Matta, *Electron Localization-Delocalization Matrices (LDMs): Theory and Applications* (Springer International Publishing AG, Cham, Switzerland, 2016).
- [3] C.F. Matta, I. Sumar, R. Cook and P.W. Ayers, Localization-delocalization and electron density-weighted connectivity matrices: A bridge between the quantum theory of atoms in molecules and chemical graph theory, In: *Applications of Topological Methods in Molecular Chemistry (Challenges and Advances in Computational Chemistry and Physics Series)*; Chauvin, R.; Silvi, B.; Alikhani, E.; Lepetit, C. (Eds.) (Springer, 2015).
- [4] I. Sumar, P.W. Ayers and C.F. Matta, 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 (2014) 190-197.
- [5] M.J. Timm, C.F. Matta, L. Massa and L. Huang, The localization-delocalization matrix and the electron density-weighted connectivity matrix of a finite graphene flake reconstructed from kernel fragments, *J. Phys. Chem. A* 118 (2014) 11304-11316.
- [6] C.F. Matta, Localization-delocalization matrices and electron density-weighted adjacency matrices: New electronic fingerprinting tools for medicinal computational chemistry, *Future Med. Chem.* 6 (2014) 1475-1479.
- [7] B. Dittrich and C.F. Matta, Contributions of charge-density research to medicinal chemistry, *Int. U. Cryst. J. (IUCrJ)* 1 (2014) 457-469.
- [8] C.F. Matta, 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 (2014) 1165-1198.

Benzoic Acid Structures with their LDMs

Unpruned and Pruned LDM for the set of benzoic acids to 3 decimal places, as well as corresponding structure. Unpruned matrix is listed first followed by the pruned matrix.



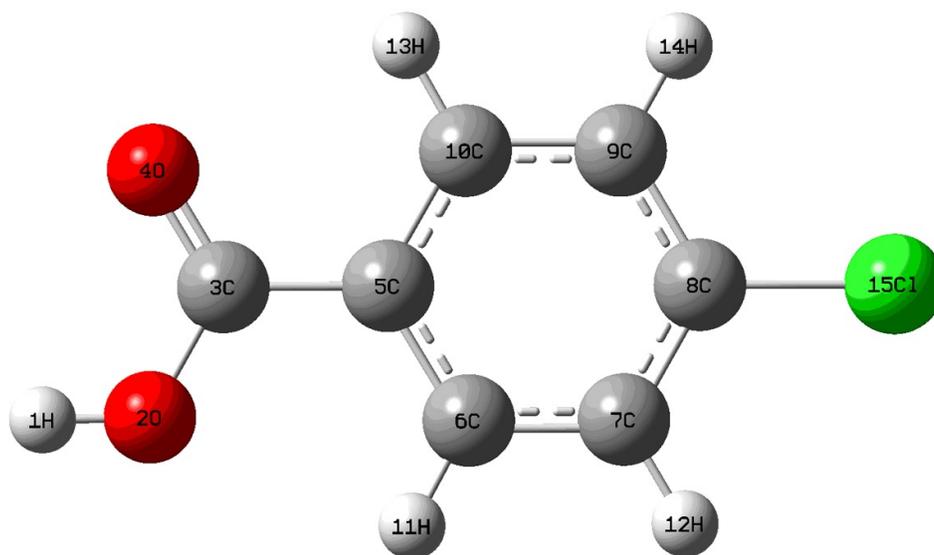
$C_7H_6O_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	H15	SUM
H1	0.074	0.318	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.415
O2	0.318	8.097	0.435	0.149	0.043	0.019	0.002	0.002	0.002	0.007	0.016	0.000	0.001	0.000	0.000	9.091
C3	0.006	0.435	2.853	0.652	0.485	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.000	4.508
O4	0.010	0.149	0.652	8.211	0.053	0.015	0.003	0.007	0.002	0.024	0.001	0.000	0.016	0.000	0.000	9.144
C5	0.004	0.043	0.485	0.053	3.913	0.670	0.036	0.047	0.036	0.667	0.021	0.005	0.021	0.005	0.002	6.009
C6	0.001	0.019	0.026	0.015	0.670	3.949	0.699	0.037	0.049	0.033	0.470	0.024	0.005	0.003	0.005	6.004
C7	0.000	0.002	0.005	0.003	0.036	0.699	3.955	0.692	0.036	0.049	0.022	0.481	0.002	0.005	0.024	6.009
C8	0.000	0.002	0.005	0.007	0.047	0.037	0.692	3.956	0.690	0.037	0.005	0.024	0.005	0.024	0.480	6.012
C9	0.000	0.002	0.005	0.002	0.036	0.049	0.036	0.690	3.954	0.701	0.002	0.005	0.022	0.480	0.024	6.008
C10	0.001	0.007	0.028	0.024	0.667	0.033	0.049	0.037	0.701	3.948	0.005	0.003	0.469	0.025	0.005	6.001
H11	0.000	0.016	0.003	0.001	0.021	0.470	0.022	0.005	0.002	0.005	0.393	0.003	0.001	0.000	0.001	0.942
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.481	0.024	0.005	0.003	0.003	0.424	0.000	0.001	0.003	0.974
H13	0.000	0.001	0.003	0.016	0.021	0.005	0.002	0.005	0.022	0.469	0.001	0.000	0.387	0.003	0.001	0.935
H14	0.000	0.000	0.001	0.000	0.005	0.003	0.005	0.024	0.480	0.025	0.000	0.001	0.003	0.423	0.003	0.972
H15	0.000	0.000	0.000	0.000	0.002	0.005	0.024	0.480	0.024	0.005	0.001	0.003	0.001	0.003	0.424	0.973
SUM	0.415	9.091	4.508	9.144	6.009	6.004	6.009	6.012	6.008	6.001	0.942	0.974	0.935	0.972	0.973	63.997



$C_8H_8O_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	C15	H16	H17	H18	SUM
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.416
O2	0.319	8.097	0.435	0.148	0.043	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.001	0.000	0.000	0.000	0.000	0.000	9.091
C3	0.006	0.435	2.856	0.651	0.488	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.000	0.000	0.000	0.000	4.514
O4	0.010	0.148	0.651	8.215	0.054	0.016	0.003	0.007	0.002	0.023	0.001	0.000	0.015	0.000	0.000	0.000	0.000	0.000	9.147
C5	0.004	0.043	0.488	0.054	3.913	0.671	0.035	0.044	0.035	0.662	0.021	0.005	0.021	0.005	0.003	0.000	0.002	0.002	6.01
C6	0.001	0.019	0.026	0.016	0.671	3.948	0.698	0.036	0.048	0.034	0.470	0.024	0.005	0.003	0.005	0.001	0.001	0.001	6.005
C7	0.000	0.002	0.005	0.003	0.035	0.698	3.957	0.679	0.035	0.047	0.023	0.478	0.002	0.005	0.027	0.007	0.006	0.006	6.015
C8	0.000	0.002	0.005	0.007	0.044	0.036	0.679	3.889	0.669	0.036	0.004	0.023	0.005	0.023	0.508	0.020	0.022	0.022	5.996
C9	0.000	0.002	0.005	0.002	0.035	0.048	0.035	0.669	3.960	0.709	0.002	0.005	0.023	0.478	0.027	0.005	0.006	0.006	6.017
C10	0.001	0.007	0.028	0.023	0.662	0.034	0.047	0.036	0.709	3.946	0.005	0.003	0.469	0.025	0.005	0.001	0.001	0.001	6.001
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.004	0.002	0.005	0.394	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.944
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.478	0.023	0.005	0.003	0.003	0.427	0.000	0.001	0.006	0.005	0.000	0.000	0.983
H13	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.005	0.023	0.469	0.001	0.000	0.389	0.003	0.001	0.000	0.000	0.000	0.937
H14	0.000	0.000	0.001	0.000	0.005	0.003	0.005	0.023	0.478	0.025	0.000	0.001	0.003	0.428	0.006	0.001	0.002	0.001	0.981
C15	0.000	0.000	0.000	0.000	0.003	0.005	0.027	0.508	0.027	0.005	0.001	0.006	0.001	0.006	3.958	0.478	0.476	0.476	5.977
H16	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.020	0.005	0.001	0.000	0.005	0.000	0.001	0.478	0.438	0.019	0.019	0.995
H17	0.000	0.000	0.000	0.000	0.002	0.001	0.006	0.022	0.006	0.001	0.000	0.000	0.000	0.002	0.476	0.019	0.431	0.019	0.985
H18	0.000	0.000	0.000	0.000	0.002	0.001	0.006	0.022	0.006	0.001	0.000	0.000	0.000	0.001	0.476	0.019	0.019	0.431	0.985
SUM	0.416	9.091	4.514	9.147	6.010	6.005	6.015	5.996	6.017	6.001	0.944	0.983	0.937	0.981	5.977	0.995	0.985	0.985	71.999

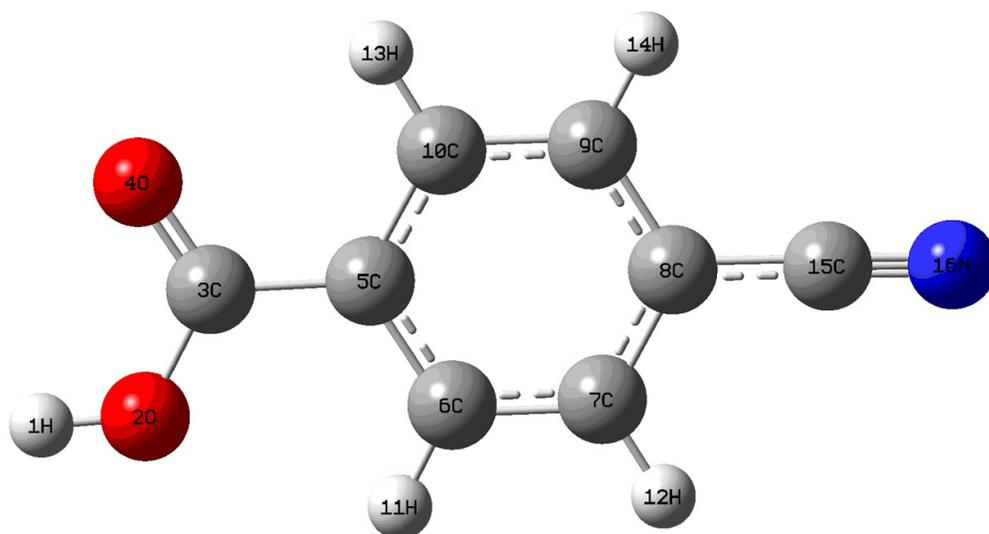
$C_8H_8O_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	CH315	SUM
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.416
O2	0.319	8.097	0.435	0.148	0.043	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.001	0.000	0.000	9.091
C3	0.006	0.435	2.856	0.651	0.488	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.001	4.514
O4	0.010	0.148	0.651	8.215	0.054	0.016	0.003	0.007	0.002	0.023	0.001	0.000	0.015	0.000	0.001	9.147
C5	0.004	0.043	0.488	0.054	3.913	0.671	0.035	0.044	0.035	0.662	0.021	0.005	0.021	0.005	0.007	6.010
C6	0.001	0.019	0.026	0.016	0.671	3.948	0.698	0.036	0.048	0.034	0.470	0.024	0.005	0.003	0.007	6.005
C7	0.000	0.002	0.005	0.003	0.035	0.698	3.957	0.679	0.035	0.047	0.023	0.478	0.002	0.005	0.046	6.015
C8	0.000	0.002	0.005	0.007	0.044	0.036	0.679	3.889	0.669	0.036	0.004	0.023	0.005	0.023	0.573	5.996
C9	0.000	0.002	0.005	0.002	0.035	0.048	0.035	0.669	3.960	0.709	0.002	0.005	0.023	0.478	0.044	6.017
C10	0.001	0.007	0.028	0.023	0.662	0.034	0.047	0.036	0.709	3.946	0.005	0.003	0.469	0.025	0.007	6.001
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.004	0.002	0.005	0.394	0.003	0.001	0.000	0.001	0.944
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.478	0.023	0.005	0.003	0.003	0.427	0.000	0.001	0.012	0.983
H13	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.005	0.023	0.469	0.001	0.000	0.389	0.003	0.001	0.937
H14	0.000	0.000	0.001	0.000	0.005	0.003	0.005	0.023	0.478	0.025	0.000	0.001	0.003	0.428	0.009	0.981
CH315	0.000	0.000	0.001	0.001	0.007	0.007	0.046	0.573	0.044	0.007	0.001	0.012	0.001	0.009	8.232	8.943
SUM	0.416	9.091	4.514	9.147	6.010	6.005	6.015	5.996	6.017	6.001	0.944	0.983	0.937	0.981	8.943	71.999

Benzoic Acid Structures with their LDMs



$C_7H_5ClO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	Cl15	SUM
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.413
O2	0.317	8.098	0.436	0.148	0.043	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.436	2.846	0.653	0.485	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.002	4.503
O4	0.010	0.148	0.653	8.208	0.054	0.015	0.003	0.006	0.002	0.024	0.001	0.000	0.015	0.000	0.002	9.142
C5	0.004	0.043	0.485	0.054	3.908	0.669	0.035	0.044	0.035	0.666	0.021	0.005	0.021	0.005	0.009	6.004
C6	0.001	0.019	0.026	0.015	0.669	3.940	0.698	0.038	0.047	0.033	0.468	0.022	0.004	0.002	0.010	5.993
C7	0.000	0.002	0.005	0.003	0.035	0.698	3.929	0.674	0.031	0.047	0.022	0.473	0.002	0.004	0.053	5.978
C8	0.000	0.002	0.005	0.006	0.044	0.038	0.674	3.866	0.672	0.039	0.005	0.023	0.005	0.023	0.554	5.956
C9	0.000	0.002	0.005	0.002	0.035	0.047	0.031	0.672	3.929	0.701	0.002	0.004	0.022	0.473	0.053	5.977
C10	0.001	0.007	0.028	0.024	0.666	0.033	0.047	0.039	0.701	3.939	0.004	0.002	0.467	0.022	0.010	5.990
H11	0.000	0.015	0.003	0.001	0.021	0.468	0.022	0.005	0.002	0.004	0.386	0.003	0.001	0.000	0.001	0.933
H12	0.000	0.000	0.001	0.000	0.005	0.022	0.473	0.023	0.004	0.002	0.003	0.401	0.000	0.001	0.012	0.948
H13	0.000	0.001	0.003	0.015	0.021	0.004	0.002	0.005	0.022	0.467	0.001	0.000	0.381	0.003	0.001	0.926
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.473	0.022	0.000	0.001	0.003	0.400	0.012	0.947
Cl15	0.000	0.001	0.002	0.002	0.009	0.01	0.053	0.554	0.053	0.01	0.001	0.012	0.001	0.012	16.479	17.198
SUM	0.413	9.091	4.503	9.142	6.004	5.993	5.978	5.956	5.977	5.99	0.933	0.948	0.926	0.947	17.198	79.999

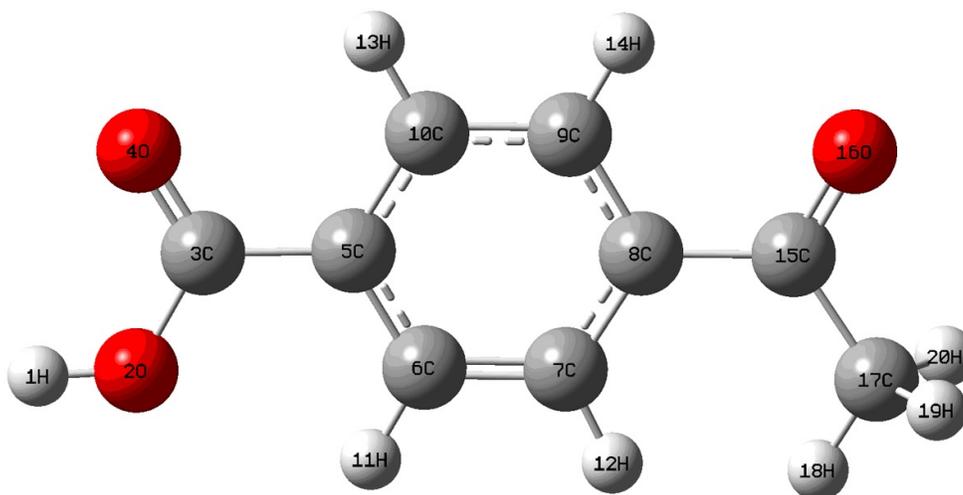
Benzoic Acid Structures with their LDMs



$C_8H_5NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	C15	N16	SUM
H1	0.073	0.315	0.006	0.009	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.410
O2	0.315	8.097	0.438	0.149	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.000	0.001	9.091
C3	0.006	0.438	2.838	0.656	0.480	0.025	0.004	0.004	0.004	0.027	0.003	0.001	0.003	0.001	0.000	0.001	4.491
O4	0.009	0.149	0.656	8.199	0.054	0.015	0.002	0.006	0.002	0.024	0.001	0.000	0.016	0.000	0.001	0.002	9.135
C5	0.004	0.043	0.480	0.054	3.907	0.669	0.036	0.043	0.036	0.666	0.021	0.005	0.021	0.005	0.004	0.011	6.004
C6	0.001	0.019	0.025	0.015	0.669	3.940	0.702	0.036	0.047	0.033	0.468	0.023	0.004	0.002	0.005	0.003	5.992
C7	0.000	0.002	0.004	0.002	0.036	0.702	3.935	0.661	0.031	0.047	0.022	0.475	0.002	0.004	0.034	0.026	5.984
C8	0.000	0.002	0.004	0.006	0.043	0.036	0.661	3.845	0.659	0.036	0.004	0.023	0.004	0.022	0.542	0.048	5.935
C9	0.000	0.001	0.004	0.002	0.036	0.047	0.031	0.659	3.935	0.704	0.002	0.004	0.022	0.475	0.034	0.026	5.984
C10	0.001	0.007	0.027	0.024	0.666	0.033	0.047	0.036	0.704	3.939	0.004	0.002	0.466	0.023	0.005	0.003	5.988
H11	0.000	0.015	0.003	0.001	0.021	0.468	0.022	0.004	0.002	0.004	0.382	0.003	0.001	0.000	0.001	0.000	0.928
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.475	0.023	0.004	0.002	0.003	0.400	0.000	0.001	0.005	0.003	0.944
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.004	0.022	0.466	0.001	0.000	0.377	0.003	0.001	0.000	0.921
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.022	0.475	0.023	0.000	0.001	0.003	0.399	0.005	0.003	0.943
C15	0.000	0.000	0.000	0.001	0.004	0.005	0.034	0.542	0.034	0.005	0.001	0.005	0.001	0.005	3.321	1.195	5.153
N16	0.000	0.001	0.001	0.002	0.011	0.003	0.026	0.048	0.026	0.003	0.000	0.003	0.000	0.003	1.195	6.776	8.097
SUM	0.410	9.091	4.491	9.135	6.004	5.992	5.984	5.935	5.984	5.988	0.928	0.944	0.921	0.943	5.153	8.097	76.000

$C_8H_5NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	CN15	SUM
H1	0.073	0.315	0.006	0.009	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.410
O2	0.315	8.097	0.438	0.149	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.438	2.838	0.656	0.480	0.025	0.004	0.004	0.004	0.027	0.003	0.001	0.003	0.001	0.002	4.491
O4	0.009	0.149	0.656	8.199	0.054	0.015	0.002	0.006	0.002	0.024	0.001	0.000	0.016	0.000	0.002	9.135
C5	0.004	0.043	0.480	0.054	3.907	0.669	0.036	0.043	0.036	0.666	0.021	0.005	0.021	0.005	0.014	6.004
C6	0.001	0.019	0.025	0.015	0.669	3.940	0.702	0.036	0.047	0.033	0.468	0.023	0.004	0.002	0.008	5.992
C7	0.000	0.002	0.004	0.002	0.036	0.702	3.935	0.661	0.031	0.047	0.022	0.475	0.002	0.004	0.060	5.984
C8	0.000	0.002	0.004	0.006	0.043	0.036	0.661	3.845	0.659	0.036	0.004	0.023	0.004	0.022	0.590	5.935
C9	0.000	0.001	0.004	0.002	0.036	0.047	0.031	0.659	3.935	0.704	0.002	0.004	0.022	0.475	0.060	5.984
C10	0.001	0.007	0.027	0.024	0.666	0.033	0.047	0.036	0.704	3.939	0.004	0.002	0.466	0.023	0.008	5.988
H11	0.000	0.015	0.003	0.001	0.021	0.468	0.022	0.004	0.002	0.004	0.382	0.003	0.001	0.000	0.001	0.928
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.475	0.023	0.004	0.002	0.003	0.400	0.000	0.001	0.008	0.944
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.004	0.022	0.466	0.001	0.000	0.377	0.003	0.001	0.921
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.022	0.475	0.023	0.000	0.001	0.003	0.399	0.008	0.943
CN15	0.000	0.001	0.002	0.002	0.014	0.008	0.060	0.590	0.060	0.008	0.001	0.008	0.001	0.008	12.486	13.250
SUM	0.410	9.091	4.491	9.135	6.004	5.992	5.984	5.935	5.984	5.988	0.928	0.944	0.921	0.943	13.250	76.000

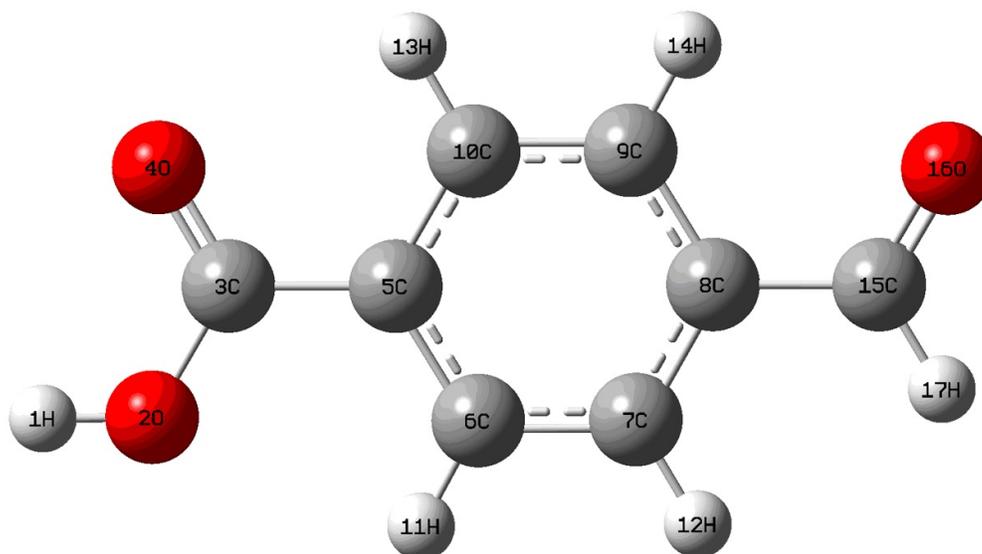
Benzoic Acid Structures with their LDMs



$C_9H_8O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	C15	O16	C17	H18	H19	H20	SUM
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.412
O2	0.317	8.096	0.437	0.149	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.091
C3	0.006	0.437	2.846	0.655	0.482	0.026	0.004	0.005	0.004	0.027	0.003	0.001	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.000	4.500
O4	0.010	0.149	0.655	8.202	0.053	0.015	0.002	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.001	0.001	0.000	0.000	0.000	0.000	9.138
C5	0.004	0.043	0.482	0.053	3.911	0.672	0.036	0.044	0.036	0.663	0.021	0.005	0.021	0.004	0.005	0.008	0.001	0.000	0.000	0.000	6.008
C6	0.001	0.019	0.026	0.015	0.672	3.945	0.699	0.035	0.048	0.033	0.469	0.024	0.004	0.002	0.004	0.002	0.001	0.000	0.000	0.000	5.999
C7	0.000	0.002	0.004	0.002	0.036	0.699	3.956	0.673	0.034	0.047	0.022	0.472	0.002	0.004	0.027	0.018	0.009	0.004	0.004	0.001	6.016
C8	0.000	0.002	0.005	0.006	0.044	0.035	0.673	3.932	0.663	0.036	0.004	0.022	0.005	0.021	0.487	0.063	0.026	0.002	0.002	0.004	6.033
C9	0.000	0.001	0.004	0.002	0.036	0.048	0.034	0.663	3.947	0.708	0.002	0.005	0.022	0.465	0.026	0.028	0.004	0.000	0.000	0.001	5.998
C10	0.001	0.007	0.027	0.023	0.663	0.033	0.047	0.036	0.708	3.945	0.005	0.002	0.468	0.022	0.005	0.003	0.001	0.000	0.000	0.000	5.996
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.022	0.004	0.002	0.005	0.390	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.938
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.472	0.022	0.005	0.002	0.003	0.412	0.000	0.001	0.003	0.001	0.010	0.005	0.005	0.001	0.971
H13	0.000	0.001	0.003	0.015	0.021	0.004	0.002	0.005	0.022	0.468	0.001	0.000	0.383	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.929
H14	0.000	0.000	0.001	0.000	0.004	0.002	0.004	0.021	0.465	0.022	0.000	0.001	0.003	0.378	0.003	0.019	0.001	0.000	0.000	0.000	0.925
C15	0.000	0.000	0.000	0.001	0.005	0.004	0.027	0.487	0.026	0.005	0.001	0.003	0.001	0.003	3.211	0.712	0.478	0.023	0.023	0.018	5.027
O16	0.000	0.000	0.001	0.001	0.008	0.002	0.018	0.063	0.028	0.003	0.000	0.001	0.000	0.019	7.12	8.135	0.064	0.012	0.012	0.012	9.092
C17	0.000	0.000	0.000	0.000	0.001	0.001	0.009	0.026	0.004	0.001	0.000	0.010	0.000	0.001	0.478	0.064	4.000	0.473	0.473	0.474	6.016
H18	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.002	0.000	0.000	0.000	0.005	0.000	0.000	0.023	0.012	0.473	0.421	0.019	0.017	0.978
H19	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.002	0.000	0.000	0.000	0.005	0.000	0.000	0.023	0.012	0.473	0.019	0.421	0.017	0.978
H20	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.018	0.012	0.474	0.017	0.017	0.408	0.955
SUM	0.412	9.091	4.500	9.138	6.008	5.999	6.016	6.033	5.998	5.996	0.938	0.971	0.929	0.925	5.027	9.092	6.016	0.978	0.978	0.955	86.001

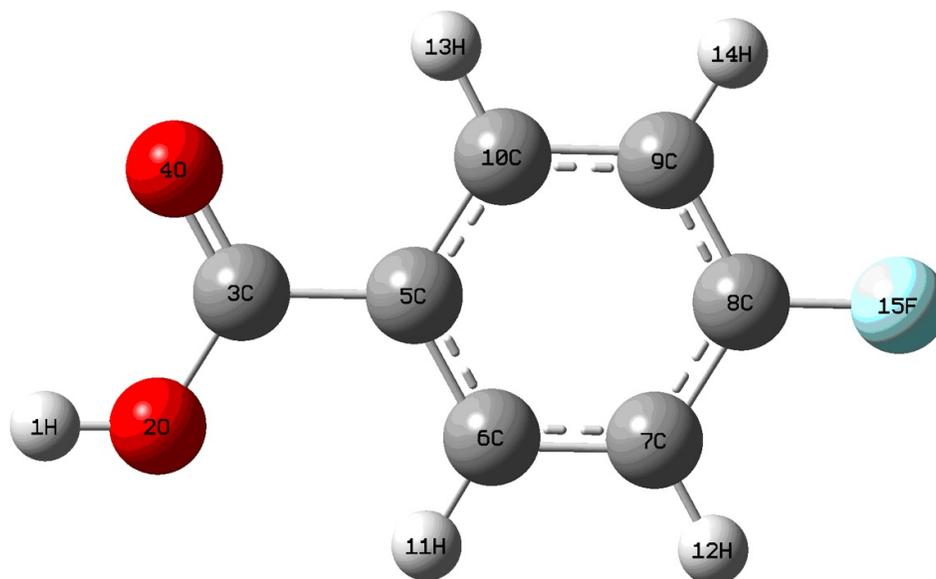
$C_9H_8O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	COCH315	SUM
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.412
O2	0.317	8.096	0.437	0.149	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.437	2.846	0.655	0.482	0.026	0.004	0.005	0.004	0.027	0.003	0.001	0.003	0.001	0.001	4.500
O4	0.010	0.149	0.655	8.202	0.053	0.015	0.002	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.015	9.138
C5	0.004	0.043	0.482	0.053	3.911	0.672	0.036	0.044	0.036	0.663	0.021	0.005	0.021	0.004	0.014	6.008
C6	0.001	0.019	0.026	0.015	0.672	3.945	0.699	0.035	0.048	0.033	0.469	0.024	0.004	0.002	0.009	5.999
C7	0.000	0.002	0.004	0.002	0.036	0.699	3.956	0.673	0.034	0.047	0.022	0.472	0.002	0.004	0.063	6.016
C8	0.000	0.002	0.005	0.006	0.044	0.035	0.673	3.932	0.663	0.036	0.004	0.022	0.005	0.021	0.586	6.033
C9	0.000	0.001	0.004	0.002	0.036	0.048	0.034	0.663	3.947	0.708	0.002	0.005	0.022	0.465	0.059	5.998
C10	0.001	0.007	0.027	0.023	0.663	0.033	0.047	0.036	0.708	3.945	0.005	0.002	0.468	0.022	0.009	5.996
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.022	0.004	0.002	0.005	0.390	0.003	0.001	0.000	0.001	0.938
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.472	0.022	0.005	0.002	0.003	0.412	0.000	0.001	0.024	0.971
H13	0.000	0.001	0.003	0.015	0.021	0.004	0.002	0.005	0.022	0.468	0.001	0.000	0.383	0.003	0.001	0.929
H14	0.000	0.000	0.001	0.000	0.004	0.002	0.004	0.021	0.465	0.022	0.000	0.001	0.003	0.378	0.023	0.925
COCH315	0.000	0.001	0.001	0.002	0.014	0.009	0.063	0.586	0.059	0.009	0.001	0.024	0.001	0.023	22.252	23.045
SUM	0.412	9.091	4.5	9.138	6.008	5.999	6.016	6.033	5.998	5.996	0.938	0.971	0.929	0.925	23.045	86.001

Benzoic Acid Structures with their LDMs



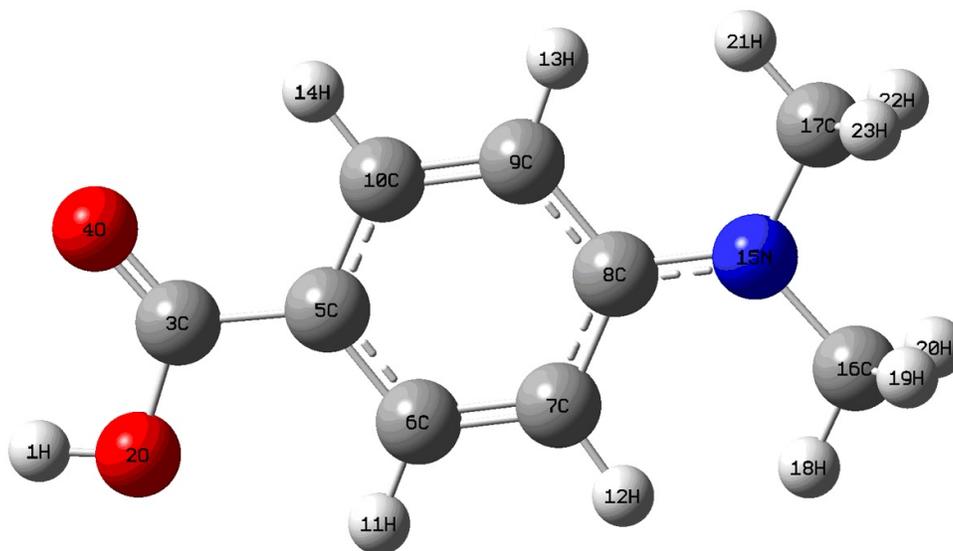
$C_8H_6O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	C15	O16	H17	SUM
H1	0.073	0.316	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.411
O2	0.316	8.096	0.438	0.150	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.000	0.001	0.000	9.091
C3	0.006	0.438	2.843	0.655	0.481	0.025	0.004	0.004	0.004	0.027	0.003	0.001	0.003	0.001	0.000	0.001	0.000	4.496
O4	0.010	0.150	0.655	8.200	0.053	0.015	0.002	0.006	0.002	0.023	0.001	0.000	0.016	0.000	0.001	0.002	0.000	9.136
C5	0.004	0.043	0.481	0.053	3.910	0.673	0.036	0.043	0.036	0.661	0.021	0.005	0.021	0.005	0.006	0.010	0.000	6.007
C6	0.001	0.019	0.025	0.015	0.673	3.944	0.698	0.035	0.048	0.033	0.469	0.024	0.004	0.002	0.005	0.002	0.001	5.996
C7	0.000	0.002	0.004	0.002	0.036	0.698	3.956	0.671	0.034	0.046	0.022	0.475	0.002	0.005	0.033	0.021	0.006	6.014
C8	0.000	0.002	0.004	0.006	0.043	0.035	0.671	3.927	0.659	0.037	0.004	0.023	0.005	0.021	0.503	0.063	0.027	6.029
C9	0.000	0.001	0.004	0.002	0.036	0.048	0.034	0.659	3.948	0.710	0.002	0.005	0.022	0.468	0.027	0.026	0.004	5.997
C10	0.001	0.007	0.027	0.023	0.661	0.033	0.046	0.037	0.71	3.944	0.004	0.002	0.467	0.023	0.005	0.002	0.001	5.993
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.022	0.004	0.002	0.004	0.387	0.003	0.001	0.000	0.001	0.000	0.000	0.934
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.475	0.023	0.005	0.002	0.003	0.414	0.000	0.001	0.006	0.002	0.007	0.967
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.005	0.022	0.467	0.001	0.000	0.381	0.003	0.001	0.000	0.000	0.926
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.005	0.021	0.468	0.023	0.000	0.001	0.003	0.384	0.003	0.014	0.001	0.930
C15	0.000	0.000	0.000	0.001	0.006	0.005	0.033	0.503	0.027	0.005	0.001	0.006	0.001	0.003	3.245	0.737	0.440	5.012
O16	0.000	0.001	0.001	0.002	0.010	0.002	0.021	0.063	0.026	0.002	0.000	0.002	0.000	0.014	0.737	8.133	0.060	9.073
H17	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.027	0.004	0.001	0.000	0.007	0.000	0.001	0.440	0.060	0.440	0.986
SUM	0.411	9.091	4.496	9.136	6.007	5.996	6.014	6.029	5.997	5.993	0.934	0.967	0.926	0.930	5.012	9.073	0.986	77.999

$C_8H_6O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	COH15	SUM
H1	0.073	0.316	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.411
O2	0.316	8.096	0.438	0.150	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.438	2.843	0.655	0.481	0.025	0.004	0.004	0.004	0.027	0.003	0.001	0.003	0.001	0.002	4.496
O4	0.010	0.150	0.655	8.200	0.053	0.015	0.002	0.006	0.002	0.023	0.001	0.000	0.016	0.000	0.003	9.136
C5	0.004	0.043	0.481	0.053	3.910	0.673	0.036	0.043	0.036	0.661	0.021	0.005	0.021	0.005	0.016	6.007
C6	0.001	0.019	0.025	0.015	0.673	3.944	0.698	0.035	0.048	0.033	0.469	0.024	0.004	0.002	0.008	5.996
C7	0.000	0.002	0.004	0.002	0.036	0.698	3.956	0.671	0.034	0.046	0.022	0.475	0.002	0.005	0.061	6.014
C8	0.000	0.002	0.004	0.006	0.043	0.035	0.671	3.927	0.659	0.037	0.004	0.023	0.005	0.021	0.593	6.029
C9	0.000	0.001	0.004	0.002	0.036	0.048	0.034	0.659	3.948	0.710	0.002	0.005	0.022	0.468	0.057	5.997
C10	0.001	0.007	0.027	0.023	0.661	0.033	0.046	0.037	0.710	3.944	0.004	0.002	0.467	0.023	0.008	5.993
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.022	0.004	0.002	0.004	0.387	0.003	0.001	0.000	0.001	0.934
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.475	0.023	0.005	0.002	0.003	0.414	0.000	0.001	0.015	0.967
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.005	0.022	0.467	0.001	0.000	0.381	0.003	0.001	0.926
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.005	0.021	0.468	0.023	0.000	0.001	0.003	0.384	0.018	0.93
COH15	0.000	0.001	0.002	0.003	0.016	0.008	0.061	0.593	0.057	0.008	0.001	0.015	0.001	0.018	14.290	15.072
SUM	0.411	9.091	4.496	9.136	6.007	5.996	6.014	6.029	5.997	5.993	0.934	0.967	0.926	0.930	15.072	77.999



$C_7H_5FO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	F15	SUM
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.413
O2	0.317	8.098	0.436	0.148	0.043	0.019	0.002	0.002	0.002	0.008	0.015	0.000	0.001	0.000	0.001	9.092
C3	0.006	0.436	2.847	0.652	0.487	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.001	4.505
O4	0.010	0.148	0.652	8.211	0.054	0.015	0.003	0.006	0.002	0.024	0.001	0.000	0.015	0.000	0.002	9.144
C5	0.004	0.043	0.487	0.054	3.909	0.669	0.035	0.042	0.035	0.666	0.021	0.005	0.021	0.005	0.008	6.004
C6	0.001	0.019	0.026	0.015	0.669	3.937	0.699	0.039	0.047	0.034	0.468	0.022	0.004	0.002	0.008	5.991
C7	0.000	0.002	0.005	0.003	0.035	0.699	3.942	0.664	0.031	0.047	0.022	0.475	0.002	0.004	0.053	5.984
C8	0.000	0.002	0.005	0.006	0.042	0.039	0.664	3.600	0.662	0.040	0.005	0.024	0.005	0.024	0.415	5.533
C9	0.000	0.002	0.005	0.002	0.035	0.047	0.031	0.662	3.942	0.702	0.002	0.004	0.022	0.475	0.053	5.984
C10	0.001	0.008	0.028	0.024	0.666	0.034	0.047	0.040	0.702	3.936	0.005	0.002	0.467	0.022	0.008	5.988
H11	0.000	0.015	0.003	0.001	0.021	0.468	0.022	0.005	0.002	0.005	0.386	0.003	0.001	0.000	0.001	0.933
H12	0.000	0.000	0.001	0.000	0.005	0.022	0.475	0.024	0.004	0.002	0.003	0.399	0.000	0.001	0.007	0.943
H13	0.000	0.001	0.003	0.015	0.021	0.004	0.002	0.005	0.022	0.467	0.001	0.000	0.380	0.003	0.001	0.925
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.024	0.475	0.022	0.000	0.001	0.003	0.398	0.007	0.941
F15	0.000	0.001	0.001	0.002	0.008	0.008	0.053	0.415	0.053	0.008	0.001	0.007	0.001	0.007	9.055	9.620
SUM	0.413	9.092	4.505	9.144	6.004	5.991	5.984	5.533	5.984	5.988	0.933	0.943	0.925	0.941	9.620	72.000

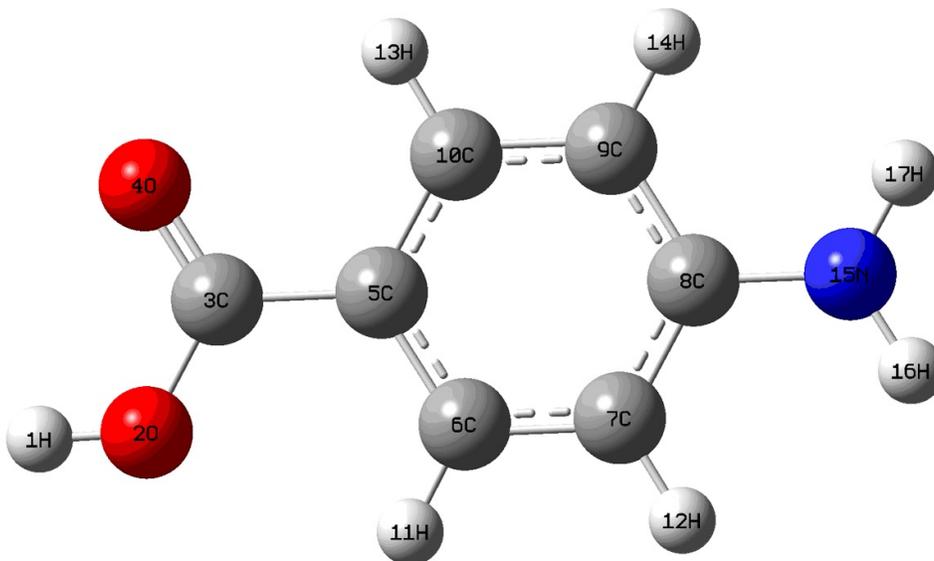
Benzoic Acid Structures with their LDMs



$C_9H_7NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	C16	C17	H18	H19	H20	H21	H22	H23	SUM
H1	0.076	0.321	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.420
O2	0.321	8.101	0.432	0.146	0.044	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.093
C3	0.006	0.432	2.867	0.645	0.501	0.026	0.006	0.005	0.006	0.028	0.002	0.001	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4.535
O4	0.011	0.146	0.645	8.227	0.055	0.015	0.004	0.006	0.002	0.023	0.001	0.000	0.000	0.005	0.015	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.156
C5	0.004	0.044	0.501	0.055	3.916	0.658	0.036	0.034	0.036	0.655	0.021	0.005	0.005	0.021	0.015	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.001	6.011
C6	0.001	0.018	0.026	0.015	0.658	3.936	0.718	0.035	0.041	0.033	0.470	0.024	0.002	0.005	0.009	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.001	5.997
C7	0.000	0.002	0.006	0.004	0.036	0.718	3.975	0.628	0.031	0.041	0.024	0.474	0.004	0.002	0.051	0.012	0.005	0.006	0.001	0.007	0.001	0.001	0.001	6.030
C8	0.000	0.002	0.005	0.006	0.034	0.035	0.628	3.575	0.626	0.035	0.005	0.022	0.022	0.005	0.542	0.015	0.015	0.004	0.004	0.003	0.003	0.004	0.004	5.589
C9	0.000	0.002	0.006	0.002	0.036	0.041	0.031	0.626	3.974	0.721	0.002	0.004	0.474	0.024	0.051	0.005	0.011	0.001	0.001	0.001	0.007	0.001	0.006	6.029
C10	0.001	0.008	0.028	0.023	0.655	0.033	0.041	0.035	0.721	3.935	0.005	0.002	0.024	0.469	0.011	0.001	0.001	0.001	0.000	0.000	0.001	0.000	0.001	5.994
H11	0.000	0.015	0.002	0.001	0.021	0.470	0.024	0.005	0.002	0.005	0.397	0.003	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.948
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.474	0.022	0.004	0.002	0.003	0.418	0.001	0.000	0.005	0.012	0.000	0.005	0.001	0.008	0.000	0.000	0.000	0.986
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.022	0.474	0.024	0.000	0.001	0.417	0.003	0.005	0.000	0.012	0.000	0.000	0.000	0.008	0.001	0.005	0.985
H14	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.005	0.024	0.469	0.001	0.000	0.003	0.392	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.941
N15	0.000	0.001	0.004	0.004	0.015	0.009	0.051	0.542	0.051	0.010	0.001	0.005	0.005	0.001	6.262	0.476	0.476	0.032	0.027	0.030	0.030	0.027	0.032	8.093
C16	0.000	0.000	0.000	0.000	0.001	0.001	0.012	0.015	0.005	0.001	0.000	0.012	0.000	0.000	0.476	3.697	0.017	0.464	0.466	0.462	0.003	0.008	0.002	5.640
C17	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.015	0.011	0.001	0.000	0.000	0.012	0.000	0.476	0.017	3.697	0.002	0.008	0.003	0.462	0.466	0.464	5.640
H18	0.000	0.000	0.000	0.000	0.001	0.001	0.006	0.004	0.001	0.001	0.000	0.005	0.000	0.000	0.032	0.464	0.002	0.436	0.018	0.019	0.001	0.000	0.000	0.983
H19	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.000	0.000	0.001	0.000	0.000	0.027	0.466	0.008	0.018	0.428	0.017	0.001	0.010	0.000	0.981
H20	0.000	0.000	0.000	0.000	0.001	0.001	0.007	0.003	0.001	0.000	0.000	0.008	0.000	0.000	0.030	0.462	0.003	0.019	0.017	0.429	0.000	0.001	0.001	0.984
H21	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.003	0.007	0.001	0.000	0.000	0.008	0.000	0.030	0.003	0.462	0.001	0.001	0.000	0.428	0.017	0.019	0.983
H22	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.027	0.008	0.466	0.000	0.010	0.001	0.017	0.428	0.018	0.983
H23	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.004	0.006	0.001	0.000	0.000	0.005	0.000	0.032	0.002	0.464	0.000	0.000	0.001	0.019	0.018	0.436	0.990
SUM	0.420	9.093	4.535	9.156	6.011	5.997	6.030	5.589	6.029	5.994	0.948	0.986	0.985	0.941	8.093	5.640	5.640	0.991	0.983	0.984	0.983	0.983	0.990	88.000

$C_9H_7NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NCH3CH315	SUM
H1	0.076	0.321	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.420
O2	0.321	8.101	0.432	0.146	0.044	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.000	0.001	0.000	9.093
C3	0.006	0.432	2.867	0.645	0.501	0.026	0.006	0.005	0.006	0.028	0.002	0.001	0.001	0.003	0.005	4.535
O4	0.011	0.146	0.645	8.227	0.055	0.015	0.004	0.006	0.002	0.023	0.001	0.000	0.000	0.005	0.005	9.156
C5	0.004	0.044	0.501	0.055	3.916	0.658	0.036	0.034	0.036	0.655	0.021	0.005	0.005	0.021	0.021	6.011
C6	0.001	0.018	0.026	0.015	0.658	3.936	0.718	0.035	0.041	0.033	0.470	0.024	0.002	0.005	0.014	5.997
C7	0.000	0.002	0.006	0.004	0.036	0.718	3.975	0.628	0.031	0.041	0.024	0.474	0.004	0.002	0.085	6.03
C8	0.000	0.002	0.005	0.006	0.034	0.035	0.628	3.575	0.626	0.035	0.005	0.022	0.022	0.005	0.593	5.589
C9	0.000	0.002	0.006	0.002	0.036	0.041	0.031	0.626	3.974	0.721	0.002	0.004	0.474	0.024	0.085	6.029
C10	0.001	0.008	0.028	0.023	0.655	0.033	0.041	0.035	0.721	3.935	0.005	0.002	0.024	0.469	0.014	5.994
H11	0.000	0.015	0.002	0.001	0.021	0.470	0.024	0.005	0.002	0.005	0.397	0.003	0.000	0.001	0.001	0.948
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.474	0.022	0.004	0.002	0.003	0.418	0.001	0.000	0.031	0.986
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.022	0.474	0.024	0.000	0.001	0.417	0.003	0.031	0.985
H14	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.005	0.024	0.469	0.001	0.000	0.003	0.392	0.001	0.941
NCH3CH315	0.000	0.002	0.005	0.005	0.021	0.014	0.085	0.593	0.085	0.014	0.001	0.031	0.031	0.001	24.398	25.287
SUM	0.420	9.093	4.535	9.156	6.011	5.997	6.030	5.589	6.029	5.994	0.948	0.986	0.985	0.941	25.287	88.000

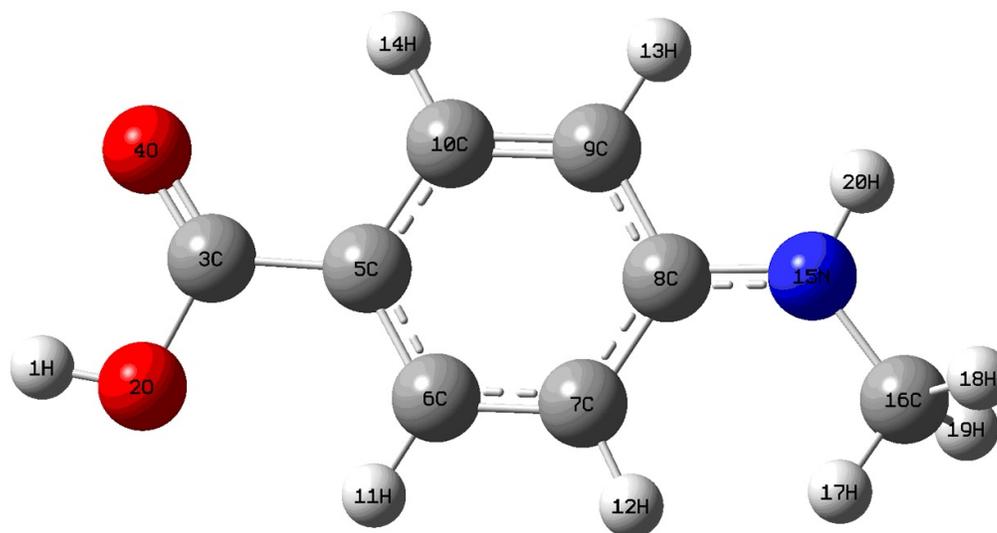
Benzoic Acid Structures with their LDMs



$C_7H_7NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	H16	H17	SUM
H1	0.075	0.320	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.418
O2	0.320	8.100	0.433	0.146	0.044	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.001	0.000	0.001	0.000	0.000	9.092
C3	0.006	0.433	2.863	0.646	0.498	0.026	0.006	0.005	0.006	0.028	0.003	0.001	0.003	0.001	0.004	0.000	0.000	4.529
O4	0.011	0.146	0.646	8.224	0.055	0.015	0.004	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.004	0.000	0.000	9.154
C5	0.004	0.044	0.498	0.055	3.914	0.660	0.036	0.038	0.036	0.656	0.021	0.005	0.020	0.005	0.015	0.001	0.001	6.009
C6	0.001	0.018	0.026	0.015	0.660	3.937	0.715	0.036	0.042	0.033	0.470	0.025	0.004	0.002	0.010	0.000	0.001	5.996
C7	0.000	0.002	0.006	0.004	0.036	0.715	3.966	0.640	0.031	0.043	0.023	0.480	0.002	0.004	0.056	0.004	0.005	6.016
C8	0.000	0.002	0.005	0.006	0.038	0.036	0.640	3.579	0.638	0.036	0.005	0.023	0.005	0.023	0.536	0.011	0.011	5.594
C9	0.000	0.002	0.006	0.002	0.036	0.042	0.031	0.638	3.966	0.718	0.002	0.004	0.023	0.479	0.056	0.005	0.004	6.016
C10	0.001	0.008	0.028	0.023	0.656	0.033	0.043	0.036	0.718	3.936	0.005	0.002	0.469	0.025	0.010	0.001	0.000	5.993
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.005	0.002	0.005	0.394	0.003	0.001	0.000	0.001	0.000	0.000	0.944
H12	0.000	0.000	0.001	0.000	0.005	0.025	0.480	0.023	0.004	0.002	0.003	0.432	0.000	0.001	0.007	0.003	0.000	0.987
H13	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.023	0.469	0.001	0.000	0.389	0.003	0.001	0.000	0.000	0.937
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.479	0.025	0.000	0.001	0.003	0.431	0.007	0.000	0.003	0.986
N15	0.000	0.001	0.004	0.004	0.015	0.010	0.056	0.536	0.056	0.010	0.001	0.007	0.001	0.007	6.533	0.420	0.420	8.080
H16	0.000	0.000	0.000	0.000	0.001	0.000	0.004	0.011	0.005	0.001	0.000	0.003	0.000	0.000	0.420	0.170	0.009	0.625
H17	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.011	0.004	0.000	0.000	0.000	0.000	0.003	0.420	0.009	0.170	0.624
SUM	0.418	9.092	4.529	9.154	6.009	5.996	6.016	5.594	6.016	5.993	0.944	0.987	0.937	0.986	8.08	0.625	0.624	72.000

$C_7H_7NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NH215	SUM
H1	0.075	0.320	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.418
O2	0.320	8.100	0.433	0.146	0.044	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.001	0.000	0.001	9.092
C3	0.006	0.433	2.863	0.646	0.498	0.026	0.006	0.005	0.006	0.028	0.003	0.001	0.003	0.001	0.004	4.529
O4	0.011	0.146	0.646	8.224	0.055	0.015	0.004	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.004	9.154
C5	0.004	0.044	0.498	0.055	3.914	0.660	0.036	0.038	0.036	0.656	0.021	0.005	0.020	0.005	0.016	6.009
C6	0.001	0.018	0.026	0.015	0.660	3.937	0.715	0.036	0.042	0.033	0.470	0.025	0.004	0.002	0.011	5.996
C7	0.000	0.002	0.006	0.004	0.036	0.715	3.966	0.640	0.031	0.043	0.023	0.480	0.002	0.004	0.065	6.016
C8	0.000	0.002	0.005	0.006	0.038	0.036	0.640	3.579	0.638	0.036	0.005	0.023	0.005	0.023	0.557	5.594
C9	0.000	0.002	0.006	0.002	0.036	0.042	0.031	0.638	3.966	0.718	0.002	0.004	0.023	0.479	0.065	6.016
C10	0.001	0.008	0.028	0.023	0.656	0.033	0.043	0.036	0.718	3.936	0.005	0.002	0.469	0.025	0.011	5.993
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.005	0.002	0.005	0.394	0.003	0.001	0.000	0.001	0.944
H12	0.000	0.000	0.001	0.000	0.005	0.025	0.480	0.023	0.004	0.002	0.003	0.432	0.000	0.001	0.010	0.987
H13	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.023	0.469	0.001	0.000	0.389	0.003	0.001	0.937
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.479	0.025	0.000	0.001	0.003	0.431	0.010	0.986
NH215	0.000	0.001	0.004	0.004	0.016	0.011	0.065	0.557	0.065	0.011	0.001	0.010	0.001	0.010	8.571	9.329
SUM	0.418	9.092	4.529	9.154	6.009	5.996	6.016	5.594	6.016	5.993	0.944	0.987	0.937	0.986	9.329	72.000

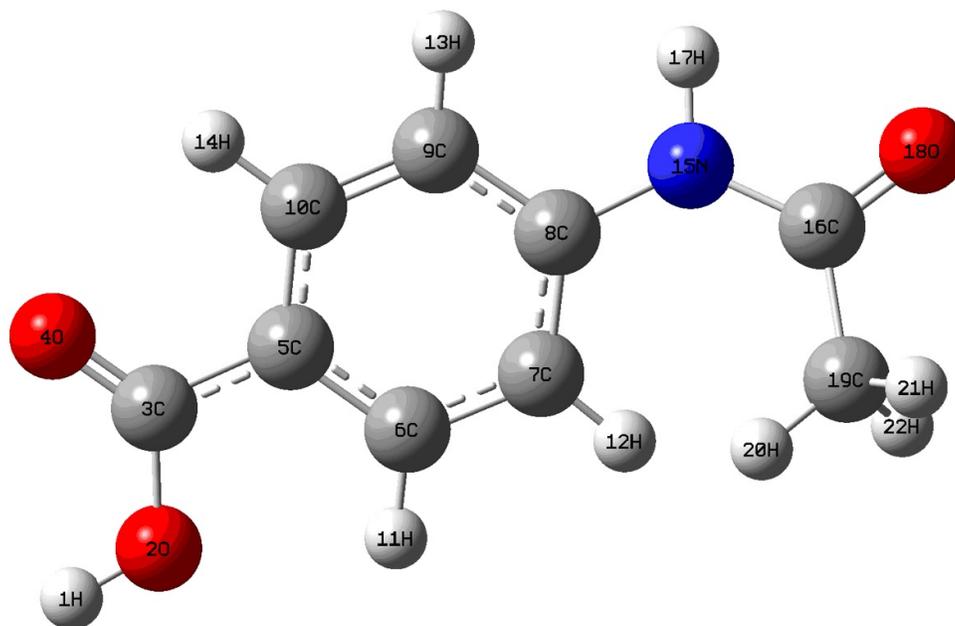
Benzoic Acid Structures with their LDMs



$C_8H_9NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	C16	H17	H18	H19	H20	SUM
H1	0.076	0.320	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.419
O2	0.320	8.101	0.432	0.146	0.044	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	9.093
C3	0.006	0.432	2.866	0.645	0.501	0.027	0.006	0.005	0.006	0.028	0.003	0.001	0.001	0.003	0.004	0.000	0.000	0.000	0.000	0.000	4.534
O4	0.011	0.146	0.645	8.226	0.055	0.016	0.004	0.006	0.002	0.023	0.001	0.000	0.000	0.015	0.004	0.000	0.000	0.000	0.000	0.000	9.156
C5	0.004	0.044	0.501	0.055	3.915	0.664	0.036	0.035	0.036	0.648	0.021	0.005	0.005	0.020	0.015	0.001	0.001	0.000	0.001	0.001	6.010
C6	0.001	0.019	0.027	0.016	0.664	3.935	0.710	0.036	0.043	0.033	0.470	0.024	0.002	0.004	0.009	0.001	0.001	0.000	0.001	0.001	5.996
C7	0.000	0.002	0.006	0.004	0.036	0.710	3.975	0.637	0.031	0.040	0.023	0.476	0.004	0.002	0.054	0.009	0.005	0.001	0.005	0.005	6.027
C8	0.000	0.002	0.005	0.006	0.035	0.036	0.637	3.568	0.626	0.035	0.004	0.022	0.023	0.005	0.539	0.015	0.002	0.004	0.003	0.011	5.578
C9	0.000	0.002	0.006	0.002	0.036	0.043	0.031	0.626	3.969	0.728	0.002	0.004	0.479	0.024	0.055	0.005	0.001	0.001	0.001	0.001	6.020
C10	0.001	0.007	0.028	0.023	0.648	0.033	0.040	0.035	0.728	3.936	0.005	0.002	0.025	0.469	0.011	0.001	0.000	0.000	0.001	0.000	5.994
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.004	0.002	0.005	0.396	0.003	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.946
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.476	0.022	0.004	0.002	0.003	0.420	0.001	0.000	0.005	0.009	0.006	0.001	0.004	0.000	0.985
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.479	0.025	0.000	0.001	0.434	0.003	0.008	0.000	0.000	0.000	0.000	0.004	0.991
H14	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.024	0.469	0.001	0.000	0.003	0.390	0.001	0.000	0.000	0.000	0.000	0.000	0.939
N15	0.000	0.001	0.004	0.004	0.015	0.009	0.054	0.539	0.055	0.011	0.001	0.005	0.008	0.001	6.405	0.482	0.032	0.031	0.034	0.414	8.106
C16	0.000	0.000	0.000	0.000	0.001	0.001	0.009	0.015	0.005	0.001	0.000	0.009	0.000	0.000	0.482	3.694	0.464	0.471	0.464	0.011	5.629
H17	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.002	0.001	0.000	0.000	0.006	0.000	0.000	0.032	0.464	0.431	0.018	0.019	0.003	0.983
H18	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.000	0.000	0.001	0.000	0.000	0.031	0.471	0.018	0.431	0.018	0.002	0.978
H19	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.003	0.001	0.001	0.000	0.004	0.000	0.000	0.034	0.464	0.019	0.018	0.437	0.001	0.990
H20	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.011	0.005	0.000	0.000	0.000	0.004	0.000	0.414	0.011	0.003	0.002	0.001	0.170	0.627
SUM	0.419	9.093	4.534	9.156	6.010	5.996	6.027	5.578	6.020	5.994	0.946	0.985	0.991	0.939	8.106	5.629	0.983	0.978	0.990	0.627	80.000

$C_8H_9NO_2$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NHCH315	SUM
H1	0.076	0.320	0.006	0.011	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.419
O2	0.320	8.101	0.432	0.146	0.044	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.000	0.001	0.002	9.093
C3	0.006	0.432	2.866	0.645	0.501	0.027	0.006	0.005	0.006	0.028	0.003	0.001	0.001	0.003	0.005	4.534
O4	0.011	0.146	0.645	8.226	0.055	0.016	0.004	0.006	0.002	0.023	0.001	0.000	0.000	0.015	0.005	9.156
C5	0.004	0.044	0.501	0.055	3.915	0.664	0.036	0.035	0.036	0.648	0.021	0.005	0.005	0.020	0.019	6.010
C6	0.001	0.019	0.027	0.016	0.664	3.935	0.710	0.036	0.043	0.033	0.470	0.024	0.002	0.004	0.012	5.996
C7	0.000	0.002	0.006	0.004	0.036	0.710	3.975	0.637	0.031	0.040	0.023	0.476	0.004	0.002	0.080	6.027
C8	0.000	0.002	0.005	0.006	0.035	0.036	0.637	3.568	0.626	0.035	0.004	0.022	0.023	0.005	0.575	5.578
C9	0.000	0.002	0.006	0.002	0.036	0.043	0.031	0.626	3.969	0.728	0.002	0.004	0.479	0.024	0.068	6.020
C10	0.001	0.007	0.028	0.023	0.648	0.033	0.040	0.035	0.728	3.936	0.005	0.002	0.025	0.469	0.014	5.994
H11	0.000	0.015	0.003	0.001	0.021	0.470	0.023	0.004	0.002	0.005	0.396	0.003	0.000	0.001	0.001	0.946
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.476	0.022	0.004	0.002	0.003	0.420	0.001	0.000	0.006	0.985
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.479	0.025	0.000	0.001	0.434	0.003	0.012	0.991
H14	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.024	0.469	0.001	0.000	0.003	0.390	0.001	0.939
NHCH315	0.000	0.001	0.004	0.004	0.015	0.009	0.054	0.539	0.055	0.011	0.001	0.005	0.008	0.001	16.495	17.313
SUM	0.419	9.093	4.534	9.156	6.010	5.996	6.027	5.578	6.020	5.994	0.946	0.985	0.991	0.939	17.313	80.000

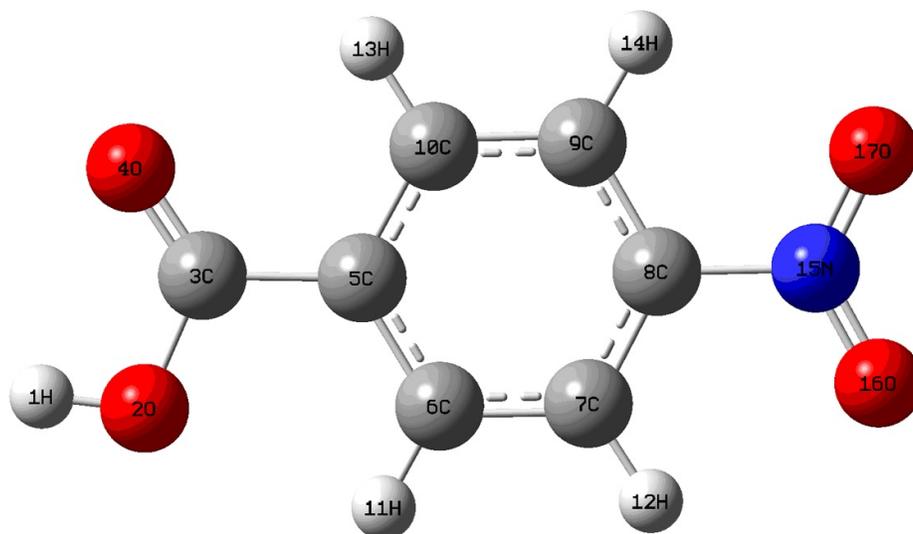
Benzoic Acid Structures with their LDMs



$C_9H_9NO_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	C16	H17	O18	C19	H20	H21	H22	SUM	
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.414
O2	0.317	8.099	0.435	0.148	0.043	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	9.092
C3	0.006	0.435	2.850	0.651	0.489	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.001	0.003	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	4.510
O4	0.010	0.148	0.651	8.212	0.054	0.016	0.003	0.006	0.002	0.023	0.001	0.000	0.000	0.015	0.002	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	9.145
C5	0.004	0.043	0.489	0.054	3.910	0.668	0.035	0.040	0.035	0.660	0.021	0.005	0.005	0.020	0.009	0.002	0.001	0.003	0.000	0.000	0.000	0.000	0.000	6.006
C6	0.001	0.019	0.026	0.016	0.668	3.938	0.703	0.036	0.045	0.033	0.469	0.023	0.002	0.004	0.007	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	5.994
C7	0.000	0.002	0.005	0.003	0.035	0.703	3.947	0.652	0.031	0.044	0.023	0.470	0.004	0.002	0.045	0.005	0.004	0.005	0.010	0.010	0.001	0.003	6.005	
C8	0.000	0.002	0.005	0.006	0.040	0.036	0.652	3.616	0.646	0.036	0.004	0.022	0.023	0.004	0.502	0.011	0.009	0.011	0.004	0.005	0.000	0.000	5.994	
C9	0.000	0.002	0.005	0.002	0.035	0.045	0.031	0.646	3.955	0.712	0.002	0.004	0.477	0.023	0.050	0.006	0.004	0.005	0.001	0.000	0.000	0.000	6.007	
C10	0.001	0.007	0.028	0.023	0.660	0.033	0.044	0.036	0.712	3.937	0.005	0.002	0.024	0.468	0.008	0.001	0.000	0.001	0.000	0.000	0.000	0.000	5.991	
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.023	0.004	0.002	0.005	0.389	0.003	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.938	
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.470	0.022	0.004	0.002	0.003	0.408	0.001	0.000	0.005	0.001	0.000	0.001	0.010	0.004	0.001	0.006	0.968	
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.477	0.024	0.000	0.001	0.419	0.003	0.009	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.972	
H14	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.004	0.023	0.468	0.001	0.000	0.003	0.383	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.929	
N15	0.000	0.001	0.002	0.002	0.009	0.007	0.045	0.502	0.050	0.008	0.001	0.005	0.009	0.001	6.413	0.491	0.397	0.134	0.037	0.005	0.005	0.005	8.127	
C16	0.000	0.000	0.000	0.000	0.002	0.000	0.005	0.011	0.006	0.001	0.000	0.001	0.000	0.000	0.491	2.918	0.009	0.659	0.471	0.019	0.019	0.023	4.636	
H17	0.000	0.000	0.000	0.000	0.001	0.001	0.004	0.009	0.004	0.000	0.000	0.000	0.003	0.000	0.397	0.009	0.156	0.011	0.004	0.000	0.001	0.000	6.000	
O18	0.000	0.000	0.001	0.001	0.003	0.001	0.005	0.011	0.005	0.001	0.000	0.001	0.000	0.000	0.134	0.659	0.011	8.212	0.059	0.010	0.012	0.010	9.136	
C19	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.004	0.001	0.000	0.000	0.010	0.000	0.000	0.037	0.471	0.004	0.059	3.986	0.471	0.472	0.471	5.998	
H20	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.005	0.000	0.000	0.000	0.004	0.000	0.000	0.005	0.019	0.000	0.010	0.471	0.415	0.016	0.018	0.975	
H21	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.005	0.019	0.001	0.012	0.472	0.016	0.402	0.017	0.948	
H22	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.005	0.023	0.000	0.010	0.471	0.018	0.017	0.416	0.971	
SUM	0.414	9.092	4.510	9.145	6.006	5.994	6.005	5.636	6.007	5.991	0.938	0.968	0.972	0.929	8.127	4.636	0.600	9.136	5.998	0.975	0.948	0.971	93.998	

$C_9H_9NO_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NHCOCH315	SUM
H1	0.074	0.317	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.414
O2	0.317	8.099	0.435	0.148	0.043	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.000	0.001	0.001	9.092
C3	0.006	0.435	2.850	0.651	0.489	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.001	0.003	0.003	4.510
O4	0.010	0.148	0.651	8.212	0.054	0.016	0.003	0.006	0.002	0.023	0.001	0.000	0.000	0.015	0.003	9.145
C5	0.004	0.043	0.489	0.054	3.910	0.668	0.035	0.040	0.035	0.660	0.021	0.005	0.005	0.020	0.016	6.006
C6	0.001	0.019	0.026	0.016	0.668	3.938	0.703	0.036	0.045	0.033	0.469	0.023	0.002	0.004	0.012	5.994
C7	0.000	0.002	0.005	0.003	0.035	0.703	3.947	0.652	0.031	0.044	0.023	0.470	0.004	0.002	0.083	6.005
C8	0.000	0.002	0.005	0.006	0.040	0.036	0.652	3.616	0.646	0.036	0.004	0.022	0.023	0.004	0.543	5.636
C9	0.000	0.002	0.005	0.002	0.035	0.045	0.031	0.646	3.955	0.712	0.002	0.004	0.477	0.023	0.067	6.007
C10	0.001	0.007	0.028	0.023	0.660	0.033	0.044	0.036	0.712	3.937	0.005	0.002	0.024	0.468	0.011	5.991
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.023	0.004	0.002	0.005	0.389	0.003	0.000	0.001	0.001	0.938
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.470	0.022	0.004	0.002	0.003	0.408	0.001	0.000	0.028	0.968
H13	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.477	0.024	0.000	0.001	0.419	0.003	0.012	0.972
H14	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.004	0.023	0.468	0.001	0.000	0.003	0.383	0.001	0.929
NHCOCH315	0.000	0.001	0.003	0.003	0.016	0.012	0.083	0.543	0.067	0.011	0.001	0.028	0.012	0.001	30.610	31.392
SUM	0.414	9.092	4.510	9.145	6.006	5.994	6.005	5.636	6.007	5.991	0.938	0.968	0.972	0.929	31.392	93.998

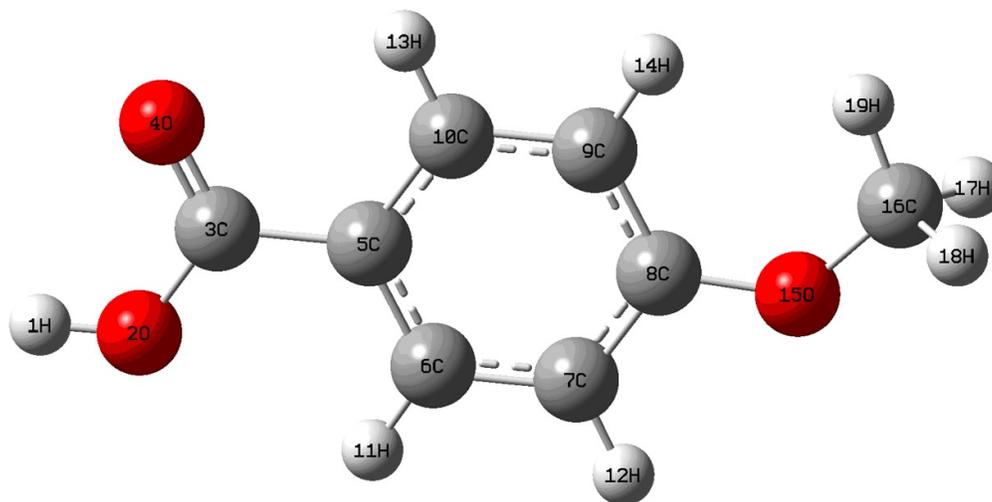
Benzoic Acid Structures with their LDMs



$C_7H_5NO_4$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	N15	O16	O17	SUM
H1	0.072	0.315	0.006	0.009	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.409
O2	0.315	8.096	0.439	0.150	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.000	0.000	0.000	9.091
C3	0.006	0.439	2.835	0.657	0.478	0.025	0.004	0.004	0.004	0.027	0.002	0.001	0.003	0.001	0.000	0.000	0.000	4.487
O4	0.009	0.150	0.657	8.196	0.054	0.015	0.002	0.006	0.002	0.024	0.001	0.000	0.016	0.000	0.001	0.001	0.001	9.133
C5	0.004	0.043	0.478	0.054	3.906	0.670	0.036	0.043	0.036	0.668	0.021	0.004	0.021	0.004	0.004	0.005	0.005	6.003
C6	0.001	0.019	0.025	0.015	0.670	3.938	0.698	0.038	0.047	0.033	0.467	0.020	0.004	0.002	0.005	0.002	0.003	5.987
C7	0.000	0.002	0.004	0.002	0.036	0.698	3.931	0.666	0.031	0.048	0.022	0.463	0.002	0.004	0.030	0.025	0.013	5.977
C8	0.000	0.002	0.004	0.006	0.043	0.038	0.666	3.766	0.664	0.038	0.004	0.021	0.004	0.021	0.424	0.047	0.047	5.796
C9	0.000	0.001	0.004	0.002	0.036	0.047	0.031	0.664	3.931	0.700	0.002	0.004	0.022	0.463	0.030	0.013	0.025	5.976
C10	0.001	0.007	0.027	0.024	0.668	0.033	0.048	0.038	0.700	3.936	0.004	0.002	0.466	0.021	0.005	0.003	0.002	5.984
H11	0.000	0.015	0.002	0.001	0.021	0.467	0.022	0.004	0.002	0.004	0.380	0.003	0.001	0.000	0.001	0.000	0.000	0.926
H12	0.000	0.000	0.001	0.000	0.004	0.020	0.463	0.021	0.004	0.002	0.003	0.364	0.000	0.001	0.003	0.018	0.001	0.907
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.004	0.022	0.466	0.001	0.000	0.375	0.003	0.001	0.000	0.000	0.918
H14	0.000	0.000	0.001	0.000	0.004	0.002	0.004	0.021	0.463	0.021	0.000	0.001	0.003	0.364	0.003	0.001	0.018	0.906
N15	0.000	0.000	0.000	0.001	0.004	0.005	0.030	0.424	0.030	0.005	0.001	0.003	0.001	0.003	4.438	0.826	0.827	6.599
O16	0.000	0.000	0.000	0.001	0.005	0.002	0.025	0.047	0.013	0.003	0.000	0.018	0.000	0.001	0.826	7.298	0.211	8.452
O17	0.000	0.000	0.000	0.001	0.005	0.003	0.013	0.047	0.025	0.002	0.000	0.001	0.000	0.018	0.827	0.211	7.297	8.451
SUM	0.409	9.091	4.487	9.133	6.003	5.987	5.977	5.796	5.976	5.984	0.926	0.907	0.918	0.906	6.599	8.452	8.451	86.000

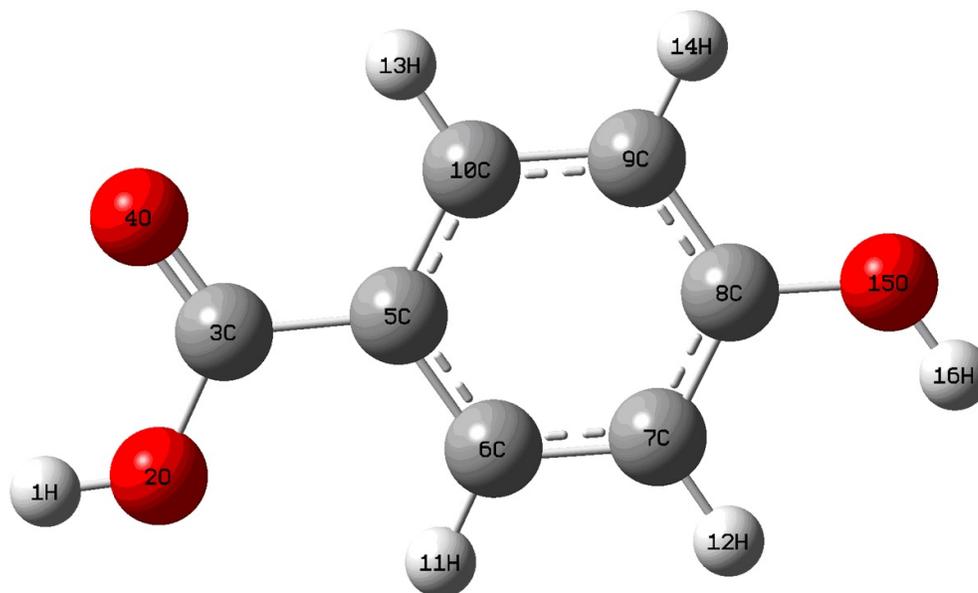
$C_7H_5NO_4$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	NO215	SUM
H1	0.072	0.315	0.006	0.009	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.409
O2	0.315	8.096	0.439	0.150	0.043	0.019	0.002	0.002	0.001	0.007	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.439	2.835	0.657	0.478	0.025	0.004	0.004	0.004	0.027	0.002	0.001	0.003	0.001	0.001	4.487
O4	0.009	0.150	0.657	8.196	0.054	0.015	0.002	0.006	0.002	0.024	0.001	0.000	0.016	0.000	0.002	9.133
C5	0.004	0.043	0.478	0.054	3.906	0.670	0.036	0.043	0.036	0.668	0.021	0.004	0.021	0.004	0.014	6.003
C6	0.001	0.019	0.025	0.015	0.670	3.938	0.698	0.038	0.047	0.033	0.467	0.020	0.004	0.002	0.010	5.987
C7	0.000	0.002	0.004	0.002	0.036	0.698	3.931	0.666	0.031	0.048	0.022	0.463	0.002	0.004	0.068	5.977
C8	0.000	0.002	0.004	0.006	0.043	0.038	0.666	3.766	0.664	0.038	0.004	0.021	0.004	0.021	0.517	5.796
C9	0.000	0.001	0.004	0.002	0.036	0.047	0.031	0.664	3.931	0.700	0.002	0.004	0.022	0.463	0.068	5.976
C10	0.001	0.007	0.027	0.024	0.668	0.033	0.048	0.038	0.700	3.936	0.004	0.002	0.466	0.021	0.010	5.984
H11	0.000	0.015	0.002	0.001	0.021	0.467	0.022	0.004	0.002	0.004	0.380	0.003	0.001	0.000	0.001	0.926
H12	0.000	0.000	0.001	0.000	0.004	0.020	0.463	0.021	0.004	0.002	0.003	0.364	0.000	0.001	0.023	0.907
H13	0.000	0.001	0.003	0.016	0.021	0.004	0.002	0.004	0.022	0.466	0.001	0.000	0.375	0.003	0.001	0.918
H14	0.000	0.000	0.001	0.000	0.004	0.002	0.004	0.021	0.463	0.021	0.000	0.001	0.003	0.364	0.023	0.906
NO215	0.000	0.001	0.001	0.002	0.014	0.010	0.068	0.517	0.068	0.010	0.001	0.023	0.001	0.023	22.761	23.501
SUM	0.409	9.091	4.487	9.133	6.003	5.987	5.977	5.796	5.976	5.984	0.926	0.907	0.918	0.906	23.501	86.000

Benzoic Acid Structures with their LDMs



$C_8H_8O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	O15	C16	H17	H18	H19	SUM	
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.417
O2	0.319	8.098	0.434	0.147	0.043	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.000	9.091
C3	0.006	0.434	2.858	0.648	0.493	0.025	0.005	0.005	0.006	0.029	0.002	0.001	0.003	0.001	0.002	0.000	0.000	0.000	0.000	0.000	4.519
O4	0.010	0.147	0.648	8.220	0.054	0.015	0.003	0.006	0.002	0.024	0.001	0.000	0.015	0.000	0.003	0.000	0.000	0.000	0.000	0.000	9.151
C5	0.004	0.043	0.493	0.054	3.913	0.655	0.035	0.039	0.036	0.670	0.021	0.005	0.021	0.005	0.012	0.001	0.000	0.001	0.001	6.008	
C6	0.001	0.018	0.025	0.015	0.655	3.939	0.719	0.037	0.043	0.033	0.469	0.023	0.005	0.002	0.009	0.001	0.000	0.000	0.000	5.996	
C7	0.000	0.002	0.005	0.003	0.035	0.719	3.952	0.636	0.031	0.046	0.023	0.475	0.002	0.004	0.057	0.005	0.001	0.001	0.001	5.998	
C8	0.000	0.002	0.005	0.006	0.039	0.037	0.636	3.539	0.651	0.037	0.005	0.022	0.004	0.023	0.466	0.014	0.004	0.003	0.003	5.496	
C9	0.000	0.002	0.006	0.002	0.036	0.043	0.031	0.651	3.970	0.700	0.002	0.004	0.023	0.473	0.055	0.011	0.001	0.006	0.006	6.023	
C10	0.001	0.008	0.029	0.024	0.670	0.033	0.046	0.037	0.700	3.935	0.005	0.002	0.468	0.023	0.008	0.001	0.000	0.001	0.001	5.991	
H11	0.000	0.015	0.002	0.001	0.021	0.469	0.023	0.005	0.002	0.005	0.392	0.003	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.941	
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.475	0.022	0.004	0.002	0.003	0.406	0.000	0.001	0.009	0.000	0.000	0.000	0.000	0.954	
H13	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.004	0.023	0.468	0.001	0.000	0.387	0.003	0.001	0.000	0.000	0.000	0.000	0.935	
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.473	0.023	0.000	0.001	0.003	0.412	0.005	0.011	0.001	0.005	0.005	0.975	
O15	0.000	0.001	0.002	0.003	0.012	0.009	0.057	0.466	0.055	0.008	0.001	0.009	0.001	0.005	7.900	0.435	0.031	0.034	0.034	9.065	
C16	0.000	0.000	0.000	0.000	0.001	0.001	0.005	0.014	0.011	0.001	0.000	0.000	0.000	0.011	0.435	3.655	0.466	0.463	0.463	5.527	
H17	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.000	0.000	0.000	0.000	0.001	0.031	0.466	0.417	0.017	0.017	0.956	
H18	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.003	0.006	0.001	0.000	0.000	0.000	0.005	0.034	0.463	0.017	0.429	0.019	0.979	
H19	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.003	0.006	0.001	0.000	0.000	0.000	0.005	0.034	0.463	0.017	0.019	0.429	0.979	
SUM	0.417	9.091	4.519	9.151	6.008	5.996	5.998	5.496	6.023	5.991	0.941	0.954	0.935	0.975	9.065	5.527	0.956	0.979	0.979	79.999	

$C_8H_8O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	OCH315	SUM
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.417
O2	0.319	8.098	0.434	0.147	0.043	0.018	0.002	0.002	0.002	0.008	0.015	0.000	0.001	0.000	0.001	9.091
C3	0.006	0.434	2.858	0.648	0.493	0.025	0.005	0.005	0.006	0.029	0.002	0.001	0.003	0.001	0.003	4.519
O4	0.010	0.147	0.648	8.220	0.054	0.015	0.003	0.006	0.002	0.024	0.001	0.000	0.015	0.000	0.003	9.151
C5	0.004	0.043	0.493	0.054	3.913	0.655	0.035	0.039	0.036	0.670	0.021	0.005	0.021	0.005	0.014	6.008
C6	0.001	0.018	0.025	0.015	0.655	3.939	0.719	0.037	0.043	0.033	0.469	0.023	0.005	0.002	0.011	5.996
C7	0.000	0.002	0.005	0.003	0.035	0.719	3.952	0.636	0.031	0.046	0.023	0.475	0.002	0.004	0.064	5.998
C8	0.000	0.002	0.005	0.006	0.039	0.037	0.636	3.539	0.651	0.037	0.005	0.022	0.004	0.023	0.490	5.496
C9	0.000	0.002	0.006	0.002	0.036	0.043	0.031	0.651	3.970	0.700	0.002	0.004	0.023	0.473	0.080	6.023
C10	0.001	0.008	0.029	0.024	0.670	0.033	0.046	0.037	0.700	3.935	0.005	0.002	0.468	0.023	0.011	5.991
H11	0.000	0.015	0.002	0.001	0.021	0.469	0.023	0.005	0.002	0.005	0.392	0.003	0.001	0.000	0.001	0.941
H12	0.000	0.000	0.001	0.000	0.005	0.023	0.475	0.022	0.004	0.002	0.003	0.406	0.000	0.001	0.010	0.954
H13	0.000	0.001	0.003	0.015	0.021	0.005	0.002	0.004	0.023	0.468	0.001	0.000	0.387	0.003	0.001	0.935
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.473	0.023	0.000	0.001	0.003	0.412	0.028	0.975
OCH315	0.000	0.001	0.003	0.003	0.014	0.011	0.064	0.490	0.080	0.011	0.001	0.010	0.001	0.028	16.788	17.505
SUM	0.417	9.091	4.519	9.151	6.008	5.996	5.998	5.496	6.023	5.991	0.941	0.954	0.935	0.975	17.505	79.999



$C_7H_6O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	O15	H16	SUM
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.416
O2	0.319	8.100	0.434	0.147	0.044	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.001	0.000	0.001	0.000	9.092
C3	0.006	0.434	2.856	0.650	0.493	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.002	0.000	4.517
O4	0.010	0.147	0.650	8.217	0.054	0.016	0.003	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.002	0.000	9.148
C5	0.004	0.044	0.493	0.054	3.911	0.668	0.036	0.040	0.035	0.658	0.021	0.005	0.020	0.005	0.012	0.000	6.006
C6	0.001	0.019	0.026	0.016	0.668	3.936	0.704	0.036	0.045	0.033	0.469	0.024	0.004	0.002	0.009	0.000	5.993
C7	0.000	0.002	0.005	0.003	0.360	0.704	3.967	0.653	0.031	0.044	0.023	0.479	0.002	0.004	0.060	0.004	6.018
C8	0.000	0.002	0.005	0.006	0.040	0.036	0.653	3.545	0.643	0.038	0.004	0.024	0.005	0.023	0.470	0.007	5.502
C9	0.000	0.002	0.005	0.002	0.035	0.045	0.031	0.643	3.952	0.714	0.002	0.004	0.023	0.475	0.057	0.004	5.995
C10	0.001	0.007	0.028	0.023	0.658	0.033	0.044	0.038	0.714	3.937	0.005	0.003	0.468	0.023	0.009	0.001	5.992
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.023	0.004	0.002	0.005	0.390	0.003	0.001	0.000	0.001	0.000	0.939
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.479	0.024	0.004	0.003	0.003	0.430	0.000	0.001	0.008	0.003	0.985
H13	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.023	0.468	0.001	0.000	0.385	0.003	0.001	0.000	0.931
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.475	0.023	0.000	0.001	0.003	0.404	0.008	0.000	0.950
O15	0.000	0.001	0.002	0.002	0.012	0.009	0.060	0.470	0.057	0.009	0.001	0.008	0.001	0.008	8.109	0.331	9.083
H16	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.007	0.004	0.001	0.000	0.003	0.000	0.000	0.331	0.081	0.432
SUM	0.416	9.092	4.517	9.148	6.006	5.993	6.018	5.502	5.995	5.992	0.939	0.985	0.931	0.950	9.083	0.432	72.000

$C_7H_6O_3$	H1	O2	C3	O4	C5	C6	C7	C8	C9	C10	H11	H12	H13	H14	OH15	SUM
H1	0.075	0.319	0.006	0.010	0.004	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.416
O2	0.319	8.100	0.434	0.147	0.044	0.019	0.002	0.002	0.002	0.007	0.015	0.000	0.001	0.000	0.001	9.092
C3	0.006	0.434	2.856	0.650	0.493	0.026	0.005	0.005	0.005	0.028	0.003	0.001	0.003	0.001	0.002	4.517
O4	0.010	0.147	0.650	8.217	0.054	0.016	0.003	0.006	0.002	0.023	0.001	0.000	0.015	0.000	0.003	9.148
C5	0.004	0.044	0.493	0.054	3.911	0.668	0.036	0.040	0.035	0.658	0.021	0.005	0.020	0.005	0.012	6.006
C6	0.001	0.019	0.026	0.016	0.668	3.936	0.704	0.036	0.045	0.033	0.469	0.024	0.004	0.002	0.010	5.993
C7	0.000	0.002	0.005	0.003	0.036	0.704	3.967	0.653	0.031	0.044	0.023	0.479	0.002	0.004	0.064	6.018
C8	0.000	0.002	0.005	0.006	0.040	0.036	0.653	3.545	0.643	0.038	0.004	0.024	0.005	0.023	0.477	5.502
C9	0.000	0.002	0.005	0.002	0.035	0.045	0.031	0.643	3.952	0.714	0.002	0.004	0.023	0.475	0.061	5.995
C10	0.001	0.007	0.028	0.023	0.658	0.033	0.044	0.038	0.714	3.937	0.005	0.003	0.468	0.023	0.010	5.992
H11	0.000	0.015	0.003	0.001	0.021	0.469	0.023	0.004	0.002	0.005	0.390	0.003	0.001	0.000	0.001	0.939
H12	0.000	0.000	0.001	0.000	0.005	0.024	0.479	0.024	0.004	0.003	0.003	0.430	0.000	0.001	0.011	0.985
H13	0.000	0.001	0.003	0.015	0.020	0.004	0.002	0.005	0.023	0.468	0.001	0.000	0.385	0.003	0.001	0.931
H14	0.000	0.000	0.001	0.000	0.005	0.002	0.004	0.023	0.475	0.023	0.000	0.001	0.003	0.404	0.009	0.950
OH15	0.000	0.001	0.002	0.003	0.012	0.010	0.064	0.477	0.061	0.010	0.001	0.011	0.001	0.009	8.853	9.515
SUM	0.416	9.092	4.517	9.148	6.006	5.993	6.018	5.502	5.995	5.992	0.939	0.985	0.931	0.950	9.515	72.000

Frobenius Distance on a per-atom basis compared with pK_a , λ_{max} , and HOMA Aromaticity measure

Table C.1: Frobenius Distance between individual atoms for only the LI to be compared directly with pK_a , the most acidic molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)

molecule	pK_a	H1	O2	C3	O4	C5	C6	C7
BANO2	3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BACN	3.55	2.83×10^{-4}	6.27×10^{-4}	3.09×10^{-3}	3.05×10^{-3}	5.44×10^{-4}	2.88×10^{-3}	3.86×10^{-3}
BACOH3	3.74	1.19×10^{-3}	2.91×10^{-4}	1.11×10^{-2}	6.74×10^{-3}	4.31×10^{-3}	7.45×10^{-3}	2.45×10^{-2}
BACOH	3.77	8.02×10^{-4}	2.49×10^{-5}	7.90×10^{-3}	3.96×10^{-3}	3.40×10^{-3}	5.96×10^{-3}	2.54×10^{-2}
BACl	3.98	1.29×10^{-3}	1.51×10^{-3}	1.13×10^{-2}	1.28×10^{-2}	1.66×10^{-3}	2.06×10^{-3}	2.47×10^{-3}
BAF	4.14	1.42×10^{-3}	2.04×10^{-3}	1.25×10^{-2}	1.51×10^{-2}	2.29×10^{-3}	2.78×10^{-4}	1.14×10^{-2}
BA	4.19	1.95×10^{-3}	4.24×10^{-4}	1.76×10^{-2}	1.58×10^{-2}	6.16×10^{-3}	1.15×10^{-2}	2.37×10^{-2}
BACH3	4.37	2.29×10^{-3}	1.08×10^{-3}	2.10×10^{-2}	1.94×10^{-2}	6.45×10^{-3}	1.05×10^{-2}	2.64×10^{-2}
BAOCH3	4.47	2.50×10^{-3}	2.20×10^{-3}	2.27×10^{-2}	2.45×10^{-2}	6.04×10^{-3}	1.91×10^{-3}	2.12×10^{-2}
BAOH	4.57	2.25×10^{-3}	3.51×10^{-3}	2.07×10^{-2}	2.14×10^{-2}	4.76×10^{-3}	1.82×10^{-3}	3.61×10^{-2}
BANH2	4.82	3.00×10^{-3}	3.73×10^{-3}	2.80×10^{-2}	2.84×10^{-2}	7.66×10^{-3}	8.47×10^{-4}	3.52×10^{-2}
BANHCOCH3	4.3	1.64×10^{-3}	2.64×10^{-3}	1.54×10^{-2}	1.65×10^{-2}	3.38×10^{-3}	5.75×10^{-4}	1.63×10^{-2}
BANHCH3	5.04	3.31×10^{-3}	4.40×10^{-3}	3.13×10^{-2}	3.08×10^{-2}	8.81×10^{-3}	2.22×10^{-3}	4.42×10^{-2}
BANCH3CH3	5.03	3.44×10^{-3}	4.33×10^{-3}	3.24×10^{-2}	3.14×10^{-2}	9.32×10^{-3}	1.65×10^{-3}	4.37×10^{-2}
molecule	pK_a	C8	C9	C10	H11	H12	H13	H14
BANO2	3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BACN	3.55	7.84×10^{-2}	3.94×10^{-3}	2.96×10^{-3}	2.20×10^{-3}	3.53×10^{-2}	2.16×10^{-3}	3.50×10^{-2}
BACOH3	3.74	1.65×10^{-1}	1.63×10^{-2}	8.47×10^{-3}	9.25×10^{-3}	4.79×10^{-2}	8.33×10^{-3}	1.44×10^{-2}
BACOH	3.77	1.61×10^{-1}	1.64×10^{-2}	7.17×10^{-3}	6.37×10^{-3}	4.97×10^{-2}	5.97×10^{-3}	2.03×10^{-2}
BACl	3.98	9.98×10^{-2}	2.49×10^{-3}	2.31×10^{-3}	5.77×10^{-3}	3.62×10^{-2}	5.86×10^{-3}	3.59×10^{-2}
BAF	4.14	1.66×10^{-1}	1.13×10^{-2}	4.90×10^{-5}	5.36×10^{-3}	3.42×10^{-2}	5.47×10^{-3}	3.39×10^{-2}
BA	4.19	1.90×10^{-1}	2.33×10^{-2}	1.14×10^{-2}	1.23×10^{-2}	5.95×10^{-2}	1.26×10^{-2}	5.92×10^{-2}
BACH3	4.37	1.22×10^{-1}	2.84×10^{-2}	1.01×10^{-2}	1.38×10^{-2}	6.27×10^{-2}	1.43×10^{-2}	6.39×10^{-2}
BAOCH3	4.47	2.28×10^{-1}	3.89×10^{-2}	1.81×10^{-3}	1.16×10^{-2}	4.18×10^{-2}	1.24×10^{-2}	4.77×10^{-2}
BAOH	4.57	2.21×10^{-1}	2.04×10^{-2}	9.40×10^{-4}	9.90×10^{-3}	6.53×10^{-2}	1.01×10^{-2}	4.04×10^{-2}
BANH2	4.82	1.87×10^{-1}	3.49×10^{-2}	3.57×10^{-4}	1.40×10^{-2}	6.76×10^{-2}	1.42×10^{-2}	6.73×10^{-2}
BANHCOCH3	4.3	1.50×10^{-1}	2.34×10^{-2}	1.08×10^{-3}	9.09×10^{-3}	4.36×10^{-2}	4.47×10^{-2}	1.92×10^{-2}
BANHCH3	5.04	1.98×10^{-1}	3.77×10^{-2}	5.92×10^{-4}	1.57×10^{-2}	5.57×10^{-2}	5.96×10^{-2}	2.64×10^{-2}
BANCH3CH3	5.03	1.91×10^{-1}	4.32×10^{-2}	1.14×10^{-3}	1.71×10^{-2}	5.36×10^{-2}	4.26×10^{-2}	2.82×10^{-2}

Table C.2: Frobenius Distance between individual atoms for only the DI to be compared directly with pK_a , the most acidic molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)

molecule	pK_a	H1	O2	C3	O4	C5	C6	C7
BANO2	3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BACN	3.55	5.31×10^{-4}	9.42×10^{-4}	2.26×10^{-3}	1.20×10^{-3}	2.56×10^{-3}	5.08×10^{-3}	1.39×10^{-2}
BACOH3	3.74	2.13×10^{-3}	2.93×10^{-3}	4.78×10^{-3}	2.05×10^{-3}	6.28×10^{-3}	5.31×10^{-3}	1.24×10^{-2}
BACOH	3.77	1.42×10^{-3}	1.94×10^{-3}	3.20×10^{-3}	1.50×10^{-3}	7.66×10^{-3}	5.51×10^{-3}	1.37×10^{-2}
BACl	3.98	2.33×10^{-3}	3.73×10^{-3}	8.80×10^{-3}	4.36×10^{-3}	7.86×10^{-3}	2.93×10^{-3}	1.36×10^{-2}
BAF	4.14	2.56×10^{-3}	4.33×10^{-3}	1.05×10^{-2}	5.18×10^{-3}	9.14×10^{-3}	3.43×10^{-3}	1.23×10^{-2}
BA	4.19	3.48×10^{-3}	4.91×10^{-3}	9.46×10^{-3}	5.11×10^{-3}	8.34×10^{-3}	5.40×10^{-3}	3.22×10^{-2}
BACH3	4.37	4.10×10^{-3}	5.76×10^{-3}	1.26×10^{-2}	6.29×10^{-3}	1.20×10^{-2}	5.94×10^{-3}	2.06×10^{-2}
BAOCH3	4.47	4.53×10^{-3}	7.17×10^{-3}	1.82×10^{-2}	8.91×10^{-3}	2.18×10^{-2}	2.64×10^{-2}	3.87×10^{-2}
BAOH	4.57	4.10×10^{-3}	6.90×10^{-3}	1.71×10^{-2}	7.93×10^{-3}	1.79×10^{-2}	8.27×10^{-3}	2.14×10^{-2}
BANH2	4.82	5.45×10^{-3}	9.05×10^{-3}	2.37×10^{-2}	1.13×10^{-2}	2.59×10^{-2}	2.12×10^{-2}	3.53×10^{-2}
BANHCOCH3	4.3	3.01×10^{-3}	5.23×10^{-3}	1.35×10^{-2}	2.25×10^{-2}	2.65×10^{-2}	7.72×10^{-3}	1.73×10^{-2}
BANHCH3	5.04	6.05×10^{-3}	9.99×10^{-3}	2.67×10^{-2}	2.50×10^{-2}	3.83×10^{-2}	1.59×10^{-2}	3.53×10^{-2}
BANCH3CH3	5.03	6.29×10^{-3}	1.02×10^{-2}	2.69×10^{-2}	2.50×10^{-2}	3.78×10^{-2}	2.55×10^{-2}	4.53×10^{-2}
molecule	pK_a	C8	C9	C10	H11	H12	H13	H14
BANO2	3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BACN	3.55	8.15×10^{-3}	1.39×10^{-2}	5.27×10^{-3}	4.28×10^{-4}	1.26×10^{-2}	4.38×10^{-4}	1.26×10^{-2}
BACOH3	3.74	8.54×10^{-3}	9.08×10^{-3}	9.78×10^{-3}	2.11×10^{-3}	9.86×10^{-3}	1.95×10^{-3}	3.37×10^{-3}
BACOH	3.77	8.18×10^{-3}	1.30×10^{-2}	1.27×10^{-2}	1.53×10^{-3}	1.27×10^{-2}	1.38×10^{-3}	6.18×10^{-3}
BACl	3.98	1.19×10^{-2}	1.35×10^{-2}	3.17×10^{-3}	1.02×10^{-3}	1.12×10^{-2}	1.14×10^{-3}	1.11×10^{-2}
BAF	4.14	4.92×10^{-3}	1.23×10^{-2}	3.76×10^{-3}	1.16×10^{-3}	1.24×10^{-2}	1.30×10^{-3}	1.23×10^{-2}
BA	4.19	3.75×10^{-2}	3.20×10^{-2}	5.51×10^{-3}	2.77×10^{-3}	1.86×10^{-2}	2.98×10^{-3}	1.86×10^{-2}
BACH3	4.37	1.52×10^{-2}	1.94×10^{-2}	1.22×10^{-2}	2.99×10^{-3}	1.56×10^{-2}	3.30×10^{-3}	1.64×10^{-2}
BAOCH3	4.47	3.28×10^{-2}	1.70×10^{-2}	5.27×10^{-3}	2.62×10^{-3}	1.29×10^{-2}	2.49×10^{-3}	1.09×10^{-2}
BAOH	4.57	2.46×10^{-2}	2.80×10^{-2}	1.73×10^{-2}	2.04×10^{-3}	1.66×10^{-2}	2.36×10^{-3}	1.32×10^{-2}
BANH2	4.82	3.66×10^{-2}	3.56×10^{-2}	2.22×10^{-2}	2.89×10^{-3}	1.75×10^{-2}	3.17×10^{-3}	1.74×10^{-2}
BANHCOCH3	4.3	3.38×10^{-2}	6.33×10^{-1}	6.29×10^{-1}	2.17×10^{-3}	8.31×10^{-3}	6.35×10^{-1}	6.28×10^{-1}
BANHCH3	5.04	5.47×10^{-2}	6.36×10^{-1}	6.30×10^{-1}	3.22×10^{-3}	1.38×10^{-2}	6.36×10^{-1}	6.28×10^{-1}
BANCH3CH3	5.03	5.98×10^{-2}	6.31×10^{-1}	6.30×10^{-1}	3.61×10^{-3}	1.18×10^{-2}	6.32×10^{-1}	6.28×10^{-1}

Table C.3: Frobenius Distance between individual atoms for only the LI to be compared directly with λ_{max} , the lowest λ_{max} molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)

molecule	λ_{max}	H1	O2	C3	O4	C5	C6	C7
BA	230	0	0	0	0	0	0	0
BAOH	255	2.94×10^{-4}	3.09×10^{-3}	3.09×10^{-3}	5.61×10^{-3}	1.40×10^{-3}	1.33×10^{-2}	1.24×10^{-2}
BANH2	288	1.04×10^{-3}	3.31×10^{-3}	1.04×10^{-2}	1.26×10^{-2}	1.50×10^{-3}	1.23×10^{-2}	1.15×10^{-2}
BAC1	240	6.68×10^{-4}	1.09×10^{-3}	6.30×10^{-3}	3.06×10^{-3}	4.51×10^{-3}	9.41×10^{-3}	2.62×10^{-2}
BACH3	240	3.38×10^{-4}	6.53×10^{-4}	3.42×10^{-3}	3.54×10^{-3}	2.91×10^{-4}	9.35×10^{-4}	2.68×10^{-3}
BAOCH3	255	5.46×10^{-4}	1.78×10^{-3}	5.05×10^{-3}	8.70×10^{-3}	1.21×10^{-4}	9.56×10^{-3}	2.49×10^{-3}
BANHCH3	303	1.36×10^{-3}	3.97×10^{-3}	1.37×10^{-2}	1.50×10^{-2}	2.65×10^{-3}	1.37×10^{-2}	2.05×10^{-2}
BANCH3CH3	315	1.48×10^{-3}	3.90×10^{-3}	1.48×10^{-2}	1.56×10^{-2}	3.16×10^{-3}	1.31×10^{-2}	2.00×10^{-2}
BANHCOCH3	275	3.12×10^{-4}	2.21×10^{-3}	2.23×10^{-3}	6.78×10^{-4}	2.78×10^{-3}	1.09×10^{-2}	7.37×10^{-3}
molecule	λ_{max}	C8	C9	C10	H11	H12	H13	H14
BA	230	0	0	0	0	0	0	0
BAOH	255	4.11×10^{-1}	2.98×10^{-3}	1.05×10^{-2}	2.43×10^{-3}	5.84×10^{-3}	2.44×10^{-3}	1.88×10^{-2}
BANH2	288	3.77×10^{-1}	1.16×10^{-2}	1.18×10^{-2}	1.65×10^{-3}	8.14×10^{-3}	1.64×10^{-3}	8.13×10^{-3}
BAC1	240	9.00×10^{-2}	2.58×10^{-2}	9.13×10^{-3}	6.56×10^{-3}	2.33×10^{-2}	6.71×10^{-3}	2.33×10^{-2}
BACH3	240	6.74×10^{-2}	5.07×10^{-3}	1.35×10^{-3}	1.46×10^{-3}	3.24×10^{-3}	1.70×10^{-3}	4.77×10^{-3}
BAOCH3	255	4.17×10^{-1}	1.56×10^{-2}	1.33×10^{-2}	7.38×10^{-4}	1.77×10^{-2}	1.62×10^{-4}	1.15×10^{-2}
BANHCH3	303	3.88×10^{-1}	1.44×10^{-2}	1.20×10^{-2}	3.36×10^{-3}	3.76×10^{-3}	4.71×10^{-2}	3.28×10^{-2}
BANCH3CH3	315	3.81×10^{-1}	1.99×10^{-2}	1.26×10^{-2}	4.74×10^{-3}	5.93×10^{-3}	3×10^{-2}	3.10×10^{-2}
BANHCOCH3	275	3.40×10^{-1}	4.91×10^{-5}	1.04×10^{-2}	3.24×10^{-3}	1.59×10^{-2}	3.22×10^{-2}	4.00×10^{-2}

Table C.4: Frobenius Distance between individual atoms for only the DI to be compared directly with λ_{max} , the lowest λ_{max} molecule's atoms are taken as the reference atoms. (Benzoic Acid Series)

molecule	λ_{max}	H1	O2	C3	O4	C5	C6	C7
BA	230	0	0	0	0	0	0	0
BAOH	255	6.21×10^{-4}	2.52×10^{-3}	8.00×10^{-3}	3.44×10^{-3}	1.42×10^{-2}	6.87×10^{-3}	3.96×10^{-2}
BANH2	288	1.97×10^{-3}	4.55×10^{-3}	1.46×10^{-2}	6.59×10^{-3}	2.20×10^{-2}	2.01×10^{-2}	5.47×10^{-2}
BAC1	240	1.16×10^{-3}	1.53×10^{-3}	1.39×10^{-3}	1.18×10^{-3}	3.58×10^{-3}	3.75×10^{-3}	2.02×10^{-2}
BACH3	240	6.21×10^{-4}	1.03×10^{-3}	3.40×10^{-3}	1.44×10^{-3}	7.07×10^{-3}	2.18×10^{-3}	1.35×10^{-2}
BAOCH3	255	1.04×10^{-3}	2.73×10^{-3}	9.12×10^{-3}	4.27×10^{-3}	1.89×10^{-2}	2.57×10^{-2}	6.02×10^{-2}
BANHCH3	303	2.57×10^{-3}	5.48×10^{-3}	1.77×10^{-2}	2.31×10^{-2}	3.55×10^{-2}	1.47×10^{-2}	5.78×10^{-2}
BANCH3CH3	315	2.81×10^{-3}	5.69×10^{-3}	1.80×10^{-2}	2.30×10^{-2}	3.51×10^{-2}	2.45×10^{-2}	6.84×10^{-2}
BANHCOCH3	275	4.74×10^{-4}	1.69×10^{-3}	5.48×10^{-3}	2.18×10^{-2}	2.50×10^{-2}	6.86×10^{-3}	4.22×10^{-2}
molecule	λ_{max}	C8	C9	C10	H11	H12	H13	H14
BA	230	0	0	0	0	0	0	0
BAOH	255	6.10×10^{-2}	4.90×10^{-2}	1.65×10^{-2}	1.20×10^{-3}	2.23×10^{-3}	1.28×10^{-3}	5.50×10^{-3}
BANH2	288	7.35×10^{-2}	5.47×10^{-2}	2.09×10^{-2}	1.32×10^{-3}	1.55×10^{-3}	1.34×10^{-3}	1.57×10^{-3}
BAC1	240	2.58×10^{-2}	2.01×10^{-2}	3.88×10^{-3}	1.86×10^{-3}	7.50×10^{-3}	1.94×10^{-3}	7.52×10^{-3}
BACH3	240	2.45×10^{-2}	2.21×10^{-2}	9.98×10^{-3}	3.71×10^{-4}	3.12×10^{-3}	6.08×10^{-4}	2.46×10^{-3}
BAOCH3	255	6.86×10^{-2}	3.99×10^{-2}	4.32×10^{-3}	1.41×10^{-3}	5.91×10^{-3}	8.17×10^{-4}	7.68×10^{-3}
BANHCH3	303	8.97×10^{-2}	6.50×10^{-1}	6.29×10^{-1}	1.65×10^{-3}	5.32×10^{-3}	6.38×10^{-1}	6.38×10^{-1}
BANCH3CH3	315	9.55×10^{-2}	6.46×10^{-1}	6.29×10^{-1}	2.06×10^{-3}	7.44×10^{-3}	6.34×10^{-1}	6.38×10^{-1}
BANHCOCH3	275	6.55×10^{-2}	6.47×10^{-1}	6.28×10^{-1}	1.72×10^{-3}	1.06×10^{-2}	6.37×10^{-1}	6.38×10^{-1}

Table C.5: Frobenius Distance between individual atoms for only the LI to be compared directly with the aromaticity measure HOMA, the highest HOMA value molecule's atoms are taken as the reference atoms. (Aromatic Series)

molecule	HOMA	C1	C2	C3	C4	C5	C6
Benzene	1.001	0	0	0	0	0	0
Anthracene(I)	0.884	6.56×10^{-2}	6.56×10^{-2}	3.06×10^{-3}	6.56×10^{-2}	6.56×10^{-2}	3.06×10^{-3}
Anthracene(O)	0.517	6.56×10^{-2}	6.56×10^{-2}	2.47×10^{-3}	9.01×10^{-4}	9.01×10^{-4}	2.47×10^{-3}
Phenanthrene(I)	0.402	6.49×10^{-2}	5.58×10^{-2}	5.58×10^{-2}	6.49×10^{-2}	2.08×10^{-3}	2.07×10^{-3}
Phenanthrene(O)	0.902	5.58×10^{-2}	6.49×10^{-2}	3.88×10^{-4}	3.67×10^{-3}	5.78×10^{-3}	5.28×10^{-3}
Naphthalene	0.779	6.37×10^{-2}	6.37×10^{-2}	3.07×10^{-3}	1.44×10^{-3}	1.44×10^{-3}	3.07×10^{-3}
Naphthacene(I)	0.774	7.06×10^{-2}	7.06×10^{-2}	2.06×10^{-3}	6.31×10^{-2}	6.31×10^{-2}	2.06×10^{-3}
Naphthacene(O)	0.325	2.67×10^{-3}	2.04×10^{-3}	6.31×10^{-2}	6.31×10^{-2}	2.04×10^{-3}	2.67×10^{-3}
Chrysene(I)	0.553	5.84×10^{-2}	5.84×10^{-2}	5.64×10^{-2}	6.77×10^{-2}	2.56×10^{-3}	5.42×10^{-3}
Chrysene(O)	0.859	6.77×10^{-2}	5.64×10^{-2}	4.00×10^{-3}	5.59×10^{-3}	3.28×10^{-3}	8.76×10^{-4}
Triphenylene(I)	0.067	5.64×10^{-2}					
Triphenylene(O)	0.93	5.64×10^{-2}	5.64×10^{-2}	8.27×10^{-3}	7.07×10^{-3}	7.06×10^{-3}	8.25×10^{-3}
Cyclohexane	-4.34	1.03×10^{-1}					

Table C.6: Frobenius Distance between individual atoms for only the DI to be compared directly with the aromaticity measure HOMA, the highest HOMA value molecule's atoms are taken as the reference atoms. (Aromatic Series)

molecule	HOMA	C1	C2	C3	C4	C5	C6
Benzene	1.001	0	0	0	0	0	0
Anthracene(I)	0.884	9.87×10^{-2}	9.87×10^{-2}	3.43×10^{-2}	9.87×10^{-2}	9.87×10^{-2}	3.43×10^{-2}
Anthracene(O)	0.517	1.51×10^{-1}	1.51×10^{-1}	1.57×10^{-1}	1.51×10^{-1}	1.51×10^{-1}	1.57×10^{-1}
Phenanthrene(I)	0.402	1.39×10^{-1}	1.51×10^{-1}	1.51×10^{-1}	1.39×10^{-1}	1.78×10^{-1}	1.78×10^{-1}
Phenanthrene(O)	0.902	7.59×10^{-2}	7.84×10^{-2}	7.77×10^{-2}	6.65×10^{-2}	6.57×10^{-2}	7.44×10^{-2}
Naphthalene	0.779	1.08×10^{-1}	1.08×10^{-1}	1.14×10^{-1}	1.05×10^{-1}	1.05×10^{-1}	1.14×10^{-1}
Naphthacene(I)	0.774	1.23×10^{-1}	1.23×10^{-1}	6.85×10^{-2}	1.21×10^{-1}	1.21×10^{-1}	6.85×10^{-2}
Naphthacene(O)	0.325	1.74×10^{-1}	1.79×10^{-1}	1.77×10^{-1}	1.77×10^{-1}	1.79×10^{-1}	1.74×10^{-1}
Chrysene(I)	0.553	1.15×10^{-1}	1.30×10^{-1}	1.39×10^{-1}	1.27×10^{-1}	1.55×10^{-1}	1.54×10^{-1}
Chrysene(O)	0.859	8.89×10^{-2}	8.74×10^{-2}	8.97×10^{-2}	8.06×10^{-2}	8.13×10^{-2}	9.19×10^{-2}
Triphenylene(I)	0.067	1.67×10^{-1}					
Triphenylene(O)	0.93	6.20×10^{-2}	6.20×10^{-2}	6.08×10^{-2}	5.01×10^{-2}	5.01×10^{-2}	6.08×10^{-2}
Cyclohexane	-4.34	3.02×10^{-1}					