Molecular Fragment and Substituent Effect Studies of Styrene Derivatives by Electron Density Shape Analysis

by © Zoltan Antal

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Faculty of Science

Memorial University of Newfoundland

May 2014

St. John's Newfoundland and Labrador

Abstract

The research work leading to the content of this thesis and to the related publications represents a novel way to analyze and characterize some aspects of the most fundamental interactions within molecules, namely through-space and through-bond interactions, which are also relevant to substituent effects. The approach and the results give quantum chemical justification for some components of the associated fundamental phenomena which are often difficult or even impossible to examine separately by experimental means. The results are obtained using well-established, thoroughly tested and verified methods of electron density shape analysis approaches, based on input electron densities obtained by various established quantum chemistry computational methods, such as Hartree-Fock, Density Functional, and Møller-Plesset methodologies. The shape analysis methods, applicable to both complete molecules and locally to molecular fragments such as functional groups, are based on two fundamental theorems of quantum chemistry: the Hohenberg-Kohn theorem and the Holographic Electron Density theorem. The first of these theorems establishes that the molecular electron density must contain all molecular properties, whereas the second of these theorems gives justification that even small molecular fragments, such as a vinyl group in our studies, can be used as a "fingerprint", fully representing all the unique intramolecular interactions within each molecule. As a consequence of these theorems, the electron density shape analysis methods, when applied to the complete electron density cloud of a target molecule, or locally to that of a molecular fragment, yield shape information which can, in principle, give highly detailed theoretical information about all molecular properties, hence, providing a suitable basis

for establishing correlations with experimentally available data about various measured chemical properties. In earlier studies, such shape analysis approaches have been used in various applied fields of chemistry, such as pharmacology and toxicology.

The present study describes the use of a series of small molecules, specifically, substituted styrenes for the analysis and shape correlations involving through-bond and through-space correlations. As a consequence of their specific structural features, the chosen molecules appear as ideal candidates to study both effects separately. In spite of their specificity, the results provide some general conclusions which are likely to contribute to a better understanding of far more complex phenomena, such as some of the components contributing to the complex folding pattern of globular proteins and more general interaction problems in the emerging field of molecular design. Even in these latter fields, these two types of interactions, through-space and through-bond, play significant roles as well. The presence of aromatic rings in these styrene derivatives is very useful as through-bond transmitting media whilst the distance between the studied molecular fragments obtained after removing the ring is also nearly optimal to be used in through-space interaction studies. The structural specificities of the model molecules have played an important role in all three publications constituting the main chapters of this thesis. Therefore, there is a strong emphasis on the effect transmitting capabilities of the aromatic ring in all 3 of these papers.

Acknowledgements

I'd like to thank my supervisor, Dr. Paul G. Mezey for his time and energy invested in my program. With a healthy blend of science and jokes he made the route leading to my degree feel like a walk in a park. Also, special thanks goes out to Peter L. Warburton, who, although officially never assumed this role, got as close as anyone could get to the position of an essential co-supervisor.

I'd like to thank my supervisory committee, to Dr. Christopher Flinn and to Dr. Paris E. Georghiou for their crucial role played in achieving my goals.

Naming everyone who helped in some way in the wonderfully organized team called the Department of Chemistry would be difficult, therefore allow me to highlight the absolutely critical assistance provided by the department's capable leadership, the beautiful ladies in the General Office, all Senior Lab Demonstrators, and the staff at the Physical Science Stores.

AceNet has a special role in providing computational resources and technical support, and I'd like to thank its representatives for their prompt and capable assistance.

Without financial support, no PhD program would have happened for me, therefore the taxpayers of Canada have my utmost gratitude. Through them, the Government of Canada can set up and finance programs such as NSERC, CFI, and CRC and even unknown international students, such as myself, can benefit from them.

The people of the Province of Newfoundland and Labrador have a special place in my heart who welcome everyone with honest warmness and give a helping hand in time of need. Through their government, they provide free healthcare for practically everyone who gets off the boat (mostly planes nowadays) and through their healthy drinking habits they left me with the memories of many-many happy days (and nights).

Two special ladies need to be mentioned, Eva Sz. Hawkins and MaryAnn Moores, without whom my stay in Newfoundland would have been empty. Their care and attention is legendary and the high levels of love and empathy in their hearts is highly addictive.

Least but not last a very special person needs to be mentioned, beautiful Zsofia. Welcoming, tolerant, patient, empathetic.

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations and Symbols

A

a	Electron Density Threshold Value
ACEnet	Atlantic Computational Excellence Network
ADMA	Adjustable Density Matrix Assembler
AFDF	Additive Fuzzy Density Fragmentation
au	Atomic Units (electron density)

B

b	Reference curvature value
B3LYP	Becke three parameter hybrid functional of Lee-Yang-Parr

C

C6	A six-membered aromatic carbon ring
cc-pVTZ	correlation consistent-polarized Valence Triple Zeta (basis set)
cc-pVDZ	correlation consistent-polarized Valence Double Zeta (basis set)

CF_3	Tri-fluoro Methyl group
CH_3	Methyl group
CHO	Aldehyde group
Cl	Chlorine
CN	Cyano group
$COCH_3$	Acetyl group
CRC	Canada Research Chair Program

D

DFT	Density Functional Theory
$\rho(r)$	Electron density at point r

E

$\pm$ E	Resonance effect

F

F	Fluorine

FSVI	Fragment Shape Variation Index

G

G(a)	Electron Density Isosurface for threshold *a*

H

H	Hydrogen
HEDT	Holographic Electron Density Theorem
HF	Hatree-Fock

I

±I	Inductive effect
IM	Isolated Molecules

L

LCAO	Linear Combination of Atomic Orbitals

LDG	Low Density Glue

M

MIDCO	Molecular Isodensity Contour
MP2	Moller-Plesset perturbation theory (second order)

N

$N(CH_3)_2$	N,N -dimethylamino group
NH_2	Amine group
NO_2	Nitro group
NSERC	Natural Sciences and Engineering Research Council of Canada

O

OCH_3	Methoxy group
OH	Hydroxyl group

P

PAH	Polycyclic Aromatic Hydrocarbon

Q

QSAR	Quantitative structure–activity relationship
QS-SM	Quantum Self-Similarity Measures

R

r	Position vector
R	CH_3 or methyl group

S

SDM	Shape Deviation Measures
SDM(A)	total shape deviation measure from the references for a given para-substituent A
SDM_{tb}	shape deviation component corresponding to the through-bond effect
SDM_{ts}	shape deviation component corresponding to the through-space effect
SG	Shape Group

SGM	Shape Group Method
SM	Similarity Matrix
S(STY,A)	shape similarity measure of vinyl groups in styrene when compared to para-A-substituted styrene
S(sty,a)	shape similarity measure of vinyl groups in styrene when compared to para-A-substituted styrene (for the corresponding two pairs of isolated molecules)
STY	Styrene

T

TB	Through-Bond
TS	Through-Space

V

VSEPR	Valence Shell Electron Pair Repulsion

List of Appendices

Chapter 2

CHAPTER 1

1.1. Introduction

Experimental research concerning intramolecular interactions often provides numerical results which collectively reflect all the complex contributions of several components, and it is often difficult to analyze the relative importance of these components separately. On the other hand, molecular modeling, using computational chemistry, can be formulated in terms of separately modeled components of complex interactions that allow the assessment of their relative importance. Computational chemistry, as a sub-branch of chemistry, uses computers to analyze chemical data obtained from experiments or to make interpretations or chemical predictions based on theory (e.g. predictions for new molecules or properties of new molecules). Specifically, choosing computational quantum chemistry as a tool for the study of details of molecular interactions provides important advantages, such as the possibility of identifying and analyzing components of interactions, which are difficult to identify and study by experiment, or even by alternative computational chemistry means. Based on quantum theory[1,2] and combined with the advancements in computational resources and computer power, a dramatic increase has been observed in the usefulness and importance of computational quantum chemistry in the past two decades, that has also motivated various new methodological advances. In fact, detailed electron density shape analysis methods owe their introduction mostly to the greatly increased possibilities and potential of computational quantum chemistry. Geometrical structures of known and unknown molecules can also be modeled in this way and a variety of computational methods exist that calculate and predict the molecular structure, properties, chemical reactions and

chemical kinetics, to name a few.[3-5] Computational chemistry has experienced an especially important and still growing role in pharmaceutical chemistry. Its influence began in the 1960s, got a stable foothold in the 1970s, experienced a large growth in the 1980s, and, by the 1990s, it became an invaluable tool, among other roles, in new drug discovery.[6]

One of the branches of computational chemistry, based on quantum mechanics (QM) methods, can be classified as either *ab-initio* (from the beginning) or *semi-empirical*. Ab-initio methods do not contain empirical parameters, whereas such parameters form the basis of semi-empirical calculations. The main branches of ab initio methods are based on the wavefunction theory at the Hartree-Fock (HF) and the post HF theory levels, and among other methods, the many-body perturbation theory (MBPT). This latter is related to the MP2 (second order Møller-Plesset perturbation theory) method.[7] Both of these methods are used in this thesis project, and an additional theory, the Density Functional Theory[8] (DFT) is also employed. This latter is represented by the "B3LYP hybrid functional" which uses the "*Becke Three Parameter Hybrid Functionals*"[9] (an exchange functional) and the correlation functional of Lee, Yang, and Parr.[10]

Chemical bonding and interactions are fundamentally dependent on the behavior of electrons. The electron density cloud around the nuclei of atoms and molecules can be determined experimentally, or it can be modeled by computational chemistry techniques. For example, the arrangement and density of electrons can be expressed in terms of the

quantum chemical wave functions describing the distribution of electrons.[7] According to the fundamental theorem of Hohenberg – Kohn[11] (often referred to as the basis of Density Functional Theory, DFT, an alternative to wave function methodologies), the non-degenerate ground state electron density contains the complete information about the molecule.

The question of substituent effects within molecules is a long debated topic of physical organic chemistry where there is a trend of attempting to improve the established models used to describe interactions between various molecular parts, such as substituents.[12-22] Substituent effects have major impact on the chemical and physical properties of molecules, including chemical reactivity, so understanding them is of utmost importance. Most earlier attempts have concentrated on experimental methods and the results included tables on "universal descriptors", such as the Hammett constants.[12] These single-, dual-, or even triple-parameter numerical scales have provided a relatively simple, quantitative, but approximate description of intramolecular interactions. However, for a more detailed analysis and interpretation, theoretical, computational chemistry methods provide alternatives, since electron clouds, in principle, contain all molecular information. Similarity indices for electronic densities, based on the works of Carbó at al., were providing safeguards in designing such tools (quantum similarity approach) based on quantum chemistry.[23-32] Electron density shape analysis is another such tool that can also be used to describe various components of molecular interactions.

Our primary concern, as well as the main topic of the thesis, is the exploration and analysis of some of the fundamental components of substituent effects in a selected set of model molecules. For this goal, finding an appropriate computer modeling tool for the analysis of molecular electron densities, as well as local, molecular fragment electron densities was essential. Local molecular fragments are sufficient: small substituents or even single atoms influence global molecular properties in a unique way as was proven by the Holographic Electron Density Theorem[33] (HEDT). The theorem states that *any small positive volume part* of the molecular electron density cloud contains the complete information about the entire molecule. Such a small part can be chosen as the vinyl group present in the model molecules used in the present study. For a numerical representation of the shape properties and shape variations of these groups, the Shape Group Method, developed earlier to find a non-ambiguous, non-visual, and non-biased way to evaluate the shape properties of molecular fragments, has been used.

The effectiveness of the shape group method to provide correlations between molecular shape features and experimentally measured properties has been proven in numerous cases.[34-60]

The main body of the thesis consists of three published scientific articles, constituting the three main chapters, namely Chapter 2, Chapter 3, and Chapter 4. These are presented in a way for the reader to discover:

a) a method which is meant to analyze intramolecular interactions by scrutinizing substituent effects within molecules;

b) a practical application of the method presented above to find predictive correlations;

c) and the extension of the method in a more general way to also include the shape modifying characteristics of pure aromatic rings as well.

The next part of the introduction focuses on a concise description of the Shape Group Method as part of the molecular shape analysis methodology. Please note that other detailed descriptions of the Shape Group methods can be found throughout the thesis, in the main research articles and also at the end of Chapter 5 (subchapter 6, Appendices). The former is an organic part of the research articles and the latter was a requirement of the thesis format. All descriptions related to the Shape Groups contain essentially the same information; therefore those readers wishing to dispense with consequent descriptions can skip the respective subsections without missing too much. Also, interested readers who wish to gain more insight into the Shape Group method should consult references [47-51, 53, 54, 57, 58], including a book written on the subject.[59]

1.2. Molecular shape analysis methodology using geometrical and topological tools

1.2.1. A brief outline of the Shape Group Methods

For a specified nuclear configuration K and electron density threshold a, the *molecular isodensity contour surface,* MIDCO, G(K,a) is defined as

$$G(K, a) = \{\mathbf{r}: \rho(K, \mathbf{r}) = a\}, \qquad (1)$$

as the set that is the collection of all points $\mathbf{r}$ of the 3D space where the molecular electronic density $\rho(K,\mathbf{r})$ is equal to the above threshold value a.

The main mathematical tools for molecular shape analysis are *shape groups.* These are algebraic groups, not directly related to point symmetry groups, nevertheless, the molecular symmetry may influence the shape groups. The shape groups are the homology groups of the collection of all MIDCO's, for the complete set of a values, obtained after truncating the MIDCO's according to some local curvature criteria. These homology groups, indeed, convey all the essential shape information about the entire fuzzy electronic cloud of the molecule.

In technical terms, when applying the shape groups, the local shape properties of the MIDCO's are specified in terms of local shape domains: for example, in terms of the locally convex, concave, or saddle-type regions of MIDCO's. In this special case, the local surface shape is characterized relative to a tangent plane.

However, simple convexity is insufficient for detailed shape characterization. A far more detailed shape description is obtained if the tangent plane is replaced by some curved reference objects T, for example, if the MIDCO is compared to a series of tangent spheres of various radii *r*. Alternatively, in some applications it is advantageous to use a series of oriented tangent ellipsoids T, especially, if shape characterization involving some reference directions is needed. These oriented tangent ellipsoids can be translated but not rotated as they are brought into tangential contact with the MIDCO surface G(K,a), in order to evaluate the relative shape of the MIDCO at this location. In general, the tangent object T (plane, sphere, or ellipsoid) may fall locally on the outside (type 0), or on the inside of the MIDCO (type 2), or it may cut into

the given MIDCO within any small neighborhood of the tangential contact point **r** of the MIDCO (type 1); that is, three types of contacts may occur. If all points **r** of the MIDCO are tested and characterized by such tangential contact with T, these differences in contact types lead to a family of local shape domains D_2, D_0, and D_1, and to a subdivision of the molecular contour surface into locally convex, locally concave, and locally saddle-type shape domains (relative to the tangent object T, that may be rather different from ordinary convexity characterization, relative to a tangent plane), denoted by the symbols D_2, D_0, and D_1, respectively. In most applications, tangent spheres are used where orientation cannot play any role, therefore one may use the curvature of the sphere, $b=1/r$, for "relative" curvature characterization. In such a case, the local "relative" shape domains D_2, D_0, and D_1, represent the *local relative convexity domains* of the MIDCO, relative to the reference curvature b. Evidently, the b=0 choice corresponds to the case of the tangent plane.

If one considers as reference objects T only tangent spheres (including the infinite radius, hence b=0 curvature tangent plane) then for each specific reference curvature b, the local shape domains D_2, D_0, and D_1 provide a complete partitioning of the MIDCO surface G(K,a). Taking all D_μ domains of a specified type $\mu = 0,1,2$, for example, taking all the locally convex domains D_2 relative to b, and excising them from the MIDCO surface G(K,a), a new, topologically more interesting object is obtained, a truncated contour surface $G(K,a,\mu)$. This truncated contour surface inherits some essential shape information from the original MIDCO surface G(K,a), in such a way that can be analyzed topologically. In fact, a topological analysis of the truncated surface $G(K,a,\mu)$ provides a shape analysis of the original MIDCO surface G(K,a). This procedure can be repeated for a whole range of reference curvature values b, leading to a detailed shape analysis of the original MIDCO surface. An important

simplification is achieved: for the whole range of possible reference curvature values b there are only a finite number of topologically different truncated MIDCO surfaces $G(K,a,\mu)$, and, in fact, their topological equivalence classes can be analyzed in any further step. The equivalence classes of these truncated surfaces can be characterized by their topological invariants, and these invariants provide a numerical shape characterization.

The main topological tools used for such shape characterization are the homology groups of equivalence classes of truncated surfaces. The homology groups of algebraic topology are topological invariants, expressing some of the most important features of the topological structure of bodies and surfaces such as MIDCO's. The ranks of these groups are the Betti numbers, themselves important topological invariants, the very numbers used in the Shape Group method for numerical shape characterization.

The Shape Groups of the original MIDCO G(K,a) are defined as the homology groups $H^p{}_\mu(a,b)$ of the truncated MIDCO surfaces $G(K,a,\mu)$. The $b^p{}_\mu(a,b)$ Betti numbers of the $H^p{}_\mu(a,b)$ shape groups form a set of numerical shape codes for the complete molecular electron density distributions, whereas the fuzzy nature of these electron densities renders more conventional shape analysis methods less than adequate. For a given MIDCO, G(K,a) of density threshold a, and for each reference curvature b, shape domain and truncation pattern μ, there are three shape groups, denoted by $H^0{}_\mu(a,b)$, $H^1{}_\mu(a,b)$, and $H^2{}_\mu(a,b)$. The upper indices are formal dimensions p of these three shape groups, (these are zero, one, and two), which collectively express the essential shape information of the MIDCO G(K,a). For these three dimensions, for each (a,b) pair of parameters, and for each shape domain truncation type μ, there are three Betti numbers, $b^0{}_\mu(a,b)$, $b^1{}_\mu(a,b)$, and $b^2{}_\mu(a,b)$.

The shape group method of topological shape description, following the spirit of the GSTE principle (Geometrical Similarity treated as Topological Equivalence), combines the advantages of geometry and topology. Whereas the local shape domains and the truncated MIDCO's $G(K,a,\mu)$ are defined in terms of geometrical classification of points of the surfaces using local curvature properties, on the other hand, the truncated surfaces $G(K,a,\mu)$ are characterized topologically by the shape groups and their Betti numbers, collected into sets of numerical shape codes.

1.2.2. Shape Group-Based Numerical Shape Codes and Measures of Molecular Similarity and Complementarity

The topological shape analysis techniques of the Shape Group method have a significant advantage: the numerical representation of shape information. Accordingly, the results of a shape group analysis can be represented by a finite set of numbers, the family of Betti numbers $b^p{}_\mu(a,b)$ for all the shape groups which occur for a given molecule. These very numbers form a numerical shape code. The shape codes themselves can be compared easily by a computer, providing a well-defined numerical measure of molecular shape similarity. Furthermore, by applying a suitable transformation, the shape codes can generate a numerical measure of shape complementarity which is important in molecular "fitting" problems. The shape codes, hence the similarity and complementarity measures, can be generated by algorithmic methods using a computer without any need for any additional information beyond the molecular electronic density. This algorithmic approach also eliminates all subjective elements of conventional, visual shape comparisons. This possibility is particularly advantageous if large sequences of molecules are to be compared, and when experimental properties are correlated with

inherent electron density features. The numerical shape codes and the derived similarity, complementarity, or other numerical shape information can be generated for practically all molecules in data banks, for example, those of drug companies. Earlier results have provided convincing evidence that in drug design and toxicology studies the local and global similarity analysis based on the shape codes are practical tools.[41,52,55,56,60]

1.2.3. Some technical details of the (a,b)-Parameter Maps as Shape Codes

The following sign convention is used: a positive b value indicates that a tangent sphere, as reference object, is placed on the exterior side of the molecular surface, whereas a negative b value indicates that the sphere is placed on the interior side of the MIDCO. The shape groups of a molecule of a fixed nuclear arrangement K, depend on two continuous parameters: on the electronic density threshold a and on the reference curvature b. A formal, two-dimensional map, called the (a,b)-map, is defined by the ranges of these two parameters, a and b, and a detailed shape characterization of the electronic density of the molecule is obtained by the shape group distribution along this map.

In most applications, the important Betti numbers conveying the chemically most relevant shape information are those of type $b^1_\mu(a,b)$ for shape groups of dimension 1, hence, the emphasis is placed on these. However, separate (a,b)-maps can be generated for each of the three types of Betti numbers, $b^0_\mu(a,b)$, $b^1_\mu(a,b)$, and $b^2_\mu(a,b)$ according to the dimension of the shape groups. In general, the Betti numbers obtained for a given pair of values of parameters a and b are assigned to the given location of the (a,b) parameter map. If a truncated contour falls apart into several pieces,

than the Betti number for each piece is considered, and these numbers are encoded into a single numerical value.

By considering a grid of a and b values within a suitable interval $[a_{min},a_{max}]$ of density thresholds and an interval $[b_{min}, b_{max}]$ of reference curvature values, the (a,b)-map of Betti numbers is usually generated in a discretized form. Since the range of these parameters often covers several orders of magnitude, it is usually advantageous to use logarithmic scales. Note that, for negative curvature parameters b the $\log|b|$ values are taken. In many recent applications of this methodology, the density threshold values a (between [0.001, 0.1 a.u.]), and [-1.0, 1.0] for the curvature b of the test spheres against which the local curvatures of the MIDCO's are compared, and a 41×21 grid is used. The logarithmic grid uses a uniform spacing between the given minima and maxima.

The values of the Betti numbers at the grid points (a,b), generate a matrix, $\mathbb{M}^{(a,b)}$ and the fuzzy electron density of the molecule is characterized by this matrix, which is used as a *numerical shape code*.

1.2.4. Generation of Shape Similarity Measures from Shape Codes

Using the matrices $\mathbb{M}^{(a,b)}$ of the (a,b)-map of Betti numbers $b^p{}_\mu(a,b)$ for a family of molecules as shape codes, numerical similarity measures can be calculated to compare these molecules. In most of the recent applications, a direct numerical comparison has been used. For two molecules, A and B, each in some fixed nuclear configuration, their shape codes $\mathbb{M}^{(a,b),A}$ and $\mathbb{M}^{(a,b),B}$, generate a numerical shape similarity measure, defined as

$$s(A,B) = m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}]/t \quad (2)$$

On the right hand side, $m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}]$ is simply the number of matches between corresponding elements of the two matrices, whereas t is the total number of elements in either matrix:

$$t = n_a n_b. \quad (3)$$

Here n_a and n_b are the number of grid divisions for parameters a and b; if the 41×21 grid is used for the ranges [0.001, 0.1 a.u.] for the density threshold values a, and [-1.0, 1.0] for the reference curvature b, then t = 861.

1.3. A brief presentation of the main Chapters

1.3.1. Chapter 2 presents a method of electron density shape analysis which focuses on separating through-bond (TB) and through-space (TS) effects clearly present in a series of target molecules and numerically tabulating them. The dominance of one of them is clearly established and specific effects such as resonance and inductive effects are identified. By identifying such relatively simple factors that influence molecular behavior, the principles for designing tools which are capable to describe more complex phenomena are available. A somewhat more distant, but still realistic application is using such tools to study the folding of globular proteins where the extensive distribution of through-space effects contributes significantly to the forces holding the whole structure together.[61]

Accordingly, Chapter 2 presents the studies of a series of *para*-substituted styrenes. Molecular fragments derived from them have the potential to provide important answers concerning TB and TS effects. Naturally, the aromatic ring can act as a through-bond transmitting media (primarily, via π-electron system) between the functional group and the target of the shape analysis, the vinyl group, situated in the *para*-position, relative to each other. The shape of the electron cloud around the vinyl group is analyzed and expressed as a numerical shape code. The shape of the vinyl group is affected by the *para*-positioned functional group via the aromatic ring. "Cutting" (computationally) out the aromatic ring and leaving the functional and vinyl groups in place, gives a tool to measure the influence of the functional group via space. In this way, numerical values can be attributed to both effects.

1.3.2. Chapter 3 is more concerned with treating the question of shape analysis results from another point of view, as well as trying to use the result for practical applications. Substituent effects are scrutinized from an aromatic electrophilic substitution (AES) point of view, as a series of di-substituted benzenes are shape-analyzed in a target area. The resulting tables can be used to predict the potential chemical behavior in an AES of known and unknown molecules having the same structure. Excellent correlations have been discovered between *ortho*-, *meta*-, or *para*-directing abilities of activating/deactivating groups and the target of the shape analysis, namely, the vinyl group. Naturally, the substituent effects in AES type reaction of these activating/deactivating groups have been thoroughly exploited experimentally, but this

theoretical method offers an electron density based, quantum chemical justification as well as a means for predictive correlations for known and new molecules.

1.3.3. Chapter 4 is mainly concerned with the shape-modifying effects of pure aromatic rings in polycyclic aromatic hydrocarbons (PAH). These molecules are great candidates because of the extensive experimental information available about them. Structure-wise, they are mostly planar molecules, with little room for conformational changes. In these PAHs, the target molecular part (for shape analysis) and the main shape modifiers are both the aromatic rings themselves. A large number of molecules contain these easily deformable (electron cloud) structures and which also act as substituent carriers, providing additional information about the total shape-modifying effects in molecules, in addition to the shape-modifying effects of functional groups. Also, in this chapter, the Shape Group Method is applied to provide information about the determination of the *Fragment Shape Variation Index* of these polycyclic aromatic hydrocarbons, a method of classification of different shape modifying molecular parts introduced in previous reportings.[62,63]

Molecular diversity is a concept often invoked in pharmaceutical and other biochemistry-related research. Whereas in the actual applications of the Shape Group method in this thesis, the concern was shape correlations with substituent effects and the through-bond and through-space components of the intramolecular interactions. Nevertheless, refocussing the shape similarity evaluation methodology from similarity to diversity is a topic that follows naturally from the approach taken.

Very briefly, the local or *fragment shape variation index (fsvi)* is defined as

$$\frac{\mathrm{fsvi(F1,F2,A1,A2)} = \mathrm{gshs(A1,A2)}}{[\mathrm{lshs(F1,F2)} + \mathrm{gshs(A1,A2)}]} \qquad (4)$$

where F1 and F2 are fragments (typically functional groups of the same stoichiometry and bond connectivity) of molecules A1 and A2, whilst *lshs* and *gshs* represent local and global shape-similarity measures involving the beforementioned fragments and molecules.[62]

1.4. Bibliography

1. P. W. Atkins and J. de Paula, *Atkins' Physical Chemistry - 8th Ed*., Oxford University Press, Oxford 2006.
2. D. A. McQuarrie, *Quantum chemistry - 2nd ed*. University Science Books, Sausalito, Ca. 2008.
3. W. J. Hehre, *Practical Strategies for Electronic Structure Calculation. Wavefunction*, Irvine, CA., 1995.
4. C. J. Cramer, *Computational Chemistry: Theories and Models*. Wiley, 2002.
5. D. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. John Wiley and Sons Ltd., 2001.
6. D. B. Boyd, *How Computational Chemistry Became Important in the Pharmaceutical Industry, in Reviews in Computational Chemistry*, Volume **23**

(eds K. B. Lipkowitz and T. R. Cundari), John Wiley & Sons, Inc., Hoboken, NJ, USA. 2007.

7. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York, 1989.

8. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules, volume 16 of International Series of Monographs on Chemistry*. Oxford, New York, 1989.

9. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, **98,** 5648-52, (1993).

10. C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, **37,** 785-89, (1988).

11. P. Hohenberg and W. Kohn, Inhomogeneous Electron Gas, *Phys. Rev.*,**136**, B864, (1964).

12. E. Anslyn, D. A. Dougherty, Modern Physical Organic Chemistry; University Science Books, 2006.

13. L. P. Hammett, The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives J. Am. Chem. Soc., 59, 96, (1937).

14. R. D. Topsom, Electronic Substituent Effects in Molecular Spectroscopy, in Progress in Physical Organic Chemistry, Volume 16, ed. R. W. Taft, Hoboken, NJ, USA, 2007.

15. R. W. Taft, Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters J. Am. Chem. Soc., 74, 2729, (1952).

16. R. W. Taft, Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters J. Am. Chem. Soc., 74, 3120, (1952).

17. R. W. Taft, Linear Steric Energy Relationships J. Am. Chem. Soc., 75, 4538, (1953).

18. W.F. Reynolds, P.G. Mezey, and G.K. Hamer, Ab initio calculations on 4-substituted styrenes: a theoretical model for the separation and evaluation of field and resonance substituent parameters Can. J. Chem., 55, 522, (1977).

19. P.G. Mezey and W.F. Reynolds, Ab initio calculations on 4-substituted benzoic acids; a further theoretical investigation into the nature of substituent effects in aromatic derivatives Can. J. Chem., 55, 1567, (1977).

20. W.F. Reynolds, T.A. Modro, P.G. Mezey, E. Skorupowa, and A. Maron, Experimental and theoretical investigation of the unusual substituent effect of the vinyl group Can. J. Chem., 58, 412, (1980).

21. W.F. Reynolds, T.A. Modro, and P.G. Mezey, A theoretical investigation of the effect of positively charged substituents on product distribution in electrophilic aromatic substitution; evidence for a dominant field effect of the positive poles, J. Chem. Soc. Perkin II, 1066, (1977).

22. L. Amat, R. Carbó-Dorca, R. Ponec, Simple Linear QSAR Models Based on Quantum Similarity Measures, J. Med. Chem., 42, 5169, (1999).

23. R.Carbó, L.Leyda, M. Arnau, How similar is a molecule to another? An electron density measure of similarity between two molecular structures Int. J. Quantum Chem., 17, 1185, (1980).

24. R. Carbó and M. Arnau, Molecular Engineering: A General Approach to QSAR, in Medicinal Chemistry Advances, ed. F.G. de las Heras and S. Vega, Pergamon Press, Oxford, 1981.

25. R. Carbó and B. Calabuig, Quantum similarity measures, molecular cloud description, and structure-properties relationships, J. Chem. Inf. Comput .Sci., 32, 600, (1992).

26. R. Carbó and B. Calabuig, Molecular quantum similarity measures and n-dimensional representation of quantum objects. I. Theoretical foundations. Int. J. Quantum Chem., 42, 1681, (1992).

27. R. Carbó and B. Calabuig. Molecular quantum similarity measures and n-dimensional representation of quantum objects. II. Practical application, Int. J. Quantum Chem., 42, 1695, (1992).

28. R. Carbó, B. Calabuig, L. Vera, and E. Besalú, Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques, in Advances in Quantum Chemistry, Vol. 25, ed. P.-O. Löwdin, J.R. Sabin, M.C. Zerner, Academic Press, New York, 1994.

29. R. Carbó, Ed., Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches, Kluwer Academic Publ., Dordrecht, 1995.

30. E. Besalú, R. Carbó, J. Mestres, and M. Solà, Foundations and Recent Developments on Molecular Quantum Similarity, in Topics in Current Chemistry, Vol. 173, Molecular Similarity, ed. K. Sen, Springer-Verlag, Heidelberg, l995.

31. R. Carbó-Dorca and E. Besalú, A general survey of molecular quantum similarity, J. Mol. Struct. (THEOCHEM), 451, 11, (1998).

32. P. Constans, L. Amat, and R. Carbó-Dorca, Toward a global maximization of the molecular similarity function: Superposition of two molecules, J. Comput. Chem., 18, 826, (1997).

33. P.G. Mezey, The Holographic Electron Density Theorem and Quantum Similarity Measures, *Mol. Phys*., **96**, 169, (1999).

34. G.A. Arteca, G.A. Heal, and P.G. Mezey, Comparison of Potential Energy Maps and Molecular Shape Invariance Maps for Two-Dimensional Conformational Problems, *Theor. Chim. Acta*, **76**, 377-390 (1990).

35. G.A. Arteca and P.G. Mezey, Analysis of Molecular Shape Changes Along Reaction Paths, *Int. J. Quantum Chem*., **38**, 713-726 (1990).

36. X. Luo, G.A. Arteca, and P.G. Mezey, Shape Analysis Along Reaction Paths of Ring Opening Reactions, *Int. J. Quantum Chem. Symp*., **25**, 335-345 (1991).

37. X. Luo, G.A. Arteca, and P.G. Mezey, Shape Similarity and Shape Stability Along Reaction Paths. The Case of the PPO & OPP Isomerization, *Int. J. Quantum Chem.*, **42**, 459-474 (1992).

38. G.A. Arteca and P.G. Mezey, Deformation of Electron Densities in Static External Fields: Shape Group Analysis for Small Molecules, *Chem. Phys.*, **161**, 1-9 (1992).

39. P.D. Walker, G.A. Arteca, and P.G. Mezey, Shape Groups of the Electronic Isodensity Surfaces of Small Molecules: the Shapes of 10-Electron Hydrides, *J. Comput. Chem.*, **14**, 1172-1183 (1993).

40. P.G. Mezey, Quantum Chemical Shape: New Density Domain Relations for the Topology of Molecular Bodies, Functional Groups, and Chemical Bonding, *Can. J. Chem.*, **72**, 928-935 (1994). (Special issue dedicated to Prof. J.C. Polanyi).

41. P.G. Mezey, *Methods of Molecular Shape-Similarity Analysis and Topological Shape Design, in "Molecular Similarity in Drug Design*, Ed. P.M. Dean, Chapman & Hall - Blackie Publishers, Glasgow, U.K., l995, pp 241-268.

42. P.G. Mezey, *Molecular Similarity Measures for Assessing Reactivity, in "Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches"*, Ed. R. Carbó, Kluwer Academic Publ., Dordrecht, The Netherlands, l995, pp 57-76.

43. P.G. Mezey, Shape Analysis of Macromolecular Electron Densities, *Structural Chem.*, **6**, 261-270 (1995).

44. P.D. Walker, G.M. Maggiora, M.A. Johnson, J.D. Petke, and P.G. Mezey, Shape Group Analysis of Molecular Similarity: Shape Similarity of Six-Membered Aromatic Ring Systems, *J. Chem. Inf. Comp. Sci.*, **35**, 568-578 (1995).

45. P.D. Walker, P.G. Mezey, G.M. Maggiora, M.A. Johnson, and J.D. Petke, Application of the Shape Group Method to Conformational Processes: Shape and Conjugation Changes in the Conformers of 2-Phenyl Pyrimidine, *J. Comput. Chem.*, **16**, 1474-1482 (1995).

46. P.G. Mezey, *Local Shape Analysis of Macromolecular Electron Densities*, in Computational Chemistry: Reviews and Current Trends, Vol.**1**", Ed. J. Leszczynski, World Scientific Publ., Singapore, 1996, pp 109-137.

47. P.G. Mezey, Group Theory of Electrostatic Potentials: A Tool for Quantum Chemical Drug Design, *J. Quant. Chem*., Quant. Biol. Symp., **12**, 113 (1986).

48. P.G. Mezey, Functional Groups in Quantum Chemistry, *Advances in Quantum Chemistry*, **27**, 163-222 (1996).

49. P.G. Mezey, The Shape of Molecular Charge Distributions: Group Theory without Symmetry, *J. Comput. Chem*., **8**, 462 (1987).

50. P.G. Mezey, *Descriptors of Molecular Shape in 3D, in "From Chemical Topology to Three-Dimensional Geometry*, Ed. A.T. Balaban, Plenum Press, New York, 1997, pp 25-42.

51. P.G. Mezey, Group Theory of Shapes of Asymmetric Biomolecules, *Int. J. Quantum Chem.*, Quant. Biol. Symp., **14**, 127 (1987).

52. P.G. Mezey and P.D. Walker, Fuzzy Molecular Fragments in Drug Research, *Drug Discovery Today (Elsevier Trend Journal)*, **2**, 6-11 (1997).

53. G. A. Arteca, V. B. Jammal, P. G. Mezey, J. S. Yadav, M. A. Hermsmeier, and T. M. Gund, Shape Group Studies of Molecular Similarity: Relative Shapes of Van der Waals and Electrostatic Potential Surfaces of Nicotinic Agonists, *J. Molec. Graphics*, **6**, 45 (1988).

54. P.G. Mezey, Topology and the Quantum Chemical Shape Concept, *Adv. in Mol. Sim.*, **2**, 79-92 (1998).

55. P.G. Mezey, Quantitative Shape - Activity Relations (QShAR), Molecular Shape Analysis, Charge Cloud Holography, and Computational Microscopy, in "*Quantitative Structure-Activity Relationships for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications*", Ed. J.D. Walker, SETAC (Society of Environmental Toxicity and Chemistry) Publ., SETAC Press, 2003, pp. 123-136.

56. P.G. Mezey, Holographic Electron Density Shape Theorem and Its Role in Drug Design and Toxicological Risk Assessment, *J. Chem. Inf. Comp. Sci*., **39**, 224-230 (1999).

57. G.A. Arteca, V.B. Jammal, and P. G. Mezey, Shape Group Studies of Molecular Similarity and Regioselectivity in Chemical Reactions, *J. Comput. Chem*., **9**, 608 (1988).

58. G.A. Arteca and P.G. Mezey , Shape Description of Conformationally Flexible Molecules: Application to Two-dimensional Conformational Problems, *Int. J. Quantum Chem*., Quant. Biol. Symp., **15**, 33 (1988).

59. P.G. Mezey, *Shape In Chemistry: An Introduction To Molecular Shape and Topology*, VCH Publishers, New York, 1993.

60. Mallakin, P.G. Mezey, Z. Zimpel, K.S. Berenhaut, B.M. Greenberg, and D.G. Dixon, Use of Quantitative Structure-Activity Relationship to Model the Photoinduced Toxicity of Anthracene and Oxygenated Anthracenes, *QSAR & Comb. Sc.*, **24**, 844-852 (2005).

61. C. Travaglini-Allocatelli, Y. Ivarsson, P. Jemth, S. Gianni, Folding and stability of globular proteins and implications for function, *Curr Op. Struct. Biol*. 19, **3**, (2009).

62. P.G. Mezey, in *Quantitative Shape—Activity Relations (QShAR), Molecular Shape Analysis, Charge Cloud Holography, and Computational Microscopy*, ed. by J.D. Walker Quantitative Structure-Activity Relationships for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications (SETAC (Society of Environmental Toxicity and Chemistry) Publ., SETAC Press, New York, 2003), pp. 123–136

63. P.G. Mezey, Molecular fragment shape variation index for functional groups and the holographic properties of electron density, *J. Math. Chem*., **50**, 926, (2012).

The following chapter is an exact copy of a published scientific article. Its format, including sub-chapters and reference list, adopts the format required by the journal in which it was published. The parameters of the article are as follows:

Authors: Zoltan Antal, Peter L. Warburton, and Paul G. Mezey

Title: Electron Density Shape Analysis of a Family of Through-Space and Through-Bond Interactions

Journal: Physical Chemistry Chemical Physics (*Phys. Chem. Chem. Phys*.)

Year, volume, and first page number: 2014, 16, 918

Note: some modifications to the original articles were necessary to successfully incorporate them into the main thesis body.

CHAPTER 2

For the multi-author nature of the first article and two-author nature of the second and third research articles, it seems warranted to name the exact contributions of each of these authors. PhD candidate Zoltan Antal carried out all research and analyzed all data resulting from such activity with the coordination of Dr. Paul G. Mezey. Also wrote the Introduction and Conclusion sections with the help of Dr. Paul G. Mezey. Having detailed knowledge on the shape analysis software used throughout the research, Dr. Peter Warburton provided crucial technical support, without which none of the research articles would have been possible.

Electron Density Shape Analysis of a Family of Through-Space and Through-Bond Interactions

Zoltan Antal, Peter L. Warburton, and Paul G. Mezey

Scientific Modeling and Simulation Laboratory (SMSL), Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, Newfoundland A1B3X7

Abstract: A family of styrene derivatives has been used to study the effects of through-space and through-bond interactions on the local and global shapes of electron densities of complete molecules and a set of substituents on their central rings. Shape analysis methods, which have been used extensively in the past for the study of molecular properties - molecular shape correlations, have shown that, in these molecules, a complementary role is played by the through-space and through-bond interactions. For each specific example, the dominance of either one of the two interactions can be identified and interpreted in terms of local shapes and the typical reactivities of the various substituents. Three levels of quantum chemical computational methods have been applied for these structures, including the B3LYP/cc-pVTZ level of density functional

methodology, and the essential conclusions are the same for all three levels. The general approach is suggested as a tool for the identification of specific interaction types which are able to modify molecular electron densities. By separately influencing the through-space and through-bond components using polar groups and groups capable of conjugation, some fine-tuning of the overall effects becomes possible. The method described may contribute to an improved understanding and control of molecular properties involving complex interactions with a possible role in the emerging field of molecular design.

2.1. Introduction

The understanding of the interactions between various parts of molecules is one of the fundamental goals of chemistry. By focusing on molecular parts, often regarded as substituents, such as alternative atoms or atomic groups replacing individual atoms in molecules, early experimental studies on substituent effects have provided some of the first, systematic models for such interactions. The way different substituents influence global molecular properties, their reactivities in some typical reactions, as well as local properties in some different, neighboring regions of the molecule, have been the topic of the early studies in physical organic chemistry.[1-11] How similar or different effects various substituents may produce have been initially investigated primarily by experimental techniques. On the other hand, among the early theoretical studies on the

relevant aspects of molecular similarity, the works based on the original quantum similarity approach of Carbó has played a central role.[11-21]

It has been clear that the electron density cloud plays a central role in transmitting the influence of substituents within a molecule, also confirmed by the fundamental theorem of density functional theory, the Hohenberg – Kohn theorem:[22] the non-degenerate ground state electron density contains the complete molecular information.

The fact that molecular electron density clouds are subject to easy deformations, *e.g.* by some nearby charged species such as an atomic ion, has raised the question: how is a local influence transmitted and how can it affect the global properties? Even in the early days of physical organic chemistry, it has been a natural assumption that any small part of the molecule, for example, even a single-atom substituent, has some influence on the global properties of the molecule. It has been shown only later, that a far stronger relation exists beyond mere "influence": any small molecular part with a positive volume, such as a local volume associated with a substituent, influences all other parts of the molecule in a *unique way*, and no other substituent can possibly reproduce the same influence! This follows from the Holographic Electron Density Theorem:[23] *any small positive volume part* of the molecular electron density cloud contains the complete information about the *entire molecule*. Consequently, a given substituent within a molecule fully "knows" what constitutes the rest of the molecule, and the mutual influence between the substituent and the rest of the molecule is *always unique*. This theorem [23] provides a strong theoretical background for looking for systematic approaches to the study of substituent effects of

local and global QSAR models involving both exhibited and latent properties,[23-26] as well as the relations between the foundations of Carbó's quantum similarity and the local-global information context of substituents.[27-39] In particular, the basic concepts of density functional theory and various practical applications, involving local chirality effects and other substituent effects have been investigated within such a framework.[40-64]

One fundamental question within the entire range of these topics is the relative importance of through-bond and through space effects, with respect to the total influence of a substituent. Although this has already been the subject of some early computational quantum chemistry investigations using relatively simple methodology available at that time,[7-10] it appears justified to address this question from the above general perspective.

Complex phenomena are often easier to understand if they are manifested in simple systems. The motivation for selecting a relatively simple set of molecules to study some of the regularities in these interactions, originates from the early advances of physical-organic chemistry, where the topic of substituent effects has played a central role, becoming one of the earliest representations of systematic rules on chemical effects within individual molecules, as reviewed by Anslyn and Dougherty.[1] The studies of Hammett,[2] Topsom,[3] and Taft[4-6] have had major influence on the interpretative aspects of physical organic chemistry, and have provided relatively simple, approximate, but quantitative (in nature) comparisons of the dominant substituent effects. Of course, it has long been recognized that the interactions of molecular parts are rather complex and no exhaustive description can be expected by any one-parameter numerical scale, or even

from two or three such parallel scales. A detailed accounting for the full extent of substituent effects has appeared to be a far too complex task without a detailed knowledge of molecular electron densities.

Nevertheless, some approximate quantum chemical models have already proven to be useful. One set of early quantum chemistry studies,[7-10] using relatively simple approximate molecular wavefunctions, has focused on the possibility of separation of through-bond and through-space effects in a certain family of model molecules where such an approximate separation could be achieved by a fortunate choice of the model molecules. Whereas the conclusions reached have been useful and have provided novel insight, the quality of the molecular wavefunction representations which could be employed in these early studies was not sufficient to provide firm answers to some of the questions raised. In view of the spectacular development of quantum chemistry computational techniques since those early studies, it appears warranted to re-visit these questions and approach the problem of through-space and through-bond interactions from a new perspective.

In 1999, Carbó-Dorca *et al.* proposed a new method of characterizing physicochemical parameters (e.g. molecular hydrophobicity, electronic effects) based on quantum self-similarity measures (QS-SM).[11] The novel approach successfully replaced the traditional parameters in classical QSAR with theoretical descriptors based on QS-SM.

Also motivated by the successes of molecular similarity studies,[11-21] the interest in a better understanding of the details of through-space and through-bond influences has

remained strong, with the ultimate aim of being able to utilize the potential influence of well-selected substituents in order to achieve a desired chemical effect.

Of course, through-bond and through-space interactions are often not easy to separate and identify individually, although it is evident that the initial stages of all bi-molecular or multi-molecular reactions primarily involve through-space interactions. Even in some mono-molecular "reactions", such as certain conformational changes, the early stages are likely to be dominated by small deformations where the potential advantages of changing the through-space interactions dominate. Hence, most changes on a potential energy hypersurface from one local minimum to another[65] are typically involving an initial change in the through-space interactions, whether the process is an actual reaction, or just a conformational change. Only at a later stage of most processes can one find that the through-bond interactions are likely to have an increased and possibly a far greater influence than the through-space interactions. In this context, the changes in the relative roles of through-space and through-bond interactions are fundamental in the understanding of most properties of potential energy hypersurfaces.[65] The same can be said about the influence of substituents on molecular shapes, starting from the classical VSEPR approaches [66, 67] to a more detailed analysis of molecular shapes.[68]

One specific family of reactions where the through-space interactions have a relatively enhanced importance even at a later stage of the reaction, are cyclo-additions,[69-73] where the actual, nearly simultaneous formation of two bonds also entails a continuing progression of through-space interactions, primarily in the region in between the two

newly formed bonds. Although the most important electron density changes occur along the newly formed bonds, as well as along some old bonds which eventually get disrupted, nevertheless, the electron density in between these bonds also changes significantly throughout the process. Some of the advanced quantum chemistry studies in this area have provided strong indications that these effects play an intricate, yet tractable role in such processes.[69-73]

One special aspect of the relations between through-space and through-bond interactions is important in macromolecular conformational studies, such as the attempts to understand the folding pattern of globular proteins. In such systems, it is believed that enhanced effects can be attributed to those low density regions of the electronic cloud which, in small molecules, have only very limited influence. By contrast, in globular macromolecules, the internal regions, even those which are not contributing much to formal chemical bonds, have more influence than conventionally assumed, simply by the sheer volume of such low density regions. Collectively, these regions, regarded as containing a Low Density "Glue" of electronic cloud,[74-77] have influence on conformational preferences of the macromolecule, since the total amount of electronic charge these low density regions represent is substantial. This is due to the significant part of the large internal volume of the macromolecule which fall in-between the local high density regions of chemical bonds. Clearly, the interactions conveyed by the Low Density Glue components involve important through-space components, although the through-bond interactions also have influence on these regions. Whereas the direct study of such regions in marcromolecules is a complex task, nevertheless, in simpler models,

such as those discussed in the present study, one may find important hints to that problem as well.

The approach employed here is in part motivated by Carbó-Dorca's quantum self-similarity measures, (QS-SM),[11-21] linking density functional theory with some more traditional approaches. This approach is based on a theory using electron density as the ultimate molecular descriptor and, as a part of it, fragment QS-SM to describe electronic substituent effects in the molecule. Carbó-Dorca's QS-SM theory also links most of the earlier approaches to the holographic electron density theorem,[23] and in this sense, it can be regarded as a fundamental component to molecular informatics using Density Functional Theory.[27-50]

In a similar spirit in 2011, Geerlings and Borgoo[78] have followed up on the notion that, in Density Functional Theory, the density functional **ρ(r)** is an information carrier and that the shape functional **σ(r)** is an equivalent, but by some technical criterion, even simpler source of information. From these information carriers, information or differences in information can be perhaps best extracted out through the Kullback–Leibler functional.[79] Also, it has been shown that the holographic electron density theorem, HEDT,[23] can be extended to the shape function, as suggested by Ayers.[80] One direct consequence of HEDT has already been pointed out by Mezey[23] that any non-zero volume, finite, arbitrary subdomain in a molecule also contains information about the overall chirality of the respective molecule, a conclusion that goes beyond the traditional "chiral center" approach. In this context, Geerlings and Borgoo[78] have used their

approach to confirm an important conclusion in a novel way: the information content of the shape function combined with the HEDT completely determines the Hamiltonian, the wave function, energy or any other observable properties, that may be regarded as a special perspective on the well-known equivalence of the electron density and the shape function as information carriers. This statement also illustrates the importance of analyzing the shapes of molecules at both global and local levels when searching for similarities between them.

2.2. Relevant aspects of shape analysis methodology

Molecular shape analysis is often regarded as the study of molecular conformations,[81] where typically, the structure of the nuclear framework is considered in the context of energy optimized molecular structures near minima of potential energy hypersurfaces.[65] However, the actual molecular shape that is most relevant for molecular interactions is the shape of the electronic density clouds,[68, 82-87] and the nuclear framework and the associated formal bond structure, often described by ball-and-stick models, is merely the skeleton that "holds" the electron density cloud, the actual molecular body, in place. It was recognized some time ago that the shape analysis of molecular electron densities provide important clues for the understanding of molecular properties and interactions. A large number of studies have been devoted to the analysis of the three-dimensional fuzzy bodies of molecular electron density clouds.[68, 82-87] Some of the tools used for this purpose involve one of the generalizations of geometry, namely topology (related but not to be confused with graph theory), where the inherently fuzzy nature of electron densities can be treated in a mathematically rigorous way. The colloquial reference to topology as

"rubber geometry" properly reflects the versatility of topology in describing flexible, mobile structures, such as vibrating molecules. The so-called Shape Group Methods, involving some elementary tools of algebraic topology, provide a group-theoretical framework for a detailed shape analysis for such fuzzy, "boundariless" objects like molecular electron densities.[68]

The molecular electron density shape analysis approach has been described and validated in detail in various publications.[68, 82-87] Here we shall give only a brief summary of the method, whereas some of the technical details, as well as the method used for fuzzy fragment density generation, are described briefly in Appendices A and B**.**

If, for a given nuclear configuration the electron density is $\rho(\mathbf{r})$ regarded as a function of the position vector **r**, then for a threshold value a, the molecular isodensity contour, MIDCO, is given as the surface $G(a) = \{r: \rho(\mathbf{r}) = a\}$. The computed MIDCOs are the primary tools of the Shape Group Method for molecular shape analysis, using *two* continuous parameters: the threshold value a for the electron density, and the parameter b defined as a reference curvature probing the local curvatures along every point of each molecular isodensity contour MIDCO G(a). At each point **r** of G(a), the local tangent plane is taken, and locally the MIDCO G(a) is regarded as a two-variable function defined over this tangent plane. The eigenvalues **v** and **w** of the second derivative (curvature) matrix (a formal Hessian matrix) of this function are the extreme local curvatures of G(a) at point **r**. If both **v** and **w** are less than the reference curvature b, (that is, if G(a) is locally convex relative to reference curvature b at point **r** then this

point **r** is formally deleted from G(a). For each (a, b) pair, one obtains a possibly truncated isocontour surface, some with several holes in it. The complete set of truncated surfaces for the whole range of available a and b values provides a detailed shape characterization of the complete molecular electron density.

Although both parameters a and b are continuous, hence there are infinitely many truncated isocontours, nevertheless, there are only a finite number of topologically different truncated isosurfaces, and their topological characterization by their one-dimensional homology groups, using the ranks of these groups, called Betti numbers, generates a finite set of integers and the shape code of the complete electron density. These shape codes, represented by an (a,b) map of Betti numbers for a finite set of a, b pairs taken from the continuous range of a and b values, can be compared directly, providing a simple, numerical shape similarity measure.

The problem of molecular electron density shape representation, a continuum problem, is reduced to a representation by a finite set of integers, preserving most of the relevant shape information. Also note that by using these shape codes, the (a,b) maps of integers, shape comparisons *do not require* any, usually ambiguous, superposition of the two molecules to be compared. In addition, the Shape Group Method is also applicable to electron densities of molecular fragments, an aspect of special relevance in our present study.

Whereas for smaller molecules, both earlier theoretical[88, 89] and experimental studies[90, 91] have shown that through-bond effects should be expected to dominate, nevertheless, even

in small molecules, but more prominently, in larger systems, the through-space effects can be significant. It appears that the larger the formal "volume" of a molecule, the more significant the through-space component becomes. Therefore, the through-space effects are expected to have an important role for both the local and the global shapes of large, globular molecules such as proteins, where, again, the rather extensive "Low Density Glue" regions[74-77] within the interior of such globular structures, falling in between the actual regions corresponding to chemical bonds, apparently have an important role in determining shape. The electron density shape aspect of the actual folding pattern and their contribution to the global stability of the macromolecule are important, even if these low density interactions are not as "focused" as formal bonding interactions. The shape aspects of this "Low Density Glue" bonding has been the subject of several studies,[74-77] and elucidation of the fundamentals of through-space effects in smaller model systems is expected to provide a better understanding of these macromolecular phenomena as well.

In the present study we shall focus on a rather simple model that allows a clear identification of at least some of the contributions to both through-bond and through-space interactions in special cases, where the model allows a formal elimination of some of the through-bond interactions, and where an electron density based similarity analysis provides numerical measures of the roles of some of these effects.

2.3. Computational Models and Methodology

How can one study the through-bond and through-space effects separately using theoretical methods only? One logical approach is based on the idea of using molecules which appear exceptionally capable of transmitting both of these effects and where the separation of the effects is also possible by disrupting the through-bond component. Aromatic rings, as formal spacing units which can contribute to formal π-system conjugation, can provide a medium for transmitting through-bond effects. On the other hand, leaving out such spacing units, the through-bond effects are disrupted, resulting in two separate molecules within a very specific distance which allows the study of the contribution to the through-space effect. Of course, within any full molecule including the spacing units, both through-bond and through-space effects will be present. In such a molecule containing an aromatic ring as the spacing unit, it is expected that the through-bond effect dominates; nonetheless, the through-space effect is also necessarily present. Therefore, the latter, after being identified separately, can be formally "subtracted" out of the effects of the complete molecule in order to obtain an approximation to a numerical or some other measure for the "pure" through-bond effect.

Figure 2.1 displays the electrostatic potential map of *para*-methoxy styrene as it appears on three molecular isodensity surfaces (MIDCOs) at three different electron density values. In the upper set of pictures, the whole molecules involve both the through-bond and through-space effects, and when the vinyl part is analyzed for shape changes, it

reflects the combination of these through-bond and through-space effects. On the other hand, the lower set of pictures represents the isolated molecules (IM), obtained from *para*-methoxy styrene, where the aromatic ring was "cut out" in order to study the "pure" through-space effect in these molecules. Naturally, the fragmentation was followed by the addition of atoms or groups of atoms to get chemically acceptable structures as IM: a hydrogen atom was added to the vinyl part and a methyl group was added to each functional group, originally also linked to a carbon atom in *para* location to the vinyl group.

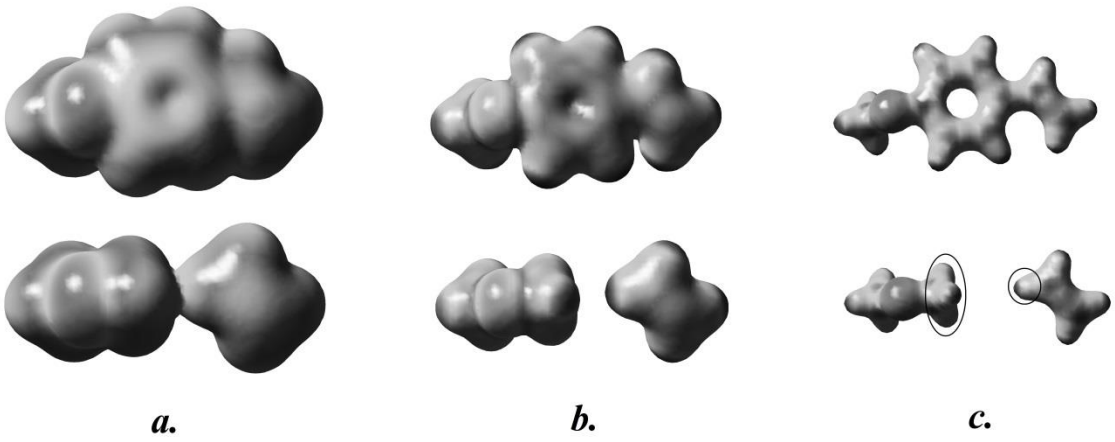


Figure 2.1. The electrostatic potential map of *para*-methoxy styrene on three different Molecular Isodensity Contour Surfaces, MIDCOs of electron density values (a): 0.001 au, (b): 0.01 au, (c): 0.1 au. The upper set of pictures represents the whole molecules, whilst the lower set their corresponding isolated molecules, IM, as fragments. Atoms or, groups of atoms were added to the fragments in order to get chemically correct structures, encircled in set (c).

In our model, the following *para* substituents to the vinyl group have been considered: Cis-VINYL, Trans-VINYL, F, OH, $N(CH_3)_2$, CH_3, CHO, CF_3, CN, H(STY), NH_2, NO_2, $COCH_3$, OCH_3, and, of course, H atom in styrene, denoted by STY in the tables.

The actual computations of the quantum chemistry results have been based on three main methodologies: HF (Hartree-Fock/cc-pVTZ) and DFT (Density Functional Theory, B3LYP model/cc-pVTZ), and MP2 (Møller-Plesset 2/cc-pVTZ) approaches (a lucid description of advantages of these methodologies can be found in ref. 92), using the Gaussian 09 program.[93] The electron density shape analysis has been performed and the numerical similarity measures have been computed using the program family [94], based on the methodology and algorithmic details described in references [82-87], and also summarized in [68]. Please note that some tables found in this thesis follow the format of the output files of the shape analysis software. Whole molecules have upper case letters (e.g.CH_3) and IMs have lower case letters (e.g.ch3). This type of notation does not necessarily follow the IUPAC notation, but facilitates easy differentiation between them.

2.4. Results and Discussion

In the next section, the focus is on the data obtained from the shape analysis of the whole molecules as well as the molecule pairs, "isolated molecules", IM, generated by formally eliminating the aromatic rings from the complete molecules. In particular, we shall investigate the results obtained for the molecules "minus" their respective central aromatic fragments in order to gain information on the through-bond interactions in these molecules.

2.4.1. The through-bond effect in *para*-substituted styrenes

2.4.1.1. Hartree-Fock/cc-pVTZ results:

The target of our analysis is the shape change in the vinyl group in the whole molecules and in the fragments, induced by the *para*-substituent of the complete molecules. The electronic cloud around the vinyl part changes as it is altered by the influences of the functional group in the *para* position and it also changes in the corresponding fragments of the molecule-pairs (IM). As mentioned, according to earlier theoretical[88, 89] and experimental results,[90, 91] through-bond effects are expected to dominate in the whole molecules and their vinyl shape altering potential should be greater than the through-space effect in the fragments. Nonetheless, this latter effect needs to be determined more precisely and expressed in some numerical way in order to get reproducible comparisons of the through-bond effect, by subtracting the through-space effects out of the total effects. For this purpose, numerical shape-similarity measures of the vinyl groups have been computed and analyzed.

The ***Hartree-Fock***[92] method, even if combined with a large correlation consistent basis set, is the computationally most economic, although the least reliable (however, as it turns out, still useful) model in our study. Table 2.1 displays the shape similarity analysis results as given by the geometries of molecules optimized by HF, showing the TB effect transmitted through the aromatic rings.

Table 2.1. Shape analysis data calculated using the HF method. The numbers represent the through-bond interaction in the whole molecules. The scale goes from 0 to 1, the former representing no similarity whilst the latter a perfect match between two molecules or parts of molecules.

	PARA FUNCTIONAL GROUP	*SIMILARITY*
1	**cis-VINYL**	0.6392
2	**trans-VINYL**	0.6563
3	**F**	0.6834
4	**OH**	0.6884
5	**$N(CH_3)_2$**	0.6897
6	**CH_3**	0.6977
7	**CHO**	0.7133
8	**CF_3**	0.7173
9	**CN**	0.7177
10	**STY**	0.7185
11	**NH_2**	0.7203
12	**NO_2**	0.7261
13	**$COCH_3$**	0.7289
14	**OCH_3**	0.7506

The numbers represent the difference between the shape analysis results obtained for the vinyl group of the whole molecules and the vinyl group of the fragments. The deviation from the ideal, perfect similarity of 1.0000 represents the shape altering effect of the pure through-bond interactions in the whole molecules.

It comes as no surprise to see both *cis*- and *trans*- vinyl groups in the top positions with the lowest measure of similarity, as they readily interact with the aromatic ring via their π-systems. A system of alternating single and double bonds throughout the whole molecule is beneficial for through-bond effects, traditionally explained by the "resonance effect". Fluorine is next in line, which, although a highly electronegative atom that

withdraws electrons from the sigma framework, has the ability to share lone pair electrons with the aromatic ring, altering the shape of the electron cloud around the vinyl group. Recalling the chemistry of fluorobenzene, it is easy to understand why fluorine has such a great effect on the analyzed vinyl part, as it greatly activates the *para*-carbon relative to its position. The next two functional groups, OH and $N(CH_3)_2$ are similar in nature as their lead atoms are highly electronegative as well as electron donating through their lone pairs. They readily interact with the aromatic ring and the vinyl group. Hartree-Fock computations have the tendency to group similar types of functional groups together such as the next four groups which share a carbon as their lead atom. Although chemically not necessarily similar, their ±I and ± E effects are buffered by the presence of the carbon. Styrene, although was expected to close the list, it appears above four other functional groups and this finding might be attributed to the limitations of HF. In this case, the functional group is represented by a single hydrogen atom and its through-bond (TB) effect is apparently overestimated. The next two functional groups are NH_2 and NO_2. Their effect on the aromatic ring is different, yet they are grouped together by all levels of theory, not only HF. Our expectations regarding the acetyl group were justified as it is placed on the bottom of the HF list, but the methoxy group, being an excellent activator which readily interacts with the aromatic ring, is underestimated by HF. Chemical intuition would place it among other groups conventionally considered similar, such as the hydroxyl and dimethyl-amino groups.

2.4.1.2. DFT results:

The ***DFT*** method applied,[92] represented by the B3LYP model, performs better and corrects some of the unexpected results of the Hartree-Fock computations (Table 2.2). One of the main differences is that DFT separates the vinyl groups; whilst *trans*-vinyl remains in the lead position, *cis*-vinyl gets pushed down to position 7. This apparent underestimation of the effect of *cis*-vinyl can be understood as soon as we analyze the performance of the MP2 model, discussed below. Fluorine is next in line, followed by the methyl group. At first glance, the position of this latter functional group may seem unexpected but, considering the fact that it may interact with the aromatic ring though hyperconjugation, could explain the results (this finding is also confirmed by the calculations resulting from MP2 geometries). Hence, in our opinion, the methyl group is in the right position on the list and not being placed there in the previous example is an apparent limitation of HF. Three groups (CHO, CN, CF_3) having carbon as their lead atom followed by NH_2 and NO_2 which are still grouped together. Styrene has been pushed down towards the end of the list, as its hydrogen atom (here the "functional group") cannot interact heavily with the aromatic ring and hence influence the electron cloud around the vinyl group. This is another improvement achieved by DFT when compared to HF. The acetyl group closes the list. According to chemical intuition, it should probably be above styrene, but the difference between the two numerical results is very minor.

Table 2.2. Shape analysis data provided by the DFT calculations. The numbers represent the through-bond interaction in the whole molecules.

	PARA FUNCTIONAL GROUP	*SIMILARITY*
1	**trans-VINYL**	0.6350
2	**F**	0.6700
3	**CH_3**	0.6761
4	**OH**	0.7538
5	**$N(CH_3)_2$**	0.7550
6	**OCH_3**	0.7567
7	**cis-VINYL**	0.7611
8	**CHO**	0.7644
9	**CN**	0.7658
10	**CF_3**	0.7663
11	**NH_2**	0.7670
12	**NO_2**	0.7699
13	**STY**	0.7740
14	**$COCH_3$**	0.7835

Table 2.3. Shape analysis data provided by the MP2 calculations. The numbers represent the through-bond interaction in the whole molecules.

	PARA FUNCTIONAL GROUP	*SIMILARITY*
1	**$N(CH_3)_2$**	0.6410
2	**cis-VINYL**	0.6465
3	**CH_3**	0.6799
4	**F**	0.6827
5	**NH_2**	0.6897
6	**NO_2**	0.6947
7	**OCH_3**	0.6984
8	**trans-VINYL**	0.6995
9	**CF_3**	0.7050
10	**$COCH_3$**	0.7090
11	**CHO**	0.7126
12	**OH**	0.7162
13	**STY**	0.7178
14	**CN**	0.7377

2.4.1.3. MP2 results:

Convergence (of the geometry optimization) problems of some molecules with the MP2[92] method were an issue and one of the "problem" molecules was $N(CH_3)_2$; nevertheless, all MP2 computations (using Gaussian 09[93]) eventually converged. Obtaining such an extreme position on the top of the MP2 list for this molecule is somewhat surprising (Table 2.3), however, in spite of the unusually difficult convergence problem, we do not suspect that a false solution is obtained. The MP2 method also separates the two vinyl groups but this time their order has been reversed. Similarly to DFT, the methyl and fluorine groups are next, but this time they are followed by NH_2 and NO_2, whilst OH gets placed to where the former pair was positioned in the case of the DFT results. Similar to the vinyl groups (for which the order is reversed), it seems that similar types of functional groups are always switched around, as if they maintain their approximate position on the list, but since different quantum chemistry models employ different computational approaches, they are placed on the list differently. The methoxy group closely follows, and functional groups (CF_3, $COCH_3$, CHO) having a carbon as their lead atom are next. Styrene and the cyano group close out the list. Once again, two similar types of groups are interchanged (as compared to DFT), namely, the acetyl and cyano groups. Their similarities are based on the fact that they are both deactivators and *meta*-directors from an electrophilic substitution point of view; also, they have a carbon as the lead atom which has an adjacent heteroatom.

2.4.2. The through-space effect

In the study of the through-space effect, the same general considerations apply for the expected performance of methodologies used, i.e. HF, DFT and MP2, as in part IVa.

Similar to the through-bond effect, the similarity numbers were collected into a similarity matrix (for some relevant information, see also Section V) which was used to extract information about the pure through-space effect. In this case, the lower right section of the matrix displayed below was the main tool, where the fragments are compared to each other (Table 2.4.).

Table 2.4. The fragment portion of the complete similarity matrix. (Also appears as the lower right section of the complete similarity matrix, see below). Here the lower case letters are used to differentiate the IM fragments from their whole molecule counterparts in a different table. Results displayed are calculated using B3LYP/cc-pVTZ geometries.

	sty	f	oh	cho	cn	nh_2	no_2	cf_3	$n(ch_3)_2$	och_3	$coch_3$	ch_3	cis vinyl	trans vinyl
sty	1	0.9814	0.9744	0.9707	0.9755	0.9783	0.9720	0.9686	0.9743	0.9579	0.9766	0.9801	0.9643	0.9749
f	0.9814	1	0.9714	0.9649	0.9767	0.9737	0.9697	0.9703	0.9749	0.9626	0.9760	0.9819	0.9696	0.9685
oh	0.9744	0.9714	1	0.9748	0.9726	0.9754	0.9662	0.9651	0.9690	0.9562	0.9678	0.9784	0.9702	0.9691
cho	0.9707	0.9649	0.9748	1	0.9702	0.9701	0.9637	0.9608	0.9731	0.9508	0.9736	0.9631	0.9630	0.9607
cn	0.9755	0.9767	0.9726	0.9702	1	0.9731	0.9744	0.9721	0.9766	0.9691	0.9743	0.9790	0.9696	0.9673
nh_2	0.9783	0.9737	0.9754	0.9701	0.9731	1	0.9649	0.9650	0.9778	0.9649	0.9671	0.9730	0.9636	0.9678
no_2	0.9720	0.9697	0.9662	0.9637	0.9744	0.9649	1	0.9680	0.9696	0.9562	0.9649	0.9708	0.9667	0.9614
cf_3	0.9686	0.9703	0.9651	0.9608	0.9721	0.9650	0.9680	1	0.9702	0.9656	0.9807	0.9685	0.9772	0.9720
$n(ch_3)_2$	0.9743	0.9749	0.9690	0.9731	0.9766	0.9778	0.9696	0.9702	1	0.9620	0.9718	0.9777	0.9654	0.9637
och_3	0.9579	0.9626	0.9562	0.9508	0.9691	0.9649	0.9562	0.9656	0.9620	1	0.9613	0.9631	0.9642	0.9567
$coch_3$	0.9766	0.9760	0.9678	0.9736	0.9743	0.9671	0.9649	0.9807	0.9718	0.9613	1	0.9725	0.9719	0.9760
ch_3	0.9801	0.9819	0.9784	0.9631	0.9790	0.9730	0.9708	0.9685	0.9777	0.9631	0.9725	1	0.9713	0.9749
cis vinyl	0.9643	0.9696	0.9702	0.9630	0.9696	0.9636	0.9667	0.9772	0.9654	0.9642	0.9719	0.9713	1	0.9673
trans vinyl	0.9749	0.9685	0.9691	0.9607	0.9673	0.9678	0.9614	0.9720	0.9637	0.9567	0.9760	0.9749	0.9673	1

As suspected because of the weak character of this effect, the variations of similarity values are not as large as they were for the through-bond effect. Most of the variations of

the numerical values are between 2-3%, the greatest variation being about 5%. Since these through-space (TS) effects manifest themselves mostly through electrostatic interactions, it is easy to see why the changes in similarity values presented above are so low (when compared to TB effects). The illustration below (Figure 2.2) presents one such fragment pair for the fragmented counterpart of *para*-methoxy-styrene. As mentioned above, after fragmentation, the structures had to be completed by the addition of a methyl group to the functional group (and a hydrogen atom to the vinyl side, respectively).

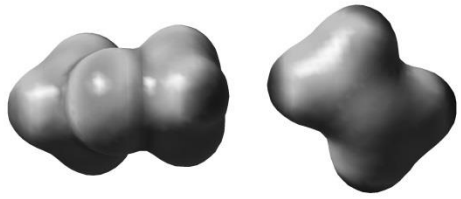


Figure 2.2. An acetone/ethene pair representing the fragmented counterpart of *para*-methoxy-styrene. The isodensity value is 0.01 a.u. and single point calculations were carried out with B3LYP/cc-pVTZ.

By doing this, the electrostatic potential between the two resulting molecules has been further decreased as the effect of oxygen is buffered by the carbon and the compensation of the charge on the oxygen is shared by two methyl groups.

In the similarity matrix, the full molecules and their fragmented counterparts are both used, and the difference between the two provides information about the through-bond effects.

Because of the abovementioned reasons, results provided by various levels of theory are very similar to each other (Table 2.5) and the similarity values differ no more than 5%, giving additional confidence in the conclusions. Furthermore, this finding also provides indication that the effects are significant enough to be reflected in HF calculations, so even HF can provide useful information.

Table 2.5. Collection of similarity data for three different levels of quantum theory.

	HF			**DFT**			**MP2**	
	PARA FUNCTIONAL GROUP			***PARA FUNCTIONAL GROUP***			***PARA FUNCTIONAL GROUP***	
		SIMILARITY			*SIMILARITY*			*SIMILARITY*
1	**och_3**	*0.9459*	1	**och_3**	*0.9579*	1	**och_3**	*0.9683*
2	**cis vinyl**	*0.9702*	2	**cis vinyl**	*0.9643*	2	**cf_3**	*0.9708*
3	**no_2**	*0.9737*	3	**cf_3**	*0.9686*	3	**$n(ch_3)_2$**	*0.9712*
4	**cn**	*0.9762*	4	**cho**	*0.9707*	4	**no_2**	*0.9736*
5	**cf_3**	*0.9795*	5	**no_2**	*0.9720*	5	**cho**	*0.9748*
6	**$n(ch_3)_2$**	*0.9813*	6	**$n(ch_3)_2$**	*0.9743*	6	**cis vinyl**	*0.9754*
7	**trans vinyl**	*0.9825*	7	**oh**	*0.9744*	7	**f**	*0.9818*
8	**f**	*0.9825*	8	**trans vinyl**	*0.9749*	8	**cn**	*0.9824*
9	**nh_2**	*0.9837*	9	**cn**	*0.9755*	9	**trans vinyl**	*0.9836*
10	**ch_3**	*0.9842*	10	**$coch_3$**	*0.9766*	10	**ch_3**	*0.9842*
11	**oh**	*0.9854*	11	**nh_2**	*0.9783*	11	**oh**	*0.9865*
12	**$coch_3$**	*0.9872*	12	**ch_3**	*0.9801*	12	**nh_2**	*0.9871*
13	**cho**	*0.9860*	13	**f**	*0.9814*	13	**$coch_3$**	*0.9871*
14	**sty**	*1.0000*	14	**sty**	*1.0000*	14	**sty**	*1.0000*

2.5. Similarity matrices (SM)

The collection of numerical similarity values, called the similarity matrix (SM), contains all similarity measures obtained for the given geometries as calculated with Hartree-Fock, B3LYP and MP2 levels of theory (the B3LYP results are presented in Table 2.6 and 2.7). There is a specific similarity matrix for every level of theory and each can be analyzed separately with the results evaluated in the conclusions section. Every SM contains four separate blocks of numbers, each representing the combinations of similarities of the "ringed" (complete) and "ringless" (IM) systems. In the SM itself, abbreviations written with upper case letters represent the ringed styrene systems and lower case abbreviations represent the ringless systems. The upper left block is the similarity map of the ringed systems, bottom left and upper right block represent the comparisons of the ringed and ringless systems, while the bottom right block shows the similarity map of the ringless systems. The upper right block mirrors the effect of the benzene ring and functional groups on the vinyl groups, therefore through-bond effects can be studied on it, whilst the bottom right block will give us information of the through-space interactions between the functional groups and vinyl groups. The upper left block is the combination of the two abovementioned effects.

If we take the simplest molecule of our system, styrene, and designate it as the reference molecule, our "functional group" will become the *para*-hydrogen of the benzene ring (relative to the vinyl group). Here, any alteration caused in the electronic cloud of the

vinyl group will be mainly the result of the benzene part influenced by a functional group. To appreciate the effect of any individual functional group, we will look at their difference relative to the reference molecule. This approach is valid for the ringless systems as well, where our reference is represented by the "ringless styrene" IM system. The combination of the two effects is given by the bottom left (or upper right) blocks and the list of substituents is in an increasing order as given below (large difference from 1 meaning increased combined effect). By studying the similarity matrices in detail, it is obvious that through-bond interactions dominate. These matrices give us useful information on the magnitude of through-space effects; therefore we have a tool to appreciate each effect on the vinyl groups and to determine their influence. There is an excellent agreement between the three levels of theory when through-space effects are described. Styrene systems compared to the reference molecule styrene show comparable similarities when their shape is analyzed. This is all reflected clearly by the similarity matrices produced by the various levels of theory (upper left blocks). Please note that two different views of the same similarity matrix can be seen below. Table 2.6 illustrates the full structure of these matrices and Table 2.7 was added for easy readability.

Table 2.6.The structure of the similarity matrix provided by B3LYP/cc-pVTZ. The TB effect highlighted in bold.

	STY	F	OH	CHO	CN	NH_2	NO_2	CF_3	$N(CH_3)_2$	OCH_3	$COCH_3$	CH_3	cis VINYL	trans VINYL
STY	1	0.8035	0.9422	0.9450	0.9292	0.9338	0.9184	0.9281	0.9172	0.9009	0.9420	0.8129	0.9025	0.7614
F	0.8035	1	0.8007	0.8016	0.7896	0.8061	0.7988	0.8058	0.7902	0.7799	0.8189	0.9456	0.7696	0.8694
OH	0.9422	0.8007	1	0.9382	0.9247	0.9498	0.8981	0.9254	0.9391	0.9217	0.9235	0.7933	0.9021	0.7667
CHO	0.9450	0.8016	0.9382	1	0.9527	0.9141	0.9261	0.9312	0.9103	0.8871	0.9375	0.8042	0.9032	0.7500
CN	0.9292	0.7896	0.9247	0.9527	1	0.9035	0.9332	0.9287	0.9104	0.8864	0.9374	0.7960	0.9141	0.7479
NH_2	0.9338	0.8061	0.9498	0.9141	0.9035	1	0.8996	0.9393	0.9250	0.9139	0.9333	0.8048	0.9065	0.7837
NO_2	0.9184	0.7988	0.8981	0.9261	0.9332	0.8996	1	0.9338	0.8888	0.8973	0.9465	0.7953	0.9069	0.7676
CF_3	0.9281	0.8058	0.9254	0.9312	0.9287	0.9393	0.9338	1	0.9004	0.9070	0.9404	0.8084	0.9160	0.7690
$N(CH_3)_2$	0.9172	0.7902	0.9391	0.9103	0.9104	0.9250	0.8888	0.9004	1	0.9308	0.9209	0.7943	0.8751	0.7576
OCH_3	0.9009	0.7799	0.9217	0.8871	0.8864	0.9139	0.8973	0.9070	0.9308	1	0.8981	0.7757	0.8879	0.7685
$COCH_3$	0.9420	0.8189	0.9235	0.9375	0.9374	0.9333	0.9465	0.9404	0.9209	0.8981	1	0.8117	0.9136	0.7685
CH_3	0.8129	0.9456	0.7933	0.8042	0.7960	0.8048	0.7953	0.8084	0.7943	0.7757	0.8117	1	0.7727	0.8587
cis VINYL	0.9025	0.7696	0.9021	0.9032	0.9141	0.9065	0.9069	0.9160	0.8751	0.8879	0.9136	0.7727	1	0.7845
trans VINYL	0.7614	0.8694	0.7667	0.7500	0.7479	0.7837	0.7676	0.7690	0.7576	0.7685	0.7685	0.8587	0.7845	1

	sty	f	oh	cho	cn	nh_2	no_2	cf_3	$n(ch_3)_2$	och_3	$coch_3$	ch_3	cis vinyl	trans vinyl
sty	0.7740	0.7728	0.7681	0.7724	0.7708	0.7709	0.7754	0.7693	0.7682	0.7679	0.7712	0.7732	0.7726	0.7694
f	0.6669	0.6700	0.6682	0.6771	0.6657	0.6688	0.6651	0.6690	0.6714	0.6561	0.6835	0.6692	0.6712	0.6710
oh	0.7602	0.7590	0.7538	0.7532	0.7576	0.7559	0.7646	0.7648	0.7520	0.7523	0.7632	0.7600	0.7588	0.7620
cho	0.7617	0.7602	0.7564	0.7644	0.7634	0.7605	0.7678	0.7675	0.7600	0.7550	0.7649	0.7617	0.7699	0.7632
cn	0.7582	0.7608	0.7602	0.7600	0.7658	0.7617	0.7699	0.7590	0.7553	0.7582	0.7559	0.7632	0.7653	0.7544
nh_2	0.7719	0.7678	0.7655	0.7662	0.7716	0.7670	0.7745	0.7690	0.7644	0.7699	0.7726	0.7705	0.7670	0.7702
no_2	0.7611	0.7603	0.7579	0.7641	0.7646	0.7624	0.7699	0.7643	0.7582	0.7600	0.7700	0.7650	0.7735	0.7656
cf_3	0.7605	0.7605	0.7547	0.7529	0.7620	0.7579	0.7684	0.7663	0.7576	0.7661	0.7670	0.7632	0.7685	0.7629
$n(ch_3)_2$	0.7573	0.7612	0.7600	0.7580	0.7655	0.7609	0.7661	0.7582	0.7550	0.7532	0.7580	0.7612	0.7591	0.7553
och_3	0.7488	0.7482	0.7465	0.7553	0.7532	0.7550	0.7596	0.7564	0.7494	0.7567	0.7612	0.7485	0.7585	0.7512
$coch_3$	0.7860	0.7845	0.7848	0.7840	0.7882	0.7849	0.7932	0.7830	0.7802	0.7816	0.7835	0.7902	0.7884	0.7828
ch_3	0.6767	0.6781	0.6739	0.6828	0.6743	0.6722	0.6743	0.6676	0.6765	0.6570	0.6858	0.6761	0.6757	0.6738
cis vinyl	0.7497	0.7526	0.7474	0.7450	0.7552	0.7447	0.7628	0.7544	0.7468	0.7576	0.7538	0.7523	0.7611	0.7456
trans vinyl	0.6347	0.6388	0.6376	0.6435	0.6427	0.6445	0.6450	0.6379	0.6473	0.6475	0.6447	0.6345	0.6469	0.6350

	sty	f	oh	cho	cn	nh_2	no_2	cf_3	$n(ch_3)_2$	och_3	$coch_3$	ch_3	cis vinyl	trans vinyl
sty	**0.7740**	0.6669	0.7602	0.7617	0.7582	0.7719	0.7611	0.7605	0.7573	0.7488	0.7860	0.6767	0.7497	0.6347
f	0.7728	**0.6700**	0.7590	0.7602	0.7608	0.7678	0.7603	0.7605	0.7612	0.7482	0.7845	0.6781	0.7526	0.6388
oh	0.7681	0.6682	**0.7538**	0.7564	0.7602	0.7655	0.7579	0.7547	0.7600	0.7465	0.7848	0.6739	0.7474	0.6376
cho	0.7724	0.6771	0.7532	**0.7644**	0.7600	0.7662	0.7641	0.7529	0.7580	0.7553	0.7840	0.6828	0.7450	0.6435
cn	0.7708	0.6657	0.7576	0.7634	**0.7658**	0.7716	0.7646	0.7620	0.7655	0.7532	0.7882	0.6743	0.7552	0.6427
nh_2	0.7709	0.6688	0.7559	0.7605	0.7617	**0.7670**	0.7624	0.7579	0.7609	0.7550	0.7849	0.6722	0.7447	0.6445
no_2	0.7754	0.6651	0.7646	0.7678	0.7699	0.7745	**0.7699**	0.7684	0.7661	0.7596	0.7932	0.6743	0.7628	0.6450
cf_3	0.7693	0.6690	0.7648	0.7675	0.7590	0.7690	0.7643	**0.7663**	0.7582	0.7564	0.7830	0.6676	0.7544	0.6379
$n(ch_3)_2$	0.7682	0.6714	0.7520	0.7600	0.7553	0.7644	0.7582	0.7576	**0.7550**	0.7494	0.7802	0.6765	0.7468	0.6473
och_3	0.7679	0.6561	0.7523	0.7550	0.7582	0.7699	0.7600	0.7661	0.7532	**0.7567**	0.7816	0.6570	0.7576	0.6475
$coch_3$	0.7712	0.6835	0.7632	0.7649	0.7559	0.7726	0.7700	0.7670	0.7580	0.7612	**0.7835**	0.6858	0.7538	0.6447
ch_3	0.7732	0.6692	0.7600	0.7617	0.7632	0.7705	0.7650	0.7632	0.7612	0.7485	0.7902	**0.6761**	0.7523	0.6345
cis vinyl	0.7726	0.6712	0.7588	0.7699	0.7653	0.7670	0.7735	0.7685	0.7591	0.7585	0.7884	0.6757	**0.7611**	0.6469
trans vinyl	0.7694	0.6710	0.7620	0.7632	0.7544	0.7702	0.7656	0.7629	0.7553	0.7512	0.7828	0.6738	0.7456	**0.6350**

	sty	f	oh	cho	cn	nh_2	no_2	cf_3	$n(ch_3)_2$	och_3	$coch_3$	ch_3	cis vinyl	trans vinyl
sty	1	0.9814	0.9744	0.9707	0.9755	0.9783	0.9720	0.9686	0.9743	0.9579	0.9766	0.9801	0.9643	0.9749
f	0.9814	1	0.9714	0.9649	0.9767	0.9737	0.9697	0.9703	0.9749	0.9626	0.9760	0.9819	0.9696	0.9685
oh	0.9744	0.9714	1	0.9748	0.9726	0.9754	0.9662	0.9651	0.9690	0.9562	0.9678	0.9784	0.9702	0.9691
cho	0.9707	0.9649	0.9748	1	0.9702	0.9701	0.9637	0.9608	0.9731	0.9508	0.9736	0.9631	0.9630	0.9607
cn	0.9755	0.9767	0.9726	0.9702	1	0.9731	0.9744	0.9721	0.9766	0.9691	0.9743	0.9790	0.9696	0.9673
nh_2	0.9783	0.9737	0.9754	0.9701	0.9731	1	0.9649	0.9650	0.9778	0.9649	0.9671	0.9730	0.9636	0.9678
no_2	0.9720	0.9697	0.9662	0.9637	0.9744	0.9649	1	0.9680	0.9696	0.9562	0.9649	0.9708	0.9667	0.9614
cf_3	0.9686	0.9703	0.9651	0.9608	0.9721	0.9650	0.9680	1	0.9702	0.9656	0.9807	0.9685	0.9772	0.9720
$n(ch_3)_2$	0.9743	0.9749	0.9690	0.9731	0.9766	0.9778	0.9696	0.9702	1	0.9620	0.9718	0.9777	0.9654	0.9637
och_3	0.9579	0.9626	0.9562	0.9508	0.9691	0.9649	0.9562	0.9656	0.9620	1	0.9613	0.9631	0.9642	0.9567
$coch_3$	0.9766	0.9760	0.9678	0.9736	0.9743	0.9671	0.9649	0.9807	0.9718	0.9613	1	0.9725	0.9719	0.9760
ch_3	0.9801	0.9819	0.9784	0.9631	0.9790	0.9730	0.9708	0.9685	0.9777	0.9631	0.9725	1	0.9713	0.9749
cis vinyl	0.9643	0.9696	0.9702	0.9630	0.9696	0.9636	0.9667	0.9772	0.9654	0.9642	0.9719	0.9713	1	0.9673
trans vinyl	0.9749	0.9685	0.9691	0.9607	0.9673	0.9678	0.9614	0.9720	0.9637	0.9567	0.9760	0.9749	0.9673	1

Table 2.7. The similarity matrix provided by the DFT model.

	STY	F	OH	CHO	CN	NH_2	NO_2	CF_3	$N(CH_3)_2$	OCH_3	$COCH_3$	CH_3	Cis VINYL	Tr VINYL
STY	1	0.8035	0.9422	0.9450	0.9292	0.9338	0.9184	0.9281	0.9172	0.9009	0.9420	0.8129	0.9025	0.7614
F	0.8035	1	0.8007	0.8016	0.7896	0.8061	0.7988	0.8058	0.7902	0.7799	0.8189	0.9456	0.7696	0.8694
OH	0.9422	0.8007	1	0.9382	0.9247	0.9498	0.8981	0.9254	0.9391	0.9217	0.9235	0.7933	0.9021	0.7667
CHO	0.9450	0.8016	0.9382	1	0.9527	0.9141	0.9261	0.9312	0.9103	0.8871	0.9375	0.8042	0.9032	0.7500
CN	0.9292	0.7896	0.9247	0.9527	1	0.9035	0.9332	0.9287	0.9104	0.8864	0.9374	0.7960	0.9141	0.7479
NH_2	0.9338	0.8061	0.9498	0.9141	0.9035	1	0.8996	0.9393	0.9250	0.9139	0.9333	0.8048	0.9065	0.7837
NO_2	0.9184	0.7988	0.8981	0.9261	0.9332	0.8996	1	0.9338	0.8888	0.8973	0.9465	0.7953	0.9069	0.7676
CF_3	0.9281	0.8058	0.9254	0.9312	0.9287	0.9393	0.9338	1	0.9004	0.9070	0.9404	0.8084	0.9160	0.7690
$N(CH_3)_2$	0.9172	0.7902	0.9391	0.9103	0.9104	0.9250	0.8888	0.9004	1	0.9308	0.9209	0.7943	0.8751	0.7576
OCH_3	0.9009	0.7799	0.9217	0.8871	0.8864	0.9139	0.8973	0.9070	0.9308	1	0.8981	0.7757	0.8879	0.7685
$COCH_3$	0.9420	0.8189	0.9235	0.9375	0.9374	0.9333	0.9465	0.9404	0.9209	0.8981	1	0.8117	0.9136	0.7685
CH_3	0.8129	0.9456	0.7933	0.8042	0.7960	0.8048	0.7953	0.8084	0.7943	0.7757	0.8117	1	0.7727	0.8587
Cis VINYL	0.9025	0.7696	0.9021	0.9032	0.9141	0.9065	0.9069	0.9160	0.8751	0.8879	0.9136	0.7727	1	0.7845
Tr VINYL	0.7614	0.8694	0.7667	0.7500	0.7479	0.7837	0.7676	0.7690	0.7576	0.7685	0.7685	0.8587	0.7845	1
sty	**0.7740**	0.6669	0.7602	0.7617	0.7582	0.7719	0.7611	0.7605	0.7573	0.7488	0.7860	0.6767	0.7497	0.6347
f	0.7728	**0.6700**	0.7590	0.7602	0.7608	0.7678	0.7603	0.7605	0.7612	0.7482	0.7845	0.6781	0.7526	0.6388
oh	0.7681	0.6682	**0.7538**	0.7564	0.7602	0.7655	0.7579	0.7547	0.7600	0.7465	0.7848	0.6739	0.7474	0.6376
cho	0.7724	0.6771	0.7532	**0.7644**	0.7600	0.7662	0.7641	0.7529	0.7580	0.7553	0.7840	0.6828	0.7450	0.6435
cn	0.7708	0.6657	0.7576	0.7634	**0.7658**	0.7716	0.7646	0.7620	0.7655	0.7532	0.7882	0.6743	0.7552	0.6427
nh_2	0.7709	0.6688	0.7559	0.7605	0.7617	**0.7670**	0.7624	0.7579	0.7609	0.7550	0.7849	0.6722	0.7447	0.6445
no_2	0.7754	0.6651	0.7646	0.7678	0.7699	0.7745	**0.7699**	0.7684	0.7661	0.7596	0.7932	0.6743	0.7628	0.6450
cf_3	0.7693	0.6690	0.7648	0.7675	0.7590	0.7690	0.7643	**0.7663**	0.7582	0.7564	0.7830	0.6676	0.7544	0.6379
$n(ch_3)_2$	0.7682	0.6714	0.7520	0.7600	0.7553	0.7644	0.7582	0.7576	**0.7550**	0.7494	0.7802	0.6765	0.7468	0.6473
och_3	0.7679	0.6561	0.7523	0.7550	0.7582	0.7699	0.7600	0.7661	0.7532	**0.7567**	0.7816	0.6570	0.7576	0.6475
$coch_3$	0.7712	0.6835	0.7632	0.7649	0.7559	0.7726	0.7700	0.7670	0.7580	0.7612	**0.7835**	0.6858	0.7538	0.6447
ch_3	0.7732	0.6692	0.7600	0.7617	0.7632	0.7705	0.7650	0.7632	0.7612	0.7485	0.7902	**0.6761**	0.7523	0.6345
cis vinyl	0.7726	0.6712	0.7588	0.7699	0.7653	0.7670	0.7735	0.7685	0.7591	0.7585	0.7884	0.6757	**0.7611**	0.6469
tr vinyl	0.7694	0.6710	0.7620	0.7632	0.7544	0.7702	0.7656	0.7629	0.7553	0.7512	0.7828	0.6738	0.7456	**0.6350**

	sty	f	oh	cho	cn	nh_2	no_2	cf_3	$n(ch_3)_2$	och_3	$coch_3$	ch_3	cis vinyl	trans vinyl
STY	0.7740	0.7728	0.7681	0.7724	0.7708	0.7709	0.7754	0.7693	0.7682	0.7679	0.7712	0.7732	0.7726	0.7694
F	0.6669	0.6700	0.6682	0.6771	0.6657	0.6688	0.6651	0.6690	0.6714	0.6561	0.6835	0.6692	0.6712	0.6710
OH	0.7602	0.7590	0.7538	0.7532	0.7576	0.7559	0.7646	0.7648	0.7520	0.7523	0.7632	0.7600	0.7588	0.7620
CHO	0.7617	0.7602	0.7564	0.7644	0.7634	0.7605	0.7678	0.7675	0.7600	0.7550	0.7649	0.7617	0.7699	0.7632
CN	0.7582	0.7608	0.7602	0.7600	0.7658	0.7617	0.7699	0.7590	0.7553	0.7582	0.7559	0.7632	0.7653	0.7544
NH_2	0.7719	0.7678	0.7655	0.7662	0.7716	0.7670	0.7745	0.7690	0.7644	0.7699	0.7726	0.7705	0.7670	0.7702
NO_2	0.7611	0.7603	0.7579	0.7641	0.7646	0.7624	0.7699	0.7643	0.7582	0.7600	0.7700	0.7650	0.7735	0.7656
CF_3	0.7605	0.7605	0.7547	0.7529	0.7620	0.7579	0.7684	0.7663	0.7576	0.7661	0.7670	0.7632	0.7685	0.7629
$N(CH_3)_2$	0.7573	0.7612	0.7600	0.7580	0.7655	0.7609	0.7661	0.7582	0.7550	0.7532	0.7580	0.7612	0.7591	0.7553
OCH_3	0.7488	0.7482	0.7465	0.7553	0.7532	0.7550	0.7596	0.7564	0.7494	0.7567	0.7612	0.7485	0.7585	0.7512
$COCH_3$	0.7860	0.7845	0.7848	0.7840	0.7882	0.7849	0.7932	0.7830	0.7802	0.7816	0.7835	0.7902	0.7884	0.7828
CH_3	0.6767	0.6781	0.6739	0.6828	0.6743	0.6722	0.6743	0.6676	0.6765	0.6570	0.6858	0.6761	0.6757	0.6738
Cis VINYL	0.7497	0.7526	0.7474	0.7450	0.7552	0.7447	0.7628	0.7544	0.7468	0.7576	0.7538	0.7523	0.7611	0.7456
Tr VINYL	0.6347	0.6388	0.6376	0.6435	0.6427	0.6445	0.6450	0.6379	0.6473	0.6475	0.6447	0.6345	0.6469	0.6350
sty	1	0.9814	0.9744	0.9707	0.9755	0.9783	0.9720	0.9686	0.9743	0.9579	0.9766	0.9801	0.9643	0.9749
f	0.9814	1	0.9714	0.9649	0.9767	0.9737	0.9697	0.9703	0.9749	0.9626	0.9760	0.9819	0.9696	0.9685
oh	0.9744	0.9714	1	0.9748	0.9726	0.9754	0.9662	0.9651	0.9690	0.9562	0.9678	0.9784	0.9702	0.9691
cho	0.9707	0.9649	0.9748	1	0.9702	0.9701	0.9637	0.9608	0.9731	0.9508	0.9736	0.9631	0.9630	0.9607
cn	0.9755	0.9767	0.9726	0.9702	1	0.9731	0.9744	0.9721	0.9766	0.9691	0.9743	0.9790	0.9696	0.9673
nh_2	0.9783	0.9737	0.9754	0.9701	0.9731	1	0.9649	0.9650	0.9778	0.9649	0.9671	0.9730	0.9636	0.9678
no_2	0.9720	0.9697	0.9662	0.9637	0.9744	0.9649	1	0.9680	0.9696	0.9562	0.9649	0.9708	0.9667	0.9614
cf_3	0.9686	0.9703	0.9651	0.9608	0.9721	0.9650	0.9680	1	0.9702	0.9656	0.9807	0.9685	0.9772	0.9720
$n(ch_3)_2$	0.9743	0.9749	0.9690	0.9731	0.9766	0.9778	0.9696	0.9702	1	0.9620	0.9718	0.9777	0.9654	0.9637
och_3	0.9579	0.9626	0.9562	0.9508	0.9691	0.9649	0.9562	0.9656	0.9620	1	0.9613	0.9631	0.9642	0.9567
$coch_3$	0.9766	0.9760	0.9678	0.9736	0.9743	0.9671	0.9649	0.9807	0.9718	0.9613	1	0.9725	0.9719	0.9760
ch_3	0.9801	0.9819	0.9784	0.9631	0.9790	0.9730	0.9708	0.9685	0.9777	0.9631	0.9725	1	0.9713	0.9749
cis vinyl	0.9643	0.9696	0.9702	0.9630	0.9696	0.9636	0.9667	0.9772	0.9654	0.9642	0.9719	0.9713	1	0.9673
tr vinyl	0.9749	0.9685	0.9691	0.9607	0.9673	0.9678	0.9614	0.9720	0.9637	0.9567	0.9760	0.9749	0.9673	1

2.6. Shape Deviation Measures (SDM)

For our current purposes of separate identification of through-bond and through-space components, the results of the shape-similarity analysis can be best analyzed if we convert the usual shape similarity measures to shape deviation measures, SDM, by the formulas discussed below.

Using styrene STY and the IM pair of styrene as references, we denote the shape similarity measure of vinyl groups in styrene STY when compared to *para*-A-substituted styrene by S(STY,A) and for the corresponding two pairs of isolated molecules by S(sty,a). Then, for a given *para*-substituent A, the total shape deviation measure from the references are defined as

$$SDM(A) = 1 - S(STY,A) \quad (5)$$

and

$$SDM(a) = 1 - S(sty,a). \quad (6)$$

The shape-deviation component corresponding to the through-bond effects is then given as

$$SDM_{tb}(A) = (1 - S(STY,A)) - (1 - S(sty,a)) = S(sty,a) - S(STY,A), \quad (7)$$

whereas the shape deviation component corresponding to the through-space effect is given as

$$SDM_{ts}(A) = SDM(A) - SDM_{tb}(A) = (1\text{-}S(STY,A)) - (S,(sty,a) - S(STY,A)) = 1\text{-}S(sty,a). \quad (8)$$

The relative magnitudes of the two components provide information on the relative roles and possible dominance of the through-bond and through-space interactions and their variations within the family of substituents studied.

The results computed for SDMs are given in Tables 2.8, 2.9, and 2.10.

Table 2.8. Shape Deviation Measures (SDM) results computed with HF/cc-pVTZ

		S(STY,A)	*S(sty,a)*	*SDM(A)*	*$SDM_{tb}(A)$*	*$SDM_{ts}(A)$*
1	**STY**	1.0000	1.0000	0.0000	0.0000	0.0000
2	**CHO**	0.9662	0.9860	0.0338	0.0198	0.0140
3	**OH**	0.9622	0.9854	0.0378	0.0232	0.0146
4	**NH_2**	0.9621	0.9837	0.0379	0.0216	0.0163
5	**CH_3**	0.9575	0.9842	0.0425	0.0267	0.0158
6	**CN**	0.9546	0.9762	0.0454	0.0216	0.0238
7	**$COCH_3$**	0.9534	0.9872	0.0466	0.0338	0.0128
8	**CF_3**	0.9505	0.9795	0.0495	0.0290	0.0205
9	**NO_2**	0.9418	0.9737	0.0582	0.0319	0.0263
10	**$N(CH_3)_2$**	0.9318	0.9813	0.0682	0.0495	0.0187
11	**OCH_3**	0.9259	0.9459	0.0741	0.0200	0.0541
12	**trans-VINYL**	0.8592	0.9825	0.1408	0.1233	0.0175
13	**F**	0.7748	0.9825	0.2252	0.2077	0.0175
14	**cis-VINYL**	0.7465	0.9702	0.2535	0.2237	0.0298

Table 2.9. SDM results computed with B3LYP/cc-pVTZ

		S(STY,A)	*S(sty,a)*	*SDM(A)*	*SDM$_{tb}$(A)*	*SDM$_{ts}$(A)*
1	**STY**	1.0000	1.0000	0.0000	0.0000	0.0000
2	**CHO**	0.9450	0.9707	0.0550	0.0257	0.0293
3	**OH**	0.9422	0.9744	0.0578	0.0322	0.0256
4	**$COCH_3$**	0.9420	0.9766	0.0580	0.0346	0.0234
5	**NH_2**	0.9338	0.9783	0.0662	0.0445	0.0217
6	**CN**	0.9292	0.9755	0.0708	0.0463	0.0245
7	**CF_3**	0.9281	0.9686	0.0719	0.0405	0.0314
8	**NO_2**	0.9184	0.9720	0.0816	0.0536	0.0280
9	**$N(CH_3)_2$**	0.9172	0.9743	0.0828	0.0571	0.0257
10	**OCH_3**	0.9025	0.9579	0.0975	0.0554	0.0421
11	**cis-VINYL**	0.9009	0.9643	0.0991	0.0634	0.0357
12	**CH_3**	0.8129	0.9801	0.1871	0.1672	0.0199
13	**F**	0.8035	0.9814	0.1965	0.1779	0.0186
14	**trans-VINYL**	0.7614	0.9749	0.2386	0.2135	0.0251

Table 2.10. SDM results computed with MP2/cc-pVTZ

		S(STY,A)	*S(sty,a)*	*SDM(A)*	*SDM$_{tb}$(A)*	*SDM$_{ts}$(A)*
1	**STY**	1.0000	1.0000	0.0000	0.0000	0.0000
2	**OH**	0.9668	0.9865	0.0332	0.0197	0.0135
3	**CHO**	0.9662	0.9748	0.0338	0.0086	0.0252
4	**CN**	0.9488	0.9824	0.0512	0.0336	0.0176
5	**NH_2**	0.9463	0.9871	0.0537	0.0408	0.0129
6	**$COCH_3$**	0.9476	0.9871	0.0524	0.0395	0.0129
7	**NO_2**	0.9236	0.9736	0.0764	0.0500	0.0264
8	**OCH_3**	0.9174	0.9683	0.0826	0.0509	0.0317
9	**CF_3**	0.9113	0.9708	0.0887	0.0595	0.0292
10	**trans-VINYL**	0.8889	0.9836	0.1111	0.0947	0.0164
11	**$N(CH_3)_2$**	0.8415	0.9712	0.1585	0.1297	0.0288
12	**F**	0.7681	0.9818	0.2319	0.2137	0.0182
13	**cis-VINYL**	0.7651	0.9754	0.2349	0.2103	0.0246
14	**CH_3**	0.7585	0.9842	0.2415	0.2257	0.0158

2.6.1. Comments of the results of shape deviation measures (SDM)

In all three types of calculations, HF, DFT, and MP2, the general trend shows a strongly increased relative weight of the through-bond contribution as the total shape deviation measure increases (from top to bottom in all three tables). The three methods provide fairly consistent general trends.

One systematic method-related observation is that the DFT method gives higher shape-modifying results at the low end of the scale, and slightly smaller shape deviation measures at the high end of the scale than the other two methods, HF and MP2. Thus, the DFT method appears to "shrink the numerical range" of calculated shape deviation measures, relative to the other two methods. Whereas this may be a more realistic scale, than one may expect for HF, however, if the trends are similar for all three methods, then MP2 and even HF may provide some increased sensitivity for shape variations, hence an easier detection of weak trends, than DFT.

When considering individual substituents, several interesting trends have been identified.

Not surprisingly, both the *cis* and *trans* vinyl groups have large effects, and both total effects are very strongly dominated by the through-bond component. Whereas some dominance is expected for such conjugation-capable groups, the extent of the overwhelming through-bond dominance is far stronger than expected. All three methods also place fluorine among the strong shape-modifiers, where, again, the through-bond component is very dominant. The OCH_3 group is also among the strong shape-modifiers, although the MP2 method gives a somewhat lower result than the other two methods.

The methyl group CH_3 is a special case. Whereas the HF results, missing some important components of the electron correlation, apparently underestimate the hyperconjugation contribution to the through-bond component of the CH_3 group, resulting in low total and low through-bond values, this is essentially corrected by the other two methods, DFT and MP2, better representing a large component of the correlation effects.

The CHO group is another special case. Although according to all three methods, the shape deviation caused by this group is relatively small, nevertheless, according to the DFT and MP2 methods, representing electron correlation, the CHO group does not follow the general trend in terms of the two components of the shape deviation measures. As shown by both DFT and MP2 calculations, the through-space contribution is greater than the through-bond contribution, and even the HF result does not show the otherwise usual large difference between these two contributions.

The OCH_3 group shows another anomaly. Whereas, according to all three methods, the shape-modifying effect of the OCH_3 group is in the medium-strong range, the through-bond and through-space components are more similar in magnitudes than in most other substituents, and for HF, actually, the through-space contribution appears far stronger than the through-bond contribution, possibly exaggerated by the lack of inclusion of electron correlation effects in HF. Nevertheless, the OCH_3 group is a strong candidate, according to all three methods, as a substituent to induce a large through-space effect, even in competition with through-bond effects. This feature of the OCH_3 group, may

serve in future molecule design projects if an enhancement of through-space effects is desired.

Another observation that may have relevance in future molecule design projects is the result that places the OH group consistently among the weak shape modifiers. Whereas the OH group, as a functional group, may well serve as a substituent for ensuring the success of many different reactions, the fact that it does not seem to induce large shape changes in the rest of the molecule, in our case, in the *para* vinyl group, may provide advantages in future molecule-design projects.

2.7. Conclusions

The conclusions of this study, some of them elaborated in some detail for specific molecules in the previous section, can be summarized as follows:

1. Both through-bond and through-space interactions can be studied by applying the local shape analysis methods and exploiting the fragment density construction method of the ADMA method.[96]
2. Shape analysis provides numerical measures to express the relations between through-space and through-bond effects and these can be compared and correlated with experimental properties and reactivity expectations.

3. Whereas through-bond interactions typically dominate, the shape similarity analysis is detailed enough so it is possible to identify and analyze through-space interactions as well, which are important in early stages of bimolecular reactions.
4. The magnitudes of shape changes, caused by both of the components of interactions, correlate well with experimental expectations of substituent effects, with additional detail provided about the two types of causes, the through-space and through-bond components of these changes.
5. Since the electronic cloud contains all information on all properties of molecules, this also applies to components of interactions, furthermore, as demonstrated by the examples studied, the interactions between functional groups are also well represented by the local electron density.
6. By a specific fragmentation and shape analysis, the interplay of local and global interactions can also be used to predict some molecular properties not expected earlier, when predictions were based only on the dominant interaction.
7. The magnitude of the relative roles of components had to be reconsidered, since through-bond interactions, in general, dominate far more than conventionally expected, even for systems of conjugated π bonds. This is likely to lead to some modifications of traditional explanations for some molecular interactions.
8. For several substituents, the analysis suggests potential uses of similar analyses in molecular design projects when the enhancement or suppression of the relative components of the interactions is desired, with strong indications that such analysis may contribute to nanostructure design.

2.8. Appendix A

2.8.1. Brief Review of Molecular Shape and Similarity Analysis by the Shape Group Method

For fuzzy objects with no precise boundaries, such as molecular electron density clouds, the shape and similarity analysis is based on the complete, infinite series of isodensity contours for the full range of electron density threshold values **a**, using curvature-based geometrical characterization for a whole range of reference curvature parameters b. By formally truncating each contour according to a curvature criterion defined by b, and sweeping through the full range of these two continuous parameters, a and b, a finite number of topological equivalence classes of truncated surface are obtained, each of which can be characterized by a finite number of integers. The ranks of their algebraic-topological homology groups, called Betti numbers are related to the number of holes in each truncated contour. This generates a numerical shape code for each, complete, fuzzy electron density cloud. For the shape similarity measures of two electron densities, a direct comparison of these shape codes is used. This shape similarity measure avoids the (often ambiguous) "optimal" superposition requirement of the molecules needed in similarity evaluation by other methods.

This approach accomplishes two things: it represents geometrical similarity as topological equivalence, an important simplification without missing essential shape information, and reduces the shape description and similarity evaluation to a comparison of numerical shape codes. A full description of the Shape Group Method

and related shape similarity evaluation can be found in ref. [68]; here we review only the essential of details found therein.

2.8.2. The Shape Group Methods

For a nuclear configuration K and density threshold a, the *molecular isodensity contour surface,* MIDCO G(K,a) is defined as

$$G(K,a) = \{ \mathbf{r} : \rho(K,\mathbf{r}) = a \} \quad (1)$$

Set G(K,a) is the collection of all points **r** where the electronic density $\rho(K,\mathbf{r})$ is equal to the given threshold value a.

The original concept of local convexity at any given point **r** of a surface is relative to "curvature" b=0, that is, relative to a tangent plane of no curvature. Accordingly, the MIDCO surface at point **r** may be regarded as a function defined over the tangent plane at **r**, allowing the two eigenvalues of the second derivative matrix, a formal Hessian matrix of this function, to be computed. The number of those eigenvalues that are less than b is denoted by μ. If b=0, then μ = 0, or 1, or 2 at point **r** corresponding to the surface being locally concave, or saddle type, or convex, respectively. However, such characterization does not give much detail and discrimination, if only b=0 is considered providing only a crude shape characterization. But if the tangent plane is replaced by a tangent sphere of radius 1/b, and the b reference curvature can be changed within a wide enough range, then the concepts of local convexity (local concavity and local saddle-shape) can be generalized as "convexity in a curved world" of curvature b, and for a range of b values, a far more detailed shape characterization is obtained. A formal negative b value indicates that the tangent sphere of radius 1/|b| is approaching the surface point

r "from the inside" of the MIDCO. That is, if a whole range of b values, including negative values is considered, than a detailed shape characterization becomes possible.

For any given b value, the tangent object T (plane or sphere) at point **r**, may fall locally on the outside, or on the inside, or it may cut into the given MIDCO surface near enough to point **r** . By carrying out this characterization for all points **r** of the MIDCO, one obtains a subdivision of the molecular contour surface into locally convex, locally concave, and locally saddle-type shape domains, denoted by D_2, D_0, and D_1, respectively, which are the *local relative convexity domains* of the MIDCO relative to the reference curvature b. Formally, this can be carried out for each density threshold a and each reference curvature b of the complete fuzzy electron density cloud.

We still have a continuum of MIDCO surfaces, one for each density threshold a, and the local curvatures of each are characterized relative to infinitely many curvature values b. We still deal with a continuum, and the problem of fuzzy object shape characterization has not been simplified.

If, however, we decide to truncate each MIDCO and eliminate either one of the three types of the D_0, or D_1, or D_2 *local relative convexity domains* D_μ, then for each truncation type $\mu = 0, 1, 2$, relative to b, we may obtain a set of new, topologically more interesting objects, a set of truncated contour surfaces $G(K,a,\mu)$ which inherit some essential shape information from the original MIDCO surface $G(K,a)$. It is important to note, however, that many of the $G(K,a,\mu)$ truncated surfaces are topologically equivalent providing an important simplification for shape analysis.

In fact, for the entire range of a and b values, we obtain only a finite number of topologically different truncated surfaces, with various number of holes in them, in various arrangements. That is, we have obtained a finite number of topological equivalence classes, which are much simpler to characterize by the tools of algebraic topology, than the continuum of the original set of infinitely many contour surfaces.

For such a topological description, each G(K,a) is characterized by its *shape groups* [68] denoted by $H^p_\mu(a,b)$, which are algebraic groups for each dimension p, p=0,1,2. These are the homology groups of a truncated version of the given MIDCO G(K,a). The truncation is determined by local shape properties expressed in terms of the reference curvature value b and by the choice of $\mu = 0, 1,$ or 2, indicating that the locally concave D_0, or the locally saddle type D_1, or the locally convex D_2 regions (relative to b) are truncated from G(K,a).

By definition,[68] *geometrical curvature conditions* lead to truncation, and all essentially equivalent truncations provide a *topological classification of shapes* expressed by the shape groups of the original MIDCO G(K,a), and the algebraic-topological homology groups $H^p_\mu(a,b)$ of the truncated surfaces $G(K,a,\mu)$. For each reference curvature b, shape domain and truncation pattern μ of a given MIDCO G(K,a) of density threshold a, there are three shape groups, $H^0_\mu(a,b)$, $H^1_\mu(a,b)$, and $H^2_\mu(a,b)$. In general, these shape groups $H^p_\mu(a,b)$ are topological invariants for each dimension p, p=0,1,2. The formal dimensions p of these three shape groups are zero, one, and two, collectively expressing the essential shape information of the MIDCO. The $b^p_\mu(a,b)$ Betti numbers are the ranks of the $H^p_\mu(a,b)$ shape groups. These ranks are integer numbers, used to generate numerical shape codes for molecular electron density distributions. Accordingly, for each (a,b) pair of parameters and for each

shape domain truncation type μ, there are three Betti numbers, $b^0_\mu(a,b)$, $b^1_\mu(a,b)$, and $b^2_\mu(a,b)$.

The outlined shape group method of topological shape description combines the advantages of geometry and topology, and follows the spirit of the GSTE principle: Geometrical Similarity is treated as Topological Equivalence. The local shape domains and the truncated MIDCO's $G(K,a,\mu)$ are defined in terms of geometrical classification of points of the surfaces using local curvature properties. In turn, the truncated surfaces $G(K,a,\mu)$ are characterized topologically by the shape groups and their Betti numbers, themselves topological invariants.

2.8.3. Numerical Shape Codes: the (a,b)-Parameter Maps

The most important Betti numbers, conveying the chemically most relevant shape information, are those of type $b^1_\mu(a,b)$ for shape groups of dimension 1, hence in most applications, only these are used.

Since the shape groups and the Betti numbers are invariant within small ranges of both parameters a and b, in practice, it is sufficient to consider only a finite set of a and b values within the chemically relevant ranges of these parameters. Since both a and b may vary several orders of magnitude and since electron density decreases exponentially at larger distances from the nuclei, it is advantageous to use a logarithmic scale for both a and b, where for negative curvature parameters b, the $\log|b|$ values are taken. In most applications, a 41×21 grid is used for the $(\log(a), \log|b|)$ pairs, where the ranges are [0.001, 0.1 a.u.] (a.u. = atomic unit) for the density threshold values a, and [-1.0, 1.0] for the reference curvature b.

The grid points of the corresponding 41×21 (a,b)-map contain the Betti numbers obtained for the grid values of the parameters a and b. For a given pair of (a,b) values, the truncated MIDCO may become disconnected, then two or more Betti numbers are obtained, and the complete set of them, encoded in a single number, is assigned to the same location of the (a,b)-map. Consequently, the (a,b)-map becomes a 41×21 matrix $\mathbb{M}^{(a,b)}$, and this matrix is a *numerical shape code* for the fuzzy electron density of the molecule.

2.8.4. Shape Similarity Measures from Shape Codes

The associated numerical shape-similarity measure s(A,B) is also based on the primary molecular property, electron density,

For a similarity evaluation of two molecules, A and B, both in some fixed nuclear configuration, we use their shape codes in their matrix forms, $\mathbb{M}^{(a,b),A}$ and $\mathbb{M}^{(a,b),B}$, respectively. The numerical shape similarity measure is defined as

$$s(A,B) = m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}] / t \qquad (2)$$

where $m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}]$ is the number of Betti number matches between corresponding elements in the two matrices, and t is the total number of elements in either matrix. With the usual 41×21 grid, where the grid divisions are $n_a = 41$ and $n_b = 21$, one has

$$t = n_a n_b = 861 \qquad (3)$$

Based on a fuzzy density fragmentation approach[96] (see also Appendix B), the same method is applicable for the description and similarity analysis of shapes of molecular fragments.

Such numerical shape similarity measures s(A,B) have been successfully tested and used in studies of Quantitative Shape Activity Relations (QShAR) and predictive correlations, as well as tools for the interpretation of local molecular properties of functional groups and molecular fragments.[97-108]

2.9. Appendix B

2.9.1. Generation of fuzzy electron densities for molecular fragments

The shape similarity comparisons of vinyl groups required the generation of fuzzy fragment densities, which are fully analogous with fuzzy densities of complete molecules. For this purpose, the AFDF (Additive Fuzzy Density Fragmentation) approach has been used, which also forms the basis of the linear scaling ADMA (Adjustable Density Matrix Assembler) macromolecular quantum chemistry method.[96]

With reference to n atomic orbitals $\varphi_i(\mathbf{r})$ (i=1,2,...,n) in an LCAO *ab initio* wavefunction of a molecule, where **r** is the position vector variable, and P is the corresponding $n \times n$ density matrix, the the electronic density $\rho(\mathbf{r})$ of the molecule is given by

$$\rho(\mathbf{r}) \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}\, \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \qquad (9)$$

Given the collection of nuclei of the molecule present in a fragment k of the molecule, then the fuzzy fragment electron density $\rho^{k}(\mathbf{r})$ of the fragment is obtained from the electron density $\rho(\mathbf{r})$ of the complete molecule, using the following definition for the elements P^{k}_{ij} of the $n \times n$ fragment density matrix:

P^{k}_{ij} = P_{ij} if both $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are AO's centered on one or two nuclei of the fragment,

= $0.5\ P_{ij}$ if precisely one of $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ is centered on a nucleus of the fragment,

= 0 otherwise,

following an approach analogous to Mulliken's population analysis, without integration to obtain charge. The fuzzy electron density of fragment k is defined as

$$\rho(\mathbf{r}) \sum_{i=1}^{n} \sum_{j=1}^{n} P^{k}_{ij}\, \varphi_i(\mathbf{r})\, \varphi_j(\mathbf{r}) \qquad (10)$$

Note that, if the nuclei of the molecule are partitioned into m mutually exclusive families to generate m fragments, then the sum of the fragment density matrices is the density matrix of the molecule, and the sum of the fragment densities is the density of the molecule, that is, the scheme is fully additive:

$$P_{ij} \sum_{k=1}^{m} P^{k}_{ij}$$

(11)

and

$$\rho(\mathbf{r}) \sum_{k=1}^{m} \rho^{k}(\mathbf{r})$$

(12)

These simple fragment additivity rules are *exact* and providing fuzzy fragment densities, allows one to use the same shape analysis and similarity methods as those developed for complete molecules.

2.10. Appendix C. Computational Methods and Resources

The initial electron density computations for the complete molecules and the IM models have been performed by the Gaussian 09 software package,[93] using geometrically optimized structures. For a pictorial representation of some of the complete molecules and IM fragment pairs, the Gaussview program was used.[93]

In order to generate the IM fragments, we have cut out the aromatic ring, followed by adding H atoms or CH_3 groups to the fragments to avoid chemically incorrect structures. The lengths of the new carbon-hydrogen bonds were set to the experimental average[95] of 109 pm and the new carbon-carbon or carbon-heteroatom distances were set to 150 pm for chemical consistency of the models. In the case of para-methoxy styrene,

the methoxy part needed a methyl group, whilst the vinyl part needed only a hydrogen atom.

After the Gaussian computations, the resulting .log files and electron density data were used as input for our shape analysis software package [94] to calculate the local shape representation of the vinyl group of each molecule, generating the numerical shape codes following the methodology described in detail in [68,82-87] (resulting in the standard four significant figures when shape similarities are expressed as similarities, matches-mismatches in the shape codes). In this way, the local shape characteristics of the vinyl groups are expressed using numbers as they are influenced by the rest of the molecule, most importantly by the functional group A in the *para* position.

Finally, the shape similarities of all the local electron densities of all vinyl groups are compared, including the shape of the full molecule vinyl group and the shape of the fragment vinyl group for the same parent molecules. As an illustration for the complete set of shape similarity results, the HF results with a large correlation consistent (cc-PVTZ) basis set are summarized in Table 2.1.

2.11. Acknowledgements

This study has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program (CRC) at the Memorial University of Newfoundland. We also thank the Atlantic Computational Excellence Network (ACEnet) Atlantic Division for computer resources and the invaluable help of the Department of Chemistry.

2.12. References

1. E. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*; University Science Books, 2006.
2. L. P. Hammett, *J. Am. Chem. Soc.*, 1937, **59**, 96.
3. R. D. Topsom, *Electronic Substituent Effects in Molecular Spectroscopy,* in *Progress in Physical Organic Chemistry*, Volume 16, ed. R. W. Taft, Hoboken, NJ, USA, 2007.
4. R. W. Taft, *J. Am. Chem. Soc.*, 1952, **74**, 2729.
5. R. W. Taft, *J. Am. Chem. Soc.*, 1952, **74,** 3120.
6. R. W. Taft, *J. Am. Chem. Soc.*, 1953, **75**, 4538.
7. W.F. Reynolds, P.G. Mezey, and G.K. Hamer, *Can. J. Chem*., 1977, **55**, 522.
8. P.G. Mezey and W.F. Reynolds, *Can. J. Chem*., 1977, **55**, 1567.
9. W.F. Reynolds, T.A. Modro, P.G. Mezey, E. Skorupowa, and A. Maron, *Can. J. Chem*., 1980, **58**, 412.
10. W.F. Reynolds, T.A. Modro, and P.G. Mezey, *J. Chem. Soc. Perkin II*, 1977, 1066.
11. L. Amat, R. Carbó-Dorca, R. Ponec, *J. Med. Chem*., 1999, **42**, 5169.
12. R.Carbó, L.Leyda, M. Arnau, *Int. J. Quantum Chem.,* 1980, **17**, 1185.
13. R. Carbó and M. Arnau, *Molecular Engineering: A General Approach to QSAR,* in *Medicinal Chemistry Advances,* ed. F.G. de las Heras and S. Vega, Pergamon Press, Oxford, 1981.
14. R. Carbó and B. Calabuig, *J. Chem. Inf. Comput .Sci*., 1992, **32**, 600.
15. R. Carbó and B. Calabuig, *Int. J. Quantum Chem.,* 1992, **42**, 1681.
16. R. Carbó and B. Calabuig. *Int. J. Quantum Chem.,* 1992, **42**, 1695.
17. R. Carbó, B. Calabuig, L. Vera, and E. Besalú, *Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques*, in *Advances in Quantum Chemistry*, Vol. 25, ed. P.-O. Löwdin, J.R. Sabin, M.C. Zerner, Academic Press, New York, 1994.

18. R. Carbó, Ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer Academic Publ., Dordrecht, 1995.
19. E. Besalú, R. Carbó, J. Mestres, and M. Solà, *Foundations and Recent Developments on Molecular Quantum Similarity*, in *Topics in Current Chemistry, Vol. 173, Molecular Similarity*, ed. K. Sen, Springer-Verlag, Heidelberg, l995.
20. R. Carbó-Dorca and E. Besalú, *J. Mol. Struct. (THEOCHEM)*, 1998, **451**, 11.
21. P. Constans, Ll. Amat, and R. Carbó-Dorca, J. Comput. Chem. 1997, **18**, 826.
22. P. Hohenberg and W. Kohn, *Phys. Rev*. 1964, **136**, B864.
23. P.G. Mezey, *Mol. Phys*. 1999, **96**, 169.
24. P.G. Mezey, *J. Chem. Inf. Comp. Sci.*, 1999, **39**, 224.
25. P.G. Mezey, *J. Math. Chem.*, 2001, **30**, 305.
26. P.G. Mezey, *J. Math. Chem.*, 2001, **30**, 299.
27. K. D. Sen, E. Besalú and R. Carbó-Dorca, *J. of Math. Chem.*, 1999, **25**, 253.
28. L. Amat, R. Carbó-Dorca, and R. Ponec, *J. Med. Chem*., 1999, **42**, 5169.
29. R Carbó-Dorca, E. Besalú, X. Gironés, *Adv. in Quantum Chem.*, 2000, **38**, 1.
30. L. Amat, E. Besalú, R. Carbó-Dorca, R. Ponec, *J. Chem. Inf. Comput. Sci.*, 2001, 41, 978.
31. X. Gironés, L. Amat, R. Carbó-Dorca, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 847.
32. R. Carbó-Dorca, E. Besalú, *J. Math. Chem.*, 2010, **49**, 836.
33. R. Carbó-Dorca and E. Besalú, *J. Comput. Chem.*, 2010, **31** 2452.
34. R. Carbó-Dorca and E. Besalú, *J. Math. Chem.*, 2010, **48**, 914.
35. E. Besalú and R. Carbó-Dorca, *J. Chem. Theory Comput.*, 2012, **8**, 854.
36. E. Besalú, R. Carbó-Dorca, *J. Math. Chem.*, 2010, **48**, 914.
37. R. Carbó-Dorca, E. Besalú, *J. Math. Chem.*, 2011, **49**, 2244.
38. R. Carbó-Dorca, *J. Math. Chem.*, 2012, **50**, 2339.

39. R. Carbó-Dorca, E. Besalú, *J. of Math. Chem.*, 2013, **51**, 660.

40. P. W. Ayers, *J. Chem. Phys.* 2000, **113**, 10886.

41. A. Ghosh, *Theor. Chem. Acc.*, 2000, **104**, 157.

42. P. Geerlings, F, De Proft, and W. Langenaeker, *Chem. Rev*., 2003, **103**,1793.

43. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, *J. Phys. Chem. A*, 2003, **107**, 11120.

44. F. De Proft, P. W. Ayers, K. D. Sen, and P. Geerlings, *J. Chem. Phys.* 2004, **120**, 9969.

45. P. Geerlings, G. Boon, C. Van Alsenoy, F. De Proft, *Int. J.Quant. Chem,* 2005, **101**, 722.

46. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, *J. Mol. Struct. (THEOCHEM),* 2005, **727**, 49.

47. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, *J. Phys. Chem. A*, 2006, **110** 5114.

48. S. Janssens, G. Boon, and P. Geerlings, *J. Phys. Chem.* A, 2006, **110** 9267.

49. P. Geerlings, F. De Proft, and P.W. Ayers, *Theor. and Comput. Chem.*, 2007, **19**, 1.

50. S. Janssens, C. Van Alsenoy, and P. Geerlings, *J. Phys. Chem*. A, 2007, **111**, 3143.

51. S. Janssens, A. Borgoo, C. Van Alsenoy, and P. Geerlings, *J. Phys. Chem.* A, 2008, **112**, 10560.

52. E. Debie, L. Jaspers, P. Bultinck, W. Herrebout, and B. Van Der Veken, *Chem. Phys. Lett.*, 2008, **450**, 426.

53. E. Debie, P. Bultinck, W. Herrebout and B. van der Veken, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3498.

54. S. Janssens, P. Bultinck, A. Borgoo, C. Van Alsenoy, and P. Geerlings, *J. Phys.*

Chem. A, 2010, **114**, 640.

55. N. Nikolova and J. Jaworska, *QSAR Comb. Sc.,* 2003, **22**, 1006.
56. V. Monev, *MATCH-Commun. Math. Comput. Chem*, 2004, **51**, 7.
57. P. J. Smith and P. L. A. Popelier, .*Org. Biomol. Chem*., 2005, **3**, 3399.
58. B. Galdino de Oliveira, R. de Cássia Maritan Ugulino de Araújo, A. Bezerra de Carvalho, and M. Neves Ramos, *Acta Chim. Slov*. 2009, **56,** 704.
59. R. O. Esquivel, N. Flores-Gallegos, J. S. Dehesa, J. Carlos Angulo, J. Antolin, S. Lopez-Rosa and K. D. Sen, *Theor. J. Phys. Chem.,* A, 2010, **114**, 1906.
60. V. V. Turovtsev and Yu. D. Orlov, *Russ. J. Phys. Chem.,* A, 2010, **84**, 965.
61. V. V. Turovtsev and Yu. D. Orlov, *Russ. J. Phys. Chem.,* A, 2010, **84**, 1174.
62. Š. Varga, *J. of Math. Chem.,* 2011, **49,** 1.
63. M. Pinsky, D. Danovich, and D. Avnir, *The J. Phys. Chem.,* C, 2010, **114**, 20342.
64. A. Zayit, M. Pinsky, H. Elgavi, C. Dryzun, and D. Avnir, *Chirality*, 2011,**23**, 17.
65. P.G. Mezey, *Potential Energy Hypersurfaces*, Elsevier, Amsterdam, 1987.
66. R.J.Gillespie and R.S.Nyholm, *Quarterly.Rev.*, 1957, **11**, 339.
67. R.J.Gillespie, *J. Chem. Educ.* 1970, **47**, 18.
68. P.G. Mezey, *Shape In Chemistry: An Introduction To Molecular Shape and Topology*, VCH Publishers, New York, 1993.
69. F. Bernardi, A. Bottoni, M. Olivucci, J. J. W. McDouall, M. A. Robb, G. Tonachini, *Theochem-J. of Mol. Struc.* 1988, **42**, 341.
70. M. Bearpark, F. Bernardi, M. Olivucci, M. A. Robb, *J. Am. Chem. Soc*., 1990, **112**, 1732.
71. F. Bernardi, M. Olivucci, M.A. Robb, *Acc. of Chem. Res.*, 1990, **23**, 405.
72. F. Bernardi, P. Celani, M. Olivucci, M. A. Robb, G. Suzzi-Valli , *J. Am. Chem. Soc*., 1995, **117**, 10531.
73. F. Bernardi, A. Bottoni, M. Olivucci, A. Venturini and M. A. Robb, *J. the Chem. Soc.-Faraday Transactions*, 1994, **90**, 1619.
74. P.G. Mezey, *Adv. Molec. Structure Res.* 1998, **4**, 115.
75. P.G. Mezey, *New Developments in Three-Dimensional Molecular Topology, in*

"Chemical Topology: Introduction and Fundamentals", Eds., D. Bonchev and D.H. Rouvray, Gordon and Breach Science Publishers, Amsterdam, 1999.

76. P.G. Mezey, *Lecture Series on Computer and Computational Sciences*, 2006, **7**, 1511.
77. P.G. Mezey, *Fuzzy Electron Density Fragments as Building Blocks in Crystal Engineering Design in "The Importance of Pi-Interactions in Crystal Engineering (Frontiers in Crystal Engineering 3)",* eds. E. Tiekink and J. Zukerman-Schpector, John Wiley & Sons., New York, 2012.
78. P. Geerlings and A. Borgoo, *Phys. Chem. Chem. Phys.*, 2011, **13**, 911.
79. S. Kullback and R. A. Leibler, *Ann. Math. Stat.*, 1951, **22**, 79.
80. P. W. Ayers, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1959.
81. Eusebio Juaristi, *Introduction to Stereochemistry and Conformational Analysis*, Wiley, New York, 1991.
82. P.G. Mezey, *J. Quantum Chem., Quant. Biol. Symp*., 1986, **12**, 113.
83. P.G. Mezey, *J. Comput. Chem*., 1987, **8**, 462.
84. P.G. Mezey, *Int. J. Quantum Chem., Quant. Biol. Symp*., 1987, **14**, 127.
85. G.A. Arteca, V.B. Jammal, P.G. Mezey, J.S.Yadav, M.A.Hermsmeier, and T.M.Gund, *J. Molec. Graphics*, 1988, **6**, 45.
86. G.A. Arteca, V.B. Jammal, and P. G. Mezey, *J. Comput. Chem*., 1988, **9**, 608.
87. G.A. Arteca and P.G. Mezey , *Int. J. Quantum Chem., Quant. Biol. Symp*., 1988, **15**, 33.
88. B. Galabov, V. Nikolova, S. Ilieva, *Chemistry - A European Journal*, 2013, **19**, 5149
89. E. G. Hohenstein, C. D. Sherrill, *WIREs Comput Mol Sci*, 2012, **2**, 304.
90. J. M. Daniel, S. D. Friess, S. Rajagopalan , S. Wendt, and R. Zenobi, *Int. J. Mass Spec*. 2002, **216**, 1.
91. S. L. Cockroft, C. A. Hunter, *Chem. Soc. Rev*., 2007, **36**, 172.
92. F. Jensen, *Introduction to Computational Chemistry*. 1999, Chichester: Wiley.
93. Gaussian 09, 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, 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

94. Shape Analysis and Similarity Program “Rhocalc” and support software. (Scientific Modeling and Simulation Laboratory, Memorial University of Newfoundland, Canada, paul.mezey@gmail.com).
95. CRC. *Handbook of Chemistry and Physics, 65th ed* ed. R. C. Weast, CRC Press, Boca Raton, 1984.
96. P.G. Mezey, *Int. J. Quantum Chem.* 1997, **63**, 39.
97. G.A. Heal, P.D. Walker, M. Ramek, and P.G. Mezey, *Canad. J. Chem*.,1996, **74**, 1660.
98. P.G. Mezey, Z. Zimpel, P. Warburton, P.D. Walker, D.G. Irvine, D.G. Dixon, and B. Greenberg, *J. Chem. Inf. Comp. Sci*., 1996, **36**, 602.
99. P.G. Mezey, *Local Shape Analysis of Macromolecular Electron Densities, in "Computational Chemistry: Reviews and Current Trends*, Vol.1", Ed. J. Leszczynski, World Scientific Publ., Singapore, 1996.
100. Z. Zimpel and P.G. Mezey, *Int. J. Quantum Chem*.,1996, **59**, 379.
101. P.G. Mezey, *Adv. in Quan. Chem.*,1996, **27**, 163.

102. P.G. Mezey, *Fuzzy Measures of Molecular Shape and Size, in "Fuzzy Logic in Chemistry"*, Ed. D.H. Rouvray, Academic Press, San Diego, 1997.

103. P.G. Mezey and P.D. Walker, *Drug Discovery Today* (Elsevier Trend Journal),1997, **2**, 6.

104. P.G. Mezey, Z. Zimpel, P. Warburton, P. D. Walker, D.G. Irvine, X.-D. Huang, D. G. Dixon, and B.M. Greenberg. *Toxicol. Chem*.,1998, **17**, 1207.

105. P.G. Mezey, *Molecular Structure - Reactivity - Toxicity Relationships, in "Soil Chemistry and Ecosystem Health"*, Ed. P.M. Huang, SSSA Publ., Pittsburgh, PA, USA, 1998.

106. P.G. Mezey, *Adv. in Mol. Sim.*,1998, **2**, 79.

107. P.G. Mezey, *Theor. and Comput. Chem*.,1999, **6**, 593, Chapter 23 in "*Pauling's Legacy: Modern Modelling of the Chemical Bond*", Eds. Z. Maksic and W.J. Orville-Thomas, Elsevier Science Publ., Amsterdam, 1999.

108. P.G. Mezey, *Quantitative Shape - Activity Relations (QShAR), Molecular Shape Analysis, Charge Cloud Holography, and Computational Microscopy, in "Quantitative Structure-Activity Relationships for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications*", Ed. J.D. Walker, SETAC (Society of Environmental Toxicity and Chemistry) Publ., SETAC Press, 2003.

The following chapter is an exact copy of a published scientific article. Its format, including sub-chapters and reference list, adopts the format required by the journal in which it was published. The parameters of the article are as follows:

Authors: Zoltan Antal and Paul G. Mezey

Title: Substituent effects and local molecular shape correlations

Journal: Physical Chemistry Chemical Physics (*Phys. Chem. Chem. Phys*.)

Year, volume, and first page number: 2014, 16, 6666.

Note: some modifications to the original articles were necessary to successfully incorporate them into the main thesis body.

CHAPTER 3

Substituent effects and local molecular shape correlations

Zoltan Antal and Paul G. Mezey

Scientific Modeling and Simulation Laboratory (SMSL), Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, Newfoundland A1B3X7

Abstract: Using a detailed electron density shape analysis methodology, a new method is proposed for studying the main components of substituent effects in a series of di-substituted benzenes, in correlation with their activating and deactivating characteristics as monitored by the induced shape changes of a local electron density cloud. The numerical measures obtained for the extent of shape changes can be correlated with known and with some unexpected effects of various substituents. The insight obtained from the shape analysis provides a theoretical, electron density based justification for some well-known trends, but it also provides new explanations for some of the unexpected features of these substituent effects.

3.1. Introduction

Systematic studies of substituent effects on aromatic rings and on the directing power of various substituents to *ortho*-, *meta*-, or *para*-positions have been among the early successes of the classical, interpretative models of organic chemistry. Energetic considerations and traditional models involving "resonance effects" and alternating charge accumulation along atoms of aromatic rings have provided attractive, although somewhat simplistic theoretical explanations for the interpretation of such substituent effects.[1-5]

These early advances [5] have been followed by a large number of more detailed studies involving various quantum chemistry methods. In this context, besides the computational methodology advances which have made possible a more rigorous basis for interpretation, the approaches to a quantum-chemical analysis of molecular similarity, suitable for numerically evaluating even small variations of intramolecular effects, have provided an important framework.[6-17] These methods have allowed a detailed, quantum chemistry based comparison of various effects of substituents within molecules, as manifested in the changes in the structures and electron distributions of molecules. The actual intramolecular effects of substituents can be monitored by investigating and using numerical measures to evaluate molecular similarities, hence the differences induced by substituent effects.

Naturally, molecular similarity problems, including those manifested by substituent effects on aromatic rings, are intimately connected to some of the fundamental questions of quantum chemistry.[19-25] In particular, advanced, interpretative density functional approaches[26-54] have also contributed to both the computational and the interpretative aspects of possible quantum chemistry studies of both global and local effects within molecules, also applicable to substituent effects.

Some of the early studies on intramolecular interactions, without explicit use of similarity measures, have focussed on the interpretation and the evaluation of the dominant energy components influencing reactivities in various molecules, involving, for example, aromatic hydrocarbons.[55-59]

However, as advancement of quantum chemistry computational methodologies, as well as interpretative approaches have progressed, more precise and more detailed studies have become possible. One of the more recent related developments[60] involves electron density shape analysis of both complete molecules and individual molecular fragments, providing a new approach for the analysis of various components of the interactions within molecules. This recent study also provides a review of some other, recent approaches as well as a comprehensive, detailed characterization of the technical aspects of the methodologies used in the shape analysis process. All of the methodological information needed to follow details of the present study are described in sufficient detail in that recent reference,[60] hence, here we shall not repeat those descriptions and only the most essential aspects of the shape analysis methods will be recalled in this paper.

3.1.1. The substituent effect problem phrased in terms of electron density shape modifying effects in a family of model molecules

In the present study, the focus is placed on local electron density shape changes in correlation with the experimentally established trends of substituent effects, as manifested by the *ortho*-, *meta*-, and *para*-directing features of a series of typical substituents on a series of model molecules containing an aromatic ring. The conclusions reached using these model compounds, a series of di-substituted benzenes, appear general enough to be useful in the interpretation of aromatic substituent effects in a wider range of aromatic compounds.

In this study, the choice of model molecules has been influenced by several factors. In order to avoid potentially confusing influences of several side-factors, compounds with a single aromatic ring and with only two substituents were examined. For the theoretical basis and technical aspects of the electron density shape analysis and similarity evaluation methodology employed, see a brief description below and a comprehensive review of all the relevant methodologies in the Appendices A, B, and C by Antal,[60] published in this same journal (not repeated here), or the detailed original literature references. Using this shape analysis and similarity evaluation methodology, the effects of various substituents are monitored by the shape changes of a molecular fragment that is common in all molecules studied ("same" only by chemical formula, but actually of a different electron density shape in each molecule, due to the various influences of the very substituent effects studied). For a better detection of the influences of the substituents, it is advantageous, if the induced shape changes in this "common" molecular

fragment are as large as possible. The vinyl group, as this "common" fragment, appears as a good choice. This is due to its relatively mobile, hence easily deformable electron density cloud related to the its π-system, where shape changes induced by the other substituents are likely to be transmitted well and even enhanced by the conjugative aromatic ring. In order to avoid the possibly confusing effects of simple steric hindrances due to crowding, the substituents studied have been placed in the *para*- position relative to the vinyl group, even those substituents which are known experimentally to favor *meta*- positions. In the latter case, the "disadvantage" of the *para*- position is expected to provide important shape-modifying effects on the vinyl group.

The simplest reference for a shape description and shape comparisons for the vinyl groups in various molecules is a choice where this group is affected only by the ring, and there's no "functional group" in the para-position. This reference is the vinyl group of styrene; therefore, technically, the formal functional group in this case is represented by the hydrogen atom. The shapes of all other vinyl groups in molecules which have a "real" functional group in the *para*- position are compared primarily to this reference vinyl group, and in each case the magnitude of shape change should directly reflect the effect of the functional group. In addition, the shapes of all vinyl groups are compared to one another as well, in order to provide additional information, especially, with respect to distinguishing between activating and deactivating groups; however, our primary concern has been the shape changes relative to the reference vinyl group in the "unsubstituted" styrene.

In Figure 3.1 the reference molecule, styrene is shown, where the "second substituent" is actually the *para*- position H atom of styrene, relative to the vinyl group. For all other molecules studied, the effect of the *para*- substituent is monitored by the electron density shape changes of the "common" vinyl group.

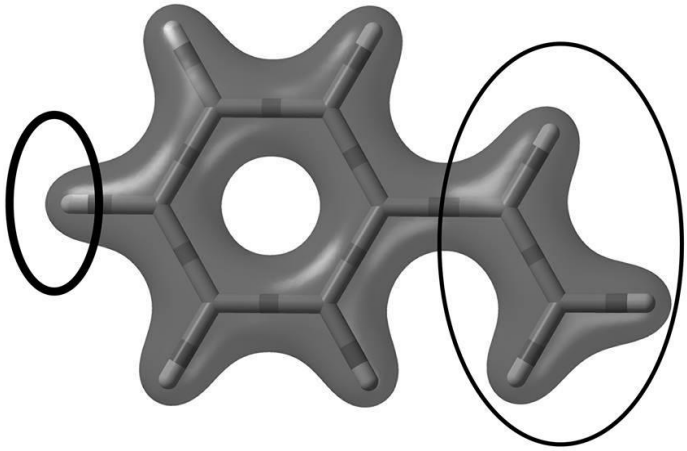


Figure 3.1. Schematic representation of the simplest of di-substituted benzenes used as reference in our study. Here the "functional group" in *para* position is hydrogen, shown in the thick circle (on the left), and the actual target of shape analysis, the vinyl group is shown in the thin circle (on the right).

Figure 3.2. presents a schematic model of styrene as the reference molecule, in comparison with *para*-nitro styrene. In styrene, the vinyl part is only influenced by the ring (in this case the *para*-functional group is a hydrogen). If, however, a *para*-functional group is added (the nitro group in the example above), the electron density of the vinyl part is affected in a unique way. Comparing the shapes of the local electron densities of these vinyl groups, and evaluating the relevant numerical shape-similarity measure, and comparing these shape similarity measures for each of the vinyl groups for the whole sequence of all *para*-substituents, the shape-influencing effects of all these substituents can be monitored. As follows from the holographic electron density theorem,[51] these shape changes are all unique, furthermore, in principle they contain information on all aspects of the differences of intramolecular substituents effects for each of the substituents studied.

The di-substituted benzenes used in our present study were *para*- substituted styrenes with the following *para*- substituents, totaling 15 functional groups:

CH_3, CHO, CF_3, *cis*-vinyl, Cl, CN, $COCH_3$, F, H (in styrene, reference), $N(CH_3)_2$, NO_2, NH_2, OCH_3, OH, *trans*-vinyl.

This series of molecules studied has covered a wide range of substituents which are able to manifest most of typical effects on the benzene ring.

One should emphasize that no two vinyl groups in this series of molecules (or any other molecules) can be exactly the same, not even by accident, as follows clearly from the holographic electron density theorem:[51-53] in the electronic ground state, any small, but positive volume part of the electron density cloud contains all information about the complete molecule. That is, an electron density fragment can be extended only in one, unique way to a complete molecule, that also implies, that no two vinyl groups, in two different molecules, can ever be exactly the same. Indeed, each vinyl group is an actual "fingerprint" of the complete molecule it belongs to, and as a consequence, it fully monitors all aspects of the intramolecular interactions generated by the actual substituent. ***This observation is the fundamental basis of the approach employed: the vinyl group shape changes necessarily reflect all aspects of the substituent effects within each molecule.***

3.2. Computational Methodology

In order to avoid potential biases of the currently used, most popular quantum chemistry methodologies as much as possible, and to obtain reliability assurances, three different levels of quantum chemical methodologies have been applied. Using the Gaussian 09 software family,[61] the optimized geometries of our target molecules were obtained using the HF (Hartree-Fock),[62] Density Functional Theory (B3LYP functional),[63] and post Hartree-Fock theory represented by the MP2 (second-order Moller-Plesset)[64] model. All of these theoretical models were combined with Dunning's correlation-consistent basis set, cc-pVTZ.[65] The resulting .log files were used as input for the electron density shape analysis software family [66]. For the interested reader, very detailed descriptions of the methodology and algorithmic details for each of the software used can be found in several references.[68-75] In particular, the local electron density approaches and the relevant shape analysis methods have been reviewed very recently by Antal,[60] and here only a very brief description will be given.

3.2.1. Relevant aspects of shape analysis methodology

Traditionally, molecular shape is often associated with molecular conformations, where the nuclear framework is considered to represent shape, and the emphasis is on energy optimized molecular structures, represented by minima of potential energy hypersurfaces.[76] In reality, molecular shape is better represented by the shape of the

electronic density clouds,[70-75] and this is especially important when considering intramolecular, (as well as intermolecular) interactions.

The shape analysis and shape similarity measures of molecular electron density clouds involve some non-trivial problems, such as the quantum mechanical uncertainty relation and the very fact that electron densities have no boundaries, and even within a semi-classical model, electron densities are fuzzy, cloud-like objects. Therefore, some of the macroscopically useful shape analysis methods of everyday objects are not applicable directly, especially, geometrical methods are of limited use. One of the generalizations of geometry that appears particularly useful is topology (connected but far more general than graph theory), where the fuzzy, "boundariless" nature of electron densities can be described with mathematical precision, in a way reminiscent to "rubber geometry" suitable for the study of shapes of flexible structures, such as molecules with mobile, vibrating structures.

The topological method developed for this purpose is the Shape Group Method, based on algebraic topology, using a group-theoretical framework for a detailed shape analysis of fuzzy molecular electron density clouds.[70-75]

We give here only a brief summary of the most essential aspects of the method. For details, the reader should consult a very recent, rather detailed and comprehensive review in the Appendices of reference [60] in this very journal (PCCP), or the original references [70-75] including a book written on the subject.[75]

The electron density shape analysis method is applicable for molecules of practically any size, however, for large molecules the computations needed for shape analysis can become rather time consuming, but more importantly, for large molecules, having far more chemical details, the interpretation of global shape differences becomes blurred. Therefore it is advantageous to focus the shape analysis on certain molecular fragments of interest. For the typical molecular fragments used in chemically relevant comparisons, the computational cost of shape analysis is only a small fragment of the cost of the quantum chemical electron density calculations. The shape analysis is based on the collection of all isodensity contours, where the "chemically most relevant" electron density range is taken with cut-off values of 0.001 and 0.1 a.u. (a.u. = atomic unit), and where the local "extreme" curvature values considered are +1.0 and -1.0, where the sign changes refer to concave or convex local shapes.

The electron density $\rho(r)$ of a given molecular conformation is regarded as a function of the position vector **r**. For a threshold value a, the molecular isodensity contour, MIDCO, is the surface $G(a) = \{\mathbf{r}: \rho(\mathbf{r}) = a\}$, that is, the collection of all points **r** where the electron density is equal to the value a. The Shape Group Method of molecular shape analysis uses *two* continuous parameters: the threshold a and parameter b, the reference curvature value against which the local curvatures along every point of each molecular isodensity contour MIDCO G(a) is tested. A local tangent plane is taken at each point **r** of G(a), then the MIDCO G(a) is viewed as a two-variable function defined over this tangent plane. This function is differentiable, and the extreme local curvatures of G(a) at point **r** are the eigenvalues **v** and **w** of the second derivative

(curvature) matrix (a formal Hessian matrix) of this function. Contour G(a) is locally convex relative to reference curvature b at point **r** if both **v** and **w** are less than the reference curvature b. If this happens, this point **r** is formally deleted from the contour G(a). If this test is performed for each point of G(a), then for each (a, b) pair, a possibly truncated isocontour surface is obtained, some with several holes in it. Taking all such possibly truncated surfaces for the whole range of available a and b values, a detailed shape characterization of the complete molecular electron density is obtained.

Note that both parameters a and b are continuous and infinitely many truncated isocontours can be obtained. However, among them there are only a finite number of topologically different types. The topological characterization of these types (topological equivalence classes) by their one-dimensional homology groups simplifies the problem considerably without losing essential shape information. The ranks of these groups, the so-called Betti numbers, are a finite set of integers, which provide a *numerical shape code* of the complete electron density cloud. The actual shape codes used are numerical codes for the Betti numbers arranged in a matrix. In most applications, a 41×21 matrix is used for the (log(a), log|b|) pairs, where the ranges are [0.001, 0.1 a.u.] (a.u. = atomic unit) for the density threshold values a, and [-1.0, 1.0] for the reference curvature b. These shape codes for a finite set of a, b pairs of values are comparable directly, and one obtains a simple, *numerical shape similarity measure,* defined as the ratio of matches between the numerical values in the corresponding two 41×21 matrices.

One should note that this shape similarity measure does not require any "superposition" of the molecules compared, hence avoiding any possible ambiguity associated with such superpositions.

Figure 3.3 presents the isodensity contours of styrene (a), *para*-fluoro styrene (b), and cis-vinyl styrene (c) at 0.015 a.u. of electron density, and their calculated shape similarity values (all relative to the reference vinyl group of styrene). The shape analysis method used provides numerical shape similarity measures, an important advantage over visual observations, as evidenced by the small, visually often ambiguous or imperceptible vinyl group shape differences shown in the figure below. These shape differences are due to the presence of the *para* functional groups in the latter two molecules, as these functional groups alter, among other shape features, the electron cloud around the vinyl group at the far end of the aromatic ring. The shape similarity values are obtained by the actual comparison of the numerical shape codes of the local vinyl electron densities in these molecules. The reference vinyl group in styrene has a self-similarity value of 1.0000. The vinyl group in *para*-fluoro styrene has a similarity value at 0.7748, and in cis-vinyl styrene, it has similarity value of 0.7476 when compared to the reference. These calculations were based on Hartree-Fock/cc-pVTZ geometries and electron densities.

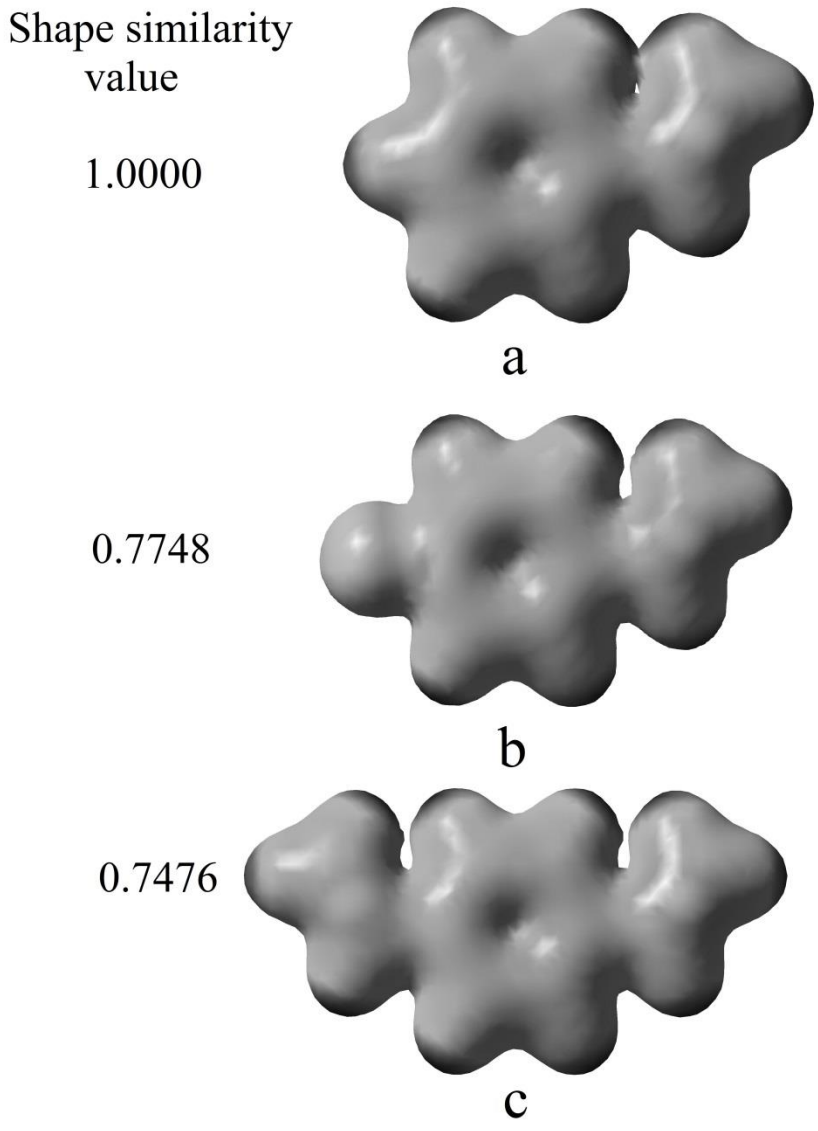


Figure 3.3. The electron density contours of a) styrene, b) *para*-fluoro styrene and c) cis-vinyl styrene at isovalue 0.015 a.u. Visual inspection provides only a very crude and hardly reproducible evaluation of shape differences and similarities among the vinyl groups; whereas the top figure, of styrene, appears rather different from the other two, the latter two are hardly distinguishable visually. By contrast, the shape analysis and the numerical similarity measures (in this case, relative to the vinyl group of styrene), provide a far more detailed and also far more reliable approach, as validated in several earlier studies.[70-75] The attachment of a functional group in *para* position alters the electronic cloud everywhere, also around the vinyl group, and these changes are monitored by the shape analysis approach. These illustrative examples use Hartree-Fock geometries and electron densities.

3.3. Results and discussion

It is well known from classical experimental data that functional groups can manifest different types of electronic effects when attached to a benzene ring. By interacting through the sigma framework and the π-system, the functional groups manifest activating/deactivating powers and *ortho*-, *meta*-, or *para*-directing abilities when different sites on the benzene ring are analyzed and the site of an electrophilic attack is studied. The relative activating/deactivating powers (typically monitored by effects of *para*-substitution) fall into weak, moderate and strong categories and a single hydrogen on the unsubstituted benzene ring is used as a reference. When a vinyl group in the un-substituted styrene is used as a reference, any shape change will be the result of the functional group itself, as transmitted through the intermediate part of the molecule.

3.3.1. The similarity matrix

Although our primary interest has been the vinyl group shape changes relative to the "unsubstituted" styrene molecule, that reflects the substituent effects relative to this "unsubstituted" molecule, nevertheless, the shape analysis and similarity evaluation has been carried out for all molecules studied, in order to provide all pair similarities as well, not only those involving the vinyl group of styrene. In these *similarity matrices* (one for each of the HF, DFT and MP2 results) one is able to compare any two vinyl groups directly, for example, one can compare vinyl group shape differences relative to the case of the most activating substituent, in our sample, the hydroxyl group.

The collection of numerical similarity values, forming a similarity matrix (SM), contains all similarity measures obtained for the given geometries as calculated with Hartree-Fock, B3LYP, and MP2 levels of theory. There is a specific similarity matrix for each level of theory (Hartree-Fock, B3LYP, MP2). Each matrix can be analyzed separately and the results form the basis of our analysis. Every SM contains a block of numbers, each representing the collection of vinyl group similarity values of all molecules with one another.

As outlined briefly above, when describing the similarity matrix and its properties, each molecule has a shape similarity value of 1.0000 to itself, representing its self-similarity. This holds also for the vinyl group of the reference molecule, hence, if all other vinyl groups are tested against this reference, then deviations from this maximum value of 1.0000 will show the shape differences from the vinyl group of this reference molecule. Large deviations from the ideal, perfect shape-match of 1.0000 mean large shape difference between the corresponding vinyl electron density clouds.

In Tables 3.1-3.3, the similarity matrices with all pairwise shape similarity values for all vinyl groups obtained at all three levels of theory are presented. The first numerical column contains the most relevant similarity values, corresponding to shape comparisons to the reference vinyl group of styrene; a choice motivated by the traditional interpretations of substituent effects, with reference to unsubstituted aromatic rings (here the vinyl group can be regarded as the "monitoring tool" of influences of the *para* substituents).

	STY	F	OH	CHO	CN	NH_2	NO_2	CF_3	$N(CH_3)_2$	OCH_3	$COCH_3$	CH_3	cis VINYL	tra VINYL	Cl
STY	1	0.7748	0.9622	0.9662	0.9546	0.9621	0.9418	0.9505	0.9318	0.9259	0.9534	0.9575	0.7465	0.8592	0.9551
F	0.7748	1	0.7596	0.7605	0.7719	0.7693	0.7608	0.7809	0.764	0.7637	0.757	0.7619	0.816	0.7942	0.7657
OH	0.9622	0.7596	1	0.9569	0.9505	0.9586	0.9331	0.9395	0.947	0.9125	0.9481	0.9686	0.743	0.861	0.9394
CHO	0.9662	0.7605	0.9569	1	0.9575	0.9481	0.9453	0.9464	0.929	0.9277	0.9545	0.957	0.7363	0.867	0.9476
CN	0.9546	0.7719	0.9505	0.9575	1	0.9499	0.9692	0.9529	0.9359	0.9317	0.9528	0.9563	0.7313	0.8575	0.9615
NH_2	0.9621	0.7693	0.9586	0.9481	0.9499	1	0.9476	0.9569	0.9387	0.9404	0.9702	0.9575	0.7526	0.852	0.9603
NO_2	0.9418	0.7608	0.9331	0.9453	0.9692	0.9476	1	0.9581	0.9203	0.9412	0.9645	0.9384	0.7401	0.8454	0.9773
CF3	0.9505	0.7809	0.9395	0.9464	0.9529	0.9569	0.9581	1	0.9255	0.9411	0.9557	0.9401	0.7433	0.8546	0.9598
$N(CH_3)_2$	0.9318	0.764	0.947	0.929	0.9359	0.9387	0.9203	0.9255	1	0.9033	0.9282	0.9593	0.7351	0.8633	0.9243
OCH_3	0.9259	0.7637	0.9125	0.9277	0.9317	0.9404	0.9412	0.9411	0.9033	1	0.9428	0.9191	0.736	0.845	0.9441
$COCH_3$	0.9534	0.757	0.9481	0.9545	0.9528	0.9702	0.9645	0.9557	0.9282	0.9428	1	0.9516	0.7561	0.8515	0.9691
CH_3	0.9575	0.7619	0.9686	0.957	0.9563	0.9575	0.9384	0.9401	0.9593	0.9191	0.9516	1	0.7343	0.8676	0.9494
cis VINYL	0.7465	0.816	0.743	0.7363	0.7313	0.7526	0.7401	0.7433	0.7351	0.736	0.7561	0.7343	1	0.7447	0.7485
tra VINYL	0.8592	0.7942	0.861	0.867	0.8575	0.852	0.8454	0.8546	0.8633	0.845	0.8515	0.8676	0.7447	1	0.8475
Cl	0.9551	0.7657	0.9394	0.9476	0.9615	0.9603	0.9773	0.9598	0.9243	0.9441	0.9691	0.9494	0.7485	0.8475	1

Table 3.1. The similarity matrix corresponding to the Hartree-Fock level of theory.

	STY	F	OH	CHO	CN	NH_2	NO_2	CF_3	$N(CH_3)_2$	OCH_3	$COCH_3$	CH_3	cis VINYL	tra VINYL	Cl
STY	1	0.7716	0.9200	0.9398	0.8972	0.9163	0.9058	0.9102	0.9088	0.8932	0.9163	0.7971	0.8743	0.7989	0.9281
F	0.7716	1	0.7832	0.7792	0.7780	0.7762	0.7678	0.7829	0.7721	0.7509	0.7792	0.9067	0.7758	0.8624	0.7791
OH	0.9200	0.7832	1	0.9159	0.9167	0.9179	0.9132	0.9150	0.9044	0.8795	0.9341	0.7916	0.8595	0.7813	0.9300
CHO	0.9398	0.7792	0.9159	1	0.9147	0.8960	0.9094	0.9265	0.8959	0.8839	0.9046	0.7901	0.8625	0.7870	0.9275
CN	0.8972	0.7780	0.9167	0.9147	1	0.9028	0.9120	0.9313	0.9067	0.8836	0.9096	0.7617	0.8716	0.7822	0.9323
NH_2	0.9176	0.7762	0.9179	0.8960	0.9028	1	0.8882	0.9133	0.9033	0.8882	0.9120	0.7782	0.8929	0.8076	0.9236
NO_2	0.9058	0.7678	0.9132	0.9094	0.9120	0.8882	1	0.9126	0.8897	0.8794	0.8938	0.7867	0.8634	0.7734	0.9042
CF3	0.9102	0.7829	0.9150	0.9265	0.9313	0.9133	0.9126	1	0.9021	0.8785	0.9055	0.7698	0.8724	0.7819	0.9347
$N(CH_3)_2$	0.9088	0.7721	0.9044	0.8959	0.9067	0.9033	0.8897	0.9021	1	0.8986	0.8721	0.7876	0.8788	0.8085	0.9084
OCH_3	0.8932	0.7509	0.8795	0.8839	0.8836	0.8882	0.8794	0.8785	0.8986	1	0.8681	0.7791	0.8361	0.7690	0.8893
$COCH_3$	0.9163	0.7792	0.9341	0.9046	0.9096	0.9120	0.8938	0.9055	0.8721	0.8681	1	0.7824	0.8680	0.7667	0.9281
CH_3	0.7971	0.9067	0.7916	0.7901	0.7617	0.7782	0.7867	0.7698	0.7876	0.7791	0.7824	1	0.7819	0.8542	0.7786
cis VINYL	0.8743	0.7758	0.8595	0.8625	0.8716	0.8929	0.8634	0.8724	0.8788	0.8361	0.8680	0.7819	1	0.8088	0.8757
tra VINYL	0.7989	0.8624	0.7813	0.7870	0.7822	0.8076	0.7734	0.7819	0.8085	0.7690	0.7667	0.8542	0.8088	1	0.7896
Cl	0.9281	0.7791	0.9300	0.9275	0.9323	0.9236	0.9042	0.9347	0.9084	0.8893	0.9281	0.7786	0.8757	0.7896	1

Table 3.2. The similarity matrix corresponding to the DFT level of theory.

		STY	F	OH	CHO	CN	NH_2	NO_2	CF_3	$N(CH_3)_2$	OCH_3	$COCH_3$	CH_3	cis VINYL	tra VINYL	Cl	
	STY	1	0.7681	0.9668	0.9662	0.9488	0.9463	0.9236	0.9398	0.9119	0.9174	0.9476	0.7585	0.7651	0.8889	0.9333	
	F	0.7681	1	0.7675	0.771	0.776	0.7866	0.7737	0.7716	0.7775	0.7573	0.7778	0.9417	0.8261	0.8011	0.7741	
	OH	0.9668	0.7675	1	0.9517	0.951	0.9499	0.9235	0.9381	0.9095	0.9144	0.9429	0.7623	0.7433	0.8829	0.9374	
	CHO	0.9662	0.771	0.9517	1	0.9586	0.9469	0.9416	0.9568	0.9048	0.9115	0.961	0.7681	0.7567	0.8893	0.9526	
	CN	0.9488	0.776	0.951	0.9586	1	0.9333	0.9498	0.9801	0.9013	0.9208	0.9516	0.7769	0.7468	0.8829	0.9631	
	NH_2	0.9463	0.7866	0.9499	0.9469	0.9333	1	0.9357	0.9267	0.9191	0.9005	0.9558	0.7725	0.7515	0.8883	0.9254	
	NO_2	0.9236	0.7737	0.9235	0.9416	0.9498	0.9357	1	0.9498	0.8942	0.8887	0.9614	0.7652	0.7273	0.8769	0.9308	
	CF3	0.9398	0.7716	0.9381	0.9568	0.9801	0.9267	0.9498	1	0.8987	0.9137	0.9491	0.7769	0.7468	0.8762	0.9683	
	$N(CH_3)_2$	0.9119	0.7775	0.9095	0.9048	0.9013	0.9191	0.8942	0.8987	1	0.8951	0.9042	0.769	0.7389	0.8757	0.8921	
	OCH_3	0.9174	0.7573	0.9144	0.9115	0.9208	0.9005	0.8887	0.9137	0.8951	1	0.9005	0.7663	0.723	0.8668	0.9036	
	$COCH_3$	0.9476	0.7778	0.9429	0.961	0.9516	0.9558	0.9614	0.9491	0.9042	0.9005	1	0.7687	0.741	0.8876	0.935	
	CH_3	0.7585	0.9417	0.7623	0.7681	0.7769	0.7725	0.7652	0.7769	0.769	0.7663	0.7687	1	0.8083	0.7667	0.7732	
	cis VINYL	0.7651	0.8261	0.7433	0.7567	0.7468	0.7515	0.7273	0.7468	0.7389	0.723	0.741	0.8083	1	0.7588	0.757	
	tra VINYL	0.8889	0.8011	0.8829	0.8893	0.8829	0.8883	0.8769	0.8762	0.8757	0.8668	0.8876	0.7667	0.7588	1	0.8753	
	Cl	0.9333	0.7741	0.9374	0.9526	0.9631	0.9254	0.9308	0.9683	0.8921	0.9036	0.935	0.7732	0.757	0.8753	1	

Table 3.3. The similarity matrix corresponding to the MP2 level of theory.

3.3.2. Numerical results

The numerical results of the vinyl group shape similarity measures are given in Figures 3.4 – 3.6. The first of these, Figure 3.4 shows the shape similarity results obtained with the Hartree-Fock method, combined with a large, Dunning type correlation-consistent basis set.[65] Figure 3.5 shows the DFT results using the B3LYP as combined also with Dunning's cc-pVTZ basis set. The last figure, Figure 3.6, shows the shape similarity results obtained by the MP2 method.

Our primary concern was the shape similarity to the reference styrene. Even at first glance, the representation and ordering of substituents is listed according to their similarity values to the reference density of the vinyl group in styrene, as correlated with their experimentally established activating power. The information present in all three figures, (3.4-3.6), closely resemble the typical representation of the Hammett type substituent effect table structure (the structure only, and not the sigma constants), with a reference molecule separating the activating and deactivating groups.

In the typical case of the Hammett tables, the reference molecule is benzene, where a single hydrogen of the benzene ring is the actual reference, and in our case, the reference is also a H atom. However, a vinyl group is also attached to the ring in the *para* position, relative to this reference hydrogen.

All three of the quantum chemical computational methods, HF, DFT, and MP2, give very similar results. The only significant differences appearing in a very few molecules

for which the calculations are noticeably different. This is possibly due to some of the specific limitations of the quantum chemical electron density computational methods used, affecting the actual input information to the common shape analysis methodology[67]. One such example is the apparent limitation of the Hartree-Fock method to adequately take into account the effects of hyperconjugation which are clearly present in the methyl group. It is likely that HF as input generates a somewhat inaccurate shape analysis and shape similarity results, making the methyl group stand out from the general trends.

Before analysing the actual numerical results, we shall consider some of the general aspects of interpreting the similarity values obtained in Figures 3.4-3.6.

Consider first the upper part of the figures, relative to the property of activation, referring to *para*- and *ortho*- directing groups.

The similarities computed are those of the shapes of vinyl groups. Clearly, a vinyl group in styrene has the shape it has because this shape belongs to the most energetically stable arrangement of this vinyl group in this molecule. If, in some other molecule the same vinyl group has a very similar shape, this indicates that not much disturbance was caused by the other substituent that would otherwise force the vinyl group to take a very different, energetically less advantageous shape. Overall, one can expect that, in the *para* position, a substituent would leave the vinyl group near to its original optimum shape it would otherwise have, if the substituent favors such an arrangement by its own preferences, that is, if a *para*- position is advantageous for the substituent too, relative to the vinyl group already there. Of course, this also means, that if only the substituent

would be present, without the vinyl group, then, by the favored directing character of the substituent, it would prefer to have the vinyl group enter the *para*- position, the very position it occupies in our model. Consequently, the more pronounced the *para*-directing trend of a substituent, the more likely it is that the shape-modification of the vinyl group is minor, that is, a high level of similarity can be expected. Indeed, the shape similarity numbers clearly back this conclusion: a higher level of shape similarity is typically obtained for a stronger *para*-directing character of the substituent.

If we turn now to the lower part of the figures, to the *meta*-directing substituents and to property of deactivation, we can see the following.

Having *meta*-directing groups in the *para*-position relative to the vinyl group, is inherently less advantageous than having *para*-directing groups there and now the various levels of *disadvantages* can be compared. A minor degree of deactivation, near to "indifference", results in a minor shape modification of the vinyl group, that is, we can expect a high level of shape similarity for weakly deactivating groups, and this is indeed the case, with the exception of F, which is a special case in most respects.

Indeed, fluorine is an odd substituent for two reasons: although the F atom has high electronegativity with a strong attachment to its electrons and with a very tight electron density cloud around it, it is not able to attract an excess amount of charge because its valence shell can be filled by a single electron when compared to the atomic ground state. Nevertheless, fluorine has a highly concentrated charge present in its tight electron density cloud. It has a major effect on the overall electron density of the molecule, the

deformation of the electron density shape extending over the entire molecule, including the *para*-position vinyl group. This explains the low shape similarity value for the vinyl group electron density obtained in the case of the F atom.

When more details of the actual numerical results are examined, several interesting trends are observed.

3.3.2.1. Hartree-Fock results

Figure 3.4 depicts the shape analysis results for Hartree-Fock theory combined with a large correlation-consistent basis set. Our reference molecule is in the middle of the list, separating the two types of directing powers into two distinct groups. On top, the activating groups are lined up from top to bottom, having the strongest activators on top and the weakest on bottom. All of these groups are *ortho*- and *para*- directing groups as established by experimental results for electrophilic aromatic substitution reactions on the substituted benzene ring.[1,4,5]

As outlined in the general comments above, the shape similarity results are affected by the influences of the groups already present on the ring. The values in the lists represent the shape difference between the vinyl electron cloud of the unsubstituted styrene and the vinyl group of the substituted one which has a functional group in *para* position (relative to vinyl), in correlation with their directing effect on some incoming electrophile. When strongly activating groups are present on the same benzene ring, the weak activating power of the vinyl groups is completely dominated by strong activators[77] such as the

hydroxyl and amino groups. These groups have lone pairs of electrons which are readily shared with the benzene ring, overcoming the possible influencing effect of the "weaker" π-electrons of the vinyl group. Therefore the shape of the electron cloud around this vinyl group has not been altered significantly as the "weak" π-electrons are "suppressed" resulting in low shape difference values for strongly activating groups. As the strength of this dominance decreases in the activating groups and the ability to interact with the benzene ring of the weak vinyl group increases, the shape similarity values will decrease as well. This trend is already noticeable in the di-methyl amino and methoxy groups as the lone pairs on these groups are being shared on both sides of the functional groups and their influence on the benzene ring decreases. The largest shape difference values are obtained with functional groups in the *para* position of equal or similar strengths, such as trans- or cis-vinyl groups, relative to our target vinyl group. As mentioned before, the methyl group, which is also a weak activating group, does not line up well with this general trend and this fact is attributed to the inability of the HF method to properly describe this rather simple looking, but by all means peculiar functional group, capable of hyperconjugation.

When considering the lower part of the list, the two "abnormal" *para*-directing, but deactivating groups stand out clearly on the list. All deactivating groups are *meta*-directing, except the halogens, therefore our analysis starts with these. The shape similarity value of fluorine falls in-between that of the trans- and cis-vinyl groups, and by recalling the experimental trends of fluoro-benzene,[77] this comes as no surprise since fluorine greatly activates the *para*-position relative to its position (naturally, the

magnitude of activation is relative to the unsubstituted benzene). Furthermore, fluorine is a rather special substituent that often does not follow the trends otherwise expected for halogens, since its high electronegativity, contrasted to its limited ability to actually add much diffuse charge to its very tight electron density cloud, results in some unexpected features (also reflected in the relatively weak acidity of HF). The ability to achieve strong activation is greatly decreased in the case of chlorine, therefore the corresponding shape similarity value is much higher (as it correlates with rates of reaction for halo-substituted benzenes, F>>Cl>>Br>>I).[77]

In this example of di-substituted benzenes, we have an activating and a deactivating group on the same ring. In these cases, the activating group dominates over the effect of the deactivating group,[77] but because of the relative positions in which these groups are placed on the benzene ring, their (incoming electrophile) directing sites coincide and the substituents reinforce each other. The *meta*-position of the deactivating group coincides with the *ortho*-position of the activating group. Since the strength of the weak activating group (vinyl) does not change, and only the strength of the deactivating group changes, the result is larger and larger shape similarity differences as we move down the list of the deactivators (as their strength increases). The similarity value of the acetyl group (weak deactivator) is slightly smaller than the value of the cyano group (stronger deactivator), but the difference is almost negligible. Weak deactivators, such as aldehyde and acetyl groups result in large shape similarity values, whilst stronger deactivators, such as the cyano, trifluoro methyl, and the nitro groups result in smaller

shape similarity values. All these similarity values are relatively large, since the dominating weak activating group cannot cause large differences in shape.

3.3.2.2. DFT results

The DFT model (Figure 3.5), represented here by the B3LYP (naturally combined with Dunning's cc-pVTZ basis set) does very well and apparently, even corrects some of the shortcomings of the HF model, by adequately describing the ability for hyperconjugation of the methyl group. By this method, the similarity value of the methyl group is correctly placed among other weak activating groups, such as the trans- and cis-vinyl groups. Another difference between the HF and DFT models is the "minor" reordering of some similarity values. For example, DFT gives a larger shape similarity value for the cis-vinyl group instead of the trans-vinyl group, but this reordering does not affect the general trends described earlier. The general trend mentioned earlier, placing of strong activators (OH, NH_2) in the group of large similarity values and smaller and smaller shape similarity values are obtained for lesser activators, as we move down the list and the activating power decreases (in the order of di-methyl amino, methoxy, the vinyl, and the methyl groups).

The very same type of difference between HF and DFT is found (although with different numerical values) for the two *para*-directing deactivating groups (halogens) and the same trends are found for the deactivating groups as well. The order of the strong deactivating groups does not correlate perfectly with the results of the HF model, but

once again this difference is minor and can be attributed to typical differences in HF and DFT methods, describing electron correlation problems differently.

3.3.2.3. MP2 results

Calculations with the post Hartree-Fock method MP2 required the most computer time of all models, but the results achieved with this model correlate very well with the experimental expectations. Each functional group lines up with the general trends discussed above, with a single functional group standing out, the methoxy group having a slightly (almost negligibly) larger similarity value than the other moderately activating functional group, the di-methyl amino group (Figure 3.6). Similarly to the HF model (and dissimilarly to the DFT model), the trans-vinyl group has a larger similarity value than the cis-vinyl group.

As a conclusion, we can state that, for all practical purposes the general trends found using all three quantum chemistry models are essentially the same.

3.3.3. Additional comments about the shape analysis results

Note that, in addition to comparisons to the vinyl group of styrene, that has been our primary concern, the similarities for every molecule pair have also been evaluated and, in more general assessments of comparisons between substituent effects and local shape variations, such complete similarity matrices can be useful. For example, in the case of the DFT similarity matrix (Table 3.2), the functional groups $–NH_2$ and $–COCH_3$ have equal shape similarity values (0.9163) when compared to the reference molecule styrene.

At first glance, this seems to be misleading since the amino group is a strong activator and an *ortho*/*para*-directing group, whilst the acetyl group is a moderately deactivating, *meta*-directing group. These results mean only that they are about equally different in shape from the reference vinyl group of styrene, and this does not imply that they should be highly similar to each other. Indeed, if the two functional groups are directly compared with each other, one will find their similarity value is far from a perfect match of 1.0000, rather, the actual similarity value obtained is 0.9120 which is actually less than their similarity values when compared to styrene. This result is in good accord with their different, activator – deactivator characteristics: their shapes are more different from each other than from the reference.

Another, potentially useful series of data that can be extracted from each similarity matrix is relevant to another choice for the reference molecule. Our choice for using styrene as reference has been based of the fact that this molecule has an "unaltered" aromatic ring with an added vinyl group, but one should not exclude other molecules as candidates for reference. For example, if we happen to choose a reference vinyl group from a molecule that is maximally deactivating within a molecular family, then (by analogy with using the "absolute zero" of the Kelvin scale instead of the Centigrade scale with some intermediate temperature as reference), some relations and correlations may appear simpler, even if such a scale is seldom if ever used in conventional interpretations of substituent effects. Nevertheless, such an "absolute" scale may have some advantages which could be explored in some future research.

Figure 3.4. The results of the HF model. The values represent similarity differences between the vinyl group of the reference molecule styrene and the vinyl group of *para* substituted styrenes. Large deviations from 1.0000 mean large shape differences between these vinyl groups.

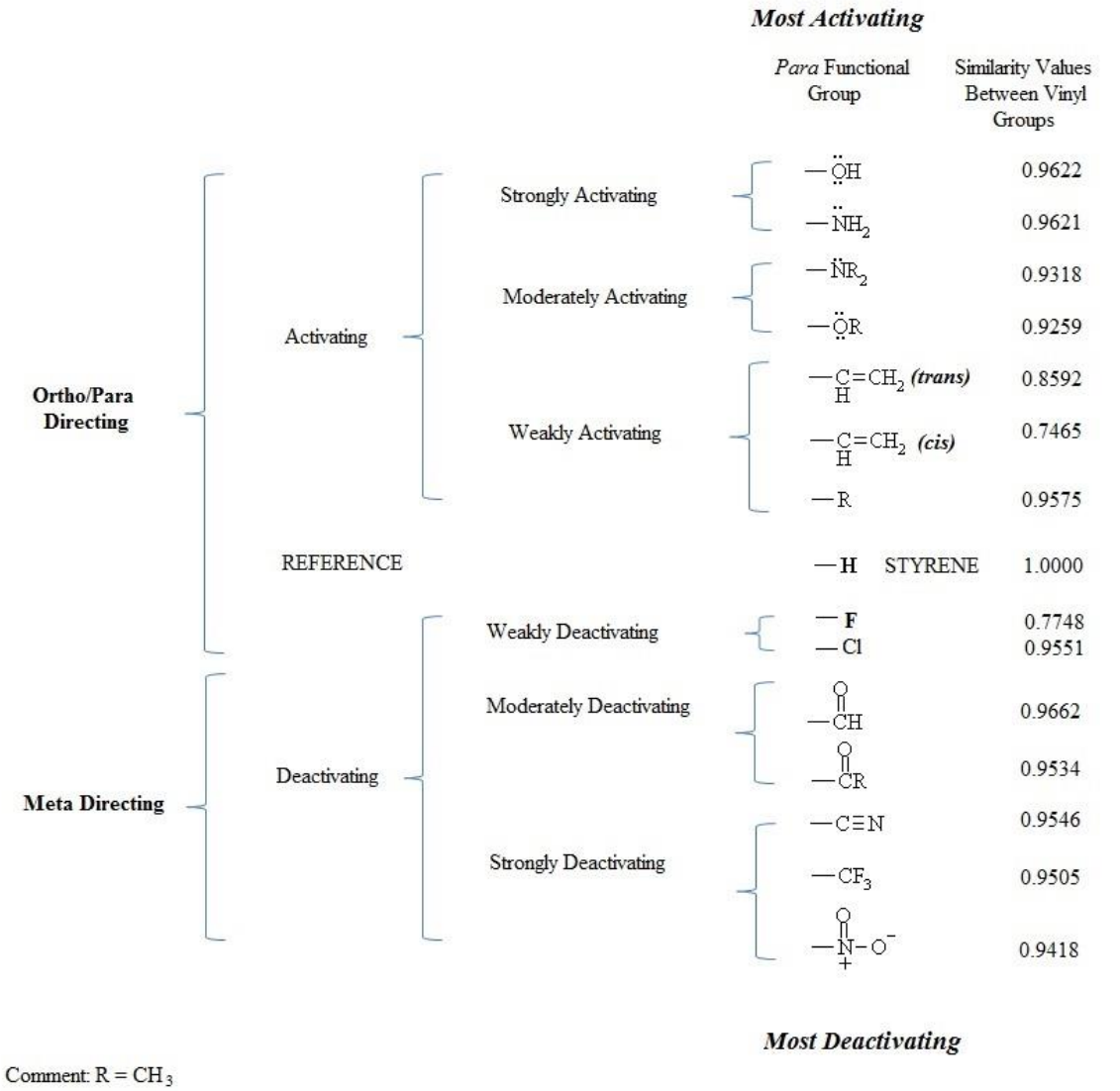


Figure 3.5. The results of the DFT model. The values represent similarity differences between the vinyl group of the reference molecule styrene and the vinyl group of *para* substituted styrenes. Large deviations from 1.0000 mean large shape differences between these vinyl groups.

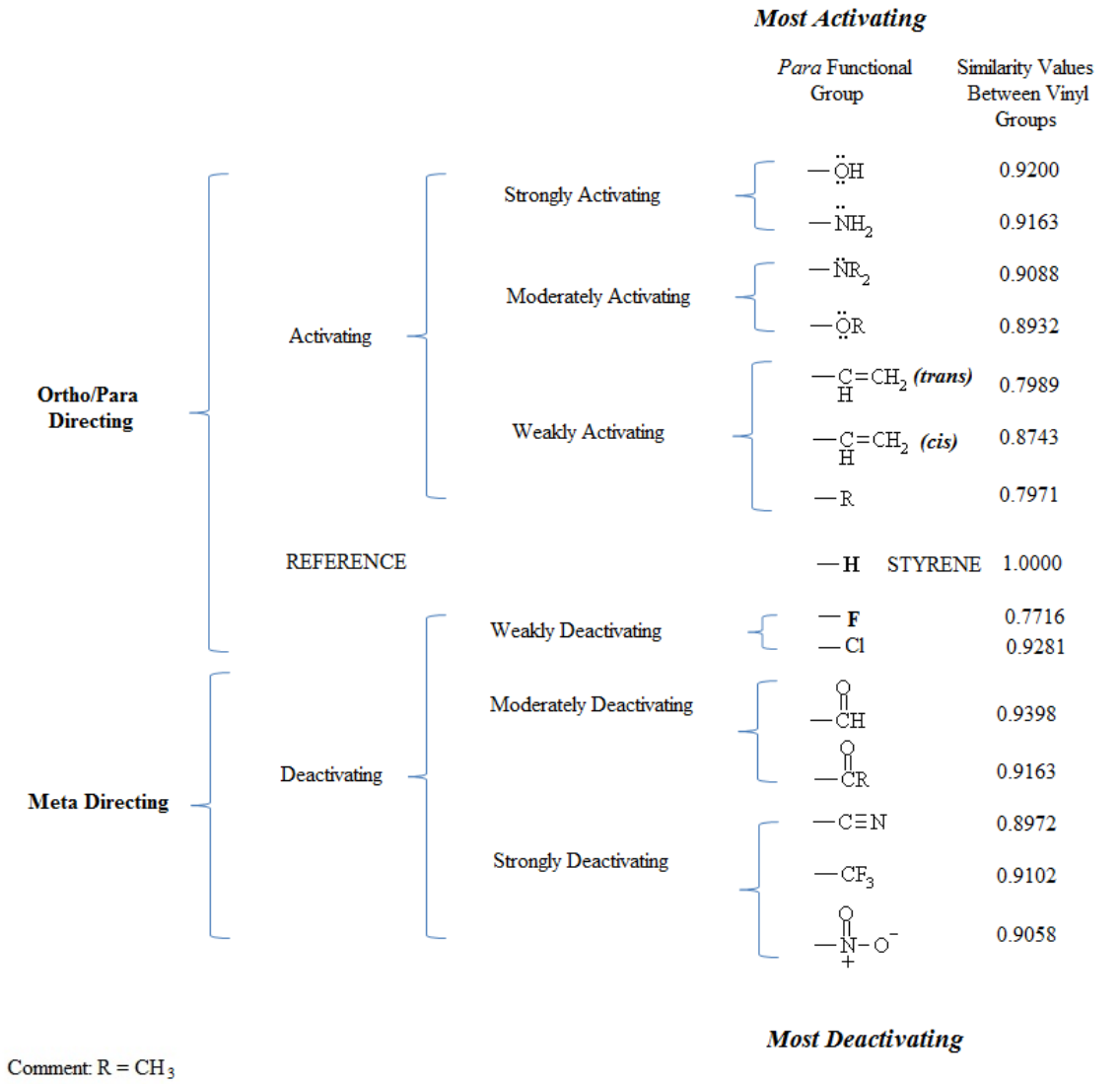
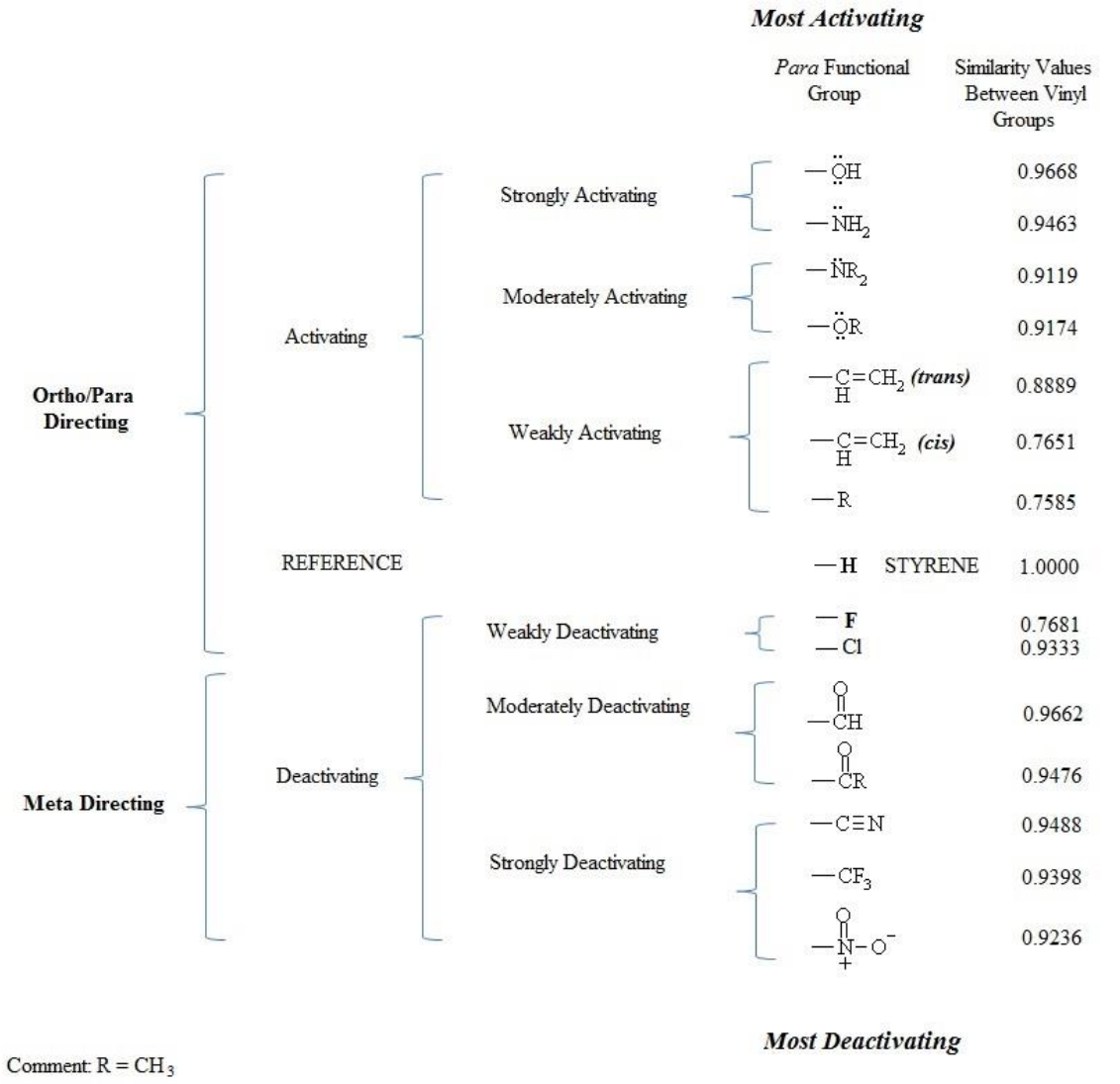


Figure 3.6. The results of the MP2 model. The values represent similarity differences between the vinyl group of the reference molecule styrene and the vinyl group of *para* substituted styrenes. Large deviations from 1.0000 mean large shape differences between these vinyl groups.



3.4. Conclusions

The shape – substituent effect correlations obtained from local electron density shape analysis and the evaluation of local shape similarity measures reflecting the effects of *para*- substituents on the shape details of a vinyl group, provide revealing correlations among the various substituent effects studied for a sequence of 15 typical substituents on aromatic rings, with various degrees of activating power and *para*- directing effects in aromatic substitution reactions.

A quantum chemical, electron density based justification is obtained for the already known, experimentally established trends in substituent effects, in particular, for the various degrees of activating and deactivating powers of substituents on aromatic rings. Besides the expected trends, some additional results provide new suggestions, for instance on how to select substituents for a desired effect, a potentially useful approach in modern, molecular design projects.

Our main problem has been to compare the magnitudes of substituent-induced shape changes, monitored by the computed similarity measures, relative to the usual reference, the ”no-further-substituent-modified” vinyl group of styrene. However, the shape similarity between any two vinyl groups associated with two different substituents can also be computed by the same technique, if this is of interest. To illustrate this, we have included complete similarity matrices for the full set of molecules (with all pair-similarities), which also provide potential suggestions for alternative scales for the study of substituent effects, such as a scale where the vinyl group of the most deactivating

molecule is used for comparison. Typically, the similarity between two vinyl groups, one belonging to an activating, the other to a deactivating substituent, is smaller than the similarity of either of these to the reference vinyl group of styrene. This is exactly what one would expect for a pair of vinyl groups, one with an activating, the other with a deactivating substituent: larger shape difference between them, than either of the their shape differences from the reference vinyl group in styrene. In general, the shape modifying aspects of such substituents are expected to be formally "opposites" on the vinyl group.

Since the method is able to generate similarity values not only with respect to a single reference molecule, but for all molecule pairs, the approach can provide information for detailed predictive correlations.

The numerical results obtained provide an electron density based scale that is more detailed and can be easily extended for any number of additional substituents, beyond the usual, experimentally derived, empirical scales used for the interpretation, and prediction of substituent effects.

In particular, if a comparison or possible prediction is needed for the potential substituent effect for a novel substituent in a series of compounds, then the method suggested here can be used to find predictive correlations, based on the already available information, and the computed shape- similarity results, including the already known, and the potentially new substituent.

3.5. Acknowledgements

This study has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program (CRC) at the Memorial University of Newfoundland. We also thank the Atlantic Computational Excellence Network (ACEnet) Atlantic Division for computer resources and the invaluable help of the Department of Chemistry.

3.6. References

1. L. P. Hammett, The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, *J. Am. Chem. Soc.*,1937, **59** , 96.

2. C. Hansch , A. Leo , S. H. Unger , K. H. Kim , D. Nikaitani , E. J. Lien, Aromatic substituent constants for structure-activity correlations, *J. Med. Chem.*, 1973, **16**, 1207.

3. A. Domenicano, P. Murray-Rust, Geometrical substituent parameters for benzene derivatives: inductive and resonance effects, *Tetrahedron Lett.*, 1979, **20**, 2283.

4. R. D. Topsom, *Electronic Substituent Effects in Molecular Spectroscopy, in Progr. in Phys. Org. Chem.*, Volume **16**, ed. R. W. Taft, Hoboken, NJ, USA, 2007.

5. R. D. Topsom, The Nature and Analysis of Substituent Electronic Effects, in Progress in Physical Organic Chemistry, Volume **12**, ed. R. W. Taft, John Wiley & Sons, New York, 2009.

6. R. Carbó, L. Leyda, and M. Arnau, How similar is a molecule to another? An electron density measure of similarity between two molecular structures, *Int. J. Quantum Chem*., 1980, **17**, 1185.

7. R. Carbó and M. Arnau, Molecular Engineering: A General Approach to QSAR, in *Med. Chem. Adv.*, F.G. de las Heras and S. Vega, Eds., Pergamon Press, Oxford, 1981.

8. R. Carbó and B. Calabuig, Quantum Similarity Measures, Molecular Cloud Descriptors and Structure-Property Relationships, *J. Chem. Inf. Comp. Sci.*, 1992, **32**, 600.

9. R. Carbó and B. Calabuig, Molecular quantum similarity measures and n-dimensional representation of quantum objects. I. Theoretical foundations, *Int. J. Quantum Chem*., 1992, **42**, 1681.

10. R. Carbó and B. Calabuig, Molecular quantum similarity measures and n-dimensional representation of quantum objects. II. Practical applications, *Int. J. Quantum Chem*., 1992, **42**, 1695.

11. R. Carbó, B. Calabuig, L. Vera, and E. Besalu, Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques, in *Adv.*

in Quantum Chem., Vol. **25**, P.-O. Löwdin, J.R. Sabin, and M.C. Zerner, Eds., Academic Press, New York, 1994.

12. R. Carbó, Ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer Academic Publ., Dordrecht, 1995.

13. E. Besalú, R. Carbó, J. Mestres, and M. Solà, Foundations and Recent Developments on Molecular Quantum Similarity, in Topics in Current Chemistry, Vol. **173**, *Mol. Sim.*, ed. K. Sen, Springer-Verlag, Heidelberg, l995.

14. R. Carbó-Dorca and E. Besalú, A general survey of Molecular Quantum Similarity, *J. Mol. Struct. (THEOCHEM)*,1998, **451**, 11.

15. P. Constans, Ll. Amat, and R. Carbó-Dorca, Toward a Global Maximization of the Molecular Similarity Function: Superposition of Two Molecules, *J. Comput. Chem.*1997, **18**, 826.

16. M. A. Johnson, G. M. Maggiora, *Concepts and applications of molecular similarity*, Wiley, Minneapolis, 1990.

17. R. Carbó-Dorca, X. Girones, and P.G. Mezey (editors) *Fundamentals of Molecular Similarity*, Kluwer Academic/Plenum Publ., New York, USA, 2001.

18. K. D. Sen, E. Besalú and R. Carbó-Dorca, A naive look on the Hohenberg–Kohn theorem, *J. of Math. Chem.*, 1999, **25**, 253.

19. Ll. Amat, R. Carbó-Dorca, and R. Ponec Simple Linear QSAR Models Based on Quantum Similarity Measures, *J. Med. Chem*., 1999, **42**, 5169.

20. R. Carbó-Dorca, E. Besalú, and X. Gironés Extended density functions, *Adv. in Quant. Chem.*, 2000, **38**, 1.

21. Ll. Amat, E. Besalú, R. Carbó-Dorca, and R. Ponec, Identification of Active Molecular Sites Using Quantum-Self-Similarity Measures, *J. Chem. Inf. Comput. Sci*., 2001, **41**, 978.

22. X Gironés, Ll. Amat, and R. Carbó-Dorca, Modeling Large Macromolecular Structures Using Promolecular Densities, *J. Chem. Inf. Comput. Sci*., 2002, **42**, 847.

23. R. Carbó-Dorca and E. Besalú, Mathematical aspects of the LCAO MO first order density function (4): a discussion on the connection of Taylor series expansion of electronic density (TSED) function with the holographic electron density theorem (HEDT) and the Hohenberg-Kohn theorem (HKT), *J. of Math. Chem.*, 2010, **49**, 836.

24. R. Carbó-Dorca and E. Besalú, Communications on quantum similarity (2): A geometric discussion on holographic electron density theorem and confined quantum similarity measures. *J. Comp. Chem.*, 2010, **31** 2452.

25. R. Carbó-Dorca and E. Besalú, A Gaussian holographic theorem and the projection of electronic density functions into the surface of a sphere**,** *J. Math. Chem.*, 2010, **48**, 914.

26. P. W. Ayers, Atoms in molecules, an axiomatic approach. I. Maximum transferability, *J. Chem. Phys.*, 2000, **113**, 10886.

27. P. Geerlings, F. De Proft, and W. Langenaeker, Conceptual Density Functional Theory, *Chem. Rev*., 2003, **103**, 1793.

28. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, Similarity and Chirality: Quantum Chemical Study of Dissimilarity of Enantiomers, *J. Phys. Chem. A*, 2003, **107**, 11120.

29. F. De Proft, P. W. Ayers, K. D. Sen, and P. Geerlings, On the importance of the "density per particle" (shape function) in the density functional theory, *J. Chem. Phys.*, 2004, **120**, 9969.

30. P. Geerlings, G. Boon, C. Van Alsenoy, F. De Proft, Density functional theory and quantum similarity, *Int. J. of Quant. Chem.*, 2005, **101**, 722.

31. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, Molecular quantum similarity of enantiomers of amino acids: a case study, *J. of Mol. Struc.: THEOCHEM*, 2005, **727**, 49.

32. G. Boon, C. Van Alsenoy, F. De Proft, P. Bultinck, and P. Geerlings, Study of Molecular Quantum Similarity of Enantiomers of Amino Acids, *J. Phys. Chem*. A, 2006, **110** 5114.

33. S. Janssens, G. Boon, and P. Geerlings, Molecular Quantum Similarity of Enantiomers: Substituted Allenes, *J. Phys. Chem. A*, 2006, **110** 9267.

34. P. Geerlings, F. De Proft, and P.W. Ayers, Chemical reactivity and the shape function, *Theor. and Comp. Chem.*, 2007, **19**, 1.

35. S. Janssens, C. Van Alsenoy, and P. Geerlings, Molecular Quantum Similarity and Chirality: Enantiomers with Two Asymmetric Centra, *J. Phys. Chem*. A, 2007, **111**, 3143.

36. S. Janssens, A. Borgoo, C. Van Alsenoy, and P. Geerlings, Information Theoretical Study of Chirality: Enantiomers with One and Two Asymmetric Centra, *J. Phys. Chem*. A, 2008, **112**, 10560.

37. E. Debie, L. Jaspers, P. Bultinck, W. Herrebout, and B. Van Der Veken, Induced solvent chirality: A VCD study of camphor in $CDCl_3$, *Chem. Phys. Lett.*, 2008, **450**, 426.

38. E. Debie, P. Bultinck, W. Herrebout and B. van der Veken, Solvent effects on IR and VCD spectra of natural products: an experimental and theoretical VCD study of pulegone, *Phys. Chem. Chem. Phys*., 2008, **10**, 3498.

39. S. Janssens, P. Bultinck, A. Borgoo, C. Van Alsenoy, and P. Geerlings, Alternative Kullback−Leibler Information Entropy for Enantiomers, *J. Phys. Chem*. A, 2010, **114**, 640.

40. P. Geerlings and A. Borgoo, Information carriers and (reading them through) information theory in quantum chemistry, *Phys. Chem. Chem. Phys*., 2011, **13**, 911.

41. N. Nikolova and J. Jaworska, Approaches to Measure Chemical Similarity - a Review, *QSAR Comb. Sci.*, 2003, **22**, 1006.

42. V. Monev, Introduction to Similarity Searching in Chemistry, *MATCH-Comm. Math. and Comp. Chem.*, 2004, **51**, 7.

43. B. Galdino de Oliveira, R. de Cássia Maritan Ugulino de Araújo, A. Bezerra de Carvalho, and M. Neves Ramos, "QTAIM Densities and Ab Initio Basis Sets: A Chemometrical Analysis of the Intermolecular Electronic Densities of the Hydrogen-Bonded Complexes C2H4O···HX (X = F, CN, NC, and CCH), *Acta Chim. Slov*., 2009, **56,** 704.

44. Š. Varga, On robust density fitting in molecules and extended systems , *J. of Math. Chem.*, 2011, **49,** 1.

45. M. Pinsky, D. Danovich, and D. Avnir, Continuous Symmetry Measures of Density Maps, *Theor. J. Phys. Chem. C*, 2010, **114**, 20342.

46. A. Zayit, M. Pinsky, H. Elgavi, C. Dryzun, and D. Avnir, A Web Site for Calculating the Degree of Chirality, *Chirality*, 2011, **23**, 17.

47. R. O. Esquivel, N. Flores-Gallegos, J. S. Dehesa, J. Carlos Angulo, J. Antolin, S. Lopez-Rosa and K. D. Sen, Phenomenological Description of a Three-Center Insertion Reaction: An Information-Theoretic Study, *Theor. J. Phys. Chem. A*, 2010, **114**, 1906.

48. V. V. Turovtsev and Yu. D. Orlov, A Quantum-Mechanical Study of the Inductive and Steric Effects for the Example of tert-Butylalkanes, *Russ. J. of Phys. Chem. A*, 2010, **84**, 965.

49. V. V. Turovtsev and Yu. D. Orlov, A Quantum-Mechanical Study of Inductive and Steric Effects in Isoalkanes , *Russ. J. of Phys. Chem. A*, 2010, **84**, 1174.

50. P. J. Smith and P. L. A. Popelier, Quantum chemical topology (QCT) descriptors as substitutes for appropriate Hammett constants, *Org. Biomol. Chem*., 2005, **3**, 3399.

51. P.G. Mezey, The Holographic Electron Density Theorem and Quantum Similarity Measures, *Mol. Phys*., 1999, **96**, 169.

52. P.G. Mezey, Holographic Electron Density Shape Theorem and Its Role in Drug Design and Toxicological Risk Assessment, *J. Chem. Inf. Comp. Sci*., 1999, **39**, 224.

53. P.G. Mezey, The Holographic Principle for Latent Molecular Properties, *J. Math. Chem.*, 2011, **30**, 299.

54. P.G. Mezey, Functional Groups in Quantum Chemistry, *Adv. in Quant. Chem.*, 1996, **27**, 163.

55. W.F. Reynolds, P.G. Mezey, W.J. Hehre, R.D. Topsom, and R.W. Taft, The Relationship Between Substituent Effects on Energy and on Charge from *ab initio*, Molecular Orbital Theory, *J. Am. Chem. Soc*., 1977, **99**, 5821.

56. W.F. Reynolds, P.G. Mezey, and G.K. Hamer, *Ab initio* Calculations on 4-Substituted Styrenes; a Theoretical Model for the Separation and Evaluation of field and Resonance Substituent Parameters, *Can. J. Chem*., 1977, **55**, 522.

57. W.F. Reynolds, T.A. Modro, and P.G. Mezey, A Theoretical Investigation of the Effect of Positively Charged Substituents on Product Distribution in Electophilic Aromatic Substitution; Evidence for a Dominant Field Effect of the Positive Poles, 1977, *J. Chem. Soc. Perkin II*, 1066.

58. P.G. Mezey and W.F. Reynolds, *Ab inito* Calculations on 4-Substituted Benzoic Acids; a Further Theoretical Investigation into the Nature of Substituent Effects in Aromatic Derivatives, *Can. J. Chem*., 1977, **55**, 1567.

59. W.F. Reynolds, T.A. Modro, P.G. Mezey, E. Skorupowa, and A. Maron, An Experimental and Theoretical Investigation of the Unusual Substituent Effect of the Vinyl Group, *Can. J. Chem*., 1980, **58**, 412.

60. Z. Antal, P.L. Warburton, and P.G. Mezey, Electron Density Shape Analysis of a Family of Through-Space and Through-Bond Interactions, *Phys. Chem. Chem. Phys*., 2014, **16**, 918.

61. Gaussian 09, 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, 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

62. A. Szabo. And N. Ostlund, *Modern Quantum Chemistry*, Dover Publication, INC., New York, 1996.

63. T. Engel, *Quantum Chemistry & Spectroscopy, 3rd ed*., Pearson Education, Inc., USA, 2013.

64. F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester, 1999.

65. H. Dunning, Thomas, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 1989. **90**. 1007.

66. Shape Analysis and Similarity Program “Rhocalc” and support software. (Scientific Modeling and Simulation Laboratory, Memorial University of Newfoundland, Canada, paul.mezey@gmail.com).

67. F. Jensen, *Introduction to Computational Chemistry*, 2nd ed., Wiley, Chichester, 2006. (Note: contains new chapters compared to first ed.)

68. Software property of the Scientific Modeling and Simulation Laboratory, Memorial University of Newfoundland, Canada, paul.mezey@gmail.com.

69. P.G. Mezey, Group Theory of Electrostatic Potentials: A Tool for Quantum Chemical Drug Design, *J. Quant. Chem.*, Quant. Biol. Symp., 1986, **12**, 113.

70. P.G. Mezey, The Shape of Molecular Charge Distributions: Group Theory without Symmetry, *J. Comput. Chem.*, 1987, **8**, 462.

71. P.G. Mezey, Group Theory of Shapes of Asymmetric Biomolecules, *Int. J. Quantum Chem.*, Quant. Biol. Symp., 1987, **14**, 127.

72. G. A. Arteca, V. B. Jammal, P. G. Mezey, J. S. Yadav, M. A. Hermsmeier, and T. M. Gund, Shape Group Studies of Molecular Similarity: Relative Shapes of Van der Waals and Electrostatic Potential Surfaces of Nicotinic Agonists, *J. Molec. Graphics*, 1988, **6**, 45.

73. G.A. Arteca, V.B. Jammal, and P. G. Mezey, Shape Group Studies of Molecular Similarity and Regioselectivity in Chemical Reactions, *J. Comput. Chem.*, 1988, **9**, 608.

74. G.A. Arteca and P.G. Mezey , Shape Description of Conformationally Flexible Molecules: Application to Two-dimensional Conformational Problems, *Int. J. Quantum Chem.*, Quant. Biol. Symp., 1988, **15**, 33.

75. P.G. Mezey, *Shape In Chemistry: An Introduction To Molecular Shape and Topology*, VCH Publishers, New York, 1993.

76. P. G. Mezey. *Potential energy hypersurfaces*, Elsevier, Amsterdam. 1987.

77. F. A. Carey and M. Giuliano, *Organic Chemistry*, 8th Ed, McGraw-Hill, 2011.

The following chapter is an exact copy of a published scientific article. Its format, including sub-chapters and reference list, adopts the format required by the journal in which it was published. The parameters of the article are as follows:

Authors: Zoltan Antal and Paul G. Mezey

Title: Molecular fragment shape variation index applied to intramolecular interaction studies

Journal: Journal of Mathematical Chemistry (*J. Math.Chem.)*

Year, volume, and first page number: 2012, 50, 942

Note: some modifications to the original articles were necessary to successfully incorporate them into the main thesis body.

CHAPTER 4

Molecular Fragment Shape Variation Index Applied to Intramolecular Interaction Studies

Zoltan Antal[1] **and Paul G. Mezey**[1, 2]

[1]Scientific Modeling and Simulation Laboratory (SMSL), Department of Chemistry and Department of Physics and Physical Oceanography, Memorial University of Newfoundland, 283 Prince Philip Drive, St. John's, NL, A1B 3X7, CANADA , paul.mezey@gmail.com

and

[2]Institute of Chemistry, Eötvös University of Budapest, HUNGARY

Key words: Fragment Shape Variation Index; Intramolecular Interactions; Holographic Electron Density Theorem; Common Structural Detail Variations; Molecular Fragmentation; Molecular Similarity; Fragment Similarity

Abstract

The fragment shape variation index approach is applied to intramolecular interactions involving C6 aromatic molecular fragments in the special case where the shape-modifying interactions are also caused primarily by other C6 aromatic fragments of the same molecule. This report is a part of a series of studies aimed at the detailed modeling of various components of intramolecular interactions among molecular fragments including aromatic ring interactions, aromatic ring and non-aromatic conjugated and non-conjugated system interactions, and more general through-space and through-bond

interactions. The ultimate purpose of these studies is a better understanding of the electron density shape modifying effects of intramolecular interactions.

4.1. Introduction

It is customary to associate chemical reactivity with various molecular fragments, and those molecular fragments which are expected to show only limited variations are expected to imply rather similar reactivities, at least in some common types of chemical processes. Reactivity, however, can be analyzed from several perspectives, for the more recent approaches the reader may consult ref. [1] and references therein. In some early theoretical studies on reactivity, especially those where local ranges of molecules have been known to have important roles in determining reactivity, some structural features and effects, called substituent effects have been studied within a common framework.[2-6] It is clear, that variations in the shape of the electron density must reflect the variations in reactivity, since, according to the Hohenberg-Kohn theorem,[7] all molecular properties are reflected in the molecular electron density. Furthermore, local shapes are also informative concerning the whole molecule, as this follows from the Holographic Electron Density Theorem[8]: any positive volume fragment of a molecular electron density cloud (in a non-degenerate ground state) contains the complete molecular information, that is, all information about the complete molecule. Hence, electron density shape analysis methods,[9] even if applied to local shapes, can provide valid information about reactivity, as well as substituent effects.

The present study will focus on the electron density fragment shape variation index for actual fragments which are not functional groups in the conventional sense, but structural units which are often carriers of actual functional groups, so their shape variation often may be regarded as secondary, triggered by other groups more directly involved in determining reactivity. Nevertheless, in the actual examples used involving primarily aromatic rings, a photochemical initiation reaction may have a specific role analogous to the role of conventional functional groups, as has been discussed in some earlier studies.[10,11,12]

The main purpose of this study is to investigate the shape variations of aromatic C6 ring fragments of molecules where the influences from the rest of the molecules are also primarily provided by similar C6 aromatic ring structures. These results are expected to serve as a guide in later studies where the electron density shape variations of aromatic C6 systems are caused by far more diverse chemical groups. Those more general investigations will attempt to elucidate, among other questions, the roles of through-space and through-bond interactions, where the through-bond interactions are accomplished by molecular fragments with rather mobile electron density clouds, such as aromatic C6 systems. In the latter case, the fragment shape variations are likely to have different character than in the present case where only the number and the mutual location of aromatic rings is a source of variations in the shapes of the electron densities of aromatic C6 systems.

When the interactions between aromatic rings are considered, polycyclic aromatic hydrocarbons, PAHs are likely candidates for studies. The experimental information and

the range of biochemical effects of PAHs have been studied extensively, and the difficulties represented by conformational variations are very limited. The essentially planar structures of PAHs, with only limited deviations from planarity for a subclass of these structures is an advantage, both for the interpretation of some of the experimental results and in the actual computational studies using quantum chemical or other molecular modeling methods. On the other hand, important insight can also be obtained if only a single aromatic C6 fragment is present in the molecule, when the local shape-modifying influences have entirely different sources, such as it is in the case of *para*-substituted styrene derivatives, the subject of a much earlier study on through-space and through-bond interactions, using earlier computational quantum chemistry methods of somewhat limited power.[2-6] We may expect that some of the earlier conclusions can be refined if the more advanced quantum chemistry methods of today are applied.

In this work, some aspects of the first of the two questions will be discussed, involving shape-modifying influences in C6 aromatic fragments if the sources of these influences are also C6 aromatic rings. Such C6 rings are special for several reasons; one aspect important in the present context is the fact that aromatic C6 rings are present in a very large number of molecules, where they have a primary role to act as carriers of functional groups. Otherwise, the C6 aromatic fragments are often considered rather stable structures and, within some conventional interpretation, their shapes are modified primarily through the influences affecting the π bond system.

4.2. An Application of the Molecular Fragment Shape Variation Index

The earlier choice for Fragment Shape Variation Index,[13] closely related to local Shape Dominance Indices, also referred to as local Shape Persistence Indices,[12] has been introduced within a rather general context, applicable with respect to several possible shape descriptors, including simple internal coordinates referring to relative nuclear positions, as well as the more detailed shape descriptors of the Shape Group Methods, as applied to the molecular electron density clouds. In this report, the latter methods, the Shape Group methods, will be applied in order to provide the input information for the determination of the Fragment Shape Variation Index and to follow the variations of the electron density shape of aromatic C6 rings, within various PAH molecules.

The steps employed in the introduction of the Shape Dominance Index[12] and the Fragment Shape Deviation Index[13] will not be repeated here, similarly, for the description of the Shape Group Methods the reader is also referred to reference [9]. Here only the actual expressions for the Fragment Shape Deviation Index are given, in a form suitable for applications in the present examples of polycyclic aromatic hydrocarbon (PAH) electron densities.

In general, for a molecular electron density of a molecule **A** with nuclear configuration **K** the notation

$$\rho(\mathbf{r}) = \rho(\mathbf{r}, A, K) \qquad (13)$$

is used, where **r** is the spatial variable.

Using isosurfaces of the molecular electron density, also called Molecular Isodensity Surfaces, MIDCOs, the distributions of various local curvature properties define a family of homology groups of algebraic topology, also called "Shape Groups" within the molecular context. The ranks of these shape groups, the so-called Betti numbers, generate a numerical shape code for each molecule or for each molecular fragment. These shape codes are given in the form of a matrix of integers M(a,b,X) for each molecule or fragment X, and for two species X_1, and X_2, a direct numerical comparison, in terms of matches and mismatches of the corresponding matrix elements generate a similarity measure denoted by

$$s_{SG}(X_1, X_2) = m[M(a,b, X_1), M(a,b, X_2)] / t, \quad (14)$$

where $m[M(a,b, X_1), M(a,b, X_2)]$ is the number of entry matches in the two shape matrices and t is the total number of entries.

Specifically, for any pair of fragments F_1 and F_2 of any pair of molecules A_1 and A_2, the possible values of the global and local electron-density shape-similarity measures $s_{SG}(A_1, A_2)$ and $s_{SG}(F_1, F_2)$ fall within the [0,1] interval:

$$0 \leq s_{SG}(A_1, A_2) \leq 1, \quad (15)$$

and

$$0 \leq s_{SG}(F_1, F_2) \leq 1. \quad (16)$$

Note that, in the present context, the shape code matrices M(a,b,A) and M(a,b,F) take on the roles of the general shape descriptors of ref. [13] as the global shape descriptor gsh(A) and the associated local shape descriptor lsh(F), respectively. An important consistency condition of the general treatment is fulfilled: one can verify easily that these two shape descriptors, M(a,b,A) and M(a,b,F) are consistent with one another: if the local molecular fragment **F** is replaced with the complete molecule **A**, then one obtains

$$lsh(A) = gsh(A) = M(a,b,A) = M(a,b,F) . \quad (17)$$

Then, applying again the general treatment of ref. [13] to the Shape Group approach, with the general comparison function taken as

$$comp(gsh(X_1), gsh(X_2)) = s_{SG}(X_1, X_2) = m[M(a,b, X_1), M(a,b, X_2)] / t, \quad (18)$$

for both cases, X = A, or X = F, where the subscript in the expression $s_{SG}(X_1, X_2)$ refers to the Shape Group method.

For any pair of local fragments F_1 and F_2 of molecules A_1 and A_2, the Shape Group induced local or *fragment shape variation index* $FSVI_{SG}$, denoted as $fsvi_{SG}$ (F_1, F_2, A_1, A_2) is defined as

$$\frac{\text{fsvi(F1,F2,A1,A2)} = \text{gshs(A1,A2)}}{[\text{lshs(F1,F2)} + \text{gshs(A1,A2)}]} \qquad (4)$$

As in the case of the general shape descriptors gsh(A) and lsh(A), the constraints on $s_{SG}(F_1, F_2)$ and $s_{SG}(A_1, A_2)$ imply that the Shape Group induced local or fragment shape variation index $FSVI_{SG}$ is also restricted to the [0,1] interval, that is, for any pair of fragments F1 and F2 of any pair of molecules A1 and A2, the inequalities

$$0 \leq fsvi_{SG}(F_1, F_2, A_1, A_2) \leq 1 \qquad (17)$$

hold.

A small local shape similarity of fragments F1 and F2 within molecules A1 and A2, when taken in comparison to the global similarities of the molecules A1 and A2, indicates a large shape variation for the fragments, and this is reflected in the fragment shape variation index $FSVI_{SG}$ as given in terms of the results of the shape group analysis.

The evaluation of the fragment shape variation index for a set of different fragment types in a series of molecules can help in the identification of various molecular regions where the effect of changes in substituents have greater or lesser effect in modifying the influence of various intramolecular interactions.

4.3. Shape Group Results and Fragment Shape Variation Index Computed for Selected Aromatic C6 Ring Fragments of 16 PAH Molecules

A series of 16 PAH molecules, listed in Table 4.1, were studied earlier,[10-12] and had their shape groups and shape code matrices computed and their numerical similarity measures calculated. The main goal of those studies was to find correlations with their aquatic toxicity with respect to the plant *Lemna gibba*. Good correlations have been obtained between their shape descriptors and experimental toxicities. Those studies revealed some useful information concerning the actual mechanism of the photoinduced toxicity of these molecules within an aquatic environment in terms of a photochemical sensitization step followed by an actual photochemical reaction. Please note that for a very detailed analysis of the individual molecules the reader should consult references [10-12] by Mezey and Zimpel, here just the most relevant data is presented which is adequate to present the FSVI.

The fragment similarities have been evaluated with reference to the central C6 ring fragment of the anthracene molecule, taking the most similar fragment form for each molecule. Using the relevant raw data from refs. [10-12], reproduced in Table 4.1, the newly computed fragment shape variation indices of the present study are also listed in Table 4.1. Also, Figure 4.1 presents the "ball and stick" model of the reference molecule anthracene (fsvi 0.500), also pyrene (0.542), and phenanthrene (0.629).

Table 4.1. Evaluation of 16 PAH molecule C6 ring Fragment Shape Variation Index with respect to the central C6 ring of anthracene (taking the most similar ring in each molecule)

Name of PAH molecule	C6 ring similarity	Molecular similarity	Fragment Shape Variation Index
anthracene	**1.000**	**1.000**	**0.500**
benzo(*a*)anthracene	**0.568**	**0.422**	**0.426**
fluoranthene	**0.386**	**0.513**	**0.571**
dibenzo[*a*,i]pyrene	**0.425**	**0.410**	**0.491**
benzo[*b*]fluorene	**0.640**	**0.436**	**0.405**
benzo[*b*]anthracene	**0.640**	**0.412**	**0.392**
benzo[*a*]pyrene	**0.471**	**0.404**	**0.462**
pyrene	**0.435**	**0.514**	**0.542**
benzo[*e*]pyrene	**0.429**	**0.409**	**0.488**
fluorene	**0.413**	**0.663**	**0.616**
benzo(*g,h,i*)perylene	**0.378**	**0.406**	**0.518**
coronene	**0.350**	**0.403**	**0.535**
phenanthrene	**0.358**	**0.606**	**0.629**
dibenz[*a,h*]anthracene	**0.659**	**0.418**	**0.388**
triphenylene	**0.342**	**0.417**	**0.549**
chrysene	**0.355**	**0.411**	**0.537**

In general, if the value of the computed fragment shape variation index is greater than 0.5, then the chosen fragment, in our case the aromatic ring fragment C6 shows greater shape variation than the shape variation of the given molecule containing the aromatic ring fragment C6, that is, the local shape change is greater than the global shape change, when the shapes of the reference molecule anthracene and the other PAH molecules studied are compared. By contrast, a computed fragment shape variation index value smaller than 0.5 indicates that the actual aromatic C6 ring fragment shows smaller shape changes than the complete PAH molecule to which the C6 fragment

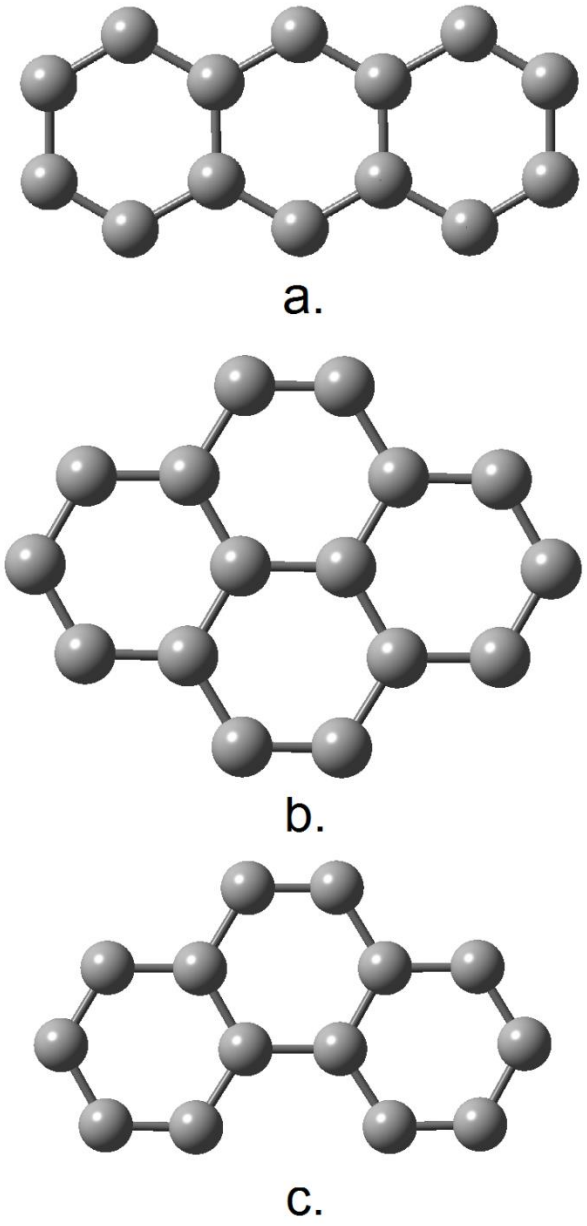


Figure 4.1. The "ball and stick" representation of anthracene, pyrene, and phenanthrene, in increasing Fragment Shape Variation Index order. The values are a) 0.500, b) 0.542, and c) 0.629.

belongs, when the local and global shapes of the anthracene molecule are used as reference in the comparisons.

The computed fragment shape variation index values indicate that both enhancement and partial suppression of fragment shape variation does occur, when compared to the global shape variation of the host molecules for these aromatic C6 fragments. Some general trends appear to emerge.

The largest values of the computed fragment shape variation index were obtained for phenanthren and fluorene, two of the relatively small PAH molecules, whereas some of the largest PAH molecules show relatively small values for the computed fragment shape variation index. This finding is in good agreement with the notion that in larger aromatic systems, the local shape variations are better balanced due to a more extensive delocalization, that is not as prominent in smaller PAH molecules.

It is interesting to note that the smallest fragment shape variation index values are obtained for two of the anthracene derivatives, dibenz[*a*,*h*]anthracene (0.388) and benzo[*b*]anthracene (0.392), where large fragment similarities, but relatively low global similarities have been obtained. Although the fact that both of these molecules are related to the reference molecule anthracene may suggest otherwise, this fact, by itself does not imply large global similarities, yet, the closer relations of the bonding structures in such cases may provide good local similarities for the respective C6 fragments. The combinations of these factors produce low fragment shape variation indices, in spite of the fact, that higher global similarity, by itself, is likely to produce higher fragment shape variation indices.

The above findings indicate that the computation of fragment shape variation indices can reveal features of intramolecular interactions which are not obvious if one is concerned with global shapes and local shapes individually, without reference to their interrelations.

4.4. Summary

Molecular fragment shape variation measures have been applied within the context of the electron density shape characterization using the Shape Group methods. Actual computations have been carried out to evaluate the fragment shape variation measures for a set of 16 polycyclic aromatic hydrocarbons, focusing on aromatic C6 fragments, where the shape-modifying influence of other molecular parts is also dominated by other aromatic rings, thereby limiting the types of influences and somewhat simplifying the analysis. These results provide reference and background for further studies, where a wider variety of fragments may produce enhanced shape changes, partly associated with through-space and through-bond interactions.

4.5. Acknowledgements

This work was supported by a Discovery Research Grant from the Natural Sciences and Engineering Research Council of Canada, and by the Canada Research Chair and Canada Foundation for Innovation projects at the Memorial University of Newfoundland, Canada.

4.6. References

1. Alejandro Toro-Labbé, Ed., *Theoretical Aspects of Chemical Reactivity*, Theoretical and Computational Chemistry, Vol.19, Elsevier, Amsterdam, 2007.

2. W.F. Reynolds, T.A. Modro, and P.G. Mezey, A Theoretical Investigation of the Effect of Positively Charged Substituents on Product Distribution in Electophilic Aromatic Substitution; Evidence for a Dominant Field Effect of the Positive Poles, *J. Chem. Soc. Perkin II*, 1977, 1066.

3. P.G. Mezey and W.F. Reynolds, *Ab inito* Calculations on 4-Substituted Benzoic Acids; a Further Theoretical Investigation into the Nature of Substituent Effects in Aromatic Derivatives, *Can. J. Chem.*1977, **55**, 1567.

4. W.F. Reynolds, P.G. Mezey, and G.K. Hamer, *Ab initio* Calculations on 4-Substituted Styrenes; a Theoretical Model for the Separation and Evaluation of Field and Resonance Substituent Parameters, *Can. J. Chem*.,1977 **55**, 522.

5. W.F. Reynolds, P.G. Mezey, W.J. Hehre, R.D. Topsom, and R.W. Taft, The Relationship Between Substituent Effects on Energy and on Charge from *ab initio* Molecular Orbital Theory, *J. Am. Chem. Soc*., 1977, **99**, 5821.

6. W.F. Reynolds, T.A. Modro, P.G. Mezey, E. Skorupowa, and A. Maron, An Experimental and Theoretical Investigation of the Unusual Substituent Effect of the Vinyl Group, *Can. J. Chem*., 1980, **58**, 412.

7. P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev*., 1964, **136**, B864.

8. P.G. Mezey, The Holographic Electron Density Theorem and Quantum Similarity Measures, *Mol. Phys.*,1999, **96**, 169.

9. P.G. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology*, VCH Publishers, New York, 1993.

10. P.G. Mezey, Z. Zimpel, P. Warburton, P.D. Walker, D.G. Irvine, D.G. Dixon, and B. Greenberg, A high-resolution shape-fragment database for toxicological shape analysis of PAHs, *J. Chem. Inf. Comp. Sci.*, 1996, **36**, 602.

11. P.G. Mezey, Z. Zimpel, P. Warburton, P.D. Walker, D.G. Irvine, X.-D. Huang, D.G. Dixon, and B. Greenberg, Use of QShAR to model the photoinduced toxicity of PAHs: Electron density shape features accurately predict toxicity. *Environ. Toxico.l Chem.*, 1998, **17**, 1207.

12. P.G. Mezey, Quantitative Shape - Activity Relations (QShAR), Molecular Shape Analysis, Charge Cloud Holography, and Computational Microscopy, in *"Quantitative Structure-Activity Relationships for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications",* Ed. J.D. Walker, SETAC (Society of Environmental Toxicity and Chemistry) Publ., SETAC Press, 2003, 123-136.

13. P.G. Mezey, Molecular fragment shape variation index for functional groups and the holographic properties of electron density, *J Math Chem*., 2012, **50**, 926.

CHAPTER 5

5.1. Summary

In this work, a new approach was presented for the identification and separation of some fundamental intramolecular effects in series of molecules. These effects cause electron density changes in a target area of these molecules which can be interpreted as a unique "echo" of the causing moiety, and classically these effects are known as through-bond and through-space effects. Quantum chemical justification was achieved for some experimental results concerning well-established substituent effects, and by applying electron density shape analysis methods and computational resources, even some fundamental reactivity patterns were identified. We can conclude, that all three quantum chemical models used (HF, DFT, and MP2) performed well, although higher levels of theory at the cost of longer computational times, performed better at "tricky" molecules or molecular fragments, such as the curious case of the simple methyl group. This apparently modest functional group would probably deserve its own computational project for a proper description of the hyperconjugative effects caused by it. The summary of the performance of individual quantum models is as follows:

HF can be used if a quick, but somewhat "rough" description of the desired effects of interactions is needed. HF theory, even combined with a fairly large basis set, fails to provide adequately accurate geometries (mostly because of the missing electron correlation energies of electrons) for some of the target molecules and therefore provides a weak basis for the shape analysis software. If more computational resources and time are invested, DFT will perform better, at least with the B3LYP functional. As always, in

the case of DFT, some fine tuning of the specific results can be achieved by choosing the correct functional, and our choice of the popular B3LYP is just one method suitable for comparison. In part due to a different representation of correlation energy in MP2, it performs very well, and if adequate computational resources are at hand, it's well worth the investment. It has to be noted at this point that some unusual failures of this model can influence the outcome of computational experiments based on this theory, as described in Chapter 2.

As a conclusion, we can state that the results achieved could be used in specific molecular design, fine tuning of some desired effect within molecules, or even identifying some of the weak, but crucial through-space interactions present in more complex structures.

Chapter 2 presents a computational method which can be used to identify and separate through-bond and through-space effects in practically any set of molecules, but it is more successful in sets in which these effects are enhanced. The abovementioned effects are difficult and sometimes even impossible to study experimentally, but using the theoretical method of electron density shape analysis, they can be assigned numerical values which greatly simplify their comparison. Experimental expectations of substituent effects, as they exert their influence on the rest of the molecule, are also largely reproduced by the shape analysis method and the magnitude of shape change reflects the influence of both through-bond and through-space components. Also, in this set of molecules, the dominance of the through-bond component is clearly indicated which has

to be taken into consideration when revisiting some of the traditional explanations for these types of molecular interactions. Nonetheless, the through-space component is clearly present and exerts its own influence as well and this also plays a crucial role when trying to design new molecular nanostructures.

Another field where these types of studies can be relevant is complementarity studies for instance. A simple, well known example is the lock-key type of enzyme activity studies. Stemming from how the shapes of molecules are described (by topologically describing their surfaces, given how many holes are generated per surface) their complements will be their exact opposites which are also topological surfaces. This new surface could be the starting point for building up a new molecular surface.

Chapter 3 presents a practical application of the method presented above. Reactivity patterns can be predicted for substituents if they are used in similar sets of molecules and when studying aromatic substitution type reactions. The activating/deactivating power and their *ortho*-, *meta*-, and *para*-directing abilities correlate excellently with the magnitude of shape change they cause in the target structure of the molecule. Quantum chemical justification has been achieved by this theoretical electron density based method of known (mostly experimentally achieved) substituent effects and the results provide novel insights as well, a crucial ingredient of new molecular design attempts. Also it's important to note that these types of studies are more thorough then mapping reactivities with electrostatic potential map studies, for instance. The justification relies on the fact the shape analysis goes through a whole

range of isodensity surfaces (the chemically most relevant intervals usually), whilst electrostatic potential maps represent a single one of these isodensity surfaces and analyzing a whole range of them can be quite time consuming.

This method could be extended to all types of molecular sets where through-bond and through-space effects need to be studied and it also can provide information for predictive correlations.

Chapter 4 deals with setting up the Fragment Shape Variation Index for molecular fragments and foundations are laid to provide a reference and background for additional studies. In these studies, partly because of the through-space and through-bond effects, a large variety of fragments could produce more enhanced shape changes.

This chapter also focuses on some properties of benzene rings, which are the actual through bond effect transmitters in the case of substituted styrenes. Here, a new way is presented to characterize aromatic C6 fragments in the specific case of polycyclic aromatic hydrocarbon molecules where the rings have a far more enhanced role than in substituted styrenes. Using the Shape Group methods, shape analysis calculations have been carried out on 16 PAHs, where the shape modifying factors were also represented by other aromatic rings. This report further enhances our understanding of effects within molecules by adding another shape modifying element in addition to the substituent effects of functional groups. Understanding the part that aromatic rings play in influencing global molecular interactions is important for the sheer number of molecules containing such stable entities and because of their ability to carry functional groups.

Also, by conducting shape analysis studies on these polycyclic aromatic hydrocarbons, some insight can be gained about the "length" of an effect a certain ring has on other rings around it. The through-bond effect "ripples" through the conjugated bonds and knowledge can be gained about the "stabilizing" effect these rings have. This however, is a topic of a future project.

6. Appendices

6.1. **Appendix A** (Note: from Chapter 2, subchapter 2.8.)

6.1.1. Brief Review of Molecular Shape and Similarity Analysis by the Shape Group Method

For fuzzy objects with no precise boundaries, such as molecular electron density clouds, the shape and similarity analysis is based on the complete, infinite series of isodensity contours for the full range of electron density threshold values a, using curvature-based geometrical characterization for a whole range of reference curvature parameters b. By formally truncating each contour according to a curvature criterion defined by b, and sweeping through the full range of these two continuous parameters, a and b, a finite number of topological equivalence classes of truncated surfaces is obtained, which can be characterized by a finite number of integers: the ranks of their algebraic-topological homology groups, called Betti numbers, are related to the number of holes in each truncated contour. This generates a numerical shape code for each, complete, fuzzy electron density cloud. For the shape similarity measures of two electron densities, a direct comparison of these shape codes is used. This shape similarity measure avoids the (often ambiguous) "optimal" superposition requirement of the molecules needed in similarity evaluation by other methods.

This approach accomplishes two things: it represents geometrical similarity as a topological equivalence, an important simplification without missing essential shape information, and reduces the shape description and similarity evaluation to a comparison of numerical shape codes. A full description of the Shape Group Method

and related shape similarity evaluation can be found in ref. [68], here, we review only the essence of details found there.

6.1.2. The Shape Group Methods

For a nuclear configuration K and density threshold a, the *molecular isodensity contour surface*, MIDCO G(K,a) is defined as

$$G(K,a) = \{ \mathbf{r} : \rho(K,\mathbf{r}) = a \} \qquad (1)$$

Set G(K,a) is the collection of all points **r** where the electronic density $\rho(K,\mathbf{r})$ is equal to the given threshold value a.

The original concept of local convexity at any given point **r** of a surface is relative to "curvature" b=0, that is, relative to a tangent plane of no curvature. Accordingly, the MIDCO surface at point **r** may be regarded as a function defined over the tangent plane at **r**, and the two eigenvalues of the second derivative matrix, a formal Hessian matrix of this function, can be computed. The number of those eigenvalues that are less than b is denoted by μ. If b=0, then $\mu = 0$, or 1, or 2 at point **r** corresponds to the surface being locally concave, or saddle type, or convex, respectively. However, such characterization does not give much detail and discrimination, that is, if only b=0 is considered, this provides only a crude shape characterization. But if the tangent plane is replaced by a tangent sphere of radius 1/b, and the b reference curvature can be changed within a wide enough range, than the concepts of local convexity (local concavity and local saddle-shape) can be generalized as "convexity in a curved world" of curvature b, and for a range of b values, a far more detailed shape characterization is obtained. A formal negative b value indicates that the tangent sphere of radius 1/|b| is approaching the surface point **r** "from the inside" of

the MIDCO. That is, if a whole range of b values, including negative values is considered, than a detailed shape characterization becomes possible.

For any given b value, the tangent object T (plane or sphere) at point **r**, may fall locally on the outside, or on the inside, or it may cut into the given MIDCO surface near enough to point **r** . By carrying out this characterization for all points **r** of the MIDCO, one obtains a subdivision of the molecular contour surface into locally convex, locally concave, and locally saddle-type shape domains, denoted by D_2, D_0, and D_1, respectively, which are the *local relative convexity domains* of the MIDCO, relative to the reference curvature b. Formally, this can be carried out for each density threshold a and each reference curvature b of the complete fuzzy electron density cloud.

We still have a continuum of MIDCO surfaces, one for each density threshold a, and the local curvatures of each are characterized relative to infinitely many curvature values b. We still deal with a continuum, and the problem of fuzzy object shape characterization has not been simplified.

If, however, we decide to truncate each MIDCO and eliminate either one of the three types of the D_0, or D_1, or D_2 *local relative convexity domains* D_μ, then for each truncation type $\mu = 0, 1, 2$, relative to b, we may obtain a set of new, topologically more interesting objects, a set of truncated contour surfaces $G(K,a,\mu)$ which inherit some essential shape information from the original MIDCO surface $G(K,a)$. It is important to note, however, that many of the $G(K,a,\mu)$ truncated surfaces are topologically equivalent, which provides an important simplification for shape analysis.

In fact, for the entire range of a and b values, we obtain only a finite number of topologically different truncated surfaces with various number of holes in them in various arrangements. That is, we have obtained a finite number of topological equivalence classes which are much simpler to characterize by the tools of algebraic topology than the continuum of the original set of infinitely many contour surfaces.

For such a topological description, each G(K,a) is characterized by its *shape groups* [68], denoted by $H^p_\mu(a,b)$, which are algebraic groups for each dimension p, p=0,1,2, specifically, these are the homology groups of a truncated version of the given MIDCO G(K,a). The truncation is determined by local shape properties expressed in terms of the reference curvature value b, and by the choice $\mu = 0, 1,$ or 2, indicating that the locally concave D_0, or the locally saddle type D_1, or the locally convex D_2 regions (relative to b) are truncated from G(K,a).

By definition[68], *geometrical curvature conditions* lead to truncation, and all essentially equivalent truncations provide a *topological classification of shapes*, expressed by the Shape Groups of the original MIDCO G(K,a), the algebraic-topological homology groups $H^p_\mu(a,b)$ of the truncated surfaces $G(K,a,\mu)$. For each reference curvature b and shape domain and truncation pattern μ of a given MIDCO G(K,a) of density threshold a, there are three shape groups, $H^0_\mu(a,b)$, $H^1_\mu(a,b)$, and $H^2_\mu(a,b)$. In general, these shape groups $H^p_\mu(a,b)$ are topological invariants for each dimension p, p=0,1,2. The formal dimensions p of these three shape groups are zero, one, and two, collectively expressing the essential shape information of the MIDCO. The $b^p_\mu(a,b)$ Betti numbers are the ranks of the $H^p_\mu(a,b)$ shape groups. These ranks are integer numbers, used to generate numerical shape codes for molecular electron density distributions. Accordingly, for each (a,b) pair of parameters and for each

shape domain truncation type μ, there are three Betti numbers, $b^0_\mu(a,b)$, $b^1_\mu(a,b)$, and $b^2_\mu(a,b)$.

The outlined shape group method of topological shape description combines the advantages of geometry and topology, and follows the spirit of the GSTE principle: Geometrical Similarity is treated as Topological Equivalence. The local shape domains and the truncated MIDCO's $G(K,a,\mu)$ are defined in terms of geometrical classification of points of the surfaces, using local curvature properties. In turn, the truncated surfaces $G(K,a,\mu)$ are characterized topologically by the shape groups and their Betti numbers, themselves topological invariants.

6.1.3. Numerical Shape Codes: the (a,b)-Parameter Maps

The most important Betti numbers, conveying the chemically most relevant shape information, are those of type $b^1_\mu(a,b)$ for shape groups of dimension 1, hence in most applications, only these are used.

Since the shape groups and the Betti numbers are invariant within small ranges of both parameters a and b, in practice, it is sufficient to consider only a finite set of a and b values within the chemically relevant ranges of these parameters. Since both a and b may vary by several orders of magnitude, and since electron density decreases exponentially at larger distances from the nuclei, it is advantageous to use a logarithmic scale for both a and b, where for negative curvature parameters b the log|b| values are taken. In most applications, a 41×21 grid is used for the (log(a), log|b|) pairs, where the ranges are [0.001, 0.1 a.u.] (a.u. = atomic unit) for the density threshold values a, and [-1.0, 1.0] for the reference curvature b.

The grid points of the corresponding 41×21 (a,b)-map contain the Betti numbers obtained for the grid values of the parameters a and b. For a given pair of (a,b) values, the truncated MIDCO may become disconnected, then two or more Betti numbers are obtained, and the complete set of them, encoded in a single number, is assigned to the same location of the (a,b)-map. Consequently, the (a,b)-map becomes a 41×21 matrix $\mathbb{M}^{(a,b)}$, and this matrix is a *numerical shape code* for the fuzzy electron density of the molecule.

6.1.4. Shape Similarity Measures from Shape Codes

The associated numerical shape-similarity measure s(A,B) is also based on the primary molecular property, electron density,

For a similarity evaluation of two molecules, A and B, both in some fixed nuclear configuration, we use their shape codes in their matrix forms, $\mathbb{M}^{(a,b),A}$ and $\mathbb{M}^{(a,b),B}$, respectively. The numerical shape similarity measure is defined as

$$s(A,B) = m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}] / t \qquad (2)$$

where $m[\mathbb{M}^{(a,b),A}, \mathbb{M}^{(a,b),B}]$ is the number of Betti number matches between corresponding elements in the two matrices, and t is the total number of elements in either matrix. With the usual 41×21 grid, where the grid divisions are $n_a = 41$ and $n_b = 21$, one has

$$t = n_a n_b = 861 \qquad (3)$$

Based on a fuzzy density fragmentation approach[96] (see also Appendix B), the same method is applicable for the description and similarity analysis of shapes of molecular fragments.

Such numerical shape similarity measures s(A,B) have been successfully tested and used in studies of Quantitative Shape Activity Relations (QShAR) and predictive correlations, as well as tools for the interpretation of local molecular properties of functional groups and molecular fragments[97-108].

6.2. Appendix B (Note: from Chapter 2, subchapter 2.9.)

6.2.1. Generation of fuzzy electron densities for molecular fragments

The shape similarity comparisons of vinyl groups required the generation of fuzzy fragment densities, which are fully analogous with fuzzy densities of complete molecules. For this purpose, the AFDF (Additive Fuzzy Density Fragmentation) approach has been used, which is also the basis of the linear scaling ADMA (Adjustable Density Matrix Assembler) macromolecular quantum chemistry method[96].

With reference to n atomic orbitals $\varphi_i(\mathbf{r})$ (i=1,2,...,n) in an LCAO *ab initio* wavefunction of a molecule, where **r** is the position vector variable, and **P** is the corresponding $n \times n$ density matrix, the electronic density $\rho(\mathbf{r})$ of the molecule is given by

$$\rho(\mathbf{r}) \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}\, \varphi_i(\mathbf{r}) \varphi_j(\mathbf{r}) \tag{9}$$

Given the collection of nuclei of the molecule present in a fragment k of the molecule, then the fuzzy fragment electron density $\rho^k(\mathbf{r})$ of the fragment is obtained from the electron density $\rho(\mathbf{r})$ of the complete molecule, using the following definition for the elements P^k_{ij} of the n × n fragment density matrix:

P^{k}_{ij} = P_{ij} if both $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are AO's centered on one or two nuclei of the fragment,

= 0.5 P_{ij} if precisely one of $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ is centered on a nucleus of the fragment,

= 0 otherwise,

following an approach analogous to Mulliken's population analysis, without integration to obtain charge. The fuzzy electron density of fragment k is defined as

$$\rho(\mathbf{r}) \sum_{i=1}^{n} \sum_{j=1}^{n} P^{k}_{ij}\, \varphi_i(\mathbf{r})\, \varphi_j(\mathbf{r}) \tag{10}$$

Note that, if the nuclei of the molecule are partitioned into m mutually exclusive families to generate m fragments, then the sum of the fragment density matrices is the density matrix of the molecule, and the sum of the fragment densities is the density of the molecule, that is, the scheme is fully additive:

$$P_{ij} \sum_{k=1}^{m} P^{k}_{ij} \tag{11}$$

and

$$\rho(\mathbf{r}) \sum_{k=1}^{m} \rho^{k}(\mathbf{r})$$

(12)

These simple fragment additivity rules are *exact* and by providing fuzzy fragment densities, allow one to use the same shape analysis and similarity methods as those developed for complete molecules.

6.3. Appendix C. (Note: from Chapter 2, subchapter 2.10.)

6.3.1. Computational Methods and Resources

The initial electron density computations for the complete molecules and the IM models have been performed by the Gaussian 09 software package[93], using geometrically optimized structures. For a pictorial representation of some of the complete molecules and IM fragment pairs, the Gaussview program[93] was used.

In order to generate the IM fragments, we have cut out the aromatic ring, followed by adding H atoms or CH_3 groups to the fragments to avoid chemically incorrect structures. The lengths of the new carbon-hydrogen bonds were set to the experimental average[95] of 109 pm and the new carbon-carbon or carbon-heteroatom distances were set to 150 pm for chemical consistency of the models. In the case of *para*-methoxy styrene, the methoxy part needed a methyl group, whilst the vinyl part only a hydrogen atom.

After the Gaussian computations, the resulting .log files and electron density data were used as input for our shape analysis software package[94] to calculate the local shape

representation of the vinyl group of each molecule, and generating the numerical shape codes, following the methodology described in detail in [68,82-87] (resulting in the standard four significant figures when shape similarities are expressed as similarities, matches-mismatches in the shape codes). In this way, the local shape characteristics of the vinyl groups are expressed with numbers as they are influenced by the rest of the molecule, most importantly by the functional group A in the *para* position.

Finally, the shape similarities of all the local electron densities of all vinyl groups are compared, including the shape of the full molecule vinyl group and the shape of the fragment vinyl group for the same parent molecules. As an illustration for the complete set of shape similarity results, the HF results with a large correlation consistent (cc-PVTZ) basis set are summarized in Table 1.

Thank You!

Note: Microsoft Power Point slide copied from the author's 6003 departmental seminar presentation.