International Symposium on
"Chemical, Physical and Mathematical Foundations of Complex Phenomena"
organized by the
International Research Unit of Advanced Future Studies, Kyoto University

Imperfect Symmetries, Symmetry Deficiency Measures, and Chirality Measures in Biochemical Processes and Their Role in Some Diseases

Paul G. Mezey

Canada Research Chair in Scientific Modelling and Simulation

Memorial University, St. John's, NL CANADA

Motivation:

Reality, involving complex phenomena, always shows imperfections when compared to ideal models.

In science we need to deal with imperfections,

such as imperfect symmetries.

We should be able to describe and measure,

how the imperfection compares to some ideal model,

and how these imperfections affect natural phenomena.

A specific, practical motivation:

Some types of symmetry deficiencies, such as various levels of chirality of L- and D- amino acids in life-forming molecules such as proteins, have strong relations to health, diseases, and medicine.

A current research collaboration with

Professor Noriko Fujii

provides both theoretical and practical motivation

Research Theme:

"Local and Global Chirality Measures of Natural Amino Acids"

Prof. Paul G. Mezey,

Memorial University, St. John's, NL CANADA

Invited by the
Kyoto University International Research Unit of Advanced Future Studies,
as Distinguished Research Professor
in charge of
"Local and Global Chirality Measures of Natural Amino Acids",
to be carried out at the
Research Reactor Institute of Kyoto University
Osaka, Japan, July 9 – Aug 8, 2016.

Host:

Prof. Noriko Fujii, Ph.D.,

Radiation Biochemistry and Biological Function,
Division of Radiation Life Science, Dept. of Radiation Life Science and Radiation Medical Science,
Research Reactor Institute, Kyoto University
2 Asashironishi Kumatori, Sennan, Osaka 590-0494, Japan
TEL:81-72-451-2496 FAX:81-72-451-2630 nfujii@rri.kyoto-u.ac.jp

An example from last week, providing strong practical motivation, the

Conference of the Japanese Society for Cataract Research,

2016 July 29-31, Morioka, Japan,

has been dealing with

very complex phenomena,

which, however, depend on

many fundamental aspects of chemistry and physics,

described by various branches of mathematics,

connecting their field to the topic of

this Kyoto International Symposium.

Conference of the Japanese Society for Cataract Research, 2016 July 29-31, Morioka, Japan

Japanese Society for Cataract Research, 2016 July 29-31, Morioka, Japan

When molecules of the eye turn into their mirror images and become misfits: Right hands in left gloves.

Paul G. Mezey

Canada Research Chair, Memorial University, St. John's, Canada

Education

09/1962 – 06/1970, MSc, PhD, Chemistry, Univ. Budapest, Hungary
09/1968 – 06/1972, MSc, Mathematics, Univ. Budapest, Hungary
05/1985, DSc Mathematical Chemistry, Univ. Saskatchewan, Canada

Scientific Carrier and Employment Carrier

07/1967 – 04/1973, Research Associate, Hungarian Acad. Sciences
05/1973 - 07/1977, Post-Doc. Fellow, University of Toronto, Canada
07/1977 - 06/2003, Prof. Chemistry, Univ. Saskatchewan, Canada
07/2003 - present, Canada Research Chair and Prof. Chemistry, Memorial University, St. John's, Canada
01/1990 – present, Editor, J. Mathematical Chemistry (Springer).

I have a cataract: not in my eyes, but above my head, this picture has been taken in front of a waterfall, a real "Cataract", so I have chosen to use this picture for this special occasion.


a and Vitreous Society
体学会総会
消印有効
特別講演
座長 岩手医科大学医学部眼科学講座 黒坂 大次郎 先生
白内障原因タンパク質の変化をとらえる化学的アプローチ；
白内障におけるホモキラリティの喪失
講師 京都大学原子炉実験所 放射線生命医科学研究本部 藤井 紀子 先生
招待講演
座長 京都大学原子炉実験所 放射線生命医科学研究本部 藤井 紀子 先生
When molecules of the eye turn into their mirror images and become misfits: Right hands in left gloves.
講師 Memorial University of Newfoundland Paul G.Mezey 先生


特別講演
座長 岩手医科大学医学部眼科学講座 黒坂 大次郎 先生
白内障原因タンパク質の変化をとらえる化学的アプローチ；
白内障におけるホモキラリティの喪失
講師 京都大学原子炉実験所 放射線生命医科学研究本部 藤井 紀子 先生
招待講演
座長 京都大学原子炉実験所 放射線生命医科学研究本部 藤井 紀子 先生
When molecules of the eye turn into their mirror images and become misfits: Right hands in left gloves.
講師 Memorial University of Newfoundland Paul G.Mezey 先生
教育講演1
座長 金沢医科大学医学部眼科学講座 佐々木 洋 先生
DNA 構造と機能を制御するケミカルバイオロジー
講師 京都大学理学研究科化学専攻生物化学分科 杉山 弘 先生
教育講演2
座長 横浜薬科大学臨床薬理学研究室 岡 美佳子 先生
代謝酵素やトランスポーターのバラツキを用い臨床試験結果を予測する
講師 横浜薬科大学臨床薬理学研究室 千葉 康司 先生
座長

I am a chemist and mathematician,

I am definitely not an expert on cataract

(although my Father, with exactly the same name, was an expert,

performing eye-surgery, operating on cataract patients).

My research is on some fundamental questions of chemistry:

global shape and local shape of electron density clouds of molecules,

approximate symmetry, symmetry deficiency,

Molecular chirality problems and chirality changes.

Since I study Fundamental Chemistry,

it is a special honour for me to talk at Kyoto University.

For many decades, I have had strong connections to Japanese Science,

especially, to Kyoto University,

with Chemistry Prof. Akitomo Tachibana and others

and it is wonderful that I may strengthen this

from a new scientific perspective of Prof. Noriko Fujii

For 8 years, I was a Foreign Member of

The Fukui Institute for Fundamental Chemistry, Kyoto,

directed by Professor Kenichi Fukui

I have the honour to have 7 publications

with Prof. Kenichi Fukui and Prof. Shigeru Arimoto as co-authors.

These papers were on various topics, but the main, official topic was

“Prebiotic Formation of Definite Sequence Biopolymers”,

a topic well fitting with the current collaboration with Prof. Noriko Fujii


井謙一博士
その人と学問
平成10年 4月10日(金)
～5月23日(土)
京都大学総合博物館


Mt. Hiei, 1994

Molecular Shape and Chirality,

The Role of Molecular Fragments in Biochemical Effects,

Local and Global Shape,

Local and Global Chirality

All molecular properties are determined by the *shape of electron density clouds,* forming the actual *"bodies" of molecules.*

Molecular shapes which are mirror images but not superimposable on each other, just as left and right hands, are called *"chiral"*.

Such *chiral molecule pairs* have equal energies and equal stabilities, but they *fit differently* into various surroundings.

The left handed variants of amino acids dominate in life forming molecules on Earth, and the right-handed forms are "misfits" in living tissues.

The purpose of this research is the

detailed understanding of processes causing

left-to-right mutations of amino acids in peptides,

some of which processes do contribute to

cataract in the eye.

This research has a focus on *shape-based*

chirality measures of the natural amino acids,

to find relations between such measures and the observed

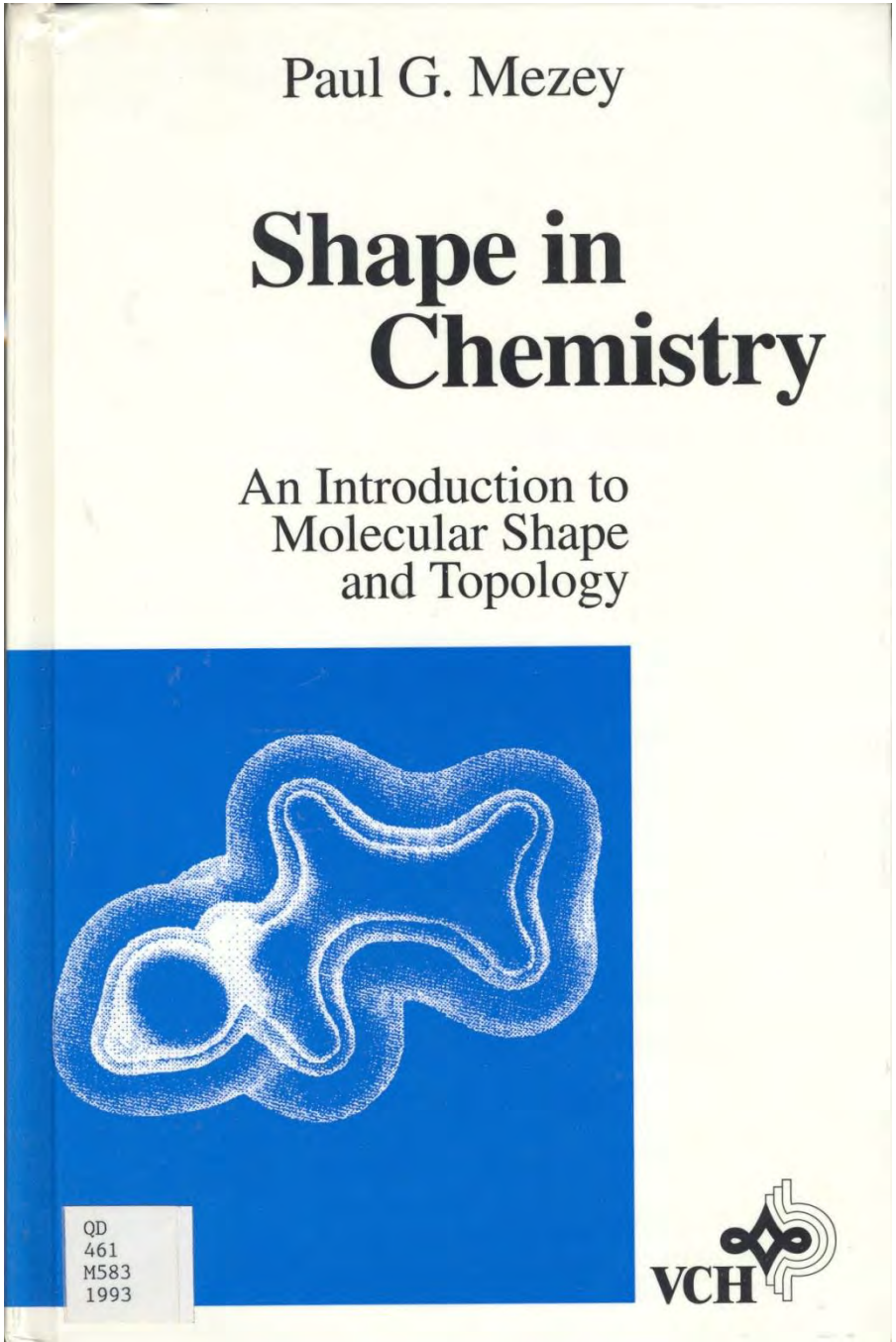
variations in the L to D, ``left to right`` changes,

we call ``enantiomerization``,

of these compounds, with often

significant biochemical, and even health effects.

Molecular Shape Analysis Methods



Shape Analysis, Paul G. Mezey

Some fundamental concepts

Molecular Shape: the shape of fuzzy electron density cloud

Molecular Similarity: similarity measures

Molecular Symmetry, perfect symmetry, imperfect symmetry

Symmetry Deficiency Measures:

The degree of similarity between imperfect and perfect symmetry

Chirality: non-superimposibility of mirror images,

Some symmetries are missing: mirror planes

and rotation-reflection symmetries of the S_{2n}-type

Chirality Measures

How can we study the shapes of molecules ?

What is a molecule?

Some atomic nuclei, and an electron density cloud.

There is nothing else in a molecule, only nuclei and electrons.

The world on that microscopic level of molecules is very different from what we are used to:

quantum mechanics plays a far more important role,

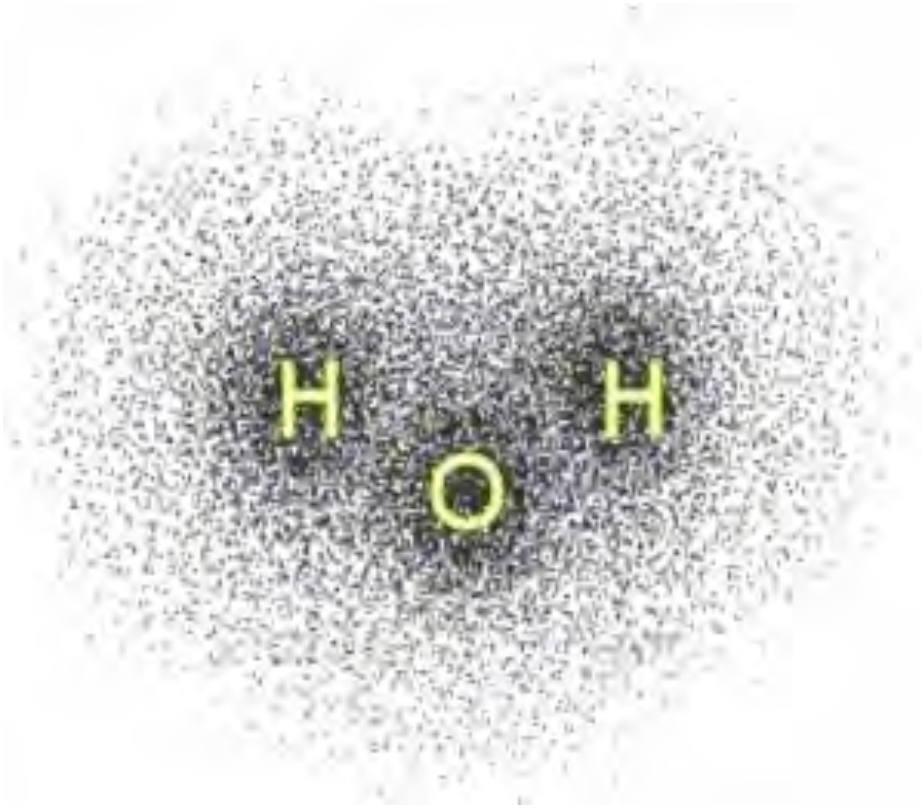
and nuclei and electrons have both particle and wave properties.

Nuclei are more like small particles, electrons are more like waves forming a fuzzy cloud.

Electrons in a molecule are best imagined as clouds.

The shape of a molecule is the shape of its electron density cloud

Think of water, the molecule of formula H_2O, or H – O – H with the bonds indicated. How does a water molecule really look like?



Let us take a look first at a sketch of a single water molecule, H_2O.

There are three positively charged centers, the nuclei of the atoms of oxygen and two hydrogens, and a fuzzy, negatively charged electron density cloud, that is thick near the nuclei and is thinning out at larger distances from the nuclei.

It is easy to describe the shape of a potato: the potato has a skin, a boundary, a given geometry, that can be depicted easily.

BUT:

How to describe the shape of a cloud, that has no boundary …..

.

The solution: use isocontours, that is, contours, where the cloud has the same density, called isodensity contours.

An isocontour surface is the collection of points where the density value is the same constant, say, constant a.

This is similar to the two-dimensional idea of describing the *shape of mountains* on a map, using lines of equal heights, *isolevels*, which are continuous lines forming loops on the map.

For the three-dimensional clouds, one has isocontour surfaces

Some constant electron density contours, “isosurfaces” of water at density values of 0.001, 0.01, 0.10, 0.30, and 0.40 a.u. (atomic units)

Atomic unit for electron density: 1 electron/bohr^3

Water, H2O

Just like a set of nested Russian wooden dolls:

open one doll, there is another inside,
open that one, there is yet another inside of that,
and so on ….

however, the electron density changes continuously,

so in this case there are infinitely many dolls !

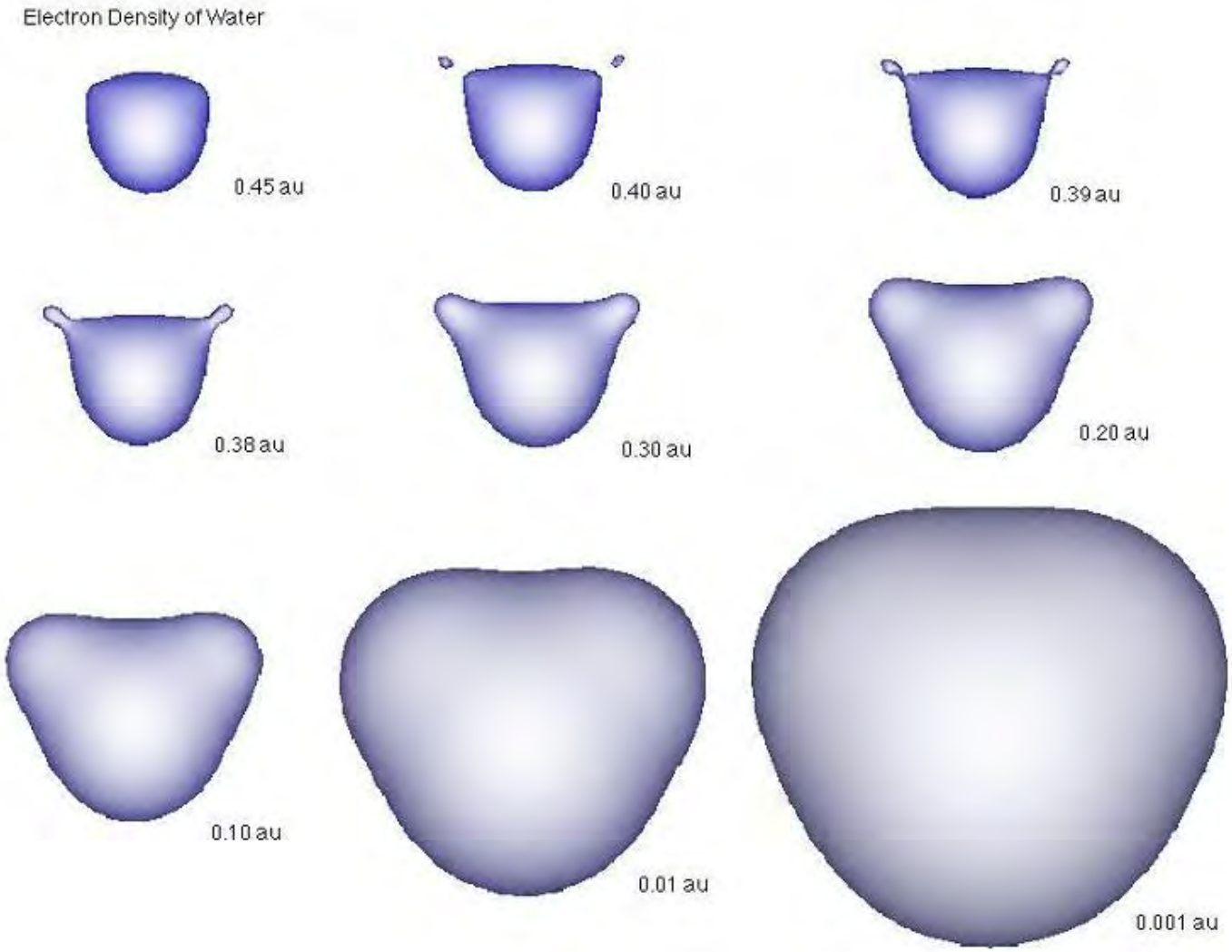
MIDCO

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

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

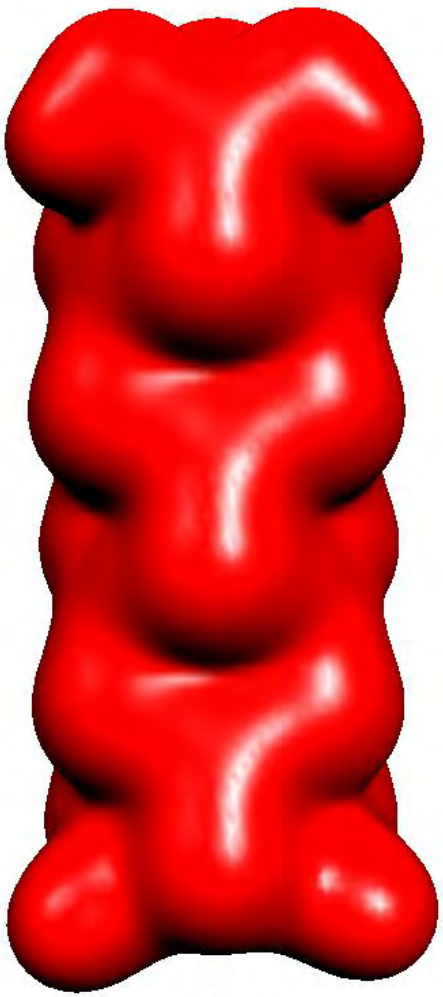
that is, as the collection of all points **r** of the 3D space where the electronic density $\rho(K,\mathbf{r})$ is equal to the threshold value a.

A *molecular isodensity contour surface,* MIDCO G(K,a) of nuclear configuration K and density threshold a is defined as G(K,a) = { r : ρ(K,r) = a }, that is, as the collection of all points r where the electronic density ρ(K,r) is equal to the threshold value a.



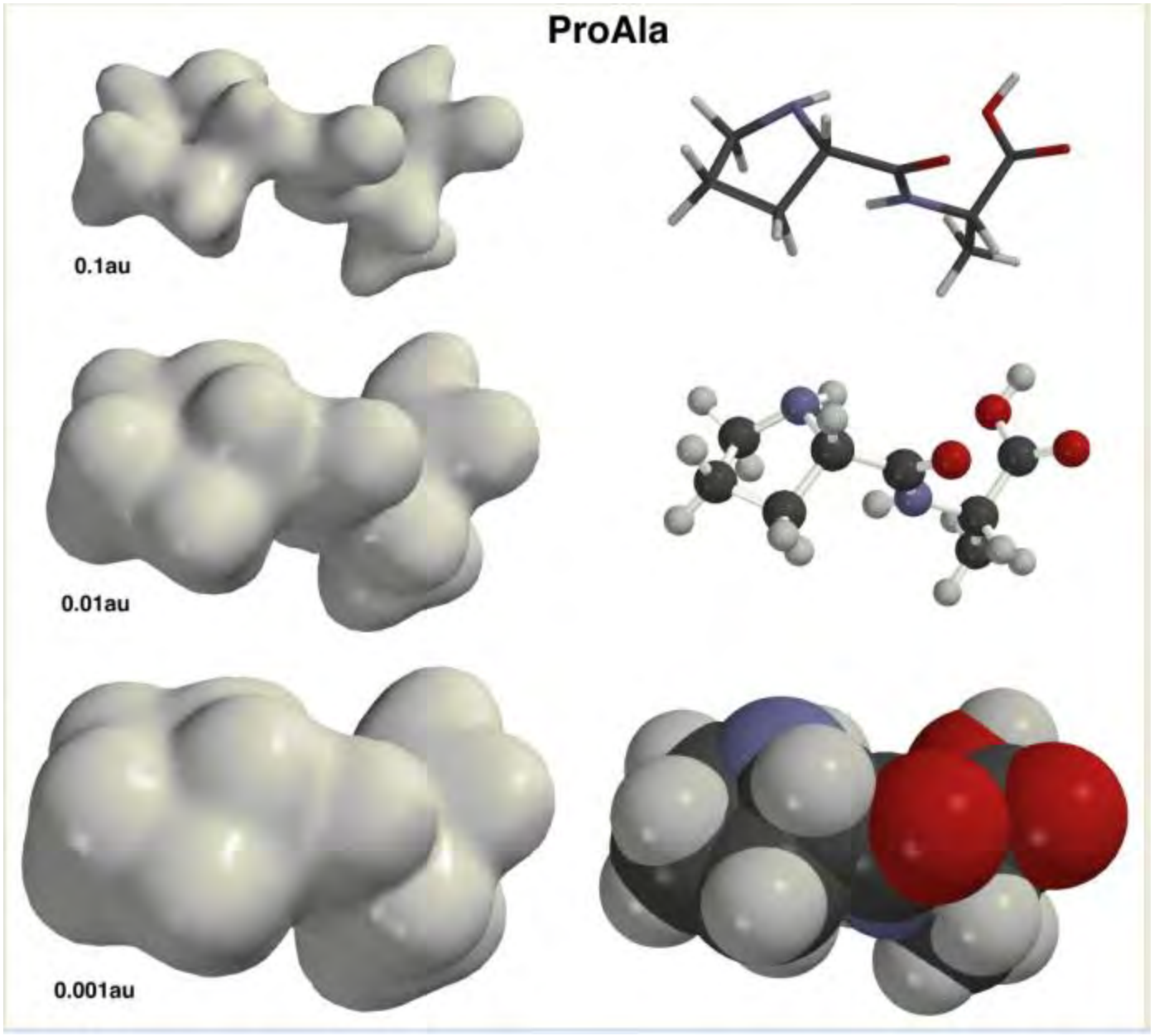
Molecular isodensity contour (MIDCO) surfaces of BNNNs

- ✓ Atomic range
- ✓ Bonding range
- ✓ Skinny molecular range
- ✓ Corpulent molecular range



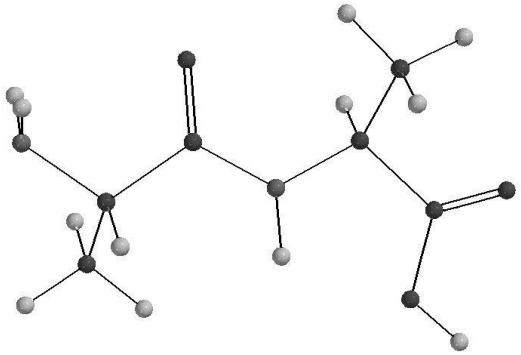
Closer to our amino acid and peptide problems:

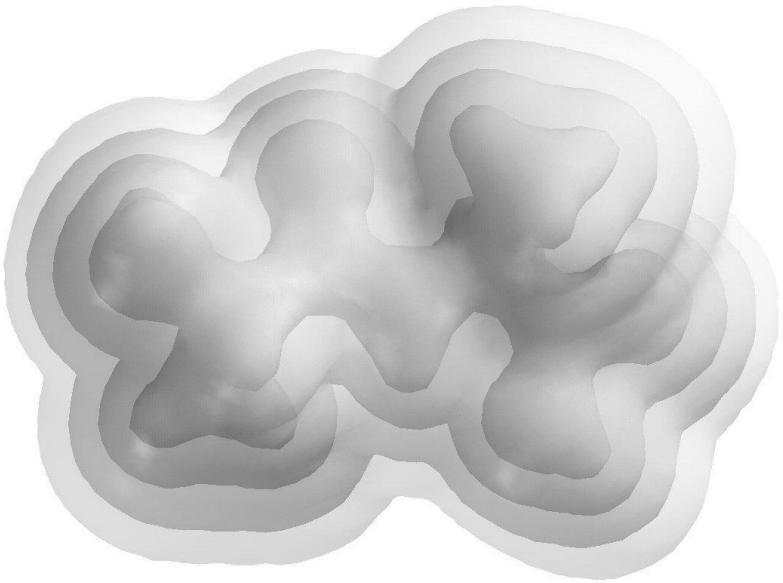
what is the shape of a dipeptide …

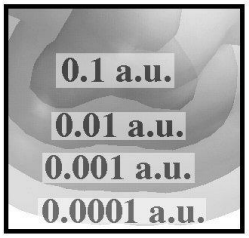

ProAla
0.1au
0.01au
0.001au

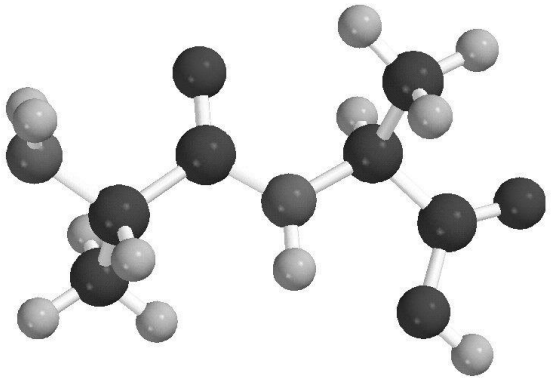
Alanylalanine, structural models and actual electron density cloud

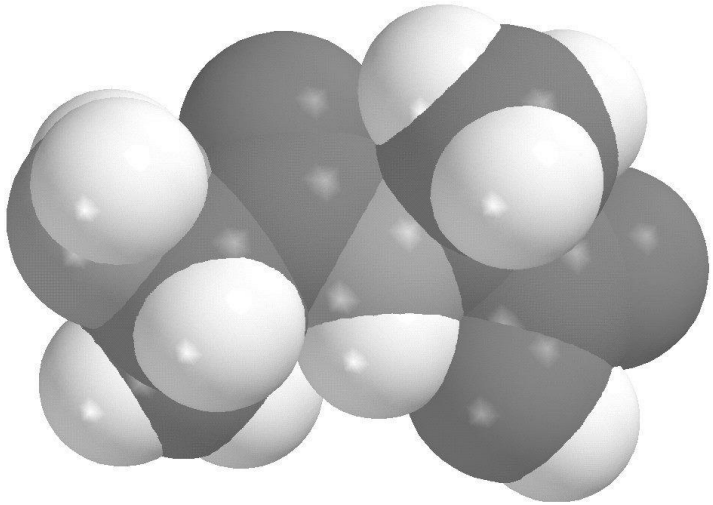
ALA-ALA



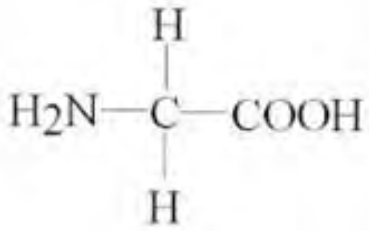




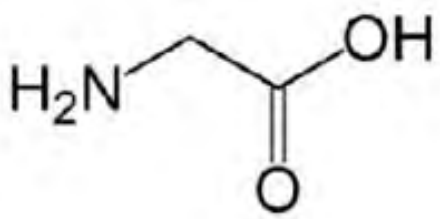


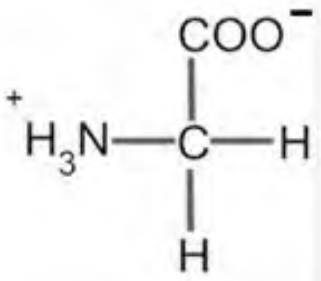


Images of Glycine Models



glycine

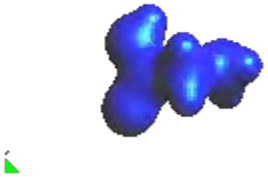


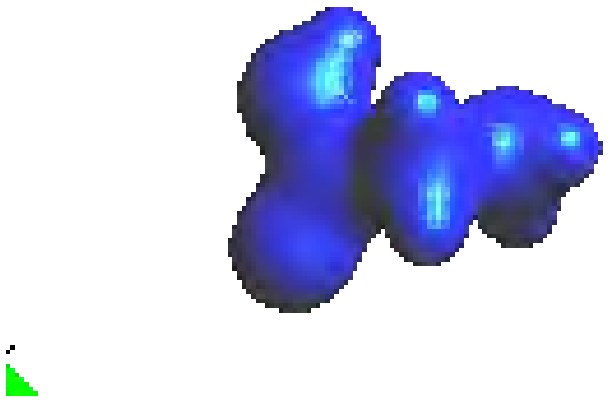


Reality: Electron Density Cloud

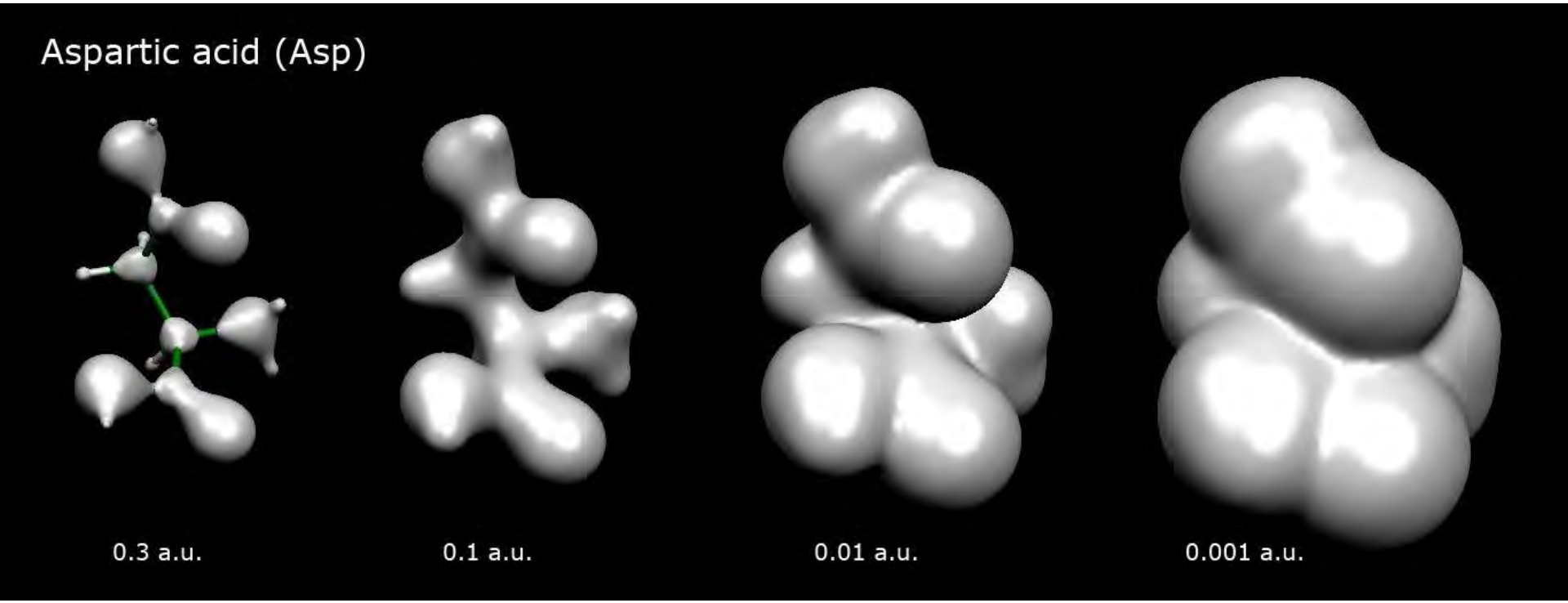
Electron Density Isocontours G(a) [connecting points of the same density value a] of Glycine

high density contours on the left, low density contours on the right



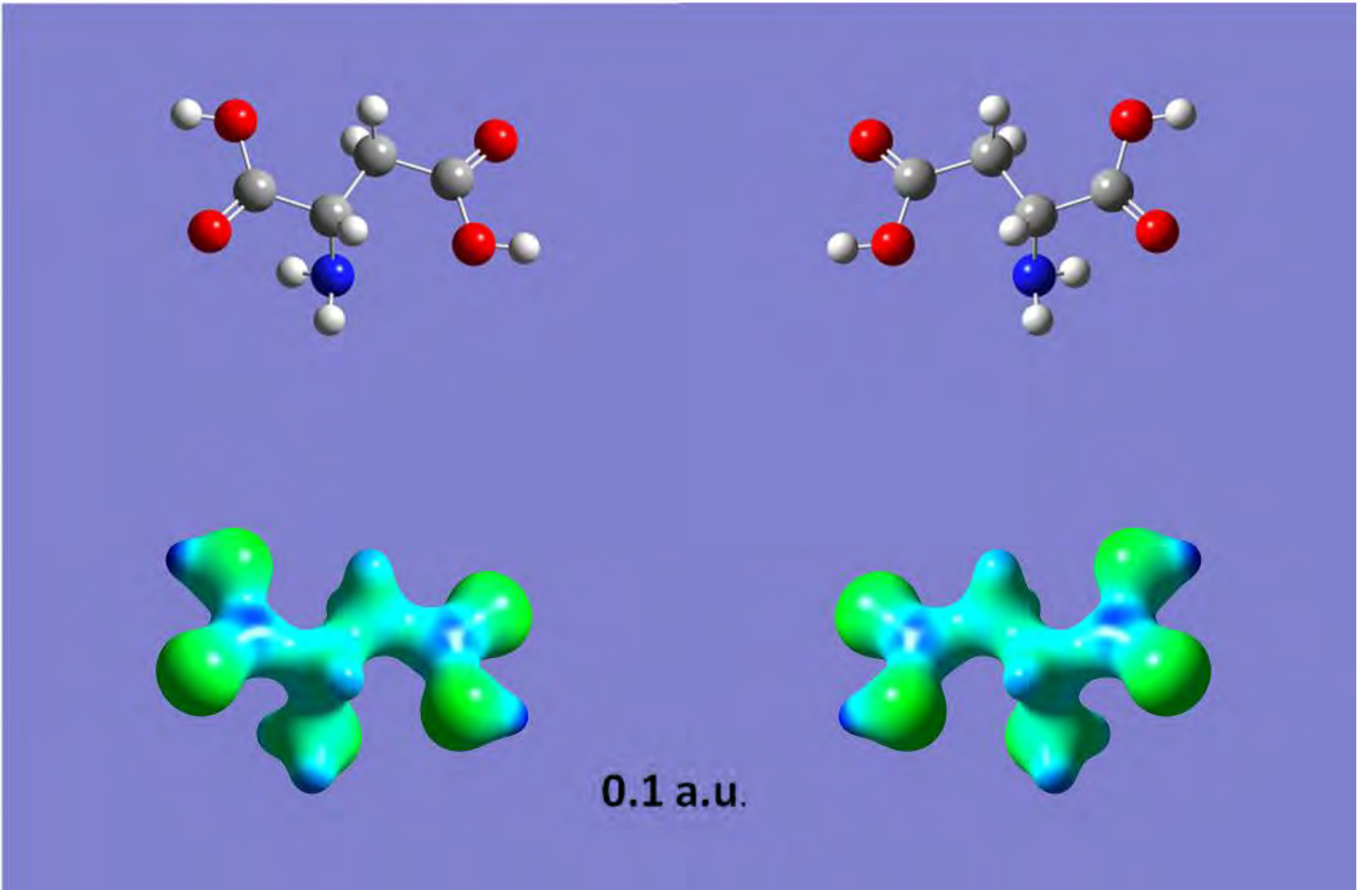


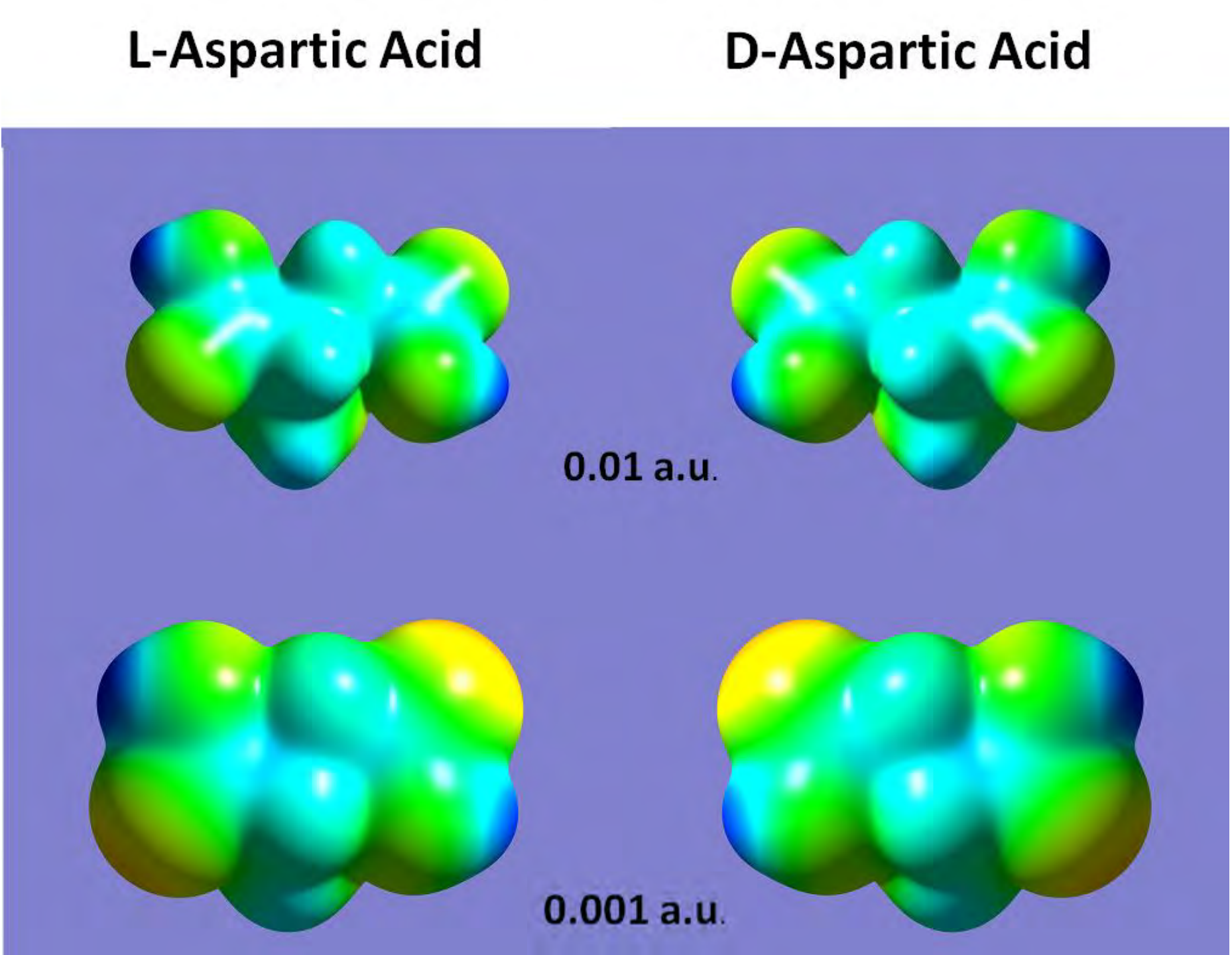
Electron density isocontours of Aspartic Acid (Asp)



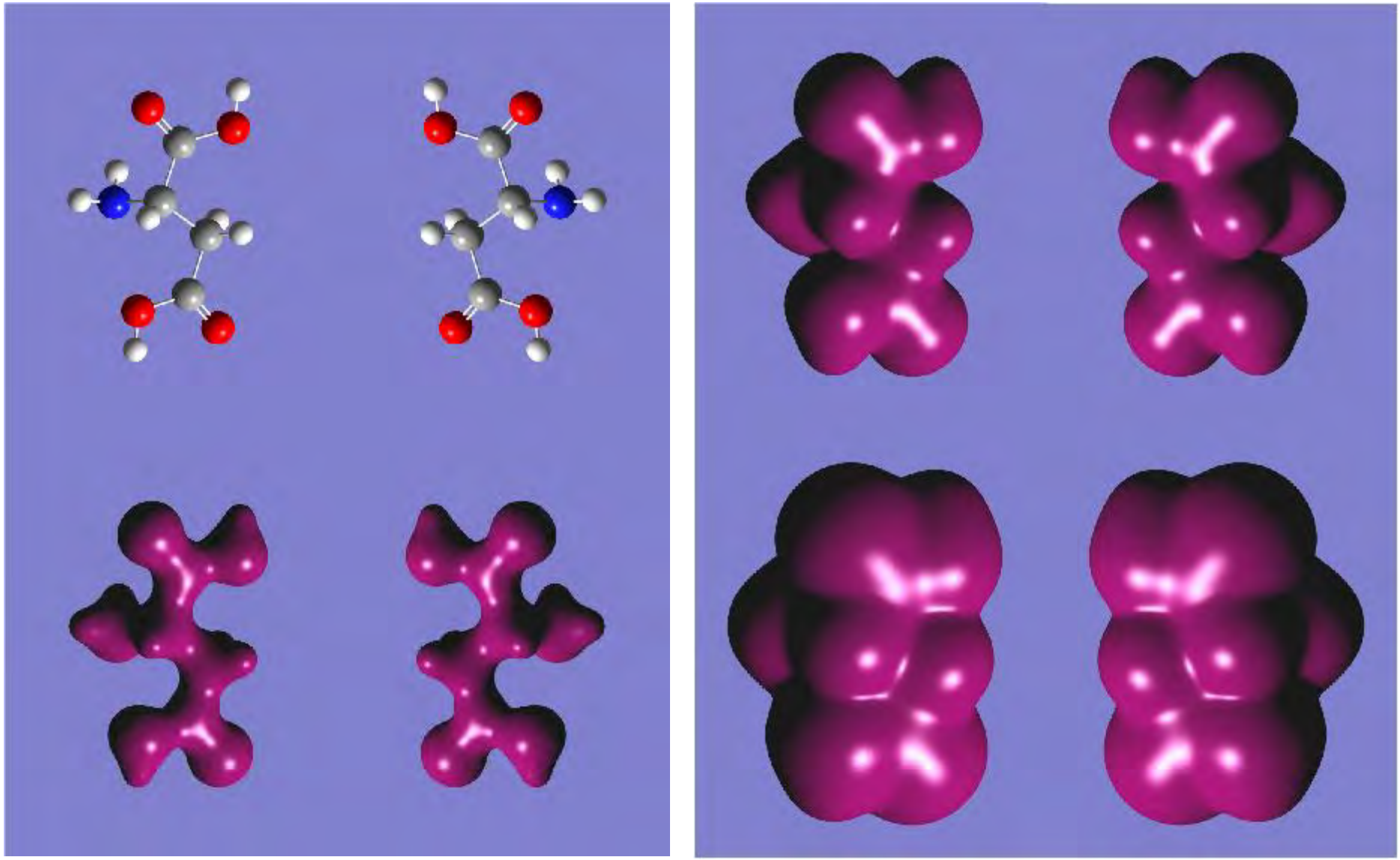
L-Aspartic Acid

D-Aspartic Acid




L-Aspartic Acid
D-Aspartic Acid
0.01 a.u.
0.001 a.u.

L-Aspartic Acid D-Aspartic Acid L-Aspartic Acid D-Aspartic Acid



Local Shape and Global Shape,
Local Symmetry and Global Symmetry,
Local Chirality and Global Chirality:

The Shape of Molecules
and the
Shape of Molecular Fragments

.

PCCP

Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2012, **14**, 8516–8522

www.rsc.org/pccp

PAPER

Natural molecular fragments, functional groups, and holographic constraints on electron densities†

Paul G. Mezey*

Received 23rd January 2012, Accepted 27th April 2012
DOI: 10.1039/c2cp40237h

One of the tools of the shape analysis of molecular electron densities, the Density Threshold Progression Approach used in Shape Group studies can also serve as a criterion for the selection of "natural" molecular fragments, relevant to functional group comparisons, reactivity studies, as well as to the study of levels of relative "autonomy" of various molecular regions. The relevance of these approaches to the fragment-based studies of large molecules, such as biopolymers and nanostructures is emphasized, and the constraints represented by the holographic electron density theorem to this and alternative recent fragment approaches are discussed. The analogies with potential energy hypersurface analysis using the Energy Threshold Progression Approach and connections to level set methods are discussed, and the common features of these seemingly distant problems are described.

Paul G. Mezey, Acc. Chem. Res. 2014, 47, 2821−2827

ACCOUNTS of chemical research

Article

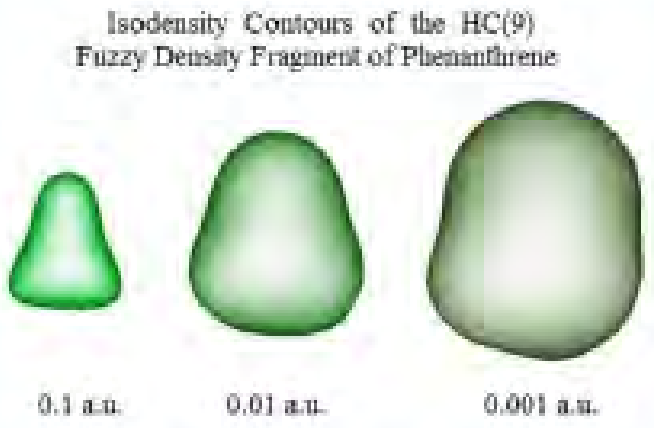
pubs.acs.org/accounts

Fuzzy Electron Density Fragments in Macromolecular Quantum Chemistry, Combinatorial Quantum Chemistry, Functional Group Analysis, and Shape−Activity Relations

Paul G. Mezey*

Scientific Modelling and Simulation Laboratory, Department of Chemistry, Memorial University of Newfoundland, St. John's, Newfoundland A1B 3X7, Canada

CONSPECTUS: Just as complete molecules have no boundaries and have "fuzzy" electron density clouds approaching zero density exponentially at large distances from the nearest nucleus, a physically justified choice for electron density fragments exhibits similar behavior. Whereas fuzzy electron densities, just as any fuzzy object, such as a thicker cloud on a foggy day, do not lend themselves to easy visualization, one may partially overcome this by using isocontours. Whereas a faithful representation of the complete fuzzy density would need infinitely many such isocontours, nevertheless, by choosing a selected few, one can still obtain a limited pictorial representation. Clearly, such images are of limited value, and one better relies on more complete mathematical representations, using, for example, density matrices of fuzzy fragment densities. A fuzzy density fragmentation can be obtained in an exactly additive way, using the output from any of the common quantum chemical computational techniques, such as Hartree−Fock, MP2, and various density functional approaches.



A Theorem on Molecular Fragments:

The Holographic Electron Density Theorem

Historical notes on the Holographic Electron Density Theorem

The Hohenberg-Kohn Theorem [1] refers to the *complete electron density*, stating that all non-degenerate, ground state molecular properties are determined by the complete electron density.

For *artificial, bounded systems* a similar result on the relation between the part and the whole has been proven earlier [2], by J. Riess and W. Münch. *Their proof was not applicable for real, boundaryless molecules.* This limitation of the result, although clearly stated by these authors, has not always been fully recognized by some later papers quoting the result.

The "Holographic Electron Density Theorem"[23] deduces complete information from the *part*, stating that any nonzero volume *part* of a molecular electron density in a non-degenerate ground state contains the *complete* information about all properties of the entire, boundaryless molecule.

That is, the complete molecular information does not require the complete electron density, and local electron density ranges already fully determine all molecular properties.

1. **P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Phys. Rev. 136, B864-B871 (1964).**
2. **J. Riess and W. Münch, "The Theorem of Hohenberg and Kohn for Subdomains of a Quantum System", Theor. Chim. Acta, 58, 295-300 (1981).**
3. **P.G. Mezey, "The Holographic Electron Density Theorem and Quantum Similarity Measures", Mol. Phys., 96, 169-178 (1999).**

A theorem on a universal relation between the part and the whole:

MOLECULAR PHYSICS, 1999, VOL. 96, NO. 2, 169–178

The holographic electron density theorem and quantum similarity measures

PAUL G. MEZEY

Institute for Advanced Study, Collegium Budapest, 1014 Budapest, Szentháromság u. 2, Hungary, and Mathematical Chemistry Research Unit, Department of Chemistry and Department of Mathematics and Statistics, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9

(*Received 26 January 1998; accepted 12 March 1998*)

How much information about the complete molecule is present in a part of the molecule? Quantum similarity measures provide comparisons between molecular electron densities based on integration over the whole space. Such integration involves boundaryless electron densities, whereas an early application of the Hohenberg–Kohn theorem to local subsystems of molecules requires these molecules to be confined to bounded, finite regions of the space. However, actual molecules have no boundaries, they are not confined to any finite region of the space. In order to find deterministic relations between local and global, boundaryless electron densities, and to classify the link between quantum similarity measures involving the full space and local subsystems, the unique extension property called the holographic property of subsystems of complete, boundaryless electron densities is established. Any nonzero volume piece of the ground state electron density completely determines the electron density of the complete, boundaryless molecule.

Electron density and the information content of molecules

The Hohenberg-Kohn Theorem, 1964: ***The molecular electron density determines the molecular energy and through the Hamiltonian, all other molecular properties.***
(Walter Kohn, Nobel Prize in Chemistry, 1992).

That is, the electron density is the fundamental information storage of molecules, the electron density actually carries the complete information about the molecule.

P.Hohenberg and W.Kohn, "Inhomogeneous electron gas", Phys.Rev.136,B864-B871(1964).

The holographic information content in *parts* of molecules:

"The Holographic Electron Density Theorem", P.G. Mezey 1999:
Any nonzero volume* part *of a molecular electron density in a non-degenerate ground state contains the* complete *information about all properties of the entire, boundaryless molecule.

P.G. Mezey, "The Holographic Electron Density Theorem and Quantum Similarity Measures", Mol. Phys., 96, 169-178 (1999).

This fundamental property of all molecules applies to all exhibited and also to all latent molecular properties.

Girona, 2005, Walter Kohn and Paul G. Mezey

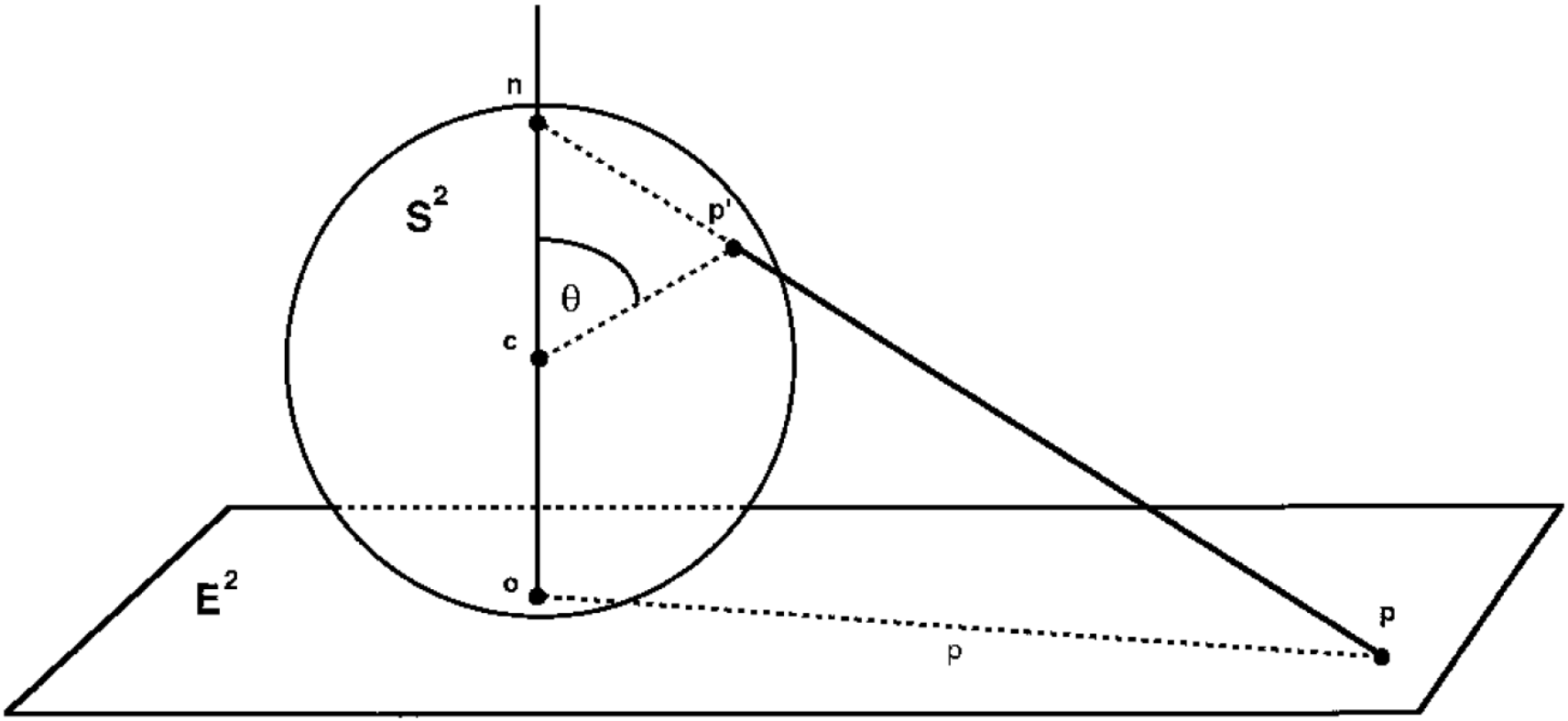


Figure 1. Three-dimensional illustration of the Alexandrov one-point compactification method, as applied to the two-dimensional plane E^2, resulting in the sphere S^2. Point **p** of distance p from the origin **o** is assigned to the unique point $\mathbf{p}'$ on the sphere S^2 using the straight line issued from the north pole **n** of sphere S^2 to point **p**. The polar angle θ of point $\mathbf{p}'$ is also indicated. A four-dimensional version of this method leads to a complete molecular electron density function that is analytical almost everywhere on a sphere S^3.

From Local Information to Global Representation:

Additive Fuzzy Density Fragment (AFDF) Approaches

Quantum Chemical Computations to Generate Exactly Additive Fuzzy Molecular Fragments

A Fragment Density Matrix Approach to Linear-Scaling Macromolecular Quantum Chemistry:

The MEDLA and ADMA Methods

The pioneer of density partitioning (he called it population analysis), R. Mulliken, near the crater of Teide, Spain, 1976

The Additive Fuzzy Density Fragmentation (AFDF) Principle.

Based on a scheme analogous to Mulliken's Population Analysis

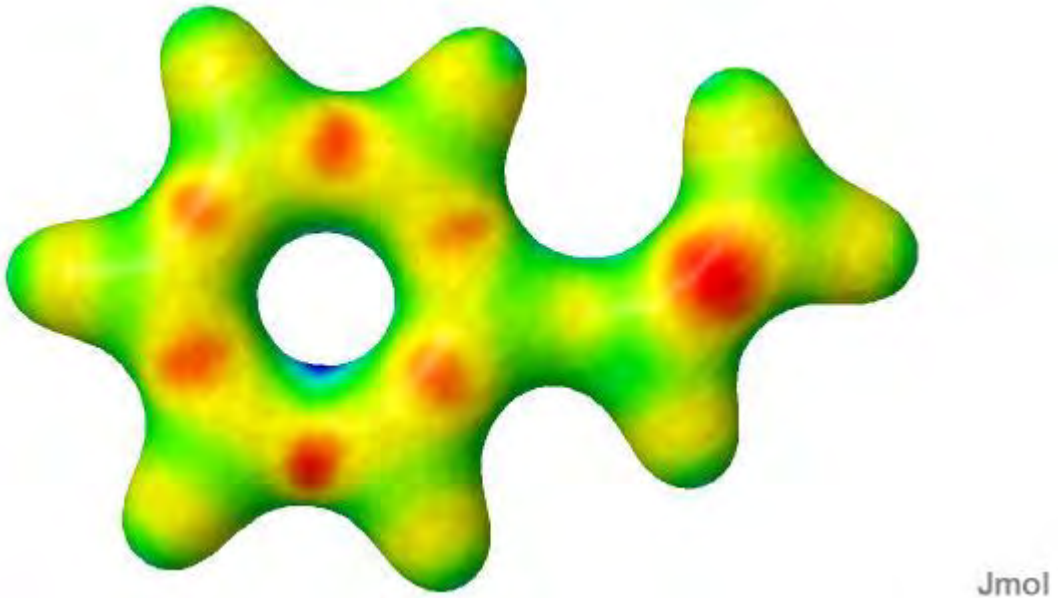
without integration.

P.D. Walker and P.G. Mezey, Molecular Electron Density Lego Approach to Molecule Building, *J. Amer. Chem. Soc.*, 115, 12423-12430 (1993).
[MEDLA reference]

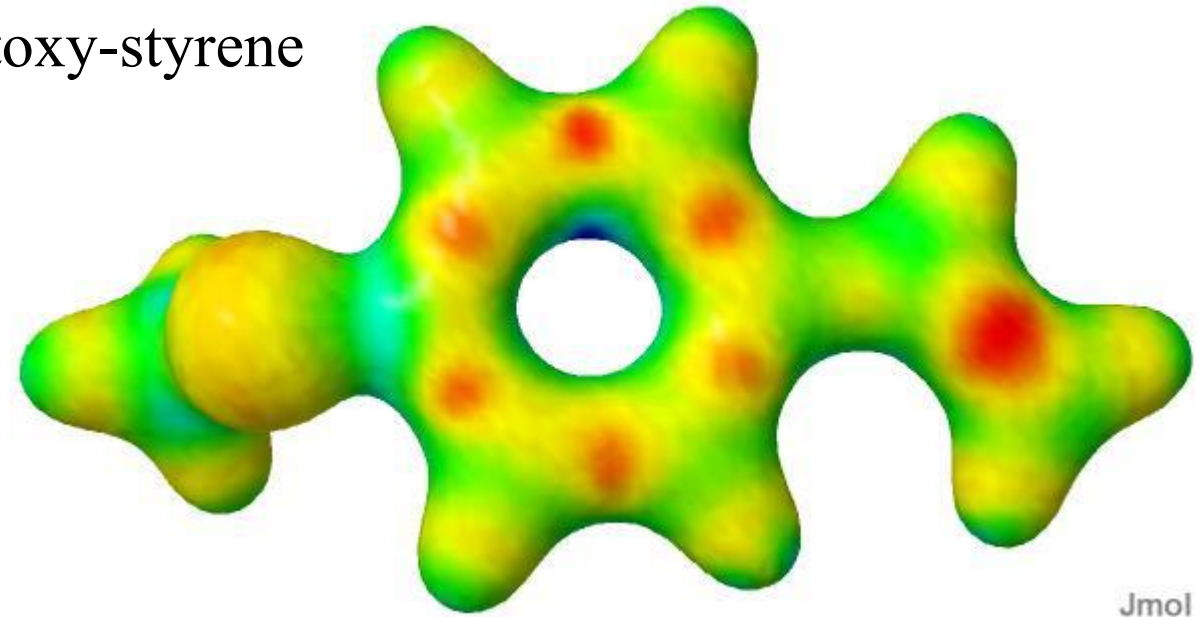
P.G. Mezey, "Macromolecular Density Matrices and Electron Densities with Adjustable Nuclear Geometries", *J. Math. Chem.*, 18, 141-168 (1995).
[ADMA reference]

First just one practical motivation for studying molecular fragments: Through-bond and through-space interactions within molecules.

Styrene



and, e.g., para-metoxy-styrene



Some practical motivation for the study of molecular fragments: Through-bond and through-space interactions within molecules.

In most cases, it is hard to study these interactions separately, but

in some special cases, it is possible.

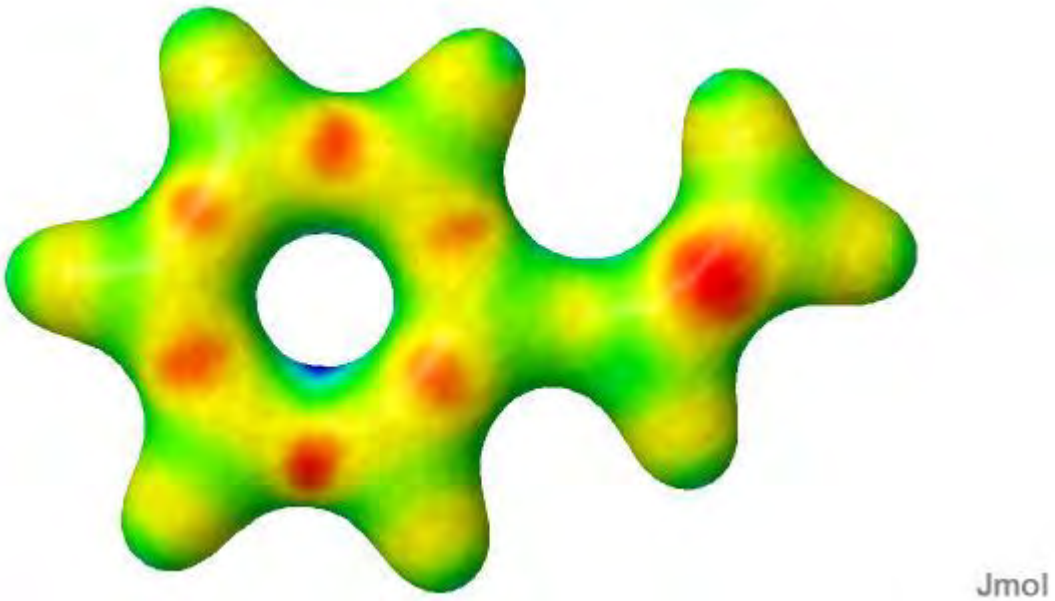
The case of para-substituted styrene molecules, two approaches:

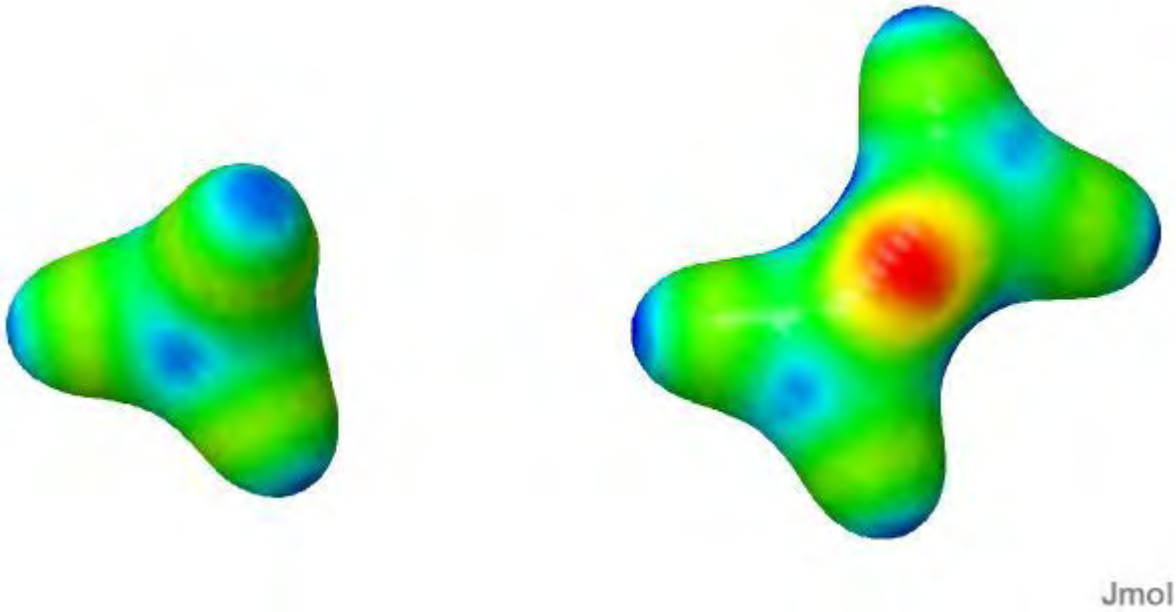
Study the fragment shape changes for the vinyl group,
as influenced by the para substituent

(a) In the complete molecule
(b) In a pair of molecules where the benzene ring is "left out"

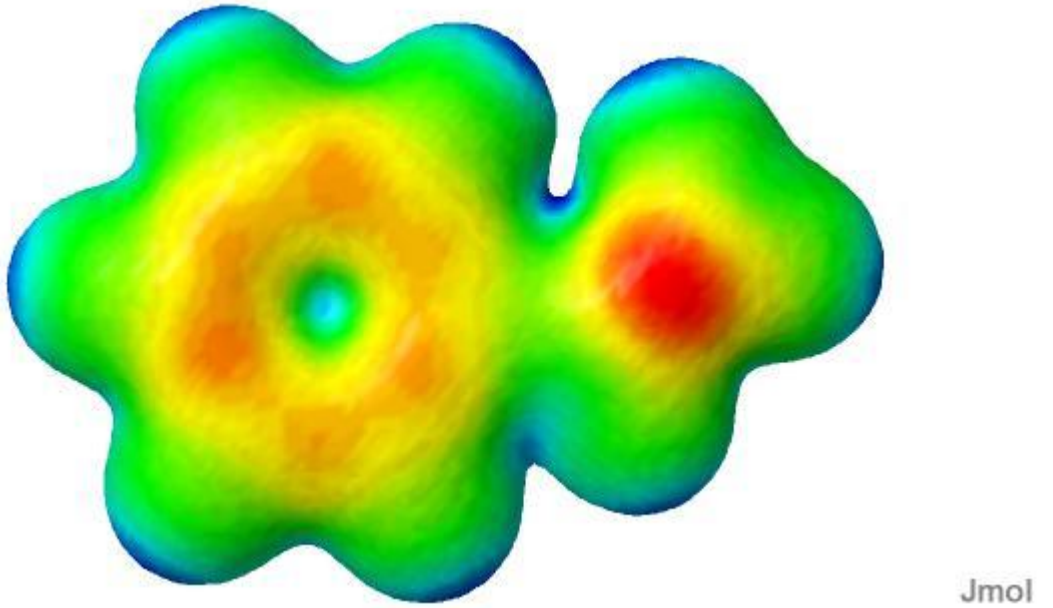
Case (a) includes both through-bond and through-space interactions
Case (b) includes only through-space interactions.

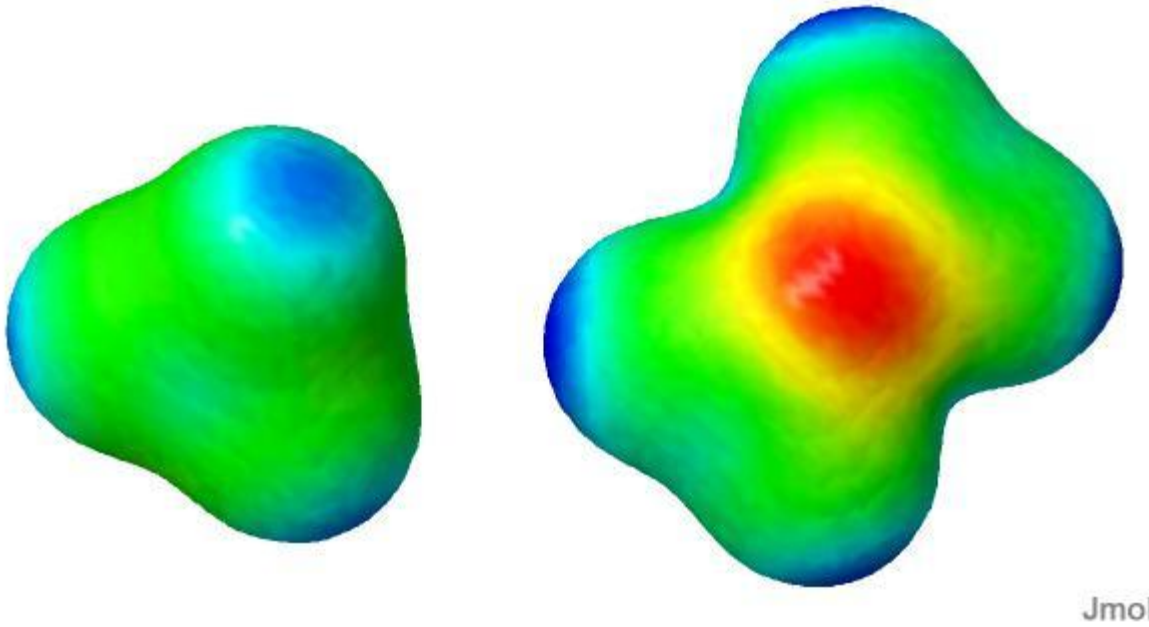
Styrene 0.1 a.u.
isodensity contours



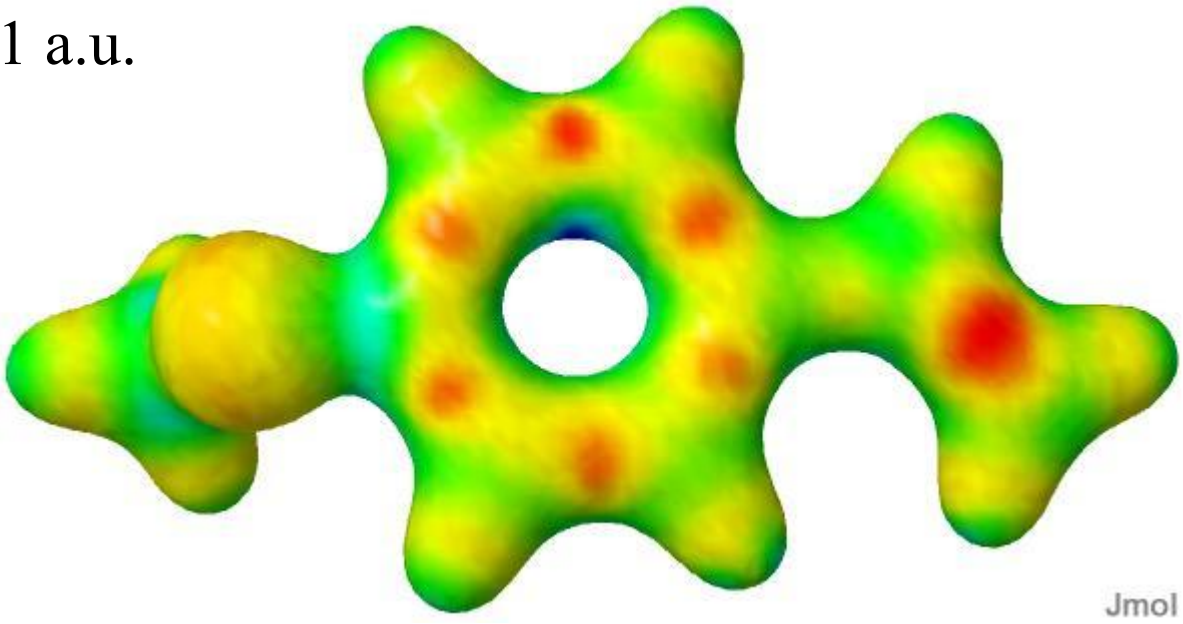


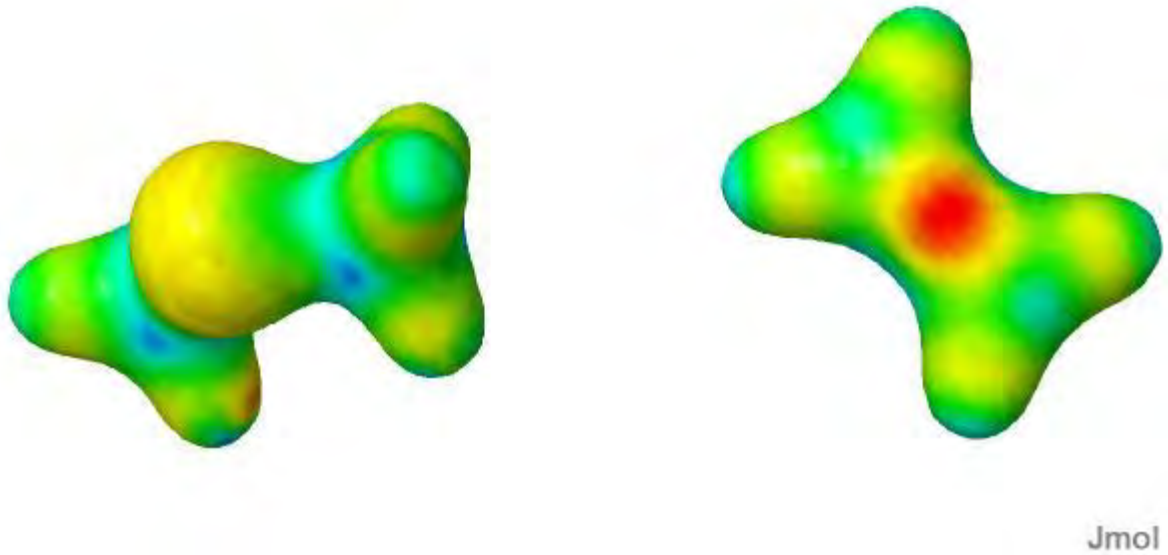
Styrene 0.01 a.u.
isodensity contours



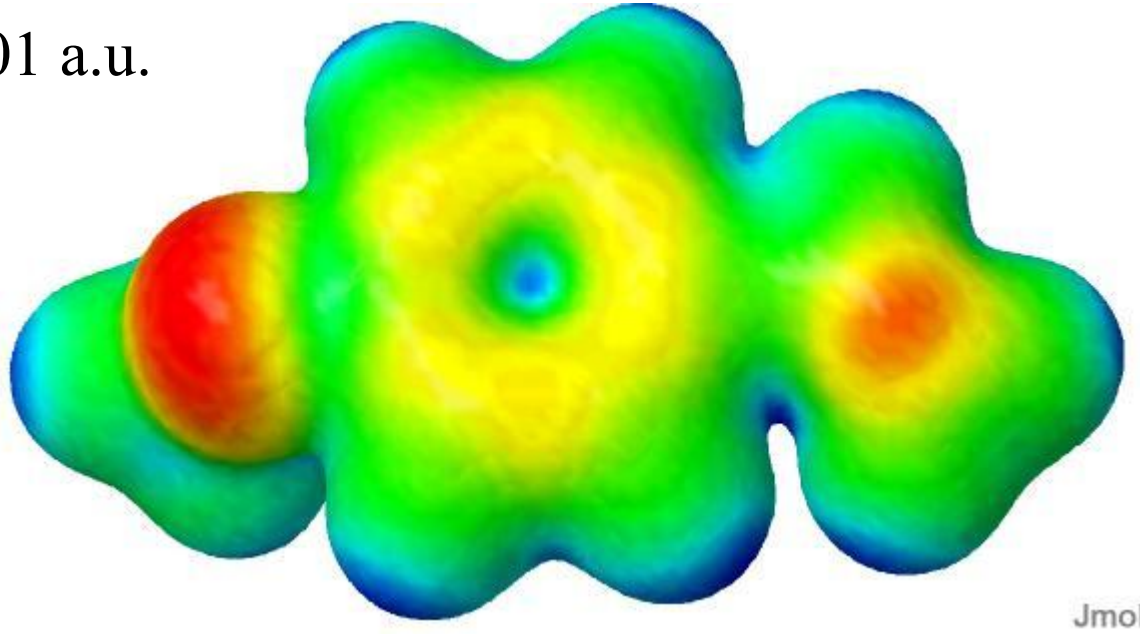


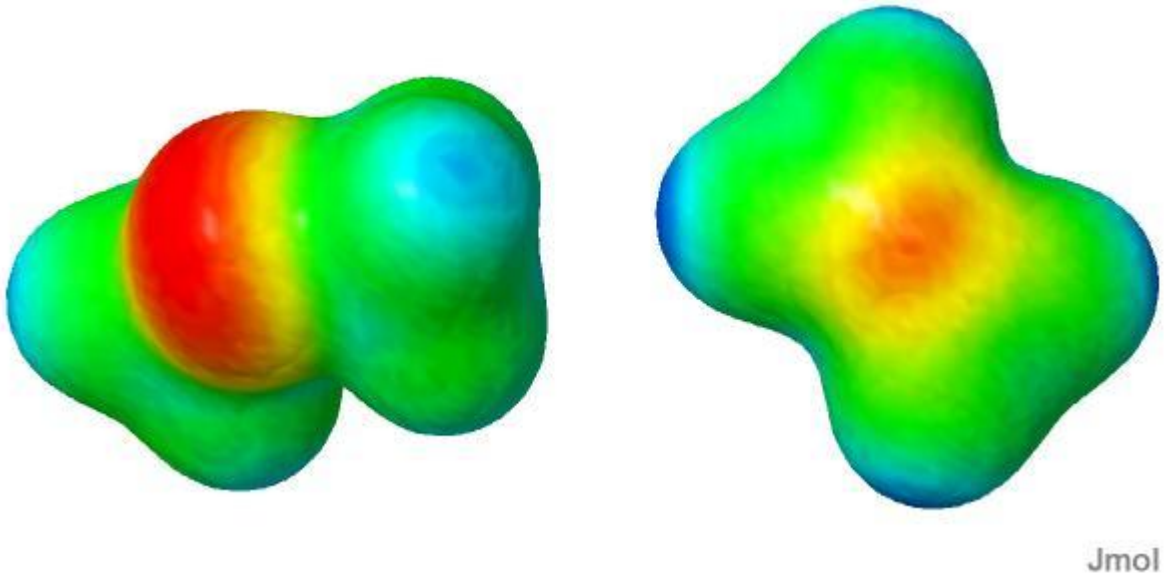
Para-metoxy-styrene 0.1 a.u.
isodensity contours





Para-metoxy-styrene 0.01 a.u.
isodensity contours





Symmetry Deficiency,

Similarity Measures and Chirality Measures

Some general approaches to

approximate symmetry

and similarity analysis

Symmetry Deficiency Measures

- **A set is an R-set if it has the symmetry element R**

- **B is an R-subset of a set A if B is a subset of A, and B is an R-set**
- **B is a maximum volume R-subset of a set A if B is an R-subset of A and if volume V(B) is maximum among all R-subsets of A**

  (Note that, while B is not necessarily unique, V(B) is).

- **C is an R-superset of a set A if C is a superset of A (A is a subset of C), and C is an R-set**
- **C is a minimum volume R-superset of a set A if C is an R-superset of A and if volume V(C) is minimum among all R-supersets of A**

  (Note that, while C is not necessarily unique, V(C) is).

- **The internal R-symmetry deficiency measure, ISD(A,R) of a set A is**

  $ISD(A,R) = 1 - V(B)/V(A)$,

  where B is a maximum volume R-subset of A.

- **The external R-symmetry deficiency measure, ESD(A,R) of a set A is**

  $ESD(A,R) = 1 - V(A)/V(C)$,

  where C is a minimum volume R-superset of A.

Similarity and Dissimilarity

Similarity Measures of Objects

Similarity as the basis of explanation and prediction

The fundamental role of similarity:

Structural similarity
as a likely indication of
functional similarity

Similar molecules are expected to have similar effects

HOW TO STUDY SIMILARITY?

All aspects of human understanding involves the concept of similarity.

We recognize things based on similarity

We understand things based on similarity

We make deductions based on similarity

We make plans based on similarity

HOW TO STUDY SIMILARITY?

All aspects of human understanding involves the concept of similarity.

There are only two possible outcomes of comparisons.

One outcome of a comparison, seldom if ever attained: equality

The other outcome of a comparison, almost always: difference

Similarity is a concept attempting to bridge these two classes.

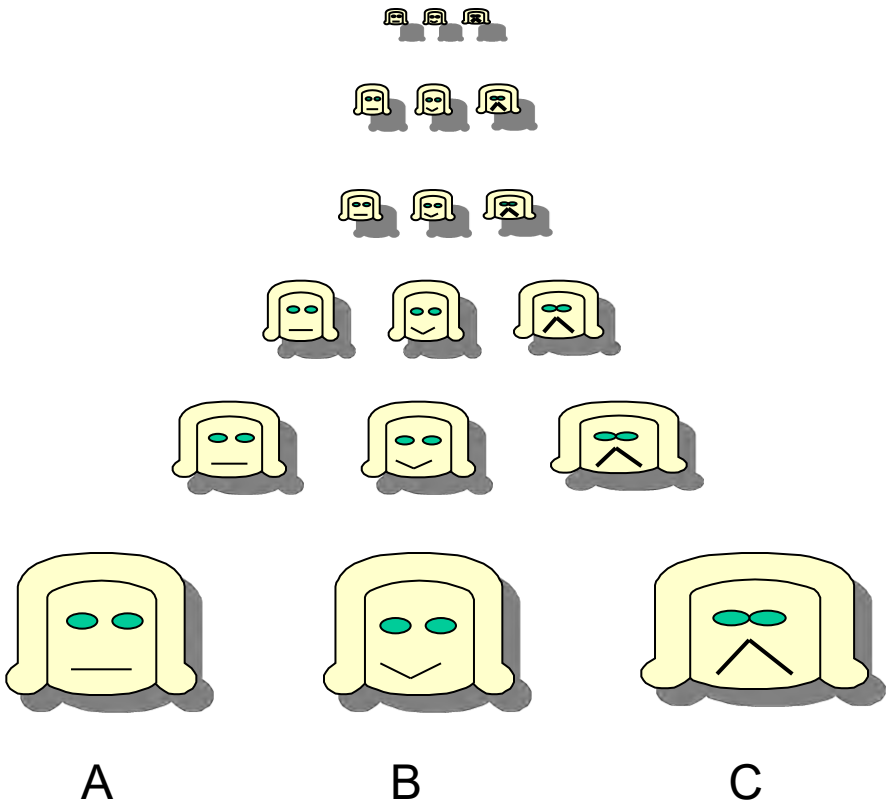
How can one quantify the degree of being different, or conversely, the degree of similarity?

SIMILARITY MEASURES

HOW TO STUDY SIMILARITY?

Resolution Based Similarity Measures

At what distance can we distinguish them?
Or, at what resolution can we distinguish them?



From these three people, C can be distinguished from the other two at a longer distance,
than the distance at which A can be distinguished from B.

Hence, A is more similar to B than either is to C.

The higher the resolution required to distinguish them, the greater the similarity.
One can use the level of resolution needed to distinguish as a similarity measure.
Resolution can be characterized by a number,
hence we have a numerical similarity measure

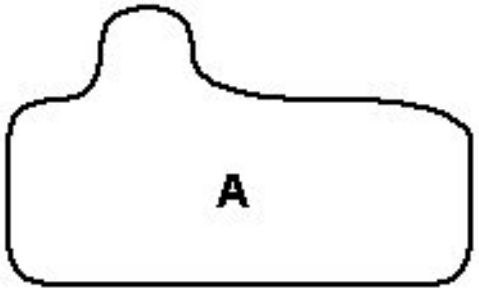
HOW TO STUDY SIMILARITY?

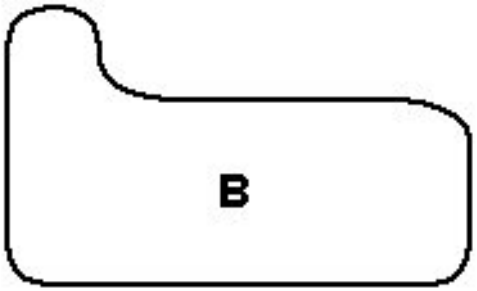
Scaling – Nesting Similarity Measures, SNSM

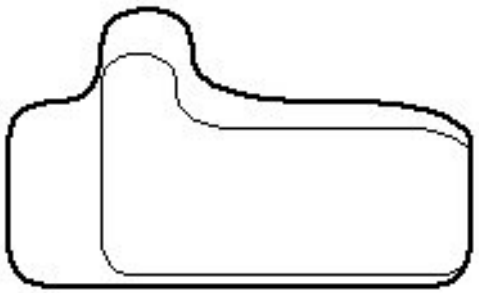
HOW TO STUDY SIMILARITY?

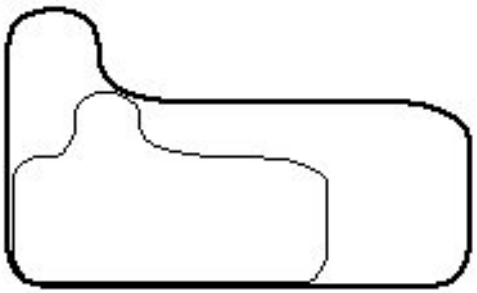
What maximal scaling factor $s_{(BA)}$ allows B to fit within A?

SCALING - NESTING SIMILARITY MEASURES (SNSM)
FOR TWO PLANAR OBJECTS A AND B









$s_{(BA)} = 0.80$

$s_{AB} = 0.74$

$s_{(AB)} = 0.68$

HOW TO STUDY SIMILARITY?

Scaling – Nesting Similarity Measures, SNSM

S(A,B) and S(B,A) are asymmetric measures, not "distance" type measures.
Many such asymmetric comparisons in science and art.
Examples:
In languages (asymmetric levels of cross-understanding,
e.g. Spanish vs. Portugese)
Similarities within hierarchies.
Replacability in roles.

However, the symmetric version of SNSM,

$$S(AB) = (S(A,B) + S(B,A)) / 2$$

can be shown to be a proper "distance" type measure,
that is, a metric, fulfilling the four criteria

$S(AB) \geq 0$,	non-negativity
$S(AB) = 0$ if and only if $A=B$	identity
$S(AB) = S(BA)$	symmetry
$S(AC) \leq S(AB) + S(BC)$	triangle inequality

HOW TO STUDY SIMILARITY?

Scaling – Nesting Similarity Measures, SNSM

P.G. Mezey, A Proof of the Metric Properties of the Symmetric Scaling-Nesting Dissimilarity Measure and Related Symmetry Deficiency Measures, Int. J. Quantum Chem., **63**, 105-109 (1997).

P.G. Mezey, Quantum Similarity Measures and Löwdin's Transform for Approximate Density Matrices and Macromolecular Forces, Int. J. Quantum Chem., **63**, 39-48 (1997).

TOPOLOGICAL APPROACH TO GEOMETRIC SIMILARITY

Geometrical Similarity as Topologial Equivalence

The GSTE principle

Molecules are not rigid, geometry is not the most useful mathematical tool for the description of their shapes.

In general, more fundamentally, for quantum mechanical objects, such as molecules, precise location constraints in geometrical models violate the Heisenberg relation.

For molecular shape characterization, a different branch of mathematics is especially useful:

TOPOLOGY

also called

"RUBBER GEOMETRY"

Some elements of topology

Topology = "rubber geometry"

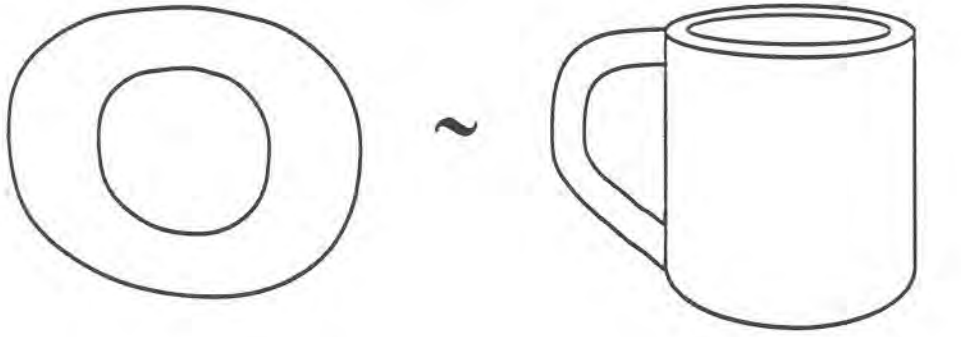
Topology = the mathematics of the essential

Topology is an excellent tool for shape analysis

Two different types of equivalences between the doughnut and the coffee cup:

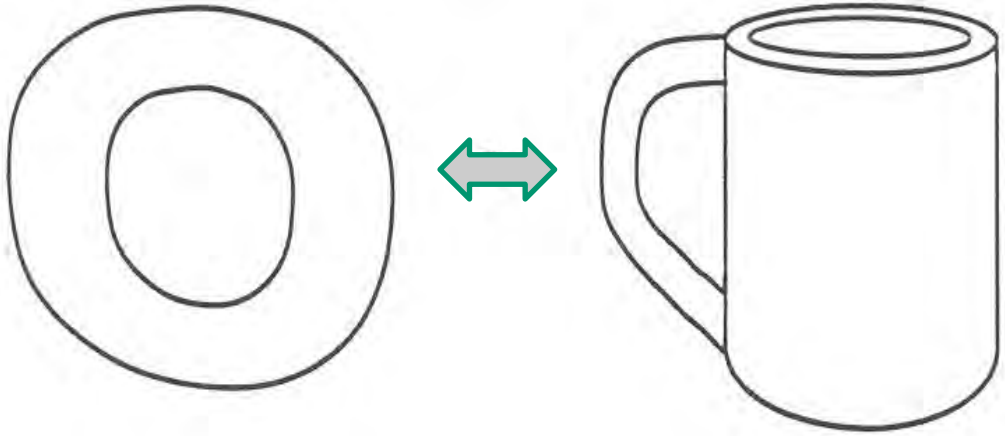
Topological equivalence

Homotopical equivalence



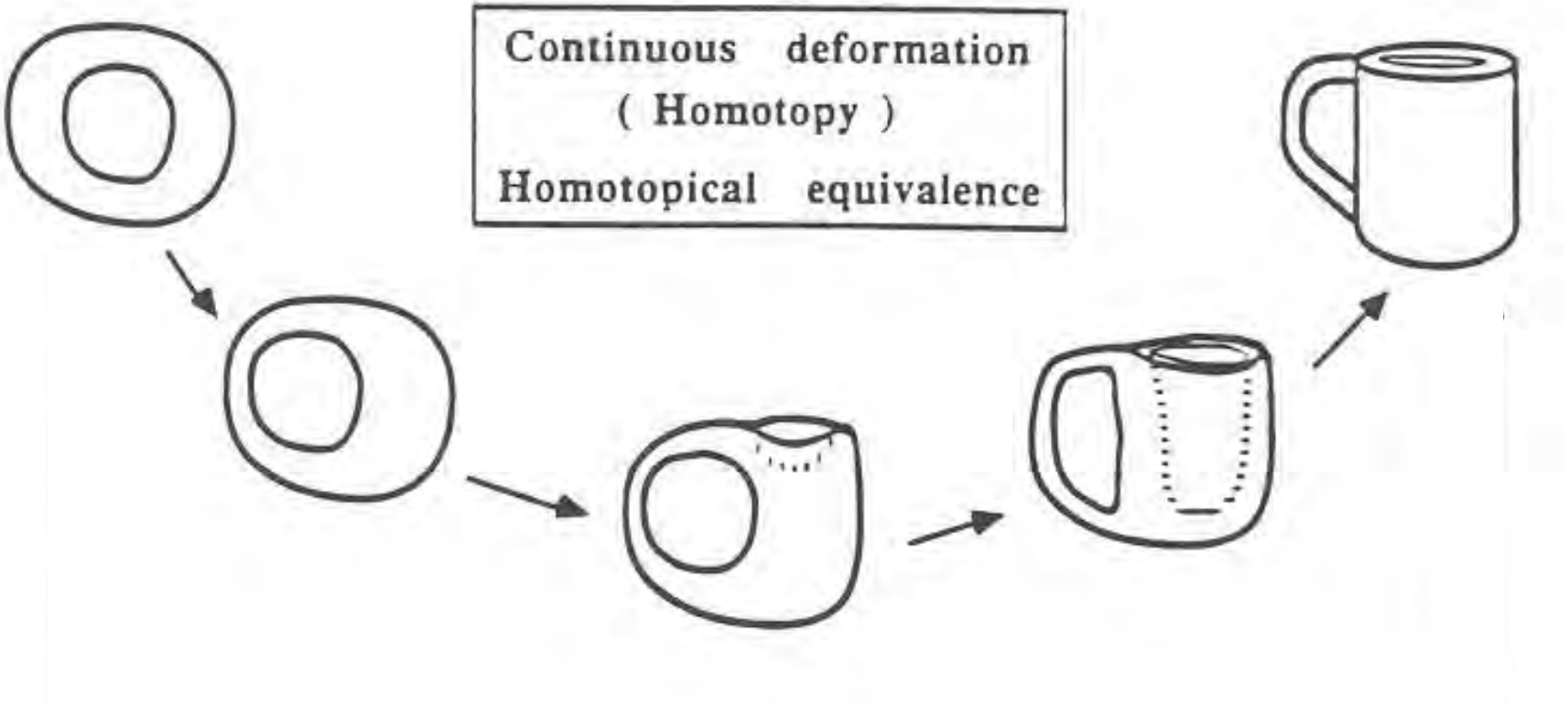
One-to-one, continuous assignment of points in both directions (Homeomorphism)

Topological equivalence



If each point of object A is assigned to a unique point of object B, and if each point of object B has a unique point of object A assigned to it, and if both this assignment and its inverse are continuous, than this assignment is a *homeomorphism.*

The doughnut and the coffee cup: for a topologist, they are the same



A continuous deformation is called a *homotopy.*
Two objects are *homotopically equivalent,* if there exists a continuous deformation converting one to the other

HOW TO STUDY SIMILARITY?

Geometrical Similarity as Topological Equivalence

The GSTE principle

Similarity Measures based on Topological Resolution

An elementary relation between generating subbases and topologies simplifies the approach to the construction of a hierarchy suitable to apply the tools of topological resolution. One may consider two generating subbases S_1 and S_2 containing families of subsets from an underlying space X, and assume the following the inclusion relation:

$$S_2 \supset S_1 \qquad (1)$$

If relation (1) holds, then the corresponding topologies $\mathbf{T}_1$ and $\mathbf{T}_2$ generated by these two subbases are comparable, and topology $\mathbf{T}_2$ is finer than topology $\mathbf{T}_1$:

$$\mathbf{T}_2 \supset \mathbf{T}_1. \qquad (2)$$

Similarity Measures based on Topological Resolution

In the simplest case, consider a countable family $\mathbb{T}$ of topologies $\mathbf{T}_i$ in the underlying set X,

$$\mathbb{T} = \{\mathbf{T}_1, \mathbf{T}_2, \ldots \mathbf{T}_i, \mathbf{T}_{i+1}, \ldots\}, \qquad (3)$$

where these topologies $\mathbf{T}_i$ are fully ordered by the stronger-weaker relation. That is, we shall assume that

$$\mathbf{T}_{i+1} \supset \mathbf{T}_i \qquad (4)$$

for any two indices i and i+1.

Similarity Measures based on Topological Resolution

As a consequence, the corresponding topological spaces $(X,\mathbf{T}_i)$ are also fully ordered, and

$$(X,\mathbf{T}_{i+1}) \supset (X,\mathbf{T}_i) \qquad (5)$$

holds for every index-pair i and $i+1$.

Similarity Measures based on Topological Resolution

The choice of family $\mathbb{T}$ of topologies is said to be actually discriminative for the given similarity analysis, if from the ordered sequence of topological spaces

$$\ldots \supset (X,\mathbf{T}_{i+1}) \supset (X,\mathbf{T}_i) \ldots \supset (X,\mathbf{T}_2) \supset (X, \mathbf{T}_1) \qquad (6)$$

there exists at least one that provides distinction among the molecules. Of course, if each topology is discriminative then the hierarchy in sequence (6) provides no variety in topological resolution for the given set of molecules

Similarity Measures based on Topological Resolution

In the above sense, the topological space $(X,\mathbf{T}_{i+1})$ describes more or at least as much detail of the underlying space X than the topological space $(X,\mathbf{T}_i)$, as implied by relation (5). Of course, if in addition to relation (5), the constraint

$$(X,\mathbf{T}_{i+1}) \text{ NE } (X,\mathbf{T}_i) \qquad (7)$$

also holds, that is, if the strict inclusion relation applies in (5), than the discrimination by $(X,\mathbf{T}_{i+1})$ is stronger than by $(X,\mathbf{T}_i)$.

If both conditions, (5) and (7) hold, then one can find functions f that map X onto itself and are $\mathbf{T}_i$ - continuous but not $\mathbf{T}_{i+1}$ – continuous. In this case, a topological description by $(X,\mathbf{T}_{i+1})$ with a higher topological resolution provides more information than a description by $(X,\mathbf{T}_i)$ with a topological resolution of lower level. In other words, $(X,\mathbf{T}_i)$ may not be fine enough to describe some details that can be captured by $(X,\mathbf{T}_{i+1})$.

Similarity Measures based on Topological Resolution

A fully topological generalization of the GSTE principle is obtained if one considers topological similarities as relations manifested in a topological equivalence according to a weaker topology.

A topological similarity at some level within the hierarchy of topologies becomes a weaker topological equivalence at some other level within the sequence. It is natural to take the strongest of these weak topologies to represent a characteristic level for the given similarity.

Similarity Measures based on Topological Resolution

The associated approach, Topological Similarity as Weaker Topological Equivalence (TSWTE), is equivalent to the ordinary Geometrical Similarity as Topological Equivalence (GSTE) approach, if the topological space (X,**T**i) used has topology **T**i as the metric topology.

Of course, then **T**i is suitable for the geometrical representation of the objects within the underlying space X, and the approach becomes geometrical. Consequently, the GSTE approach can be regarded as a special case of the more general TSWTE approach.

Similarity Measures based on Topological Resolution

The TSWTE approach is a natural byproduct of any topological hierarchy, and is the simplest if the topological spaces involved are fully ordered. However, if only a partial order exists within a topological family, the approach is still applicable.

Similarity Measures based on Topological Resolution

[1] Mezey, P.G. (1991) The Degree of Similarity of Three-Dimensional Bodies; Applications to Molecular Shapes, J. Math. Chem., **7**, 39-49.

[2] Mezey, P.G. (2000) Shape-Similarity Relations Based on Topological Resolution, J. Math. Chem., **27**, 61-69.

[3] Mezey, P.G. (1993) Shape in Chemistry: An Introduction to Molecular Shape and Topology , VCH Publishers, New York.

[4] Mezey, P.G. (1997) Quantum Chemistry of Macromolecular Shape, Internat. Rev. Phys. Chem., **16**, 361-388.

[5] Mezey, P.G. (1997) Shape in Quantum Chemistry. In Conceptual Trends in Quantum Chemistry, Vol. 3, Calais. J.-L.; Kryachko, E.S. (eds.), Kluwer Academic Publ., Dordrecht, The Netherlands, pp 519-550.

[6] Mezey, P.G. (2000) Topological Methods of Molecular Shape Analysis: Continuum Models and Discretization, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **51**, 267-278.

[7] Mezey, P.G., Ponec, R., Amat, Ll., and Carbo-Dorca, R. (1999) Quantum Similarity Approach to the Characterization of Molecular Chirality, Enantiomer, **4**, 371-378.

THE SPECIFIC PROBLEMS OF MOLECULAR SIMILARITY

Molecules are fuzzy, they have no boundaries.

How to study their shapes, their similarities,

their approximate symmetries,

and their level of chirality?

MIDCOs

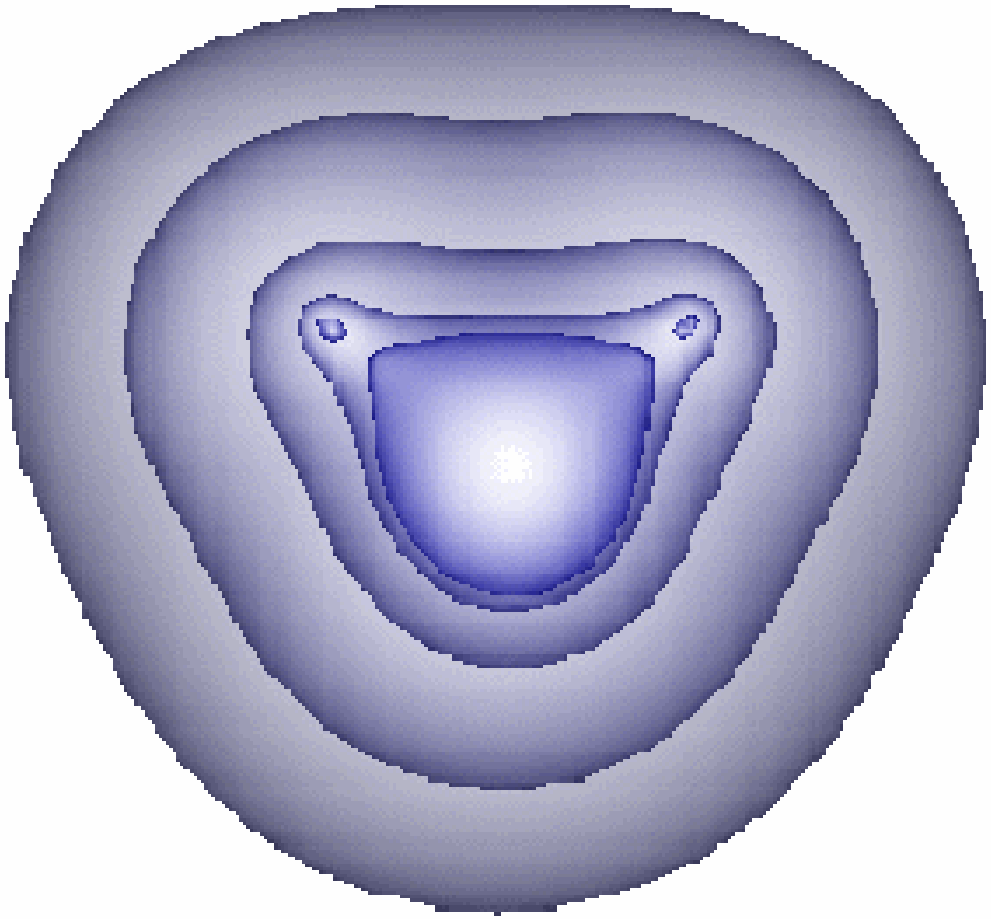
A *molecular isodensity contour surface,* MIDCO G(K,a) of nuclear configuration K and density threshold a is defined as

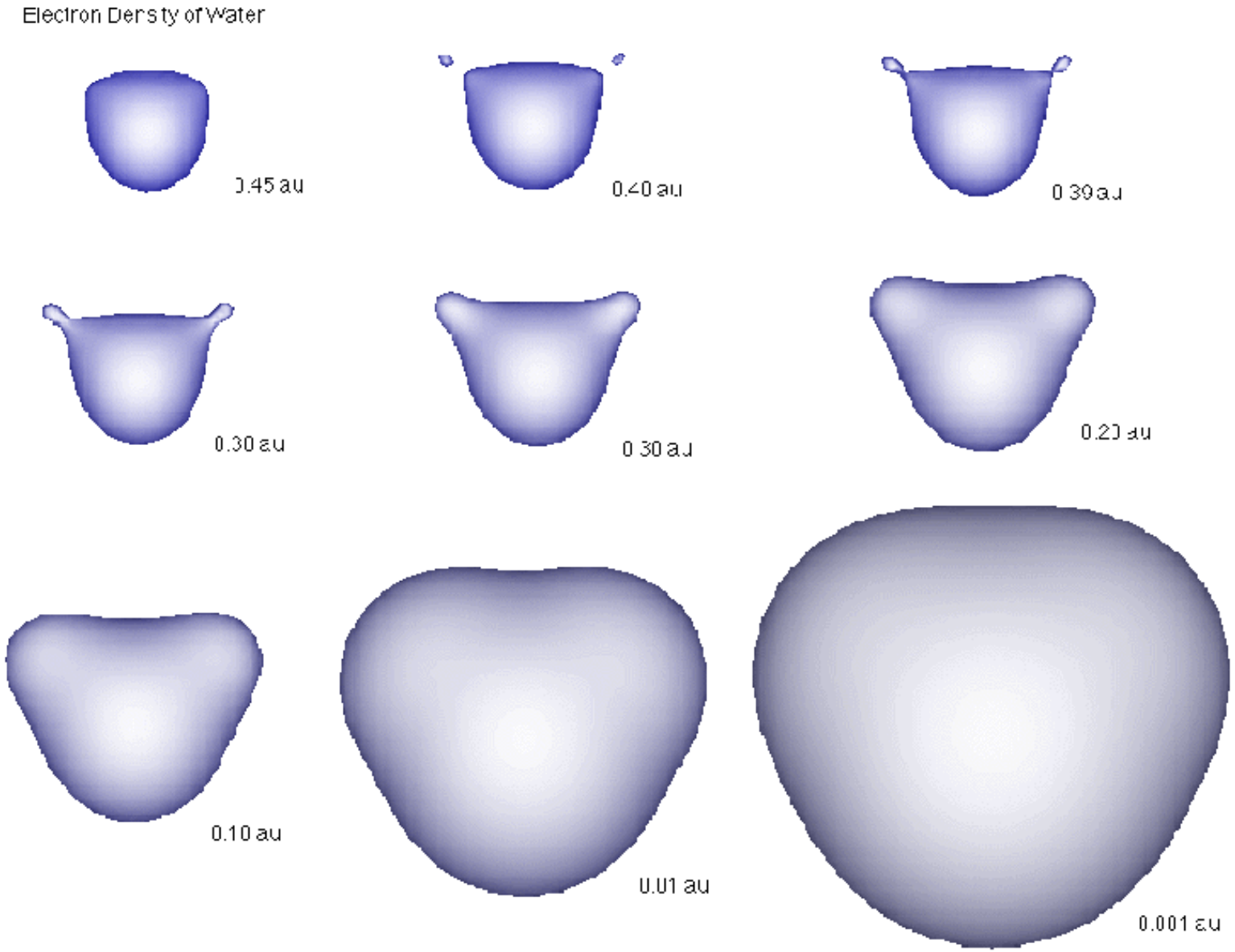
$$G(K,a) = \{ \mathbf{r} : \rho(K,\mathbf{r}) = a \},$$

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

The *shape groups* are algebraic groups, not related to point symmetry groups, although the presence of symmetry may influence the shape groups. In technical terms, the shape groups are the homology groups of truncated objects, where the truncation is determined by local shape (specifically, local curvature) properties.

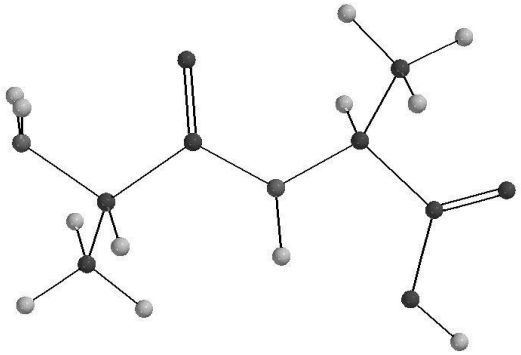
Some constant electron density contours, "isosurfaces" of water at density values of 0.001, 0.01, 0.10, 0.30, and 0.40 a.u. (atomic units)

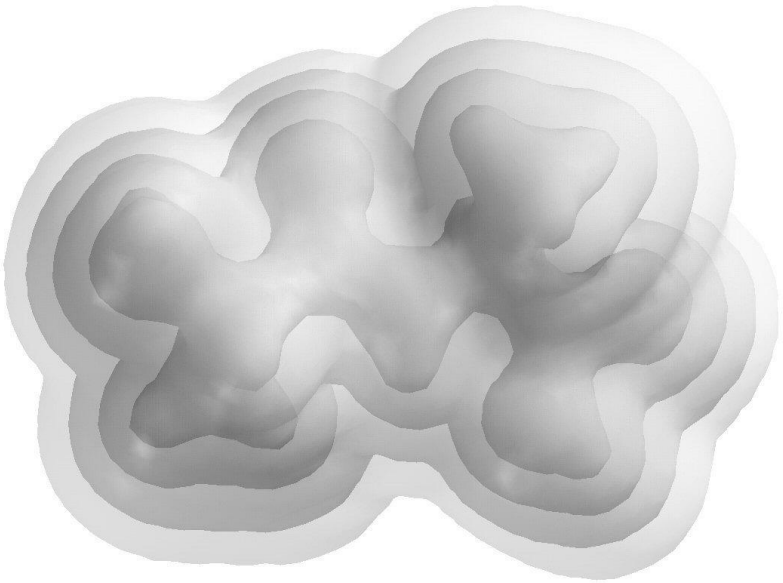


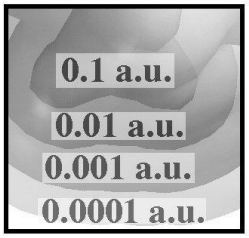

Electron Density of Water
0.45 au
0.40 au
0.39 au
0.30 au
0.30 au
0.20 au
0.10 au
0.01 au
0.001 au

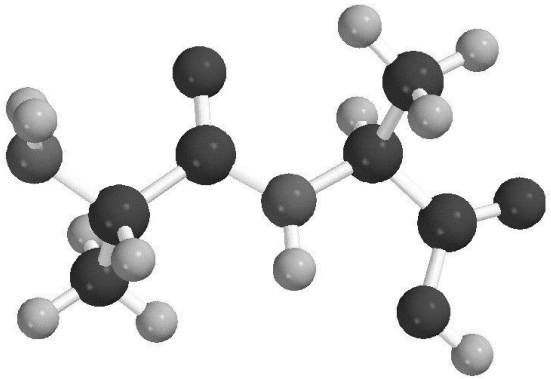
Alanylalanine, structural models and actual electron density cloud

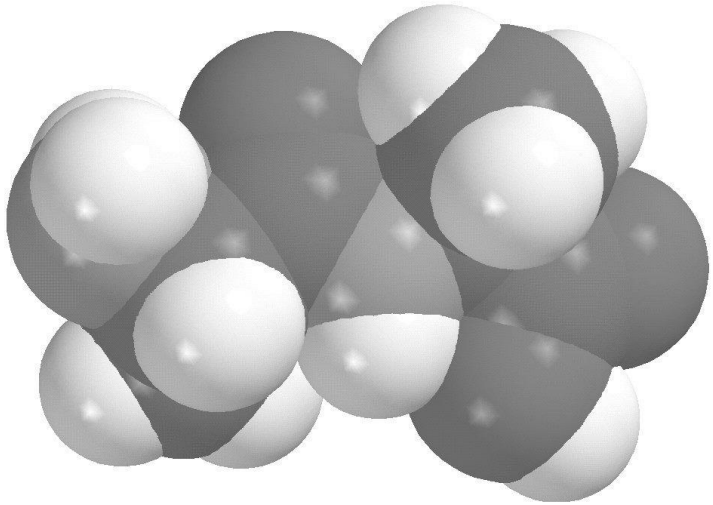
ALA-ALA











The similarity between the shapes of any two objects, such as the shapes of two faces or the shapes of two molecules, can be evaluated by the *Shape Group Method*, using an algorithm and computer programs (see some details later).

The shape difference between the L and D forms provides a numerical measure for the Degree of Chirality:

a small value for the only slightly chiral objects,

and a large value for very chiral objects.

For the natural amino acids, the *molecular fragment-based local chirality measures*, compared to *global chirality measures of the complete molecules*, are expected to show correlations with relative rates of enantiomerization, which, in turn, *might provide new insight into the mechanisms of such, often harmful changes of amino acids,*

such as those leading to cataract formation.

Some fundamental ideas about shape

A very short review of concepts:

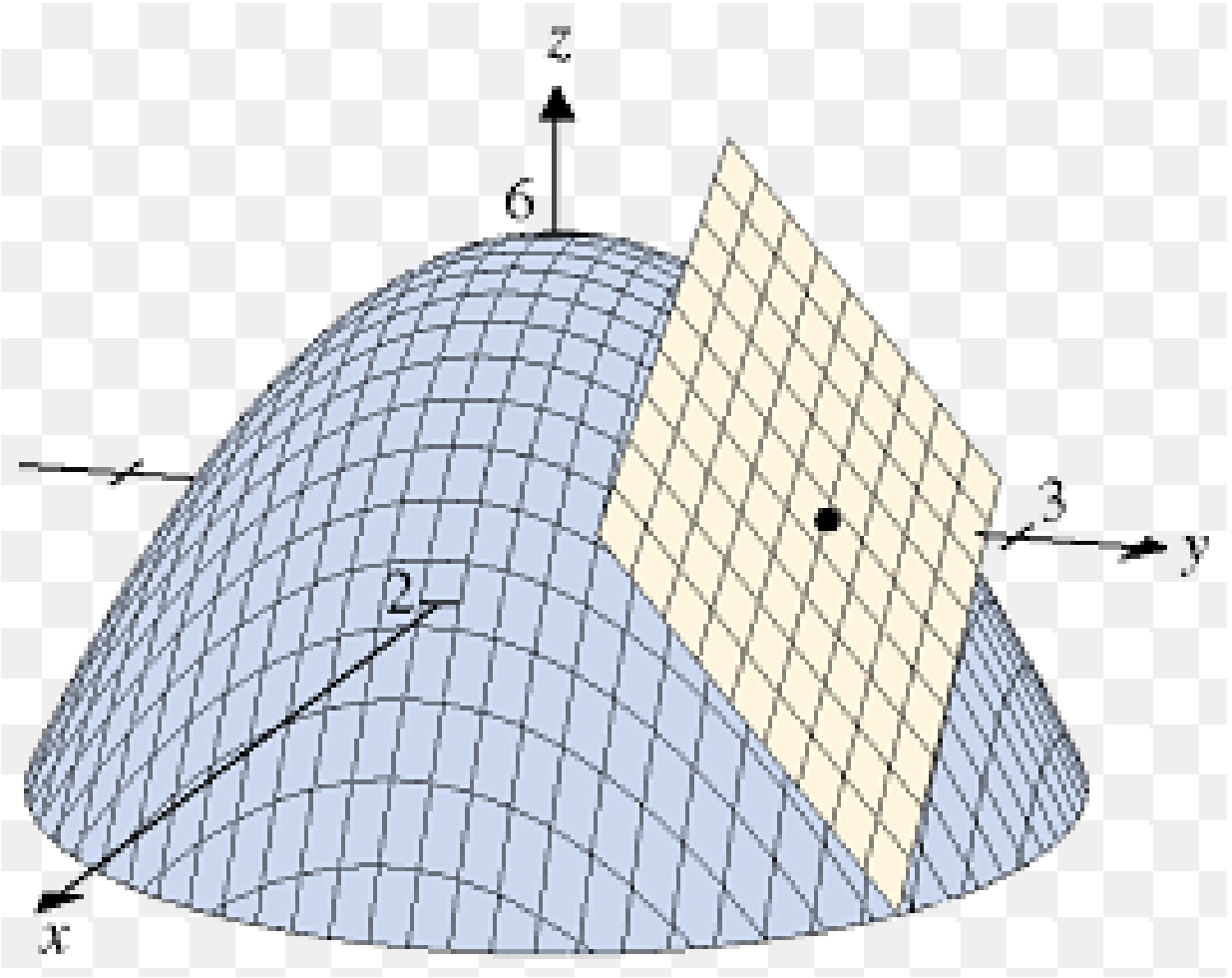
Tangent plane to a surface:
the surface and the tangent plane touch at a point, where they are "locally parallel"

Convex surface region: like a "bump"

Concave surface region: like a "cave", a "dip"

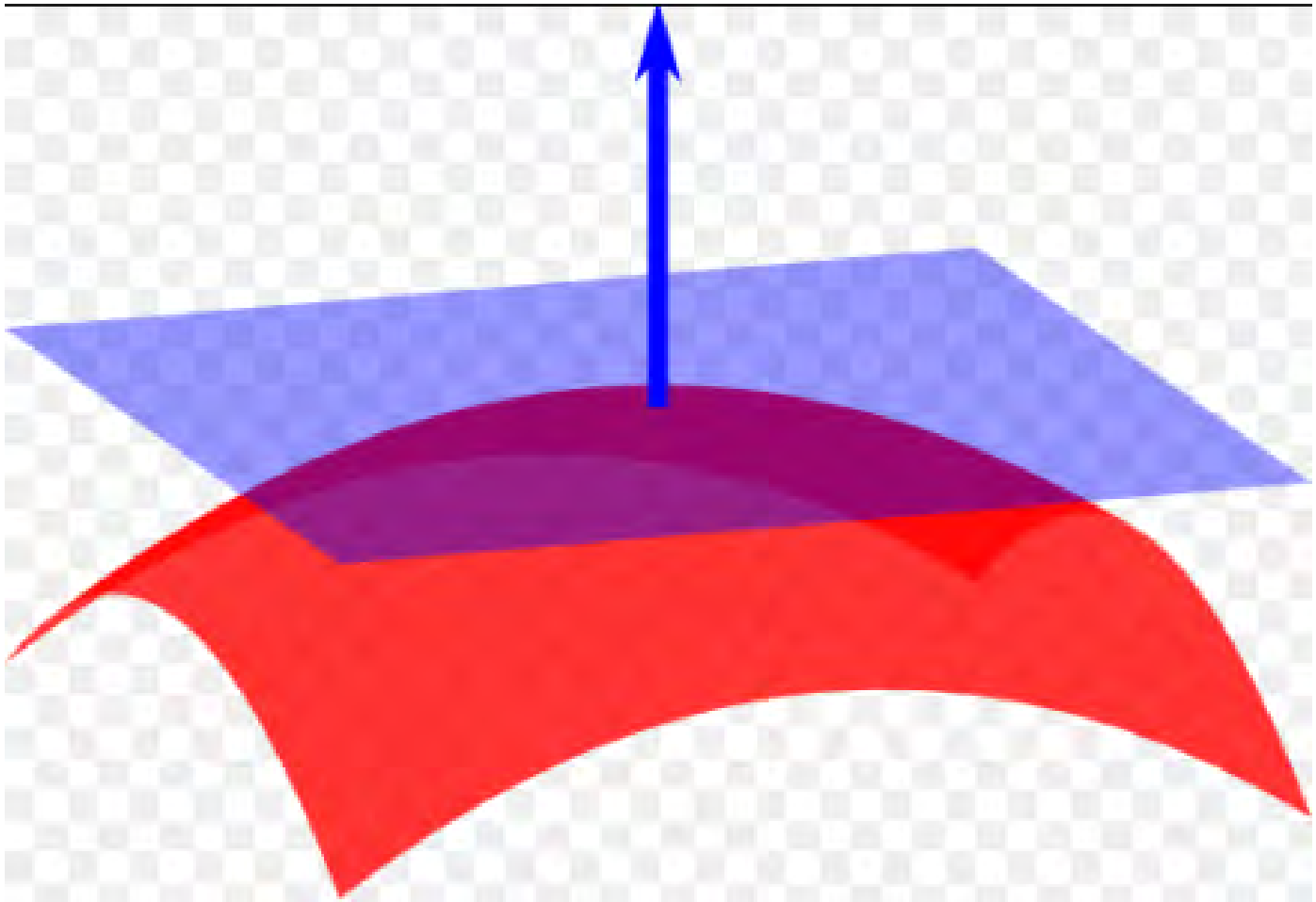
Saddle surface region: like a "saddle" on a horse

A tangent plane **(white) at some point (black dot) of a surface (blue) is "touching" the surface at this point, and the tangent plane and the surface are locally "parallel" there**

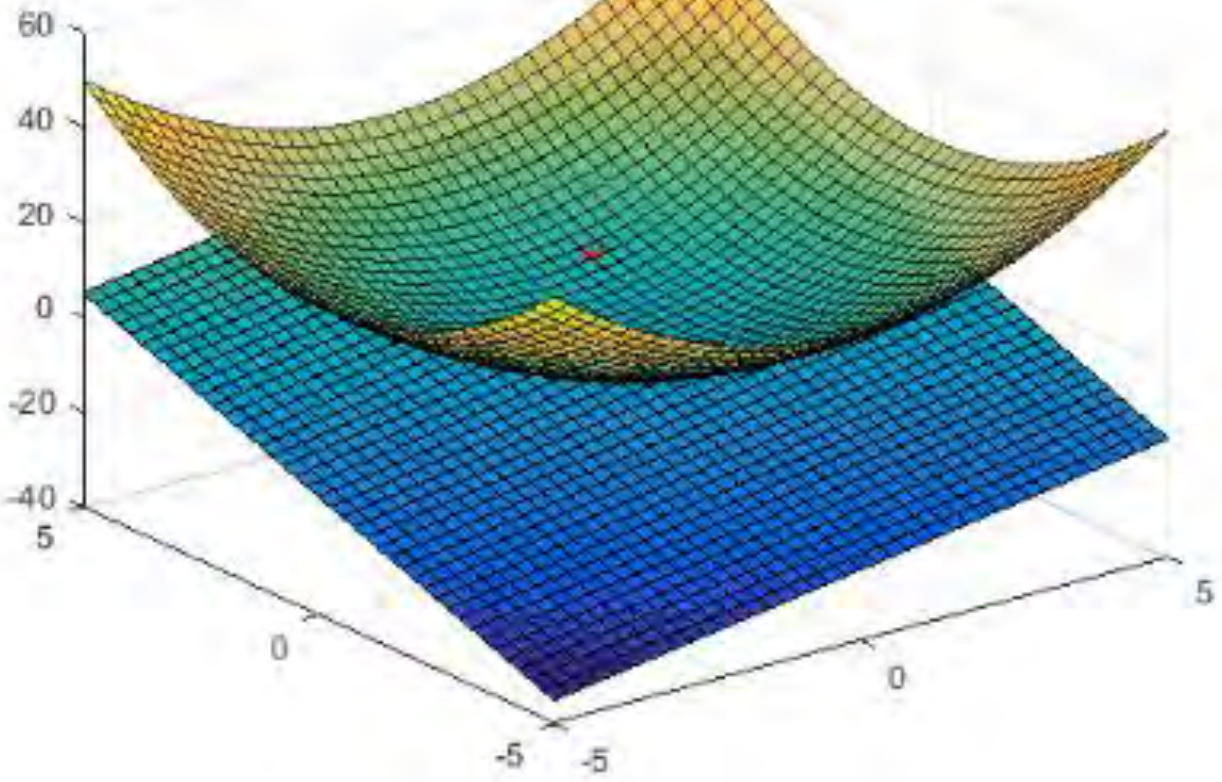


A convex surface **is like a "bump", something "round"**

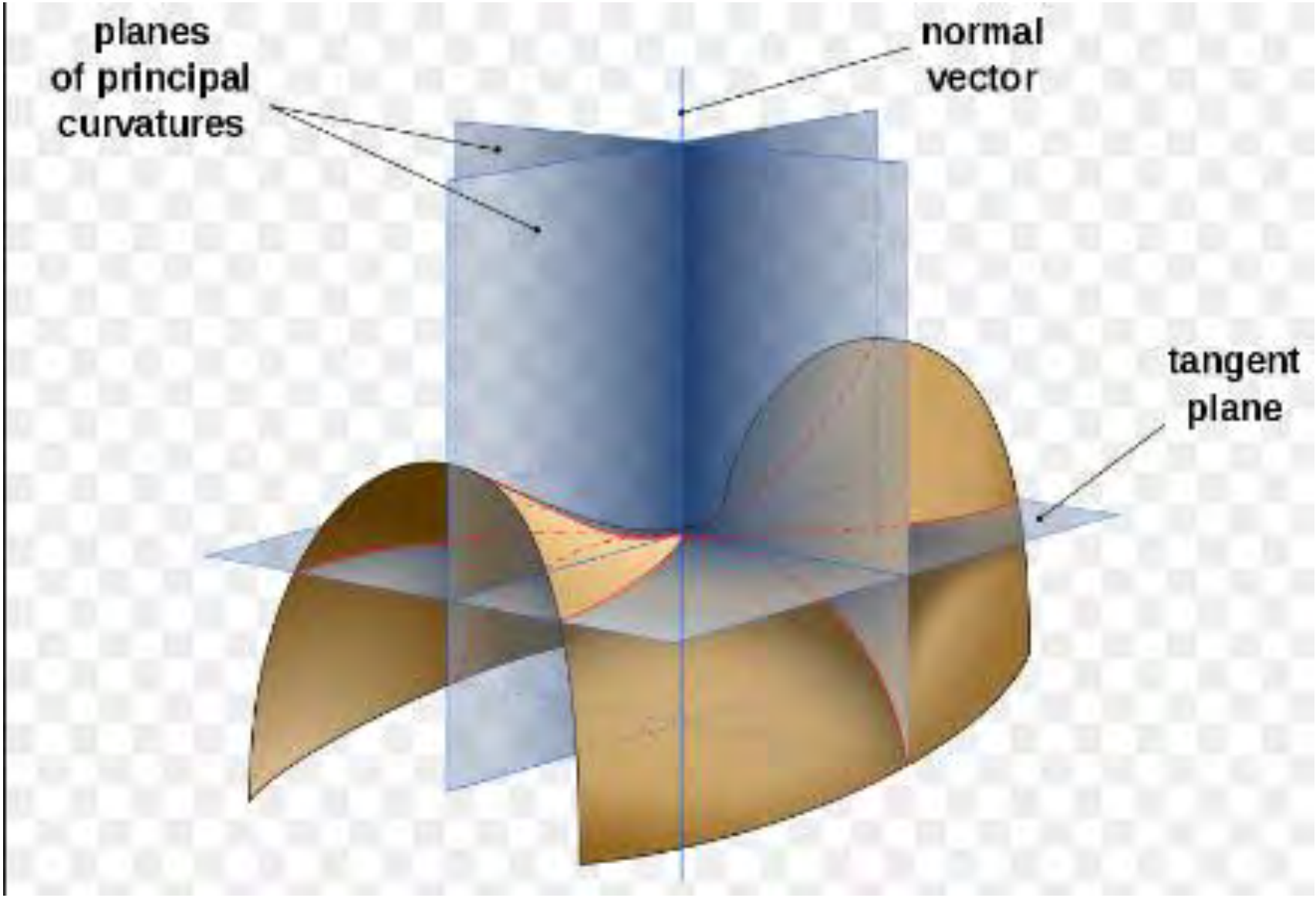
A convex surface (red) and the tangent plane (blue) touching it at the point where the arrow stands



A concave surface **can surround something, like a "cave", (origin of the word *concave*). Concave surface (yellow) and the tangent plane (blue) touching at the red point.**



Saddle surface "on a brown horse", and the tangent plane (horizontal) that cuts into it.



Parabolic points:

border points of convex, saddle, and concave surface regions.

Shape analysis, for faces, for molecules,
for the shapes of tumors, for the shapes of flowers
Same approach applies, as in Molecular Shape Analysis

My starting point:

The mathematician Felix Klein, of Erlangen, Germany, tried to define mathematically, what is beautiful,

he tried to develop a mathematical model for beauty,

using lines of parabolic points on surfaces of statues

As a starting point, I have used his ideas.
who used differential geometry to analyze shapes

A more modern approach
using the tools of topology:

Algebraic topology and differential topology

On replicas of statues, Felix Klein

marked the lines separating

convex,

***concave,* and**

saddle type regions

of the surfaces, and studied the

patterns so obtained.

Handwork of Felix Klein,

on a copy of the bust of the Belvedere Apollo,

to be found at the

University of Göttingen, Mathematics Library

Shape Group Methods, SGM,
Paul G. Mezey

Felix Klein was not happy with the results, he thought that nothing useful came out, and stopped the project.

I have read about it, and I have found it fascinating:

If this is not good for beauty,

it still can be good for shape analysis!

One can apply the same ideas to molecular surfaces,

for example, to Molecular IsoDensity COntour surfaces, MIDCO's.

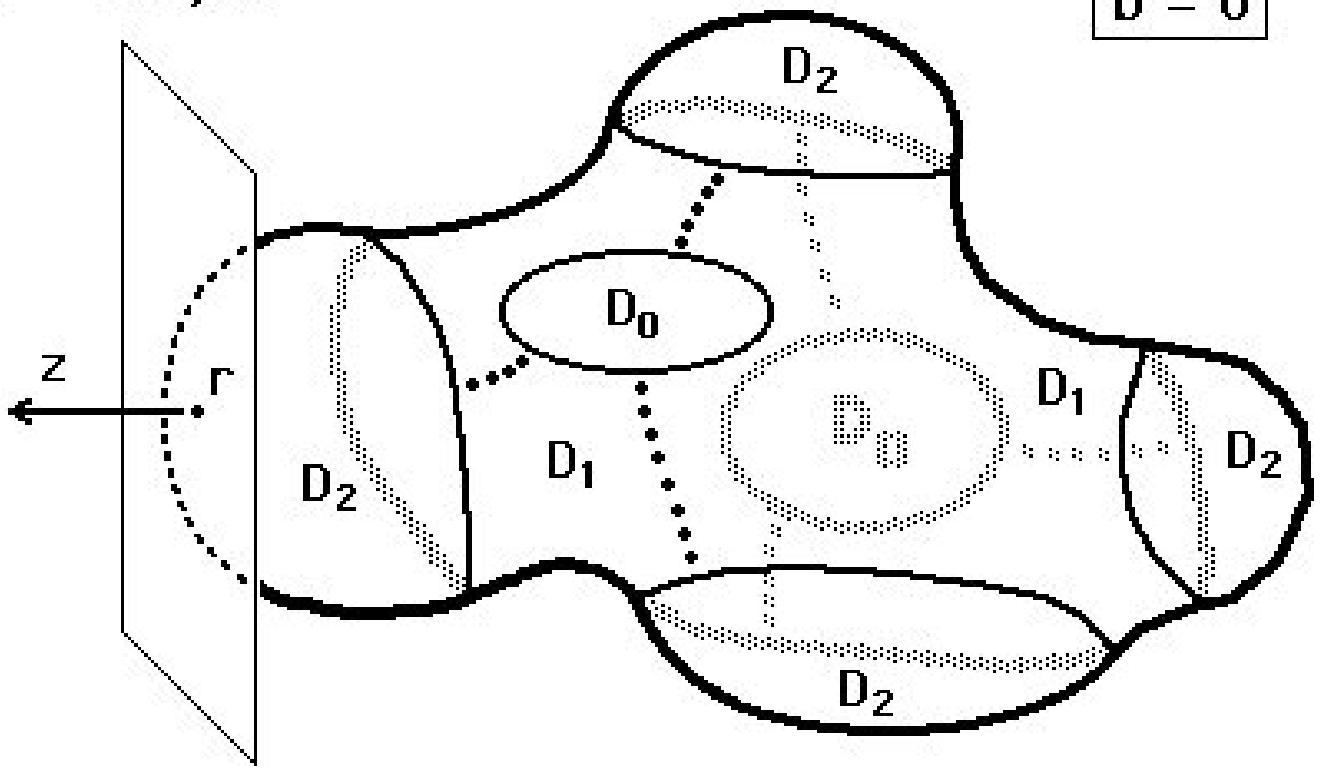
For generalization, an alternative, equivalent approach is more feasible, where local convexity is considered:

The locally *convex*, *saddle-type*, and *concave* regions of any given surface, these regions are labelled D_2, D_1, D_0, respectively.

Contour surface
G(a)

Tangent plane
$^2R(r) = {}^2S(r,0)$

b = 0



Convexity domain partitioning of surfaces

Locally ***convex*** regions D_2,

Locally ***saddle-type*** regions D_1,

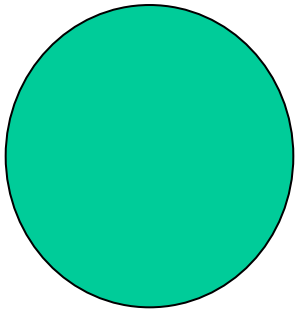
Locally ***concave*** regions D_0.

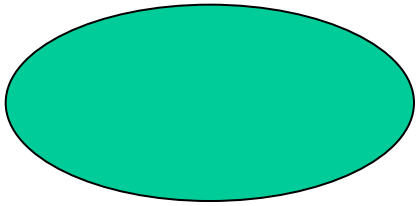
The parabolic points form the boundary lines of these regions.

The indices 2, 1, and 0 refer to the number of negative eigenvalues of the local Hessian matrix at each point r, where the surface is regarded as a function defined over the local tangent plane, taken as horizontal.

Does Felix Klein's idea provide a sufficiently detailed shape description?

Are the shapes of European and American footballs the same?





Not really!

But, by the criteria of *local convexity* and domain separation by *parabolic points*, these balls are indistinguishable: just a single D_2 region in both cases!

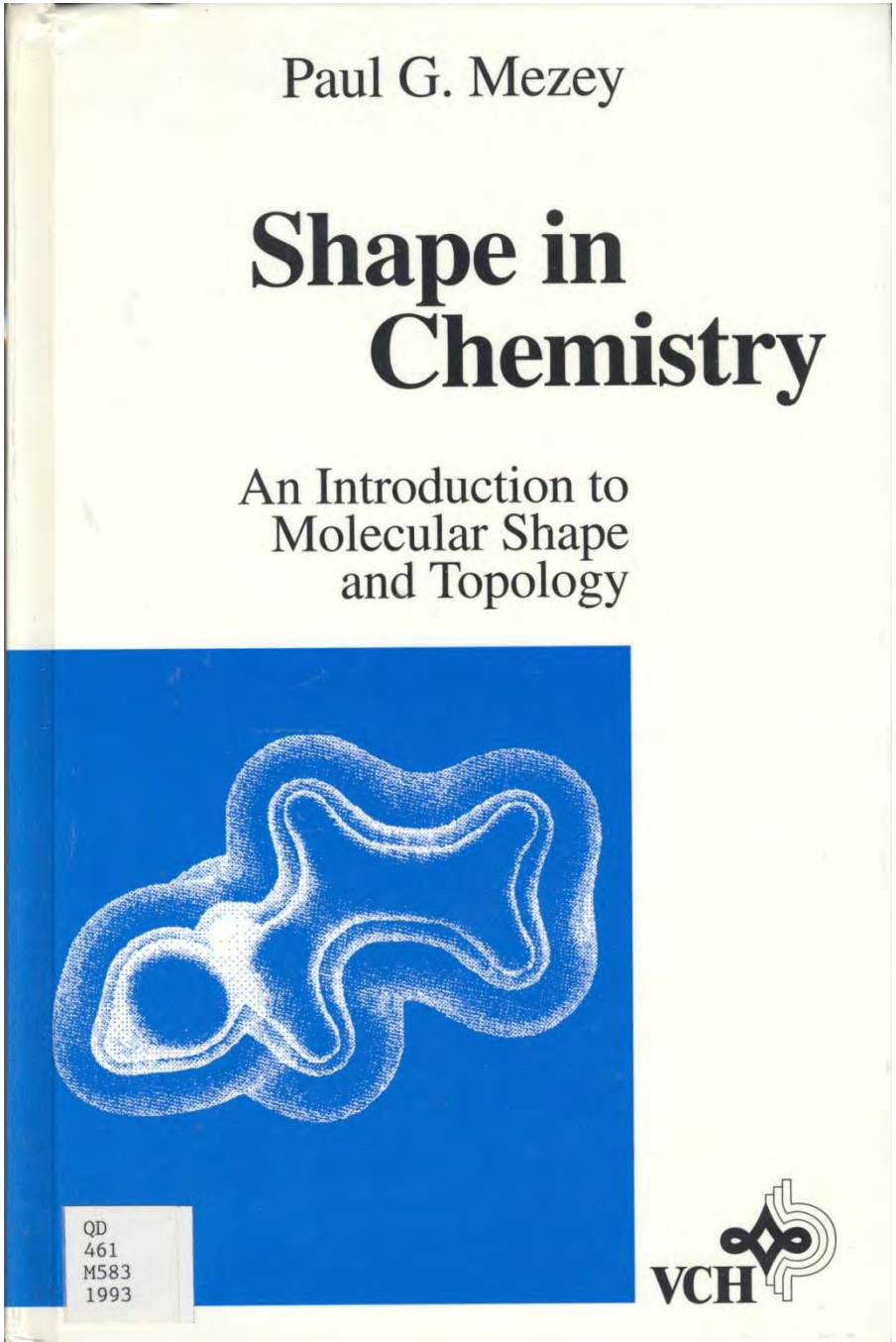
Not satisfactory, one needs finer tools.

Convexity in a Curved Universe …..

Shape Group Method

.

Molecular Shape Analysis Methods



Shape Analysis, Paul G. Mezey

Convexity is generalized with reference to tangent planes:

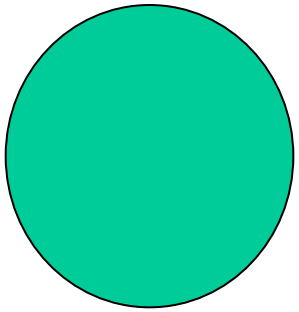
For ordinary convexity: is the object locally more curved, or less curved than the tangent plane, or cuts into the tangent plane?

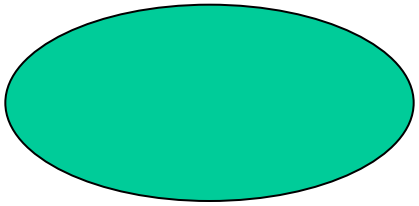
How could one define convexity in a curved Universe?

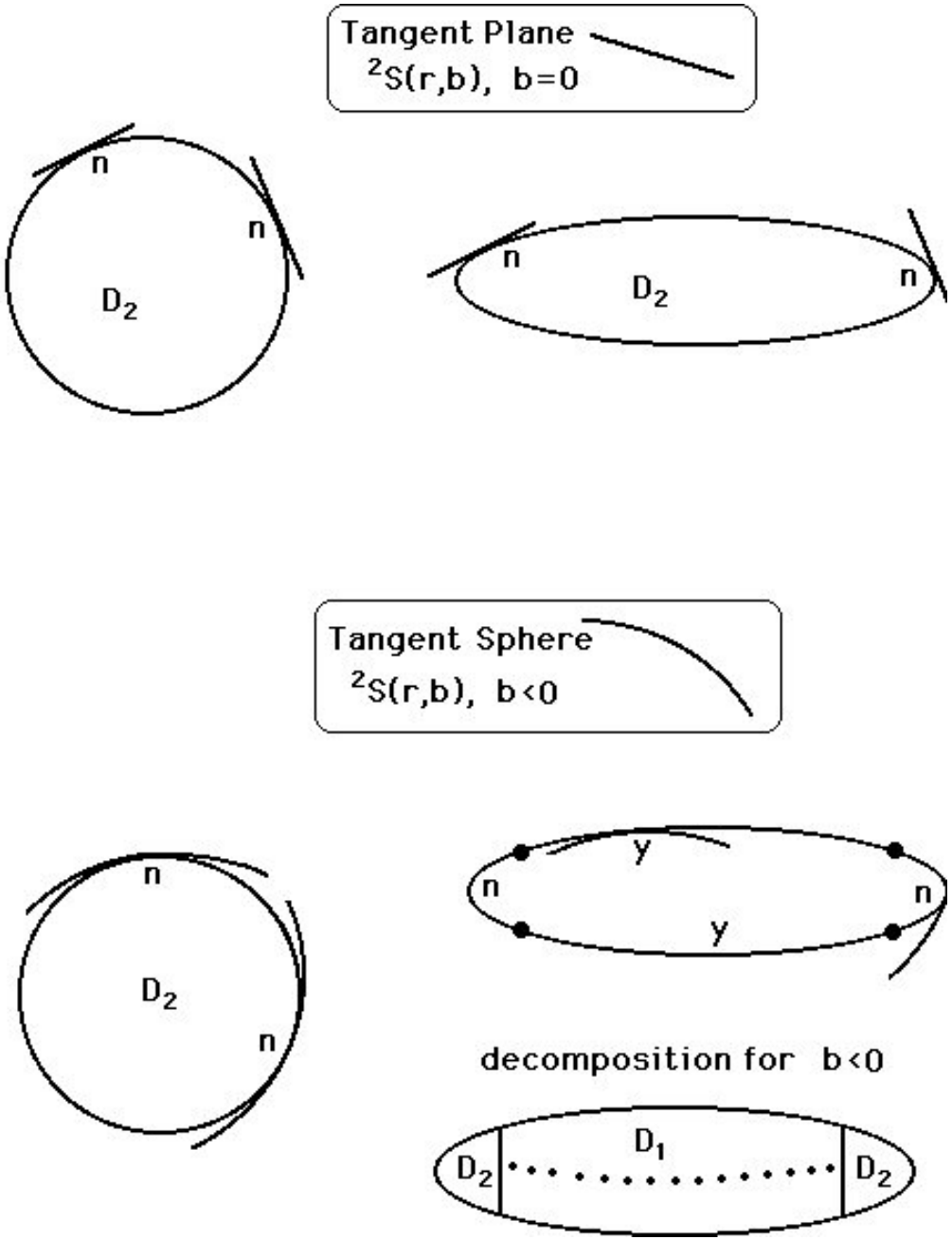
In a curved Universe, convexity is defined relative to tangent sphere:
is the object locally more curved, or less curved than the
tangent sphere, or cuts into the tangent sphere?

The local curvature of the tangent sphere of radius r is a reference, that defines relative convexity, relative to this reference curvature, b = 1/r

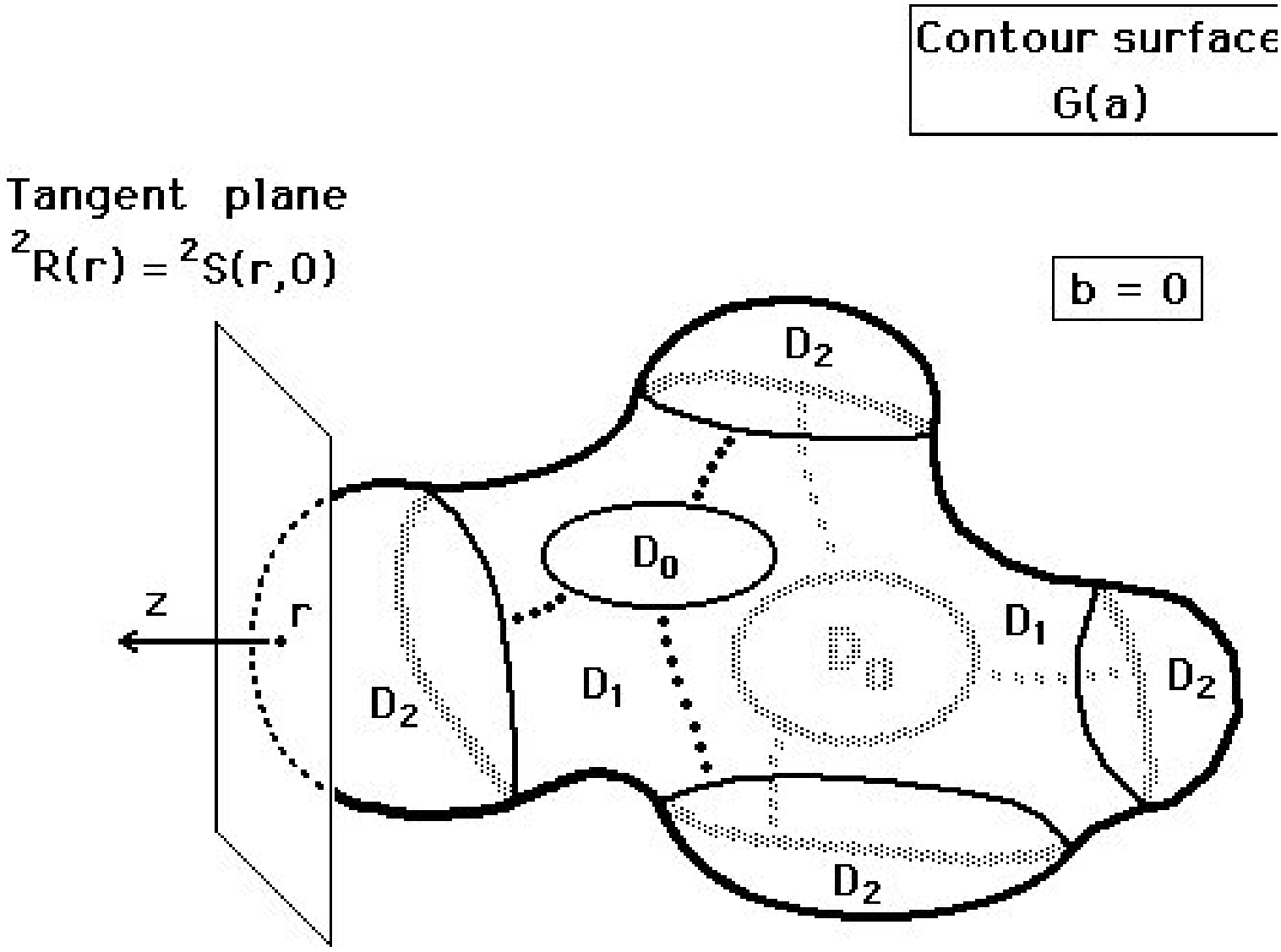
The shapes of European and American footballs are now distinguishable by this approach.



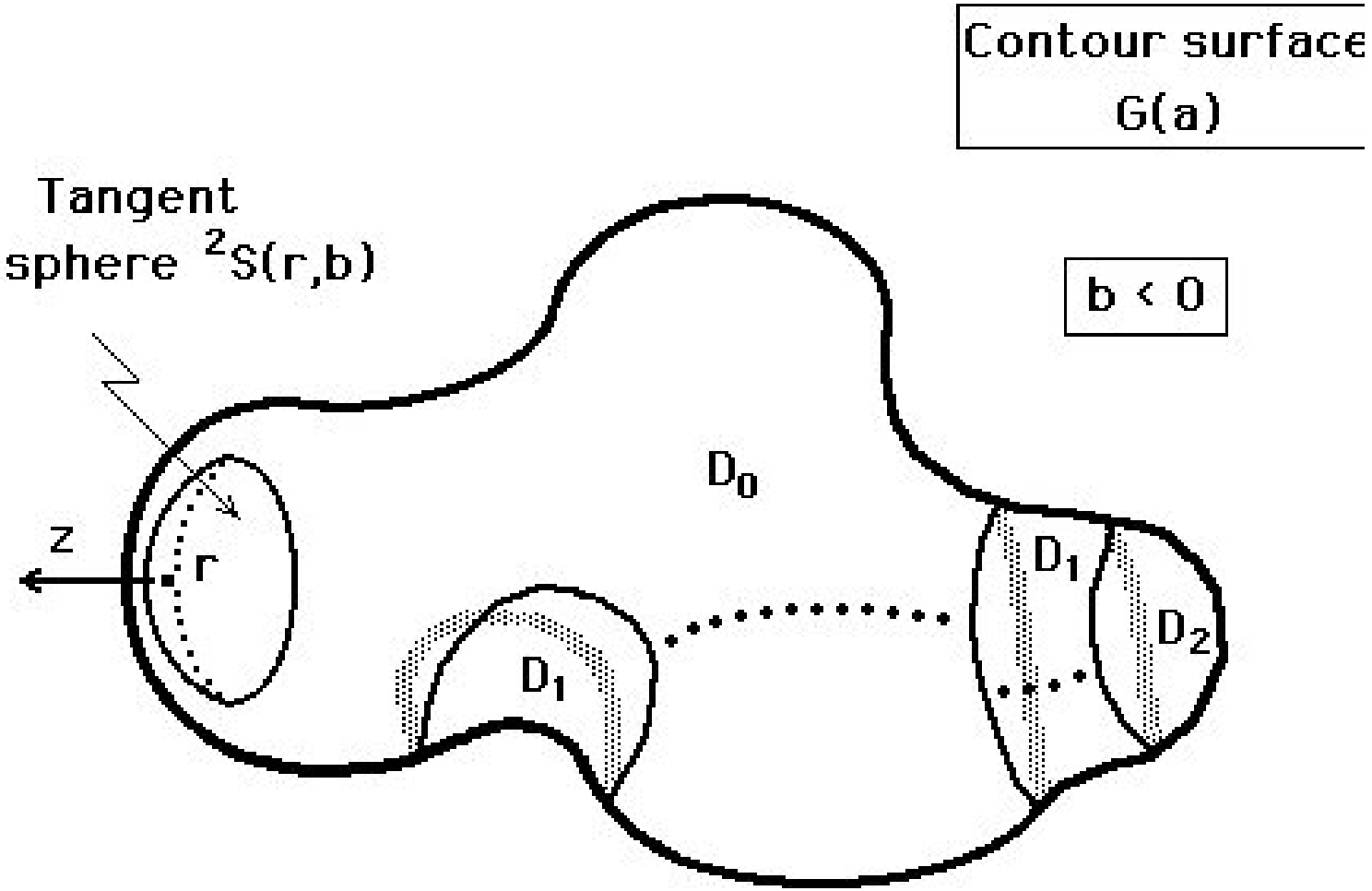



Tangent Plane
$^2S(r,b)$, b=0
n
n
D_2
n
D_2
n
Tangent Sphere
$^2S(r,b)$, b<0
n
D_2
n
y
n
n
y
decomposition for b<0
D_1
D_2
D_2

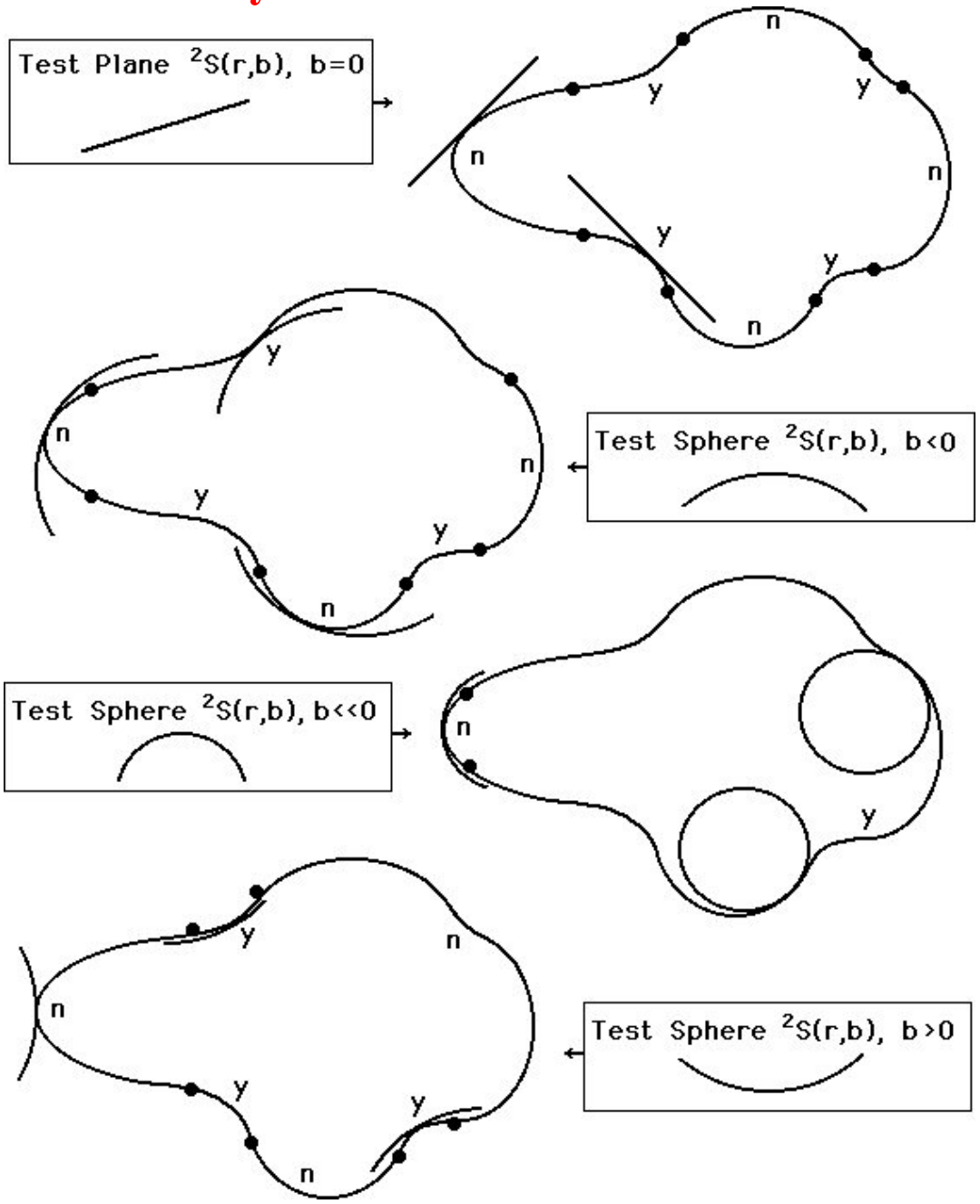
Ordinary local convexity:
convexity relative to a tangent plane of curvature b=0



Relative convexity (convexity in a curved universe): convexity relative to a reference curvature b



Relative convexity (convexity in a curved universe): convexity relative to a reference curvature b



Relative convexity domain partitioning of surfaces

Convexity in a curved Universe, relative to some reference curvature b, where b = 1/r, with r the radius of the tangent sphere

Locally *convex* regions *relative to curvature* b, D_2,

Locally *saddle-type* regions *relative to curvature* b, D_1,

Locally *concave* regions *relative to curvature* b, D_0.

At each point **r,** the indices 2, 1, and 0 refer to the number of eigenvalues of the local Hessian which are ***less then*** b, where the surface is regarded as a function defined over the local tangent plane, if this plane is taken as horizontal.

Shape Groups

The electron densities of the molecules and all their isocontours are calculated by quantum chemistry methods.

For the chemically important range (0.001 a.u. - 0.1 a.u) of density threshold values a of all isocontours, and for the chemically important range (- 10 .... + 10) of all tangent sphere curvatures b, the software determines the

locally convex D2,
locally concave D0,
and saddle regions D1,

of all isocontours.

Shape Groups

For most *small changes* of a and b values, *only the sizes* of these D0, D1, and D2 regions change, but *their topological pattern remains the same* on the isosurfaces, however, for *some changes* of a and b values, *even the topological pattern of these regions changes.*

The main observation: **for the entire range of patrameters a and b,** ***only a finite number of topologically different patterns of domains exist.***

These patterns can be characterized topologically, using the homology groups of algebraic topology.

The Shape Groups **are the** ***one-dimensional homology groups*** **of these patterns.**

Shape Group Method

For compete details see P.G. Mezey Shape in Chemistry, An Introduction to Molecular Shape and Topology, VCH New York 1993

Summary of terminology and notations:

Electron density value at the contour surface a

Curvature value of tangent sphere b

A shape pattern:

a relative arrangement of all convex, saddle, and concave regions, without any concern about their size.

For the whole range of a and b values, there are only a finite number of such (topologically equivalent) ***shape patterns.***

Using the tools of Topology, these patterns define the so-called homology groups on each and every one of the whole range of isodensity surfaces

The 1-dimensional homology groups are the *Shape Groups* of the complete electron density clouds, fully describing their shape.

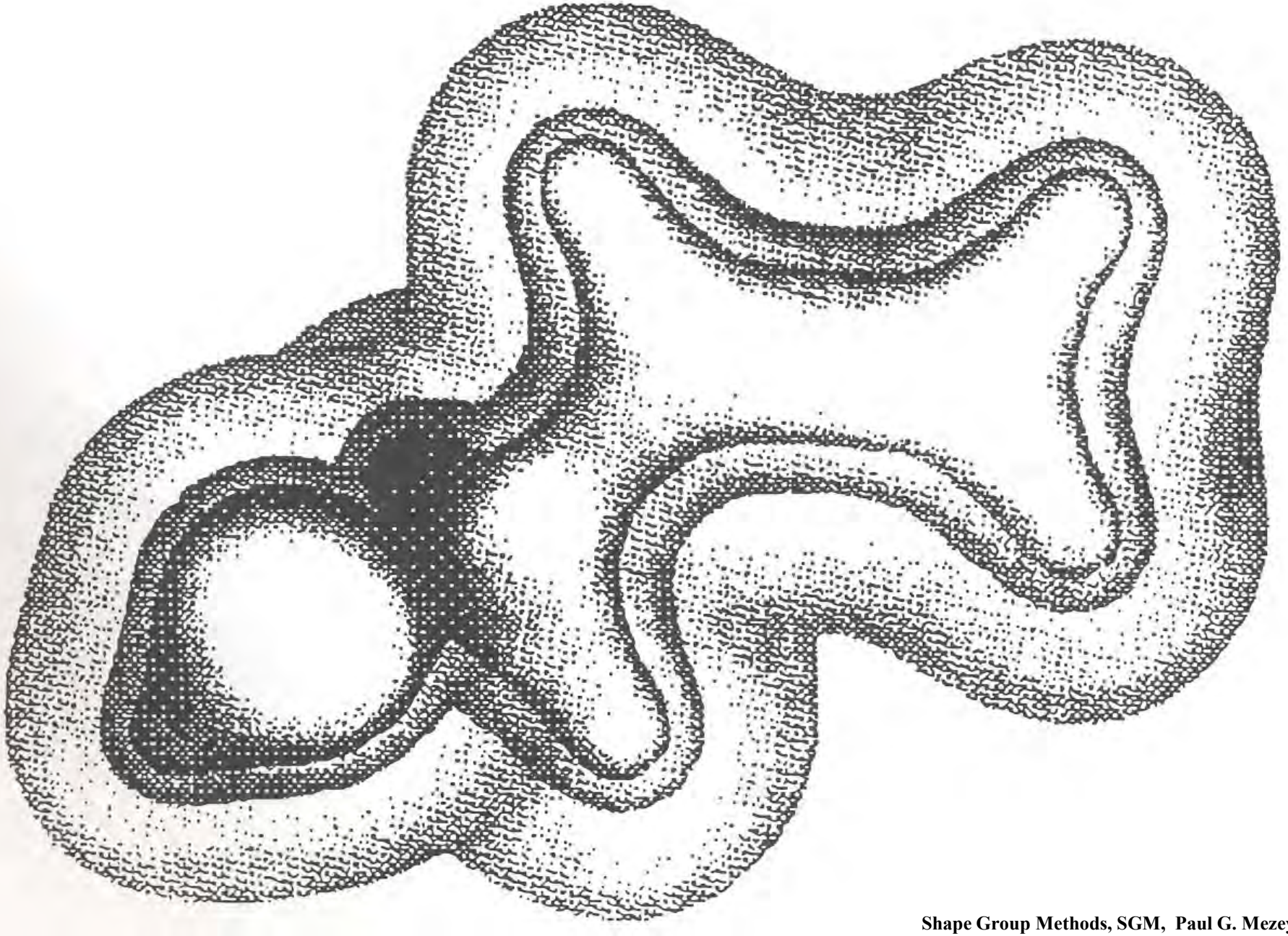
The ranks of these Shape Groups are the Betti numbers, they depend on the a, b values, generating a numerical shape table, we call the (a,b)-map, that *is used as a numerical shape code.*

These shape codes can be compared as number sequences, providing *numerical shape similarity measures.*

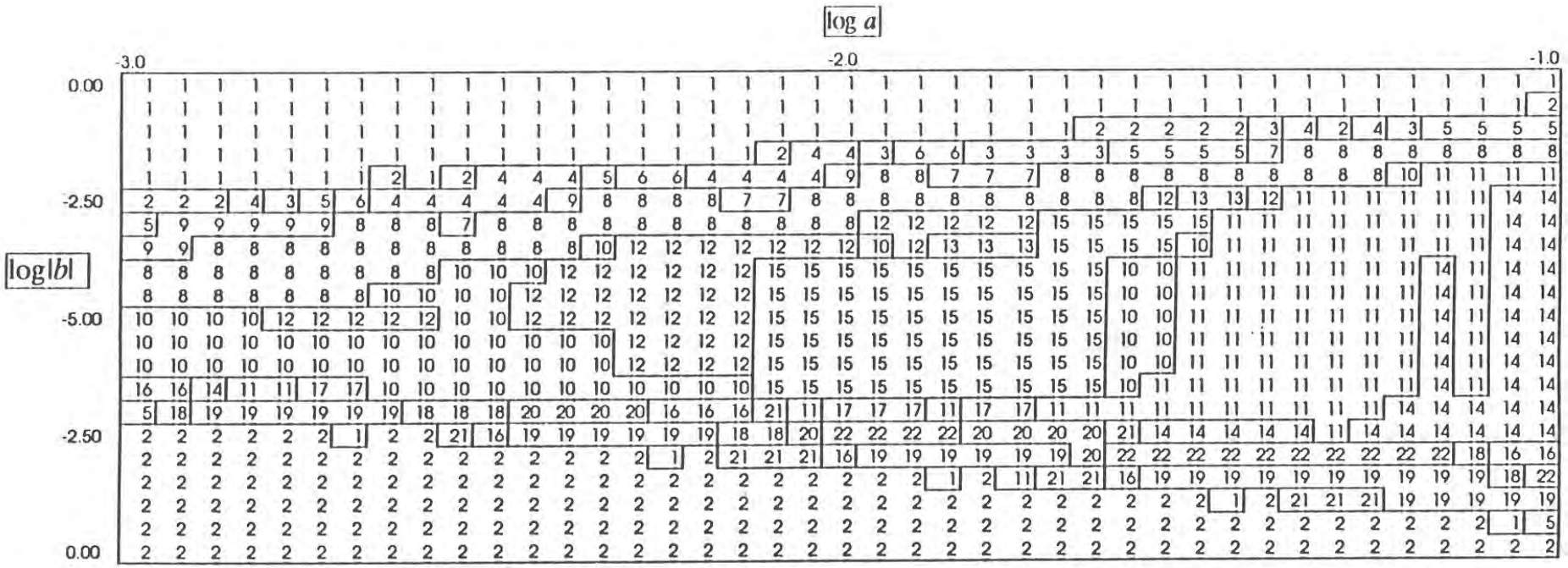
These shape similarity measures can be adopted as *chirality measures,* for both *complete molecules or molecular fragments.*

In L - D transformations, these chirality measures, determined as the global measure for the whole molecule, or the local measure for a molecular fragment, show characteristic changes, indicating in which stage are the *local or the global chirality propeties dominant.*

Allyl alcohol, three nested isocontours, of 0.2, 0.1, and 0.01 a.u.



Global (a,b)-map, the global shape code for Allyl alcohol



Legend	
1 = -2	12 = 004
2 = 0	13 = 0004
3 = 000	14 = 4
4 = 00	15 = 0003
5 = -1	16 = 7
6 = 0000	17 = 03
7 = 0002	18 = 8
8 = 002	19 = 6
9 = 02	20 = 9
10 = 003	21 = 5
11 = 3	22 = 10

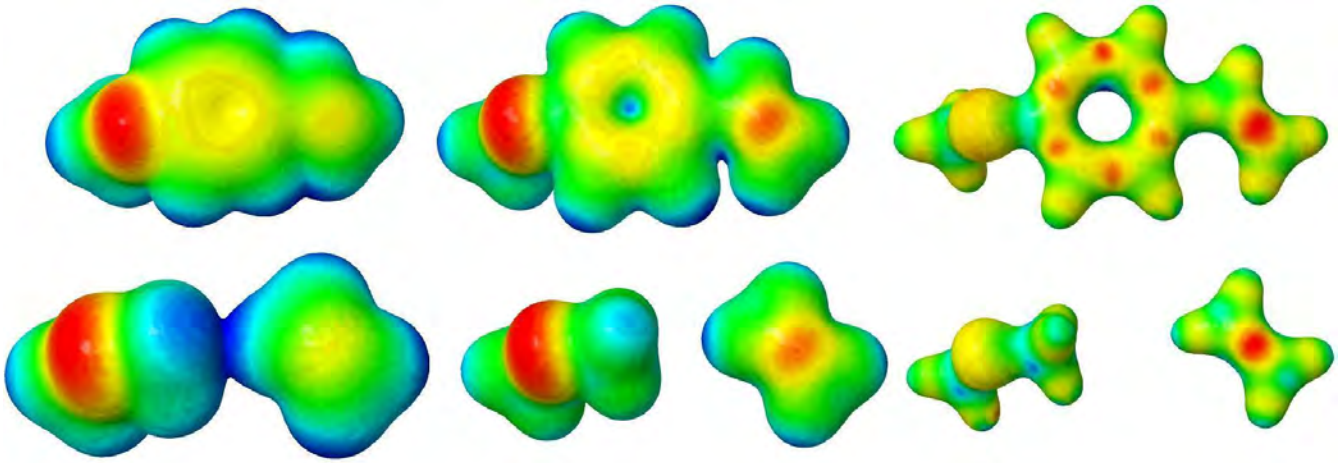
How Much Molecular Fragments Know and What Are They Good For ?

Complementary roles and separation of main components of interaction effects: through-bond and through-space effects

- Complex phenomena are easier to understand if they are manifested
- in simple systems, such as styrene and its derivatives.

Both through-bond and through-space (ringed system)

Through-space only ("ringless" system)



0.001 a.u. 0.01 a.u. 0.1 a.u.

- HF/cc-pVTZ, B3LYP/cc-pVTZ, MP2/cc-pVTZ for geometry calculation, followed by the shape code generation for the vinyl part.
- Approach for "through-space" only:
  - Get rid of the aromatic ring
  - Complete the fragments to be chemically correct
  - SP calculations/MO calculations – shape code for vinyl part.

Phys. Chem. Chem. Phys., 2014, **16**, 918–932

PCCP

PAPER

Electron density shape analysis of a family of through-space and through-bond interactions

Cite this: *Phys. Chem. Chem. Phys.*, 2014, **16**, 918

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

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 property–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.

Received 18th September 2013,
Accepted 8th November 2013

DOI: 10.1039/c3cp53954g

www.rsc.org/pccp

Shape Similarity

- A numerical **measure of shape similarity** of any two electron densities is obtained by comparing the (a,b) maps (the shape codes of the Betti numbers), taking the ratio of matches against the total number of elements of the shape codes(861).
- “1.0” is a perfect match and “0.0” means no similarity.
- For a family of molecules, these shape similarity measures are collected into a **Similarity Matrix.**

Similarity Matrix (DFT)

	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

- All possible pair similarities are present.
- Extractable information:
  - Directly: Through-space (TS) effects
  - Indirectly: Through-bond (TB) effects
  - Correlations, by choosing an appropriate reference molecule
  - One Similarity Matrix for every level of theory and results can be analyzed separately.

Through-Bond effect

DFT SM results displayed.

The numbers will help us understand the magnitude of the TB effect, the smaller the similarity value, the larger the TB effect.

Strong interactions via mobile *pi*-system should result in low similarity values and vice versa

Dominant effect, it is expected to be large, compared to TS.

	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

Through-Space effect

Appropriate part of SM is analyzed to get the TS effect.

Not very dramatic, the greatest variation less than 5%, but at that distance it is expected to be low.

Mostly electrostatic interaction.

The addition of atoms when completing the fragments further decreased the effect of pure functional group.

Good agreement between models when this effect is being studied.

Small, important because they help calculate the TB effect and after all, it is a direct measure.

MP2

PARA FUNCTIONAL GROUP

		SIMILARITY
1	och_3	*0.9683*
2	cf_3	*0.9708*
3	$n(ch_3)_2$	*0.9712*
4	no_2	*0.9736*
5	cho	*0.9748*
6	cis vinyl	*0.9754*
7	f	*0.9818*
8	cn	*0.9824*
9	trans vinyl	*0.9836*
10	ch_3	*0.9842*
11	oh	*0.9865*
12	nh_2	*0.9871*
13	$coch_3$	*0.9871*
14	sty	*1.0000*

Shape Deviation Measures

- A system devised to better display hard to see results.
- It's a measure that shows right away the dominance and strength of an effect of a functional group.
- 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.

		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

MP2 SDM (above)

Phys. Chem. Chem. Phys., 2014, **16**, 6666–6678

PCCP

PAPER

Substituent effects and local molecular shape correlations

Cite this: *Phys. Chem. Chem. Phys.*, 2014, **16**, 6666

Zoltan Antal and Paul G. Mezey*

Received 9th December 2013,
Accepted 10th February 2014

DOI: 10.1039/c3cp55192j

www.rsc.org/pccp

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 disubstituted benzenes, in correlation with their activating and deactivating characteristics as observed 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.

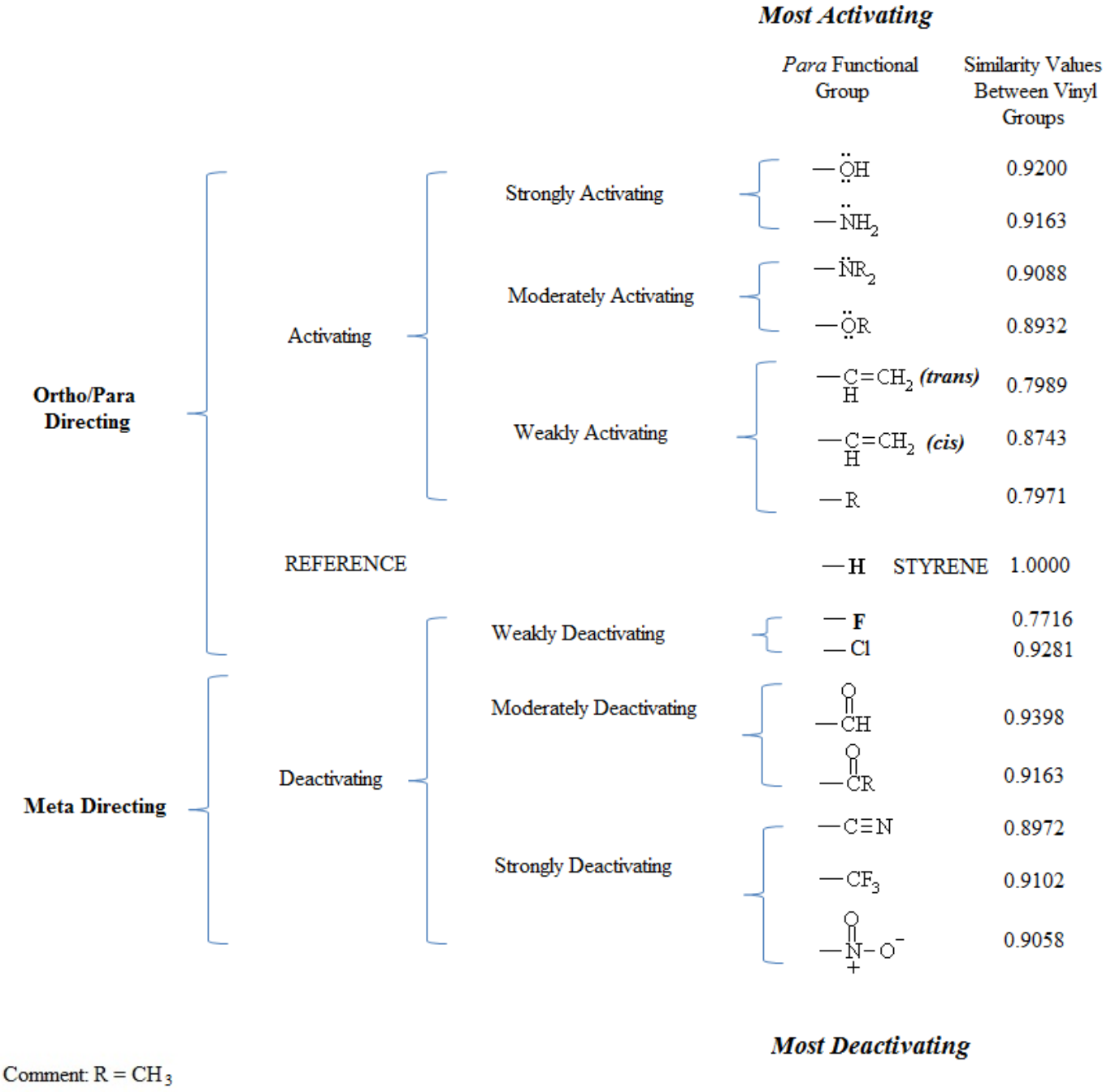
Shape Similarity Study of Substituent Effects

- Approaching the method from an organic chemist's point of view: *ortho*-, *meta*-, and *para*-directing substituents can be activators and deactivators in Aromatic Electrophilic Substitution type of reactions.
- "This ring ain't big enough for the both of us !" - two competing functional groups on one ring – *di*-substituted benzenes instead of para-substituted STY
- Functional groups modify the shape of the rest of the molecule, and these shape changes can be correlated with the type and strength of activation/deactivation.

- Styrene is reference – only the functional group is different in every case → shape change is the consequence of the functional group.
- What results are expected? - energetically most stable arrangement and its outcome.
- More activating/para directing the functional group → less influence on vinyl shape.
- More deactivating/meta directing → more influence on vinyl shape.

Styrene and *Para-Nitro* Styrene (above)

- Hammett tables for typical functional groups and similarity values
- Results correlate well with experimental data.
- NH2 and COCH3 results are the same! Does this mean that they look the same ? (DFT)
- Activator – deactivator characteristics.



Fragment Shape Variation Index

Various molecular fragments and substituents are capable of different extent of shape changes, and it is useful to view their actual shape changes in the context of the variability of such shape changes.

For a substituent A, a small shape change may be more significant than a large shape change for substituent B, if the shape variability of A is much less than that of B. Example: small/compact fluorine and large/"bulgy" sulfur.

Shape changes can be fully understood if one looks at local and global shape changes and at their interrelation too.

The concept of **Shape Variation Index** has been introduced in order to address this question.

Aromatic ring comparison and FSVI

Polycyclic Aromatic Hydrocarbons, 1st carcinogen discovered (benzopyrene).

Aromatic ring shape similarity data correlates well with aquatic toxicity results.

FSVI: indicates the "importance" of change by looking at the whole picture instead of individual components.

Local and global shapes are interrelated → some features might not be obvious if treated separately.

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
Fluoranthen	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
Phenanthren	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

The similarity between the shapes of any two objects, such as the shapes of two faces or the shapes of two molecules, can be evaluated by the *Shape Group Method*, using computer programs (see some details later).

The shape difference between the L and D forms provides a numerical measure for the Degree of Chirality:

a small value for the only slightly chiral objects,

and a large value for very chiral objects.

For the natural amino acids, the *molecular fragment-based local chirality measures*, compared to *global chirality measures of the complete molecules*, are expected to show correlations with relative rates of enantiomerization, which, in turn, *might provide new insight into the mechanisms of such, often harmful changes of amino acids,*

such as those leading to cataract formation.

Functional Groups

Quantum Chemical Functional Groups

A simple electron density criterion has been introduced for the identification and quantum chemical definition of a generalized concept of functional groups [10], based on the following simple principle:

if a group of atomic nuclei within a molecule is separated from the rest of the nuclei by an isodensity contour surface, then these nuclei and an associated fuzzy part of the electron density cloud possess some degree of "autonomy" within the molecule.

This "autonomy" might be of a rather limited degree but it is still reminiscent to the autonomy of two molecules placed in the vicinity of each other, close enough that some electron density overlap occurs between them, but not close enough for any chemical reaction and noticeable rearrangement of the bonding pattern to occur. In both cases, there exist molecular isodensity contours, MIDCO's, separating the set of nuclei of a quasi or truly autonomous molecular entity (either that of a molecular component of limited autonomy, or those of two individual molecules).

Due to their characteristic reactivity, only modestly modified by their surroundings, it is natural to expect from ***chemical functional groups within a molecule to possess such isodensity contours.*** Conversely, one may regard fuzzy molecular components with such isocontours as generalized functional groups [10].

An Approach to Functional Groups in Quantum Chemistry

P.G. Mezey, *Quantum Chemical Shape: New Density Domain Relations for the Topology of Molecular Bodies, Functional Groups, and Chemical Bonding,*

Canad. J. Chem., **72**, 928-935 (1994). (Special issue dedicated to Prof. J.C. Polanyi).

P.G. Mezey, *Functional Groups in Quantum Chemistry,*

Advances in Quantum Chemistry, **27**, 163-222 (1996).

P.G. Mezey, *Local Electron Densities and Functional Groups in Quantum Chemistry,* in "Topics in Current Chemistry, Vol. 203, "Correlation and Localization", Ed. P.R. Surjan, Springer-Verlag, Berlin, Heidelberg, New York, 1999, pp. 167-186.

G. M. Maggiora and P.G. Mezey, *A Fuzzy Set Approach to Functional Group Comparisons Based on and Asymmetric Similarity Measure,*

Int. J. Quant. Chem, **74**, 503-514 (1999).

P.G. Mezey, K. Fukui, and S. Arimoto, *A Treatment of Small Deformations of Polyhedral Shapes of Functional Group Distributions in Biomolecules,*

Int. J. Quant. Chem., **76,** 756-761 (2000).

P.G. Mezey, *Computer-Aided Drug Design: Some Fundamental Aspects*

J. Mol. Mod., **6**, 150-157 (2000).

Some background:

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

Density Domains and Functional Groups

an Approach Based on Molecular Isodensity Contours, MIDCOs

Density Domain DD(K,a) is a domain enclosed by a MIDCO C(K,a) :

$DD(K,a) = \{ \mathbf{r} : \mathbf{r}(K,\mathbf{r}) \geq a \}.$

The density domain and the fuzzy electron density fragmentation approaches have been suggested for a quantum chemical criterion and representation of formal ***functional groups***.

P.G. Mezey, *Quantum Chemical Shape: New Density Domain Relations for the Topology of Molecular Bodies, Functional Groups, and Chemical Bonding,*

Canad. J. Chem., **72**, 928-935 (1994). (Special issue dedicated to Prof. J.C. Polanyi).

P.G. Mezey, *Functional Groups in Quantum Chemistry,*

Advances in Quantum Chemistry, **27**, 163-222 (1996).

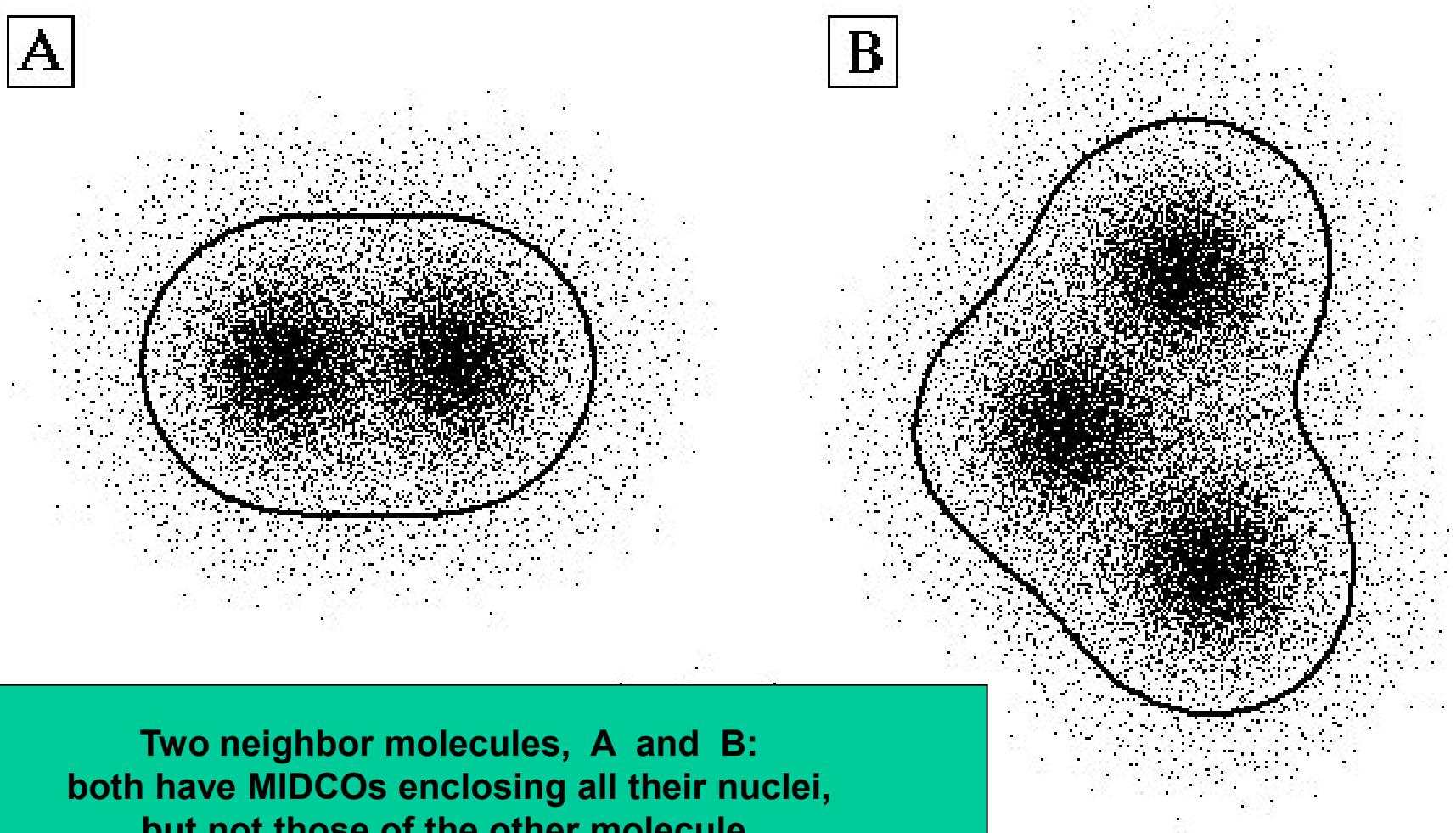
Density Domains and Functional Groups

an Approach Based on Molecular Isodensity Contours, MIDCOs

A useful analogy: consider two molecules near one another.

As long as these molecules have separate identity, each must have some Density Domain containing all the nuclei of the molecule, but none of the nuclei of the other molecule. Separate identity is manifested by such density domains.

Functional groups and limited autonomy within molecules



Two neighbor molecules, A and B:
both have MIDCOs enclosing all their nuclei,
but not those of the other molecule.
This indicates some degree of *autonomy* for both molecules

Density Domains and Functional Groups

an Approach Based on Molecular Isodensity Contours, MIDCOs

Consider now a *single molecule* and one of its connected density domains DD(K,a) and the nuclei enclosed by it. The very fact that this subset of the nuclei of the molecule is separated from the rest of the nuclei by the boundary C(K,a) of the density domain DD(K,a) indicates that these nuclei, together with the local electronic density cloud surrounding them, represent a sub-entity of the molecule, with limited autonomy, and some degree of individual identity.

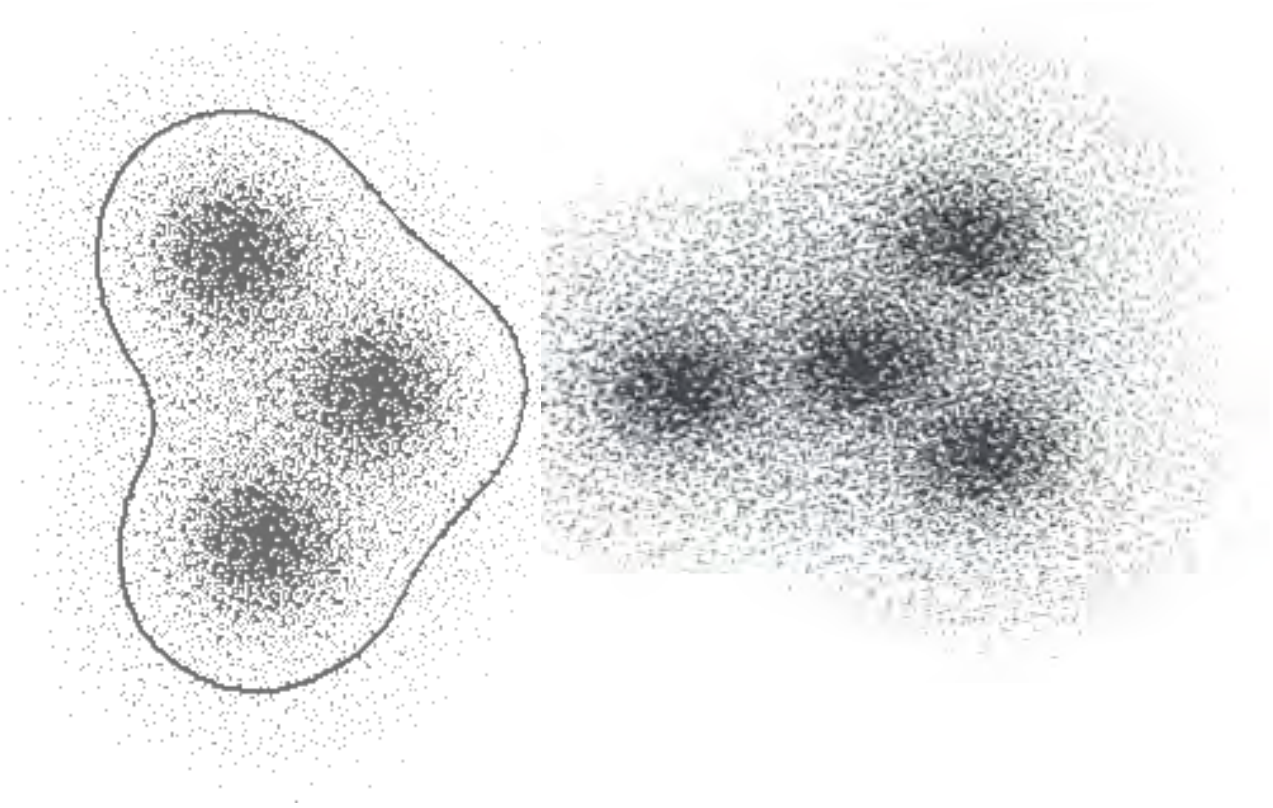
It is natural to regard such a density domain DD(K,a) as a criterion and the fuzzy density fragment for the nuclei within DD(K,a) as a representative of a formal ***functional group***.

Functional groups and limited autonomy within molecules:

use the same condition as for two separate molecules.

Take a single molecule:

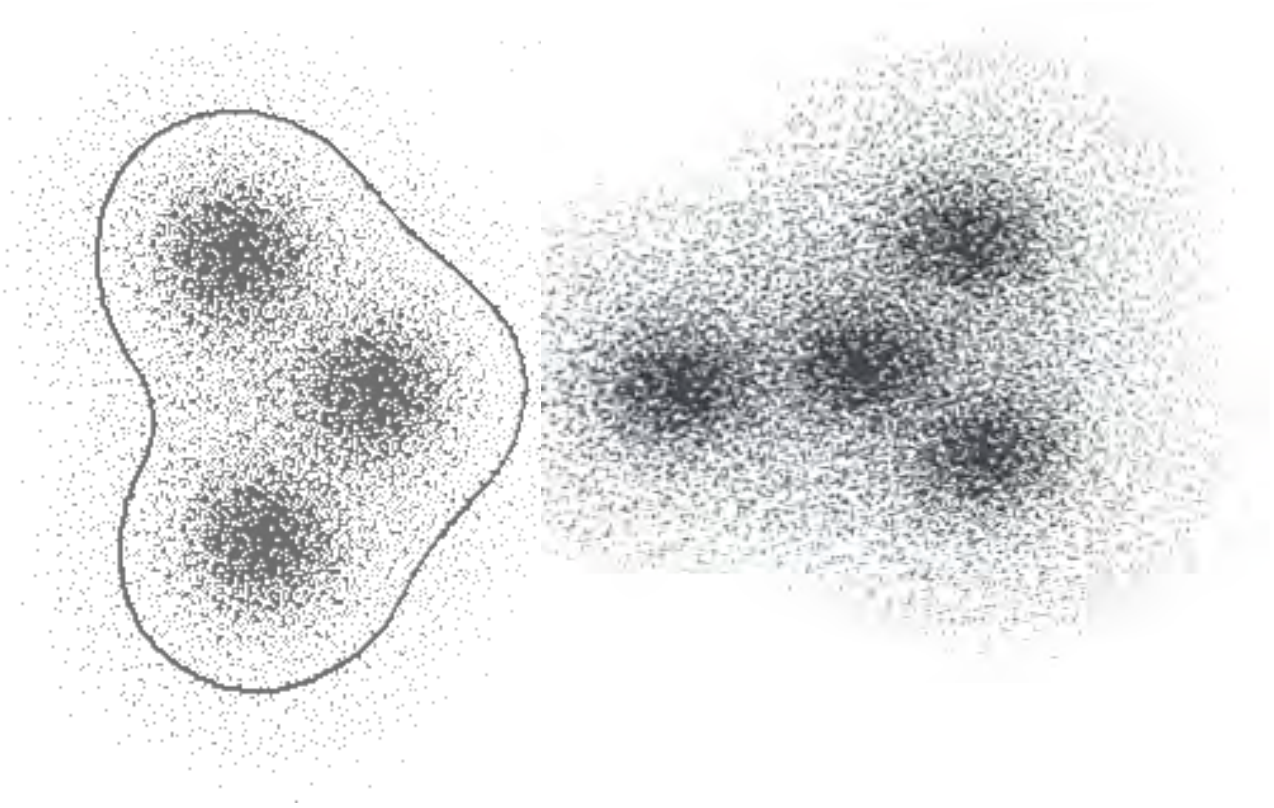
The existence of a MIDCO separating a set of nuclei from the rest of the nuclei of the molecule, indicates a *local, limited autonomy* of a Functional Group



Functional groups and limited autonomy within molecules,

using the same condition as for two separate molecules.

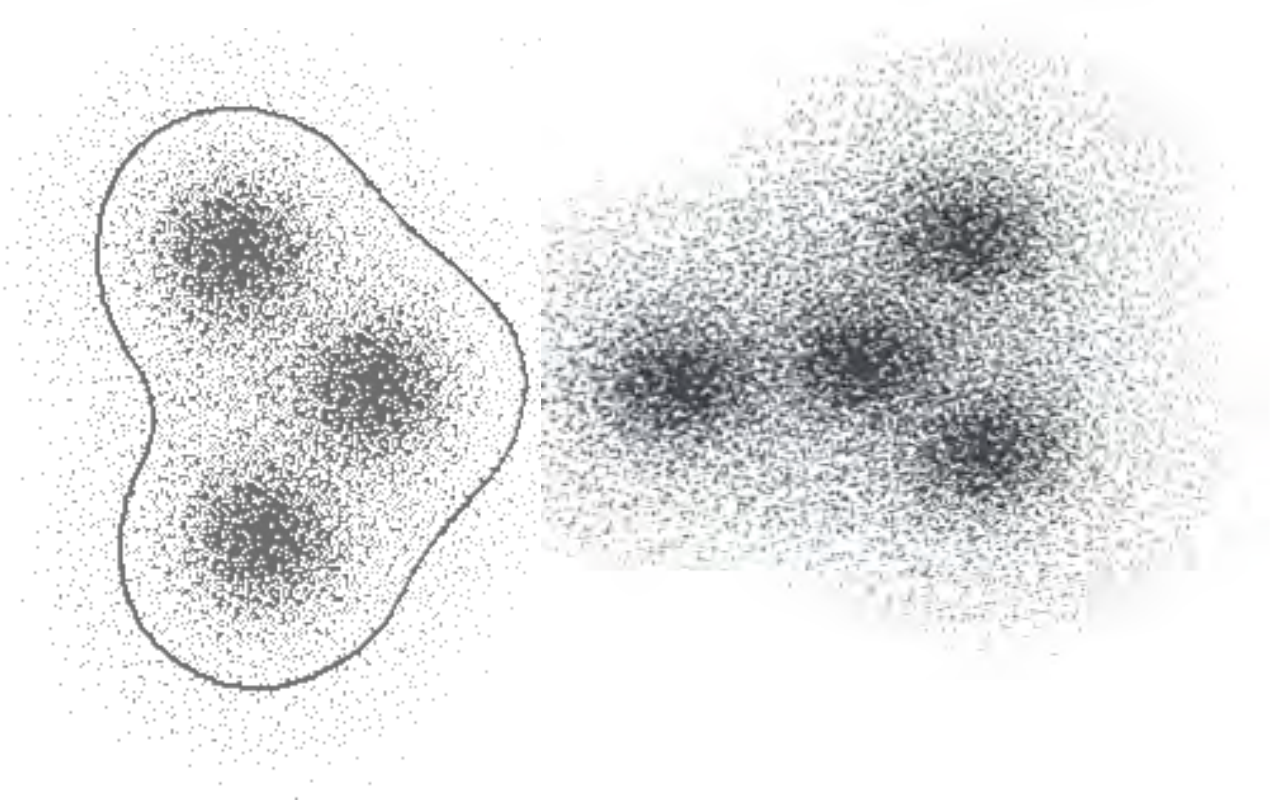
Fuzzy functional group: **the fuzzy ADMA fragment for the nuclei of DD(K,a)**



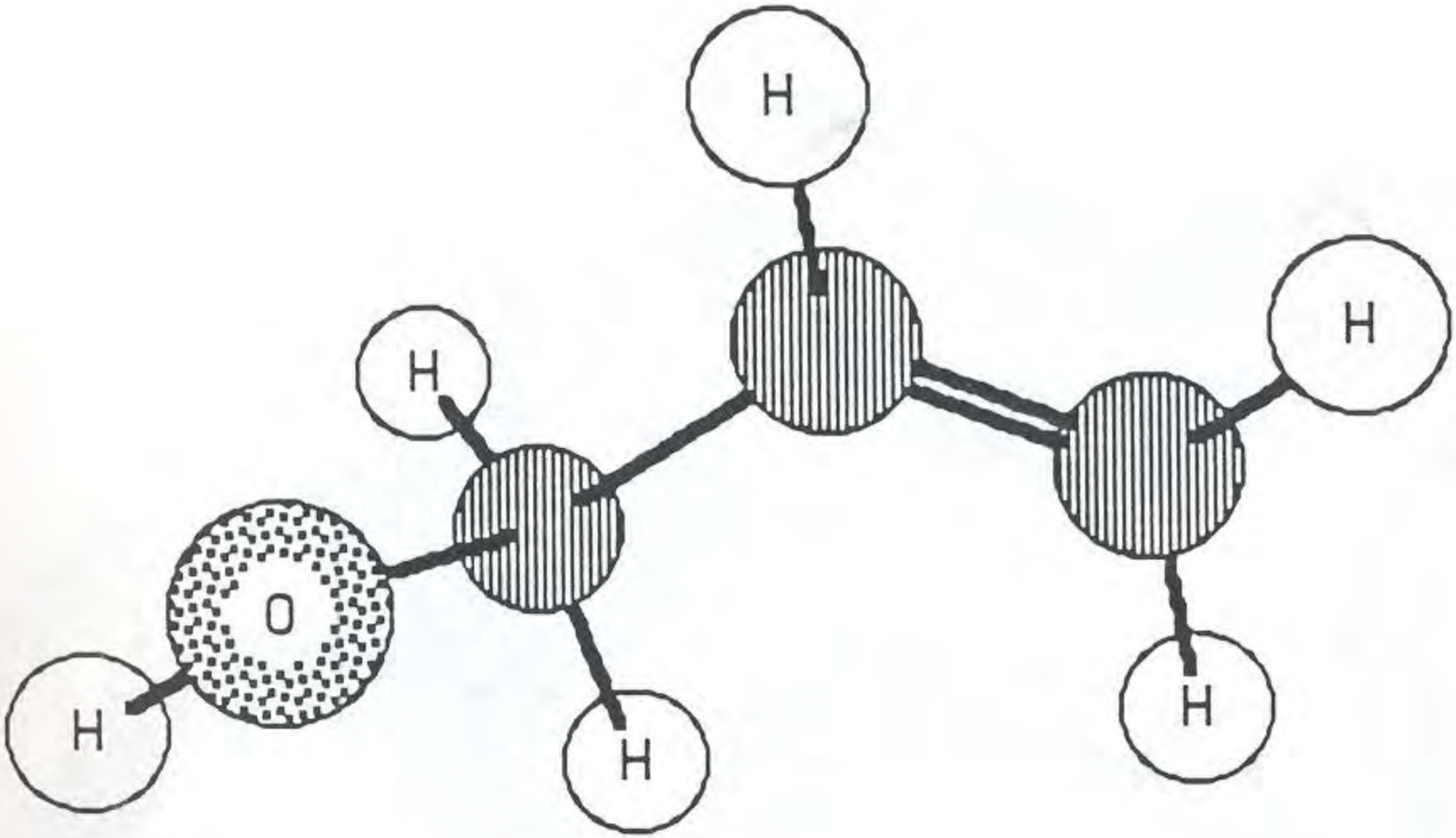
Functional groups and limited autonomy within molecules,

using the same condition as for two separate molecules.

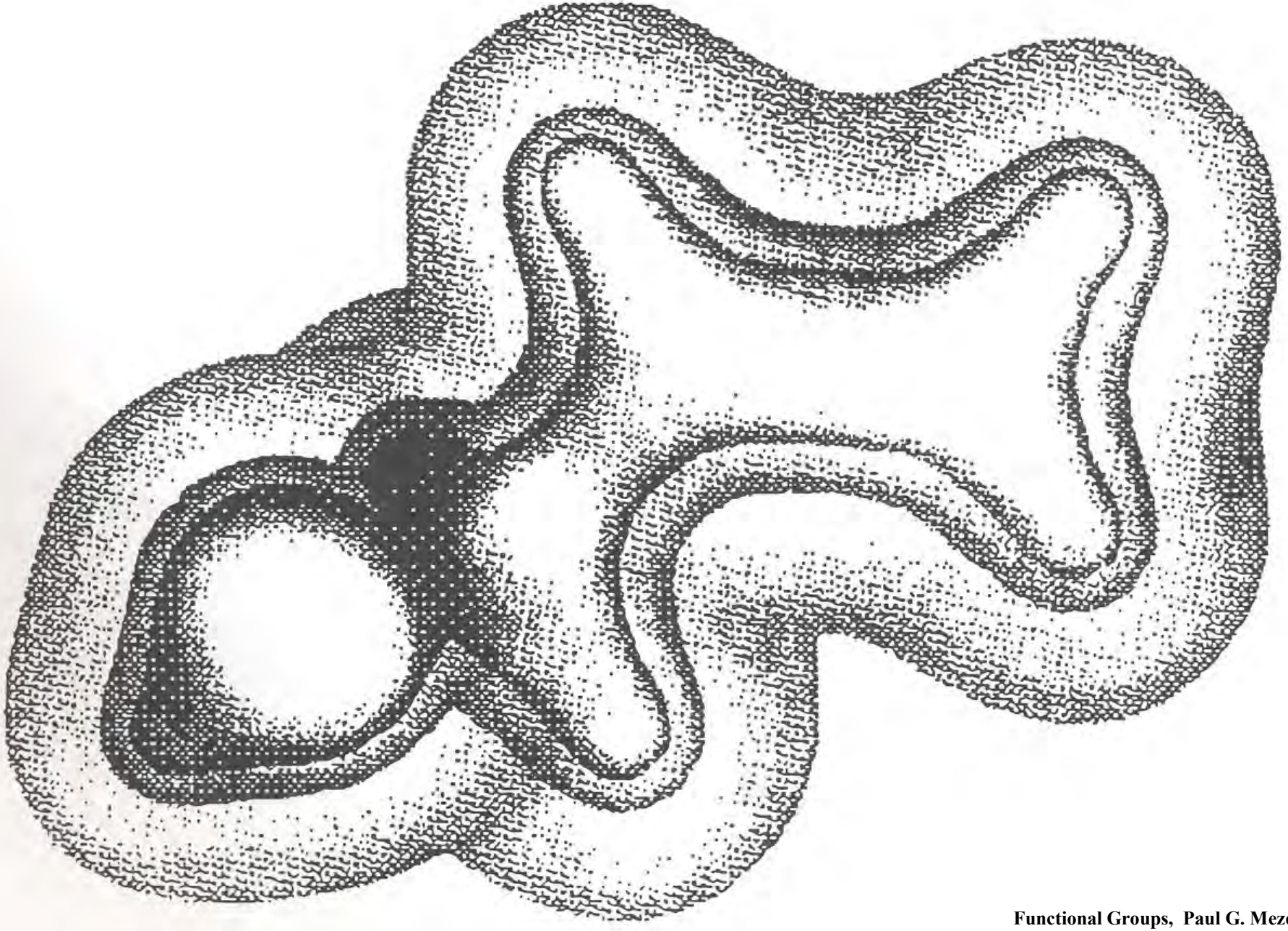
This criterion does not directly address another aspect of functional groups: reactivity, although it is a natural expectation that functional groups show characteristic reactivity properties.



Allyl alcohol, traditional structural formula



Allyl alcohol, three nested isodensity contours, of 0.2, 0.1, and 0.01 a.u.



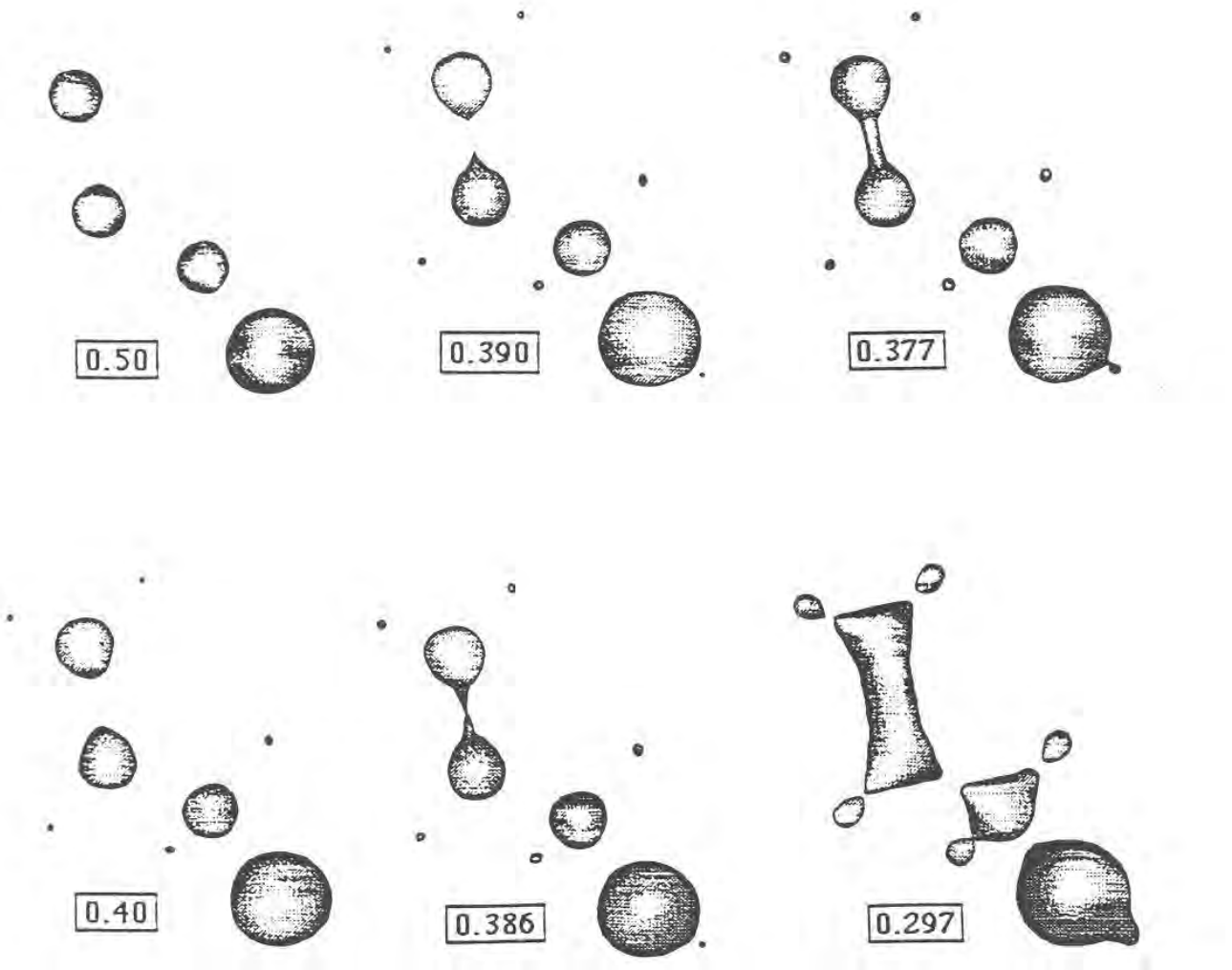


Figure 2.5 Some of the high density threshold density domains of the most stable conformation of allyl alcohol, CH_2=CH-CH_2-OH, as calculated with the GAUSSIAN 90 and GSHAPE 90 programs, using a 6-31G* basis set.

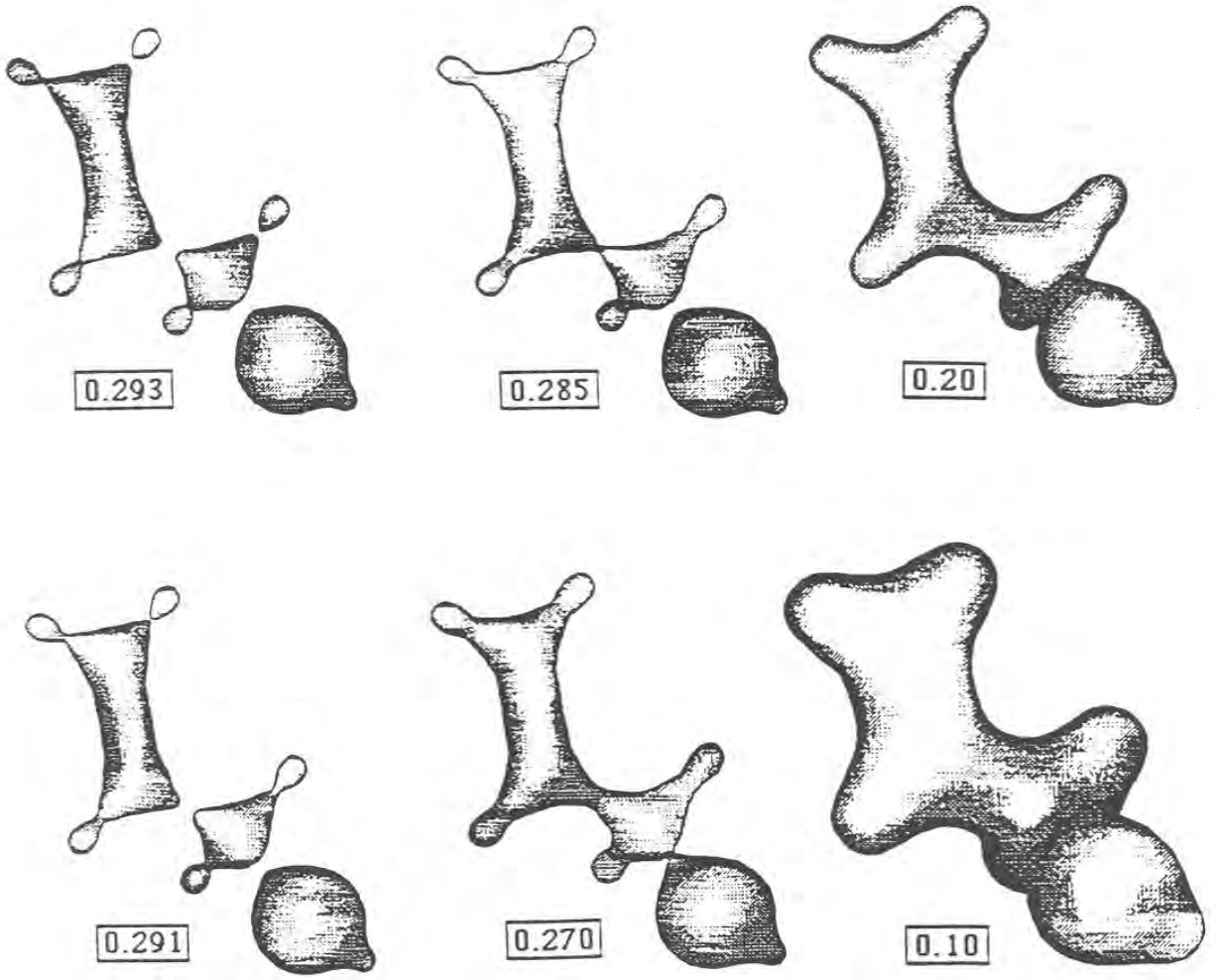
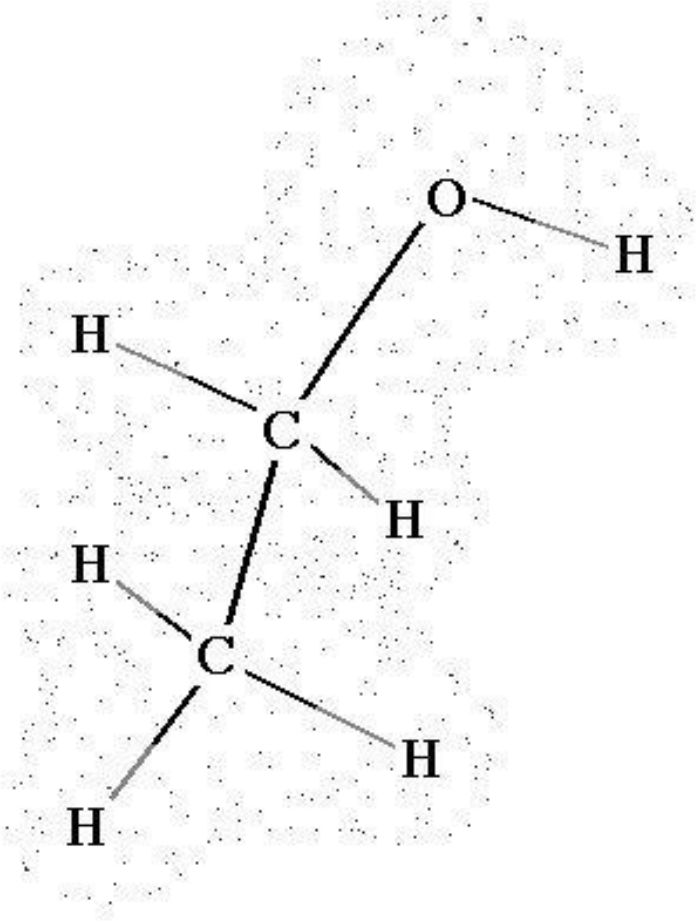
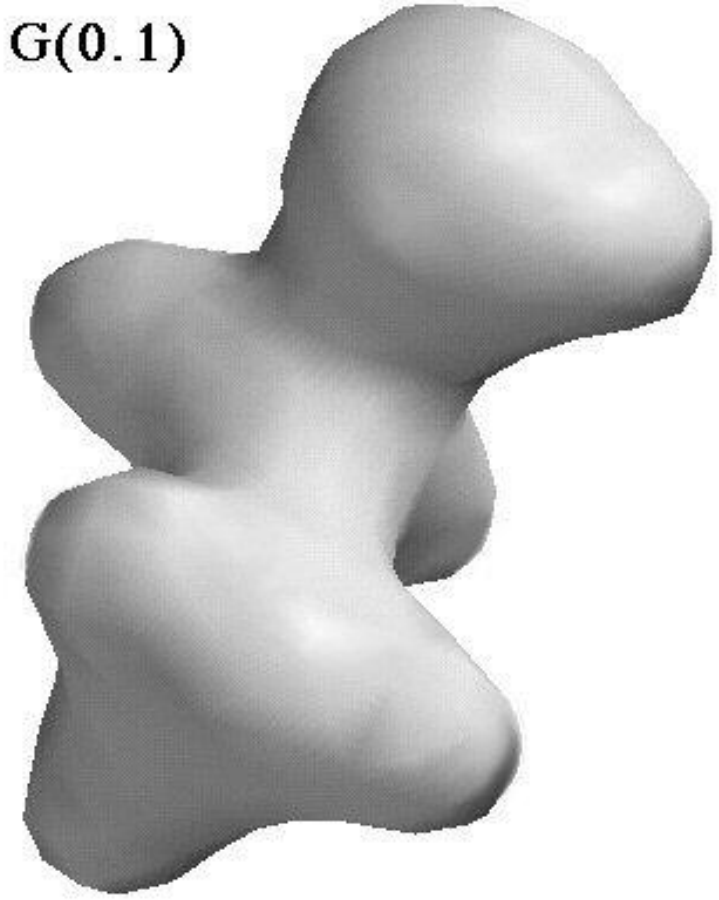


Figure 2.6 Some of the low density threshold density domains of the most stable conformation of allyl alcohol, CH_2=CH-CH_2-OH, as calculated with the GAUSSIAN 90 and GSHAPE 90 programs, using a 6-31G* basis set.

Functional groups and limited autonomy within molecules: example of ethanol (ethyl alcohol).





No methyl group in ethanol

Example:

According to the quantum chemical definition of functional groups, the ethanol molecule contains the

-CH2-CH3 and

-OH

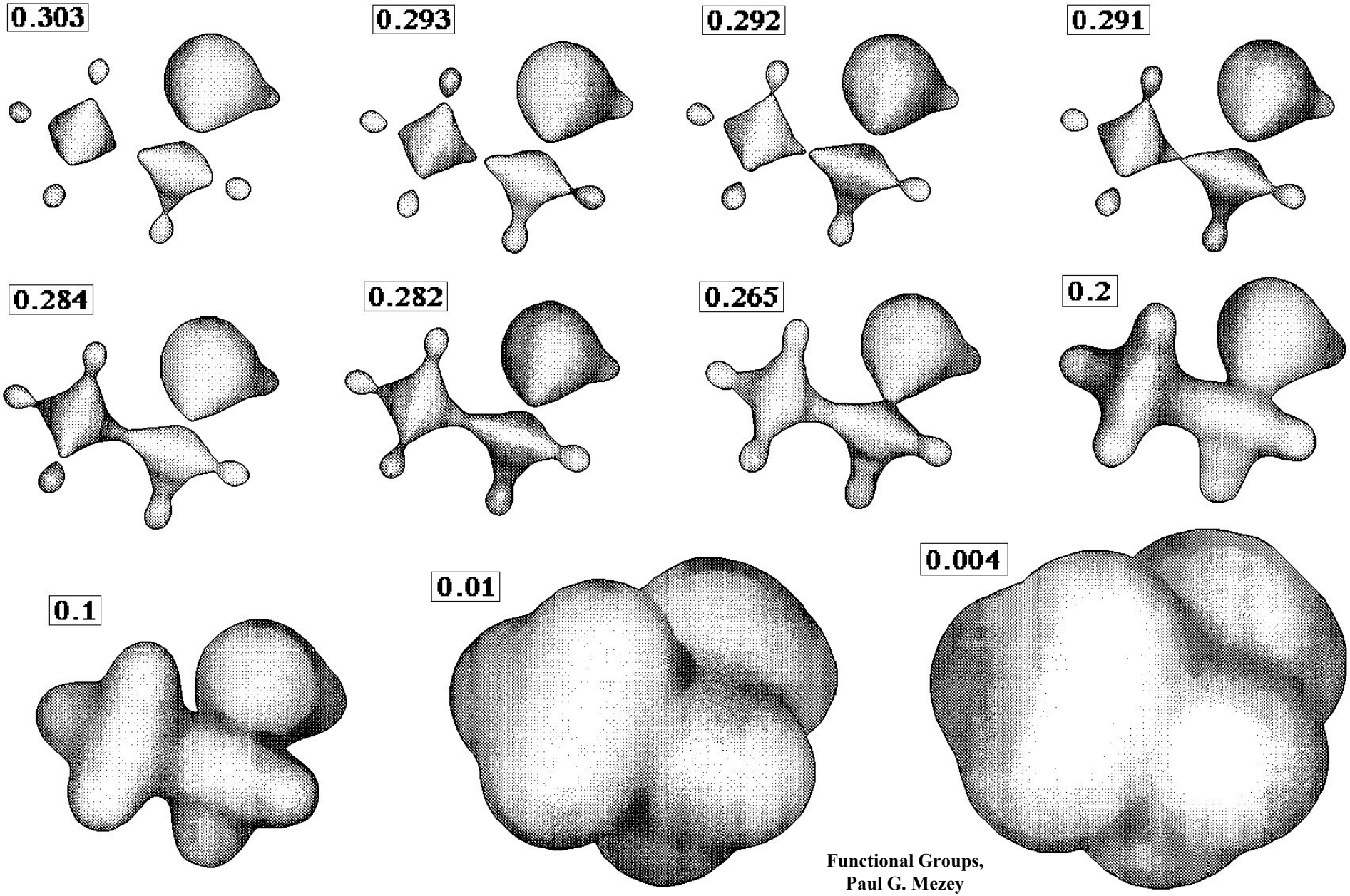
functional groups, but not the

-CH3

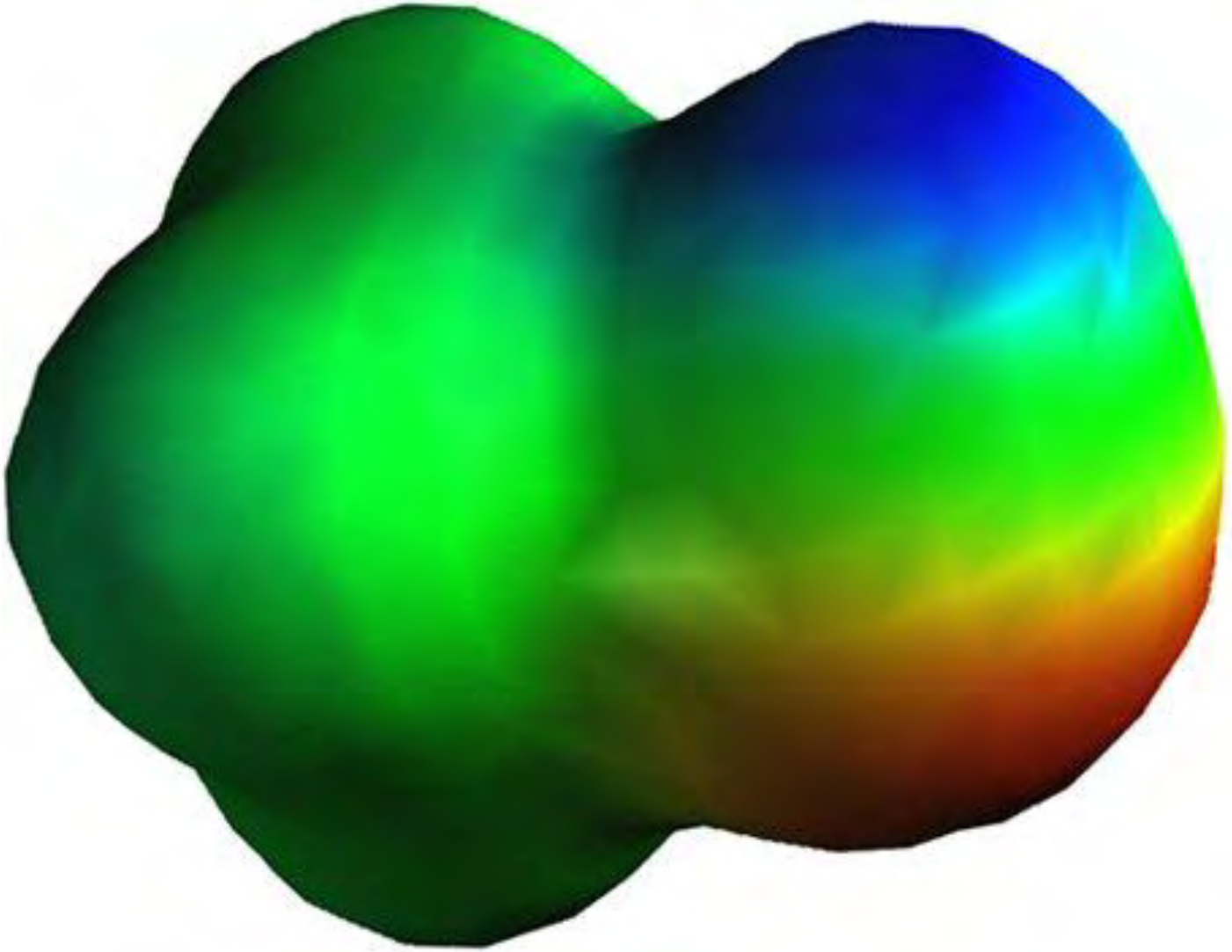
functional group.

There is no methyl group as functional group in ethanol !

In ethanol, (ethylalcohol), CH3CH2OH, there is no methyl group, as functional group !



Functional groups and limited autonomy within molecules: example of methanol, CH3OH.



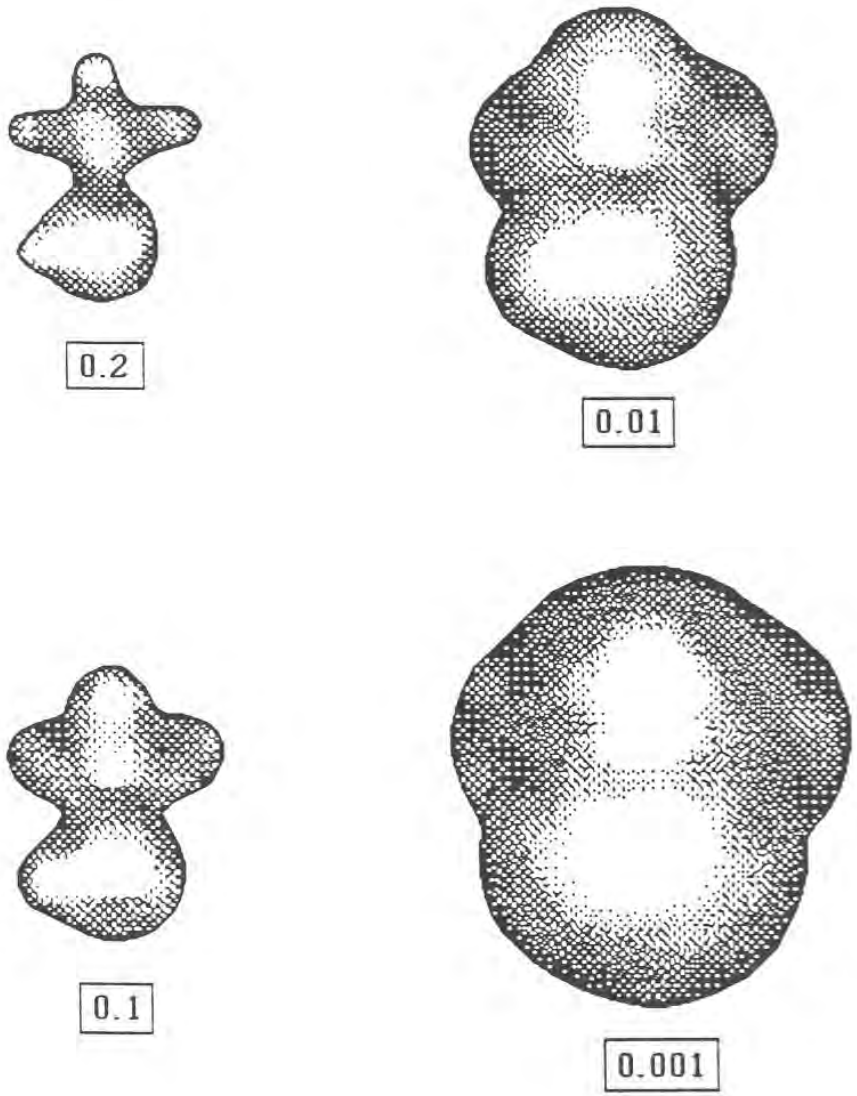


Figure 2.1 Four MIDCO's, $G(a_1)$, $G(a_2)$, $G(a_3)$, and $G(a_4)$ of the methanol molecule CH_3OH are shown for the contour density values $a_1 = 0.20$, $a_2 = 0.10$, $a_3 = 0.01$, and $a_4 = 0.001$, respectively, as calculated with the GAUSSIAN 90 [253] and GSHAPE 90 [254] programs, using a 6-31G* Gaussian basis set. In all figures, the density threshold values a are given in atomic units.

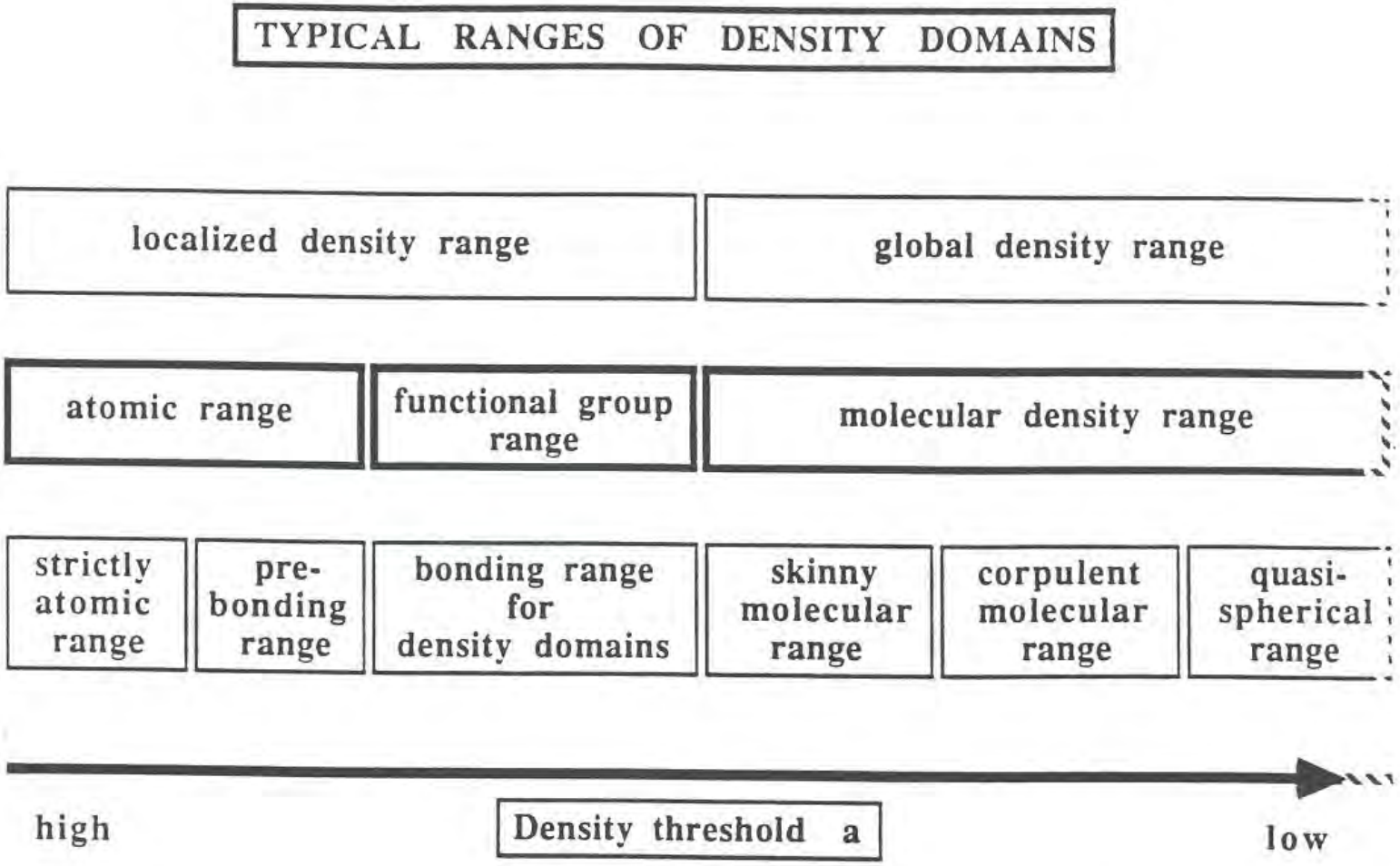


Figure 2.7 Classification of density domains according to ranges of the density threshold a.

Summary: Density Domains and Functional Groups

Density Domain DD(K,a) :

$DD(K,a) = \{ \mathbf{r} : \mathbf{r}(K,\mathbf{r}) \geq a \}$

is the domain enclosed by a MIDCO $C(K,a) = \{ \mathbf{r} : \mathbf{r}(K,\mathbf{r}) = a \}$.

The density domain approach has been suggested for a quantum chemical representation of formal functional groups. (P.G. Mezey, Canad. J. Chem., **72**, 928-935 (1994). (Special issue dedicated to Prof. J.C. Polanyi), and P.G. Mezey, *Functional Groups in Quantum Chemistry,* Advances in Quantum Chemistry, **27**, 163-222 (1996).

Consider a single connected density domain DD(K,a) and the nuclei enclosed by it. The very fact that this subset of the nuclei of the molecule is separated from the rest of the nuclei by the boundary C(K,a) of the density domain DD(K,a) indicates that these nuclei, together with the local electronic density cloud surrounding them, represent a sub-entity of the molecule, with individual identity.

It is natural to regard this density domain DD(K,a) as a criterion and a representative of a formal functional group.

Fuzzy functional group: the fuzzy ADMA fragment for the nuclei of DD(K,a)

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

The local and global information content of molecules

- **How different two, formally identical functional groups can be?**
- **How do local ranges of molecules influence global properties?**
- **How are global molecular properties reflected in local ranges of molecules?**
- **The Holographic Principle of Electron Densities .**

Functional group polyhedra

Take a set of functional groups from a molecule M, and define a unique point for each. Take the convex hull of these points, this defines a functional group polyhedron FGP(M) for the molecule M

This polyhedron provide a simplified representation of some essential geometrical aspects of the molecule M.

The symmetry and chirality properties of an FGP(M) polyhedron can be studied directly, leading to a "low resolution" description, and an FGP(M) polyhedron can be compared directly to another FGP(M') polyhedron.

Whereas the original input information comes from quantum chemistry, some of the tools for the above purposes are provided by topology and discrete mathematics.

P.G. Mezey, K. Fukui, and S. Arimoto, *A Treatment of Small Deformations of Polyhedral Shapes of Functional Group Distributions in Biomolecules,*
Int. J. Quant. Chem., 76, 756-761 (2000).

J. Phys. Chem. A **2003,** *107,* 11120−11127

Similarity and Chirality: Quantum Chemical Study of Dissimilarity of Enantiomers

Greet Boon,[†] Christian Van Alsenoy,[‡] Frank De Proft,[†] Patrick Bultinck,[§] and Paul Geerlings*,[†]

Free University of Brussels, Faculteit Wetenschappen, Eenheid Algemene Chemie, Pleinlaan 2, B-1050 Brussels, Belgium, University of Antwerp, Department of Chemistry, Universiteitsplein 1, B-2610 Antwerp, Belgium, and Ghent University, Department of Inorganic and Physical Chemistry, Krijgslaan 281 (S-3), B-9000 Gent, Belgium

Received: July 4, 2003; In Final Form: October 6, 2003

In the analysis of molecular similarity and chirality of enantiomers performed in this paper, we introduce a new local similarity index based on the Hirshfeld partitioning. In the framework of conceptual density functional theory and considering the enantiomers of the halomethane CHFClBr and of the amino acids alanine and leucine, this index is used to investigate the dissimilarity of chiral molecules. Furthermore, we illustrate Mezey's holographic electron density theorem.

ELSEVIER

Journal of Molecular Structure: THEOCHEM 727 (2005) 49–56

THEO CHEM

www.elsevier.com/locate/theochem

Molecular quantum similarity of enantiomers of amino acids: a case study

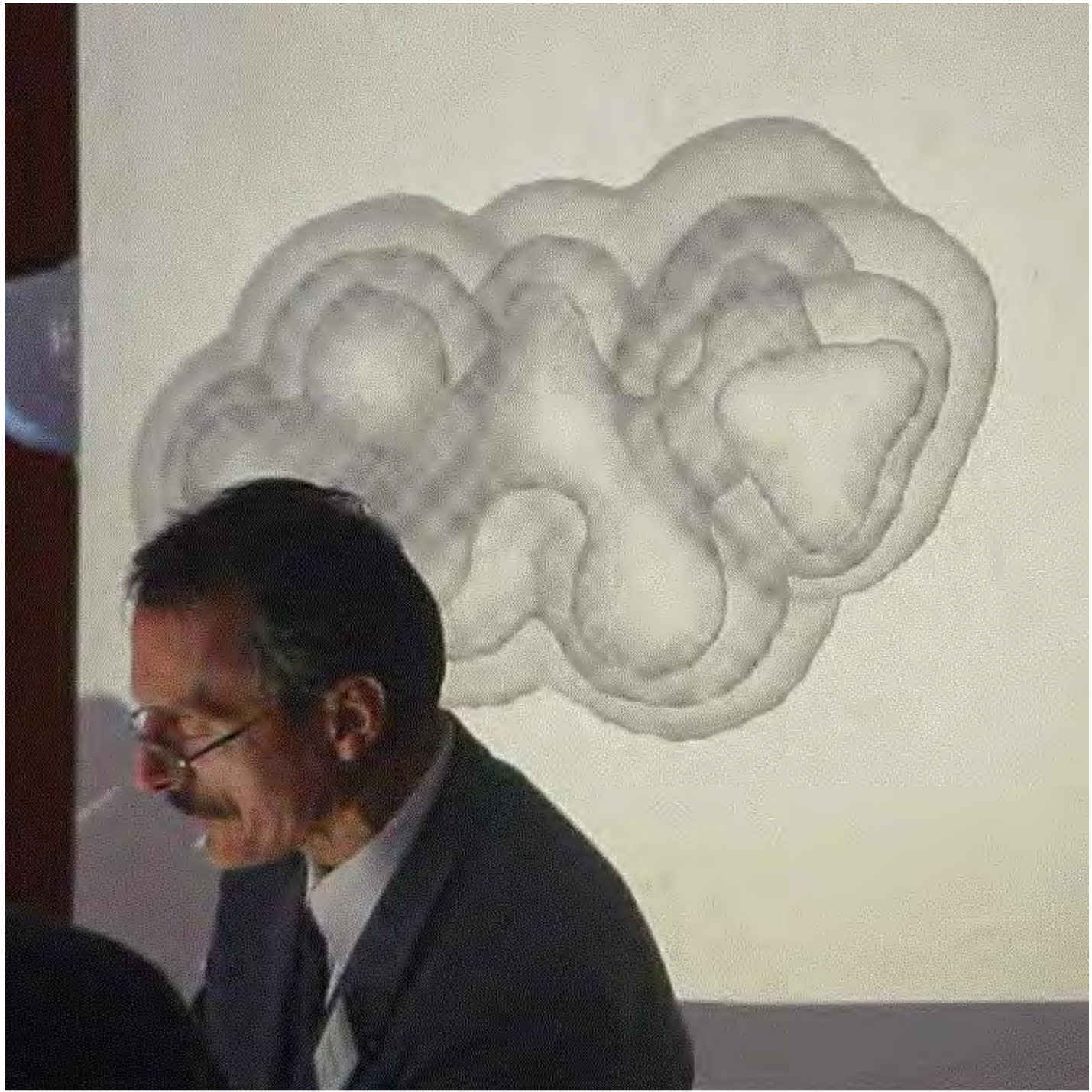
Greet Boon[a], Christian Van Alsenoy[b], Frank De Proft[a], Patrick Bultinck[c], Paul Geerlings[a,*]

[a]*Faculteit Wetenschappen, Eenheid Algemene Chemie (ALGC), Free University of Brussels (VUB), Pleinlaan 2, B-1050 Brussels, Belgium*
[b]*Department of Chemistry, University of Antwerp (UA), Universiteitsplein 1, B-2610 Antwerp, Belgium*
[c]*Department of Inorganic and Physical Chemistry, Ghent University (UGent), Krijgslaan 281 (S-3), B-9000 Gent, Belgium*

Received 26 October 2004; accepted 18 January 2005
Available online 24 June 2005

Abstract

In this paper, methodological in nature, molecular quantum similarity is evaluated for enantiomers in the case of molecules showing conformational flexibility, proposing the use of a Boltzmann weighted similarity index. As a case study, the conformers of the enantiomers of the amino acids Alanine and Serine were examined. The second aim is, next to studying global indices, the evaluation of local similarity using our earlier proposed local similarity index based on the Hirshfeld partitioning, in order to further quantify the consequences of Mezey's Holographic Electron Density Theorem in chiral systems.



Molecules are
on my mind

or

the "halo"
of a
Saint Chemist