Complementing the Shape Group Method: Assessing Chirality

A Thesis submitted to the

College of Graduate Studies and Research

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the

Department of Chemistry

University of Saskatchewan

By

Peter Lloyd Warburton

Spring 2004

COPYRIGHT

Acknowledgements

Many people have guided me during the development of this thesis, and so, I would be remiss if I did not acknowledge their significant contributions to this work.

I must first thank the members of my Advisory Committee, Drs. Katie Mitchell, David Palmer, Ron Steer, and Stephen Urquhart for their willingness to participate in an advisory capacity on a research project that was, for the most part, quite different from their own research experience. Their ability to constructively criticize and analyze the work pushed me towards refining not only my methods, but also the manner in which I present the results. Additionally, Dr. Paul Mezey, as my supervisor, must be thanked for his generosity in financial support. As well, he allowed me the independence to work on this thesis in my own way, which led to a renewed sense of excitement about research that, for a time, I had lost.

Next, I would like to make special mention of Brenda Weenk, Dwight Reynaud, and Alexis Dahl, all of whom I enjoyed working with during my time in Chemistry Stores.

My time in Saskatoon has been made very special by a group of people who made the good times good and the bad times bearable. Thanks to Jason, Val, Don, Shawn, Kevin, Tanya, Carole, Aaron, Darren, Maley, Sophie, and the rest of the "Tall Canadians."

Finally, I would like to give special recognition to the efforts of Dr. Thomas Exner. Our continual conversations, the ideas he suggested, the coffee we drank, and, especially, his friendship, made this thesis much better than it would have been had he not been a member of the group.

For my parents, Keith and Pat, and my sister, Kathryn.

"So... when are you going to be done?"

Abstract

The Shape Group method is a powerful tool in the analysis of the shape of molecules, and in the correlation of molecular shape features to molecular properties in Quantitative Shape-Activity Relationship (QShAR) studies. However, the main disadvantage inherent in the method is that mirror image molecules are considered to be "exactly" similar. As such, the method requires a complementary chirality measure to allow for complete analysis where chirality is involved.

In this work, two methods of creating chirality measures to complement the Shape Group method are presented. The first is based upon the assigning of handedness values to each array point of the computer file that contains specific property information and uses the parallels between a lattice animal inscribed in a Jordan curve, and the array points inscribed in an isodensity contour. Each array point can then be treated as a face-labelled cube, which is often a chiral object that can have an assigned handedness value. Grouping of these handedness values allows for the creation of chirality measures.

In the second method, the Shape Group method is applied to electron density representations created by subtracting one fragmentary electron density from others and analysing the shape similarities of the resultant difference densities.

With both methods, chirality information that is already embedded within the shape descriptions of electron density representations is emphasized.

The Shape Group method and the developed chirality measures are then used to simply correlate the shape and chirality of the stereogenic carbon of molecules to optical rotation and rotational strengths of various classes of molecules.

Table of Contents

List of Figures

List of Tables

I INTRODUCTION

The Shape Group method[1-47] has been established as a useful tool for chemists in several different areas related to molecular shape analysis. Different theoretical aspects of the Method have been developed to assist in the study of molecules in several different representative forms, including van der Waals surfaces,[2-13] electrostatic potential surfaces,[13-16] electron isodensity contour surfaces[16-29] and nuclear potential surfaces.[30]

Regardless of the molecular representation choice, the Shape Group method can be used in many varied applications. It has been used in the analysis of shape changes due to conformational variations.[2,3,21-23,31-33] Additionally, the Method has been used in the analysis of molecular shape in external electric fields,[34] shape change and Hammond Postulate studies of chemical reactions,[9-11,29,35] density domain and functional group identity studies,[12,36,37] macromolecular shape analysis and protein folding studies,[31,38-40] Quantitative Shape-Activity Relationship (QShAR) studies.[26,27] Its potential role in drug design has also been discussed.[41]

However, one drawback of the Shape Group method is the inability to differentiate between mirror image forms of molecules. By the nature of the Method, where molecular shapes are determined by analyzing the topological equivalents of spheres with holes based upon the curvature domain membership of points on the surface, such differentiation is mathematically impossible. Mirror image compounds will have the same shape and will be identified as "exactly" similar to each other.

However, the chemistry of mirror image compounds, as well as a physical property, e.g. the rotation of plane-polarized light by chiral molecules,[48] show that this "exact" similarity is a mathematical construct that does not represent reality.

The use of chirality measures based upon other molecular similarity measures[49-60] would be an ideal choice for obtaining the supplementary chirality information of molecules, except that all such measures only define the "amount" of chirality, and do not differentiate between mirror image pairs. In terms of Shape Group method analysis of chemical properties, it is this potential differentiation that would be the most important feature. Additionally, a simple measure based upon information gleaned from existing files used in Shape Group work would be preferred.

As a first attempt, initial results utilizing Shape Group analysis of absolute difference densities as a means of measuring chirality of olfactophore molecules were successful in correlating shape features of molecules with perceived smell.[61] These absolute difference densities were obtained by taking the absolute values point-by-point of the files created by subtracting the electron density of one molecule from another. However, while the results were encouraging, the absolute difference density method has many deficiencies. Of these deficiencies, it is the complexity of the shapes generated by taking the absolute values of difference densities that is the most notable. Such complex shapes, the result of contour-in-contour features created by the absolute value mathematics used, lead to small shape similarity numbers, allowing for the determination of only "coarse" chirality information.

Secondly, the complex shapes are often computationally difficult to deal with, in terms of computational effort.

While these other deficiencies still exist, the application of the Shape Group method to difference density representations as opposed to absolute difference density representations should reduce the effects of the complex shape problems. Utilizing difference densities, where both positive and negative valued contours are considered independently, accomplishes a simplification of the shapes to be analyzed by avoiding the contour-in-contour problem.

With the deficiencies of the difference density methods in mind, an attempt at the development of a second type of chirality measure to complement the Shape Group method seemed of potential value. As stated previously, the chirality measure needed to have several properties. First, the measure should be able to distinguish between mirror image forms of molecules, by assigning a sign to each member of the pair based on some inherent, consistent property of the electron density. Secondly, the measure should be continuous, with possible measures from negative to positive numbers. This requirement demands an achiral molecule be assigned a chirality measure of zero. Third, the measure should ideally be obtained from the pre-existing computer files used in the Shape Group analysis of the electron densities of molecules. Chirality measures based upon array point handedness assignment fulfill these requirements.

The development of array point handedness chirality measures is possible due to the array-based nature of computer files to store information. Since such arrays of information resemble lattice animals,[44,45,62-70] any measure that assesses the

differences in chirality of lattice animals will be applicable to the measurement of chirality array-based electron density files.

In this work, the measure of the chirality of a lattice animal is accomplished by placing the animal into a *gradient-labelled space*. Each cell of the lattice animal is then assigned a handedness value based upon where the gradient vector intersects the cell wall. When the handedness values of all the cells are collected, a chirality measure can be defined by the number of cells of each handedness type.

Since isodensity contours in electron density files define an array-based lattice animal in a complicated gradient-labelled space, the developed technique is almost directly transferable to a Shape Group method complementary measure of chirality. Because the gradient-labelled space is generally more complex than that of the lattice animal case, the handedness of an array point cell is determined by the gradient vector defined by the face-label differences mapped onto the faces of each array point "cell" of the file.

With the Shape Group method and the two developed complementary chirality measures, it is possible to simply correlate the magnitudes of optical rotations of amino acids[71,72] and trisubstituted 2,2'-spirobiindanes,[73-76] as well as the magnitudes of the rotational strengths of methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives,[77,78] to the shape features and chirality measures of the stereocentres of the molecules.

II THEORY

II.1 The Shape Group Method

The Shape Group method has an extensive and rigorously defined theoretical basis, which would be too lengthy to be presented here in full. For a more complete picture of the Shape Group method, many of the original works have extensive theory sections.[1-47]

The principle of the Shape Group method is based upon defining a truncated surface into homology groups and then encoding the information of the homology groups into easily compared codes. Homology is a specific mathematical term to describe an equivalence of certain features of a surface. However, discussion of these features requires that certain terms first be defined.

A *p-face* is a domain C(p,i) on the surface of dimension p. In a three-dimensional space, which will be used for all further examples, a surface is a two-dimensional construct. Within the surface, three types (n types in an n-dimensional space) of domains can be defined. These are 0-faces (points on the surface), 1-faces, (lines upon the surface), and 2-faces, which are areas upon the surface. In Figure II.1.1 we see a simple example of a small portion of a surface, denoted as G. In the Figure, the 0-faces are denoted by the lower case letters "w" to "z", the 1-faces are denoted by "a" to "e", and the 2-faces are "A" and "B".

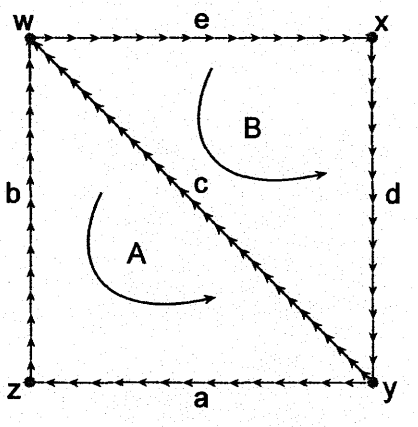


Figure II.1.1 A simple example of p-faces on a small portion of a surface.

A *p-chain* (c^p)is a linear combination of p-faces with the general form

$$c^p = \sum_i u_i C(p,i) \qquad (II.1.1)$$

where u_i denotes the *i*th coefficient acting upon the *i*th p-face. Such p-chains can be subjected to mathematical manipulation such as addition or multiplication to form a group C^p. An example, for instance, of a 1-chain is described by c+d+e in Figure II.1.1.

Each p-face can be assigned an *orientation*. For convenience, this orientation takes an arbitrary value of +1 or –1. For all 0-faces it is generally always convenient to assign an orientation of +1. The orientation of a 1-face is denoted graphically by an arrowhead upon the line and the orientation of a 2-face is graphically given by a clockwise or a counterclockwise arrow within the area (Figure II.1.1). The numerical orientation for 1- and 2-faces is arbitrary, as long as consistency is maintained.

The orientation of a p-face is important in that it provides a means of determining the *incidence number* of the relationship between a p-face and a (p-1)-face. The incidence number n_{ij} between the *i*th p-face and the *j*th (p-1)-face may take the values of +1 if the two share a common orientation and the (p-1)-face is a

boundary element (defined below) of the p-face, or -1 if they have opposing orientations and the (p-1)-face is a boundary element of the p-face. If the (p-1)-face is not a boundary element of the p-face, then the incidence number between the two faces is zero. Formally, the incidence number is given by

$$n_{ij}(p-1)=\begin{cases}0\\+1.\\-1\end{cases} \qquad (II.1.2)$$

In terms of a graphical treatment of the problem, a 1-face is considered to have an incidence number of +1 with a 0-face if the arrowhead of the 1-face points towards the 0-face and –1 if it points away. The incidence number of a 2-face to a 1-face is dependent on whether the arrowheads of each face pass each other going in the same direction (+1) or in opposing directions (-1). In Figure II.1.1, the incidence of 2-face "B" with 1-faces "c", "d", and "e" would be -1 in all cases because the arrow of the 2-face moves in an opposing direction to those of each of the 1-faces. The 1-faces "a" and "b" are not incidental to "B", and the incidence numbers for those two 1-faces with "B" would be zero.

A *boundary* $\Delta C(p,i)$of the *i*th p-face is a (p-1)-chain that bounds the p-face, separating it from all other p-faces. For example, the three sides of "B" in Figure II.1.1 form the boundary of the area inside the triangle. Mathematically, it is possible to define the boundary using the incidence numbers of the p-face with all *j* defined (p-1)-faces, such that

$$\Delta C(p,i)=\sum_j n_{ij}(p-1)C(p-1,j), \qquad (II.1.3)$$

where C(p-1,j) is the *j*th (p-1)-face. The boundary of a p-chain Δc^p can also be described as a linear combination of the boundaries of the p-faces that are the elements of the p-chain

$$\Delta c^p = \sum_i u_i \Delta C(p,i). \qquad (II.1.4)$$

Therefore, the boundary of "B" in Figure II.1.1 is the 1-chain given by -c-d-e.

As a secondary function of the definition of boundary, the boundary of a boundary ($\Delta\Delta C(p,i)$) is zero. Consider again the case of "B" where the boundary of the area is the chain of three lines. The boundary of the boundary of the area would be the 0-chain defined by the points that define the end of the lines. Since each point (0-face) is involved in bounding two of the lines, the net result is that incidence number contributions for the 0-faces will always cancel each other, leaving $\Delta\Delta C(p,i)$ as zero.

The definition of boundary can be further explored by defining a *p-cycle*. A p-cycle is a p-chain with a zero boundary ($\Delta c^p = 0$). All such p-cycles form a subgroup Z^p of C^p. Therefore, the 1-chain -c-d-e is a 1-cycle because it is a boundary, and therefore the boundary of -c-d-e is zero, fulfilling the definition of a p-cycle.

A p-cycle is a *bounding p-cycle* if the p-chain c^p that defines the p-cycle is the boundary of a (p+1)-chain. The 1-chain -c-d-e is a boundary, and a p-cycle, so it is a bounding p-cycle. Bounding p-cycles form a subgroup B^p of all p-cycles Z^p, which, as mentioned previously, is a subgroup of all p-chains C^p.

With these defined terms, two p-chains can be described as *homologous* if their difference is a bounding p-cycle

$$c_1^p - c_2^p = c_b^p, \qquad (II.1.5)$$

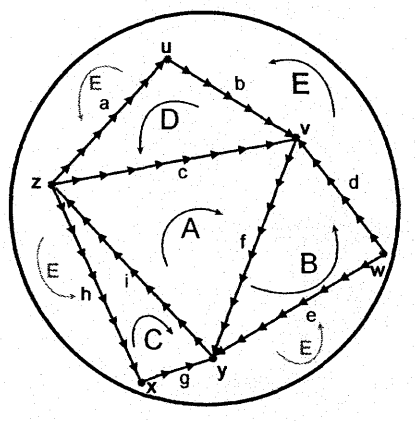
or in correct notation $c_1^p \sim c_2^p$. The 1-chain a+b-2c-d-e is homologous to the 1-chain a+b-c because the difference of the chains is -c-d-e, which is the bounding 1-cycle of 2-face "B".

The *p-dimensional integer homology group* H^p is the set of all homology equivalence classes $[c^p]$ such that $H^p = Z^p - B^p$.

Homology groups can be described using *Betti numbers* b_p which are the ranks of the homology groups H^p. In a three-dimensional space with a two-dimensional surface, there are three Betti numbers, b_0, b_1 and b_2, used to describe the homology groups of dimension 0, 1, and 2. To calculate Betti numbers requires information regarding the number of each type of p-face (α_p), as well as the rank r_p of the incidence matrix $\underline{I}_p$, which is the matrix of incidence numbers between all p-faces and all (p+1)-faces. The Betti numbers are

$$\begin{aligned} b_0 &= \alpha_0 - r_0 \\ b_p &= \alpha_p - r_p - r_{p-1} \end{aligned} \quad . \qquad (II.1.6)$$

Consider a surface in a three-dimensional space. The surface should be topologically equivalent to a sphere with no holes. To analyze the homology groups of the surface, the surface can be arbitrarily subdivided into 2-faces. The 2-chain of all the 2-faces equals the whole of the surface. Each 2-face must therefore be bounded by a chain of 1-faces, and each 1-face must be bounded by a chain of 0-faces. Figure II.1.2 is provided as an example of a topological sphere subdivided into ($\alpha_2 = 5$) 2-faces (A to E). The 2-faces are bounded by ($\alpha_1 = 9$) 1-faces (a to i), which are in turn bounded by ($\alpha_0 = 6$) 0-faces (u to z).



$\underline{I}_0$ =

	u	v	w	x	y	z
a	+1	0	0	0	0	-1
b	-1	+1	0	0	0	0
c	0	+1	0	0	0	-1
d	0	+1	-1	0	0	0
e	0	0	-1	0	+1	0
f	0	-1	0	0	+1	0
g	0	0	0	-1	+1	0
h	0	0	0	+1	0	-1
i	0	0	0	0	-1	+1

rank=5

$\underline{I}_1$ =

	a	b	c	d	e	f	g	h	i
A	0	0	+1	0	0	+1	0	0	+1
B	0	0	0	+1	-1	+1	0	0	0
C	0	0	0	0	0	0	-1	-1	-1
D	-1	-1	+1	0	0	0	0	0	0
E	+1	+1	0	-1	+1	0	-1	-1	0

rank=4

Figure II.1.2 A topological sphere subdivided into 2-, 1-, and 0-faces. The incidence matrices $\underline{I}_0$ and $\underline{I}_1$, as well as the rank of the incidence matrices are also shown.

In the Figure, the 0-faces are assumed to have an orientation of +1. For the 1- and 2-faces, the orientation is given graphically with arrowheads. The orientation for E is repeated in grey in several places for ease of determining incidence matrices. The incidence matrix $\underline{I}_0$ gives the incidence numbers between the 1-faces and the 0-faces, while $\underline{I}_1$.gives the incidence numbers between 2-faces and 1-faces.

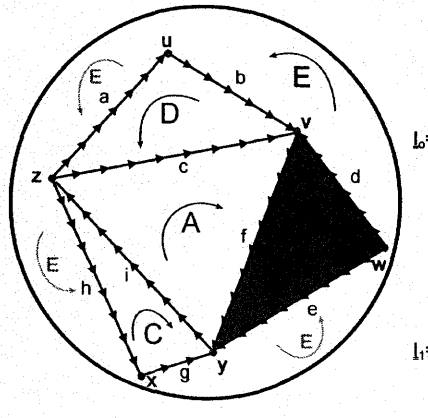
As is shown in the Figure, the rank r_0 of $\underline{I}_0$ is five, while the rank r_1 of $\underline{I}_1$ is four. Not shown, by default, is r_2, which would be the rank of $\underline{I}_2$. However, this incidence matrix has no meaning, as there are no 3-faces, so r_2 equals zero.

Utilizing Equation II.1.6, one can calculate the Betti numbers for the topological sphere in the Figure. The zeroth Betti number, b_0, is the number of 0-faces (six) minus r_0 (five), resulting in $b_0 = 1$. The first Betti number, b_1, is the number of 1-faces (nine) minus r_1 (four) minus r_0 (five), resulting in $b_1 = 0$. Finally, the second Betti number b_2 is the number of 2-faces (five) minus r_2 (zero) minus r_1 (four), resulting in $b_2 = 1$.

The Figure and calculated Betti numbers emphasize the point made by the Poincaré index theorem for a closed surface, that $b_p = b_{n-p}$ where n indicates the

dimension of the surface. In the specific example, since $b_2 = b_0$, we know the surface is closed (there are no holes).

In Figure II.1.3, the case of the same topological sphere is presented where 2-face "B" has been truncated, creating a hole in the sphere.



$\underline{I}_0$ =

	u	v	w	x	y	z
a	+1	0	0	0	0	-1
b	-1	+1	0	0	0	0
c	0	+1	0	0	0	-1
d	0	+1	-1	0	0	0
e	0	0	-1	0	+1	0
f	0	-1	0	0	+1	0
g	0	0	0	-1	+1	0
h	0	0	0	+1	0	-1
i	0	0	0	0	-1	+1

rank=5

$\underline{I}_1$ =

	a	b	c	d	e	f	g	h	i
A	0	0	+1	0	0	+1	0	0	+1
C	0	0	0	0	0	0	-1	-1	-1
D	-1	-1	+1	0	0	0	0	0	0
E	+1	+1	0	-1	+1	0	-1	-1	0

rank=4

Figure II.1.3 A topological sphere with one hole created by truncating the 2-face "B".

In Figure II.1.3 the incidence matrix $\underline{I}_0$ does not change from Figure II.1.2 as the truncation does not include the 1-faces that bounded "B". However, $\underline{I}_1$ changes, as the incidence numbers for "B" no longer have meaning. This truncation has not changed the rank of the matrix, though, as r_1 is still four. When the Betti numbers are calculated, it is found that $b_0 = 6 - 5 = 1$, $b_1 = 9 - 4 - 5 = 0$, and $b_2 = 4 - 0 - 4 = 0$. Since $b_2 \neq b_0$, the surface is not a closed surface according to the Poincaré index theorem. As the surface was created with a hole, this is known to be true. Effectively, a second Betti number b_2 of zero indicates a two-dimensional surface has holes.

Figure II.1.4 shows the sphere of Figure II.1.2 with 2-faces "B" and "D" truncated, forming two holes in the surface. There are two holes because the

common 0-face "v" has not been truncated and so the truncated regions are separately bound.

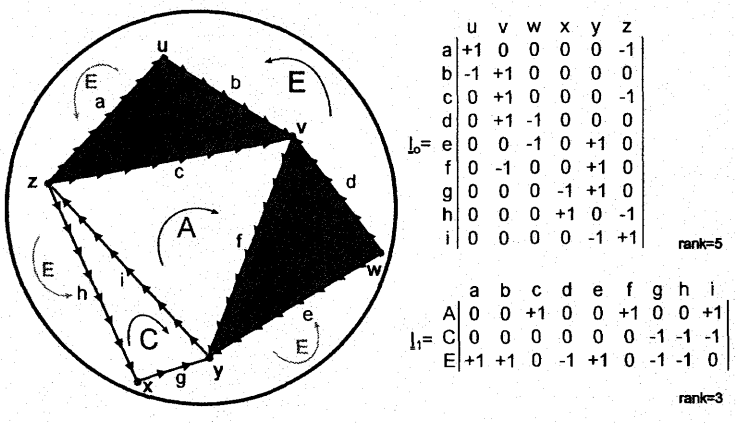


Figure II.1.4 A topological sphere with two holes created by truncating the 2-faces labelled "B" and "D".

In Figure II.1.4, the number of 2-faces decreases to three. Once again, the incidence matrix $\underline{I}_0$ changes, but this time, the rank does change so that $r_2 = 3$. In this case, the zeroth Betti number is still one, as the number of 0-faces and the rank of $\underline{I}_0$ have not changed. The first Betti number b_1 is $9 - 3 - 5 = 1$, and has changed from the previous Figure due to the change in r_1. Finally $b_2 = 3 - 0 - 3 = 0$. Again, b_2 being zero shows the surface is not closed, and has holes. However, it is b_1 that is of interest. In the case where a surface has holes, as indicated by the second Betti number, it has been mathematically shown that the first Betti number b_1 is the number of holes in the surface minus one. Additionally, the zeroth Betti number can be considered to be the number of distinct surface "pieces".

In terms of molecular shape analysis, the surface in the 3-dimensional space can be defined in many differing ways. The most commonly used is to define the surface as an electronic isodensity contour, at some given isodensity value, *a*. The value of the isodensity contour can take meaningful values between zero and

infinity. However, at large values of *a*, the isodensity contour surface G(a) is found to be tight around each of the nuclei defining the structure of the molecule, resulting in N distinct spheres, where N is the number of nuclei. Additionally, at very low values of *a*, the isodensity contour encloses all of the nuclei within the space, but does so in such a manner that the isodensity contour is effectively spherical in shape. Through previous studies, it has been found that most effective shape information of isodensity contours lies in between isodensity contour values of 10^{-3} and 10^{-1} electrons per bohr cubed ($e^-/bohr^3$). Here, one bohr (also symbolized as a_o) is approximately 52.9 picometres.

To analyze the shape of an isodensity contour in terms of homology groups also requires that the surface be truncated in some consistent manner, so the shape may be described in terms of the "number of holes." One possible method to accomplish this is to divide the surface into differing domains of points based on the curvature of the surface relative to some reference value, b^{-1}. This reference can be thought of as the reference sphere with a radius of b^{-1}. If the value of *b* is positive, the surface of the reference sphere would be found to curve inside the surface. If *b* is negative, the reference sphere curves outside the surface. If *b* takes on the value of zero, then the reference curvature is that of the plane tangent to the surface point of interest.

Mathematically, the curvature of the surface point of interest can be fully described (for a 2-dimensional surface) by defining two orthogonal vectors that form the basis of the plane tangent to the surface point. These two vectors, combined with the electron density gradient vector (which will be orthogonal to both tangent

plane vectors) define a local coordinate frame for the three dimensional space, with the origin of the local frame at the point of interest. For some specific choice of orthogonal tangent plane vectors (the eigenvectors), the 2x2 Hessian matrix description of the tangent plane will have eigenvalues h_1 and h_2 that describe the surface curvature relative to the tangent plane.

To describe the surface curvature relative to a reference curvature only requires a comparison of the eigenvalues of the Hessian matrix to the value of *b*. If $h_1 \leq h_2 < b$ the surface point is locally convex compared to the reference curvature, and can be said to belong to the domain $D_2(b)$. If the reference curvature value is less than one eigenvalue and greater than the other ($h_1 < b < h_2$) then the surface is locally saddle ($D_1(b)$) at the point of interest. Finally, if *b* is less than both eigenvalues, then the surface is locally concave, and belongs to the domain $D_0(b)$.

The curvature value *b* can take values from negative infinity to positive infinity. However, at large absolute magnitudes of *b*, the reference sphere is a very tight, almost point-like object, and so the local curvature domain will be the same for all points. Since little shape differentiation occurs between molecules in this case, the values for $|b|$ are commonly taken at zero (a tangent plane), as well as between one bohr and 10^{-5} bohr. At one bohr, the reference sphere has a "tight" curvature, while at 10^{-5} bohr the curvature is "gentle." In electron density representations of molecules, gentle curvatures are especially seen in the non-bonding regions around heavier atoms. Tight curvatures are seen in the almost cusp-like regions where the electron density contributions of two spatially separate groups merge in a manner similar to the sphere overlap in a van der Waals representation of the molecule.

Once an isodensity contour and a reference curvature are chosen, and all points of the surface are assigned into curvature domain groups, certain specific domain groups of the surface can be truncated to create holes in the surface. Since the isodensity contour is subsequently topologically equivalent to a sphere with holes, the first Betti number can be used to assign a shape code to the surface.

Figure II.1.5 gives an example of the process of assigning shape to an isodensity contour (a = 0.050 e^-/bohr^3) of R-alanine at a given reference curvature (b = –0.19 bohr).

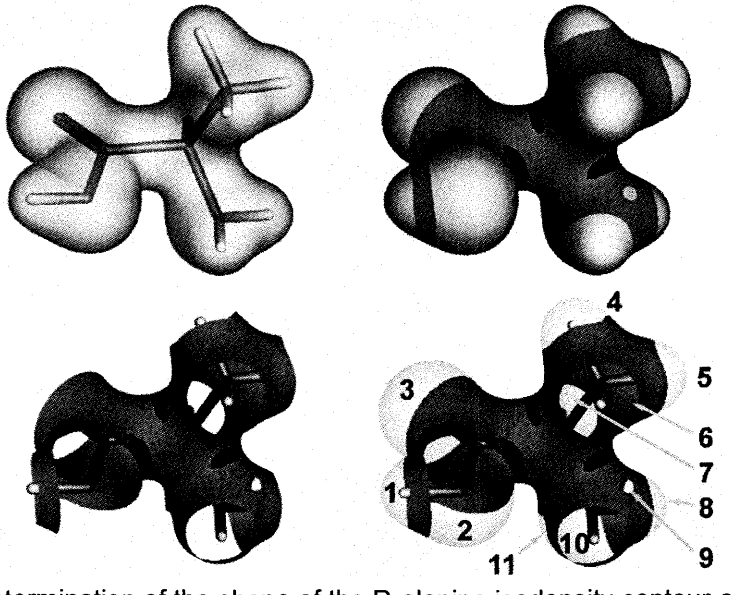


Figure II.1.5 Determination of the shape of the R-alanine isodensity contour a = 0.050 e^- /bohr^3 compared to a reference curvature of a sphere with radius 1/(-0.19 bohr).

The upper left corner of the Figure shows the isodensity contour of the molecule, while the structure is shown inside the isocontour. The upper right corner shows the division of the surface into curvature domains. The lightest grey areas are all locally convex (D_2) compared to the reference curvature, while dark grey areas are locally saddle (D_1) and black are the locally concave (D_0) regions of the surface. In the lower left corner, all of the locally convex regions have been

truncated. Through the created holes, the internal molecular structure can be seen. The lower right corner gives an alternate view of the truncated surface with the truncated regions slightly visible to show where holes can be created. Each of the eleven separate holes is distinctly identified and numbered.

The shape of the surface presented in Figure II.1.5 can be described by the first Betti number. Since this number is the number of holes minus one, the first Betti number for this surface would be ten. This information can be encoded to give the *shape code* of the surface.

However, the example in Figure II.1.5 is relatively simple, in that the truncated surface is a single piece, and therefore can be described by the first Betti number. In the cases where surfaces are, by truncation, subdivided into two or more individual pieces, a single Betti number cannot describe the surface. In these cases, each individual piece is topologically equivalent to a sphere with at least one hole in it, as this first hole is what separates the piece from the other pieces. Other holes may also exist in the piece. Therefore each separate piece can be assigned a first Betti number. The shape of the surface then is described by an ordered list of these first Betti numbers. This list is known as the shape ID vector.

An ordered list of Betti numbers can be unwieldy to compare, however, and so the list can be further modified by creating a *shape ID number* c'(*a*,*b*) based on a prime number encoding scheme of the ID vector. If the ordered Betti number list (the *shape ID vector*) is ordered from the largest magnitude Betti number to the smallest, then

$$c'(a,b) = 2^{B(1)+1} \times 3^{B(2)+1} \times \ldots \times P_{k+1}^{B(k)+1}. \qquad (II.1.7)$$

The shape ID number is determined by taking the second prime number (two) to the power of the largest Betti number (B(1)) of the ID vector plus one. This is multiplied by the third prime number (three) to the power of the second largest Betti number plus one. Multiplications continue until the *k* Betti numbers of the shape vector are used as powers of the series of prime numbers. The shape ID number is unique to a given shape ID vector and can be decoded into the vector by factorization and reordering by magnitude.

Such an encoding scheme can readily make comparable the shapes of two different electron density representations, *at the same values of a and b.* However, the choice of values of isodensity *a* and reference curvature *b* is arbitrary, and different combinations of these values will result in different shape codes for the same molecule. Total shape characterization of a molecule is, therefore, achieved by a collection of shape codes at various combinations of values. Forty-one isodensity contours G(*a*) are chosen throughout the range 10^{-3} e^-/bohr^3 $\leq a \leq$ 10^{-1} e^- /bohr^3. Each isodensity contour is analysed at one of ten reference curvatures in the range 10^{-5} bohr $\leq |b| \leq$ 1 bohr (b can be positive or negative) plus that of a plane (b = 0), to give 21 separate reference curvatures. The net result of all the shape analyses is an 861 member (41x21) matrix of shape codes called an (*a,b)*-map.

The shapes of molecules can then be compared with a single value called the *shape similarity index* *S*(X,Y) based upon the member comparisons *L*(*a,b*) of the (*a,b*)-maps. For the molecules X and Y, the *shape equivalence* (Δ) of the molecules at isodensity *a* and curvature *b* is given by

$$\begin{aligned} \Delta(X,Y,a,b) &= 1 \quad \textit{if} \quad L(X,a,b) = L(Y,a,b) \\ \Delta(X,Y,a,b) &= 0 \quad \textit{if} \quad L(X,a,b) \neq L(Y,a,b) \end{aligned} \qquad (\textit{II.1}.8)$$

and the shape similarity index is

$$S(X,Y) = \sum_{k=1}^{N_a} \sum_{l=1}^{N_b} \Delta(X,Y,a_k,b_l) \Big/ N_a N_b \,, \qquad (II.1.9)$$

where N_a and N_b are the number of differing values chosen for *a* and *b* (historically forty-one and twenty-one). Therefore, the shape similarity index is the sum of the number of common shape codes ($\Delta(X,Y,a,b) = 1$) between the two molecules, divided by 861, the number of maximum possible comparisons. This means the shape similarity index takes values in the interval between zero (no shape code similarity between the molecules, i.e. total dissimilarity) and one (all shape code comparisons are equal, i.e. total similarity of molecules).

II.2 Lattice Animals

One tool used in the description of the similarity of two objects within an n-dimensional space is the inscribed lattice animal or polycube. To clarify, a lattice animal is the specific case of a two-dimensional polycube, otherwise known as a polyominoe.[62] Lattice animals have been used in various applications noted elsewhere.[44,63-69]

Polycubes are constructs within the Cartesian space, $\mathbf{R}^n$, based on the face-wise connection of building units called n-cubes. The resultant polycubes are more easily directly compared to each other than are curves or surfaces within the n-dimensional space.

II.2.1 n-Cubes

Within an n-dimensional space, an n-cube is a specific construct based on a set of fixed properties. These properties can be described in terms of vertices, edges and faces. Once defined, the n-cube encloses a volume within the space.

The creation of an n-cube is based on a few simple properties:

(1) All edges are line segments of the same length that are bounded by a vertex on each end.

(2) Each vertex is a bound to n different edges. Therefore, there are 2^n vertices in an n-cube. All edges that share a common vertex as a bound must be mutually orthogonal to each other, and therefore the vertex group of edges will span the dimension of the space. The total number of edges will be the dimension times the number of vertices divided by two, or $n2^n/2 = n2^{n-1}$.

(3) There are 2n faces, one each for the negative and positive Cartesian directions in the space. Each face is an (n-1)-dimensional object.

In the general case, we have assigned names to n-cubes in 2-, 3-, and 4-dimensional spaces. These n-cubes are known as squares, cubes and hypercubes. If the length of an edge is taken to be some arbitrary unit vector denoted by u, then the volume (or area) of the space contained within the bounds defined by the vertices, edges and faces will be one unit to the *n*th power.

Lattice animals and polycubes are collections of face-connected n-cubes. As the term lattice animal suggests, a discussion of the concept of a lattice in terms of n-cubes is useful.

II.2.2 Lattices

A simple definition of a lattice is a regular, repeating arrangement of points within a space. Any space of dimension greater than two can host a lattice, but chemistry deals mostly with two-dimensional (chemistry on surfaces) and three-dimensional (e.g. crystal) lattices.

Within the space, the smallest number of points that can be repeated to create the entire lattice is often described as the unit cell. By translating the unit cell repeatedly in each of the Cartesian directions by distances equal to that of the unit cell edge, the entire lattice is created. Therefore, to create an entire lattice only requires the definition of the unit cell type and edge length.

Because an n-cube is very specifically defined, it can be used as a unit cell. In this case, the unit cell length is defined, and all angles between edges must be 90°. Therefore, once an n-cube is specified, its representation can be specified even more simply within the space.

For every n-cube there is a unique point that is an equal distance from all of the vertices. This point can be referred to as the cell centre. This point has an important property in that the shortest line segment between the cell centre and any point within a given face will be perpendicular to that face and will have a length of one-half the unit length. This means a translation of the cell centre by one unit length in any coordinate direction will create the next lattice cell. Therefore the position of each cell can be described completely by the coordinates of the cell centre. If the unit cell length is arbitrarily defined within the space as being one, and

the first cell is centred on the origin of the space, then any subsequent cell will have integer coordinates in each of the Cartesian directions.

II.2.3 Creating Lattice Animals and Polycubes

A polycube does not consist of the entire lattice within the space, but is some sub-lattice within the space. The creation of a polycube is accomplished based on three allowed types of interaction between cells. All contact between cells must be based on commonalities, in that each cell must share a vertex, edge, face, or any combination of these with its neighbour cells. However, further restrictions are applied to simplify the creation of polycubes.[45] Two n-cubes (C and C') within the polycube are connected if and only if:

(1) When C and C' have a common edge contact, they must have a common face contact, or there must be a cell C" that has a common face contact with both C and C'.

(2) When C and C' have a common vertex contact, they must also have a common edge, or there must be two face-connected cells C" and C"' that have C face-connected to C" and C' face connected to C"'.

(3) The polycube is topologically equivalent to the body it is representing.

Within a given dimension, and number of cells in the polycube, there will be a finite number of polycubes of differing shape. As the number of cells increases, the number of different polycubes increases. Figure II.2.1 shows the distinctly shaped 2-dimensional lattice animals consisting of two to five cells.

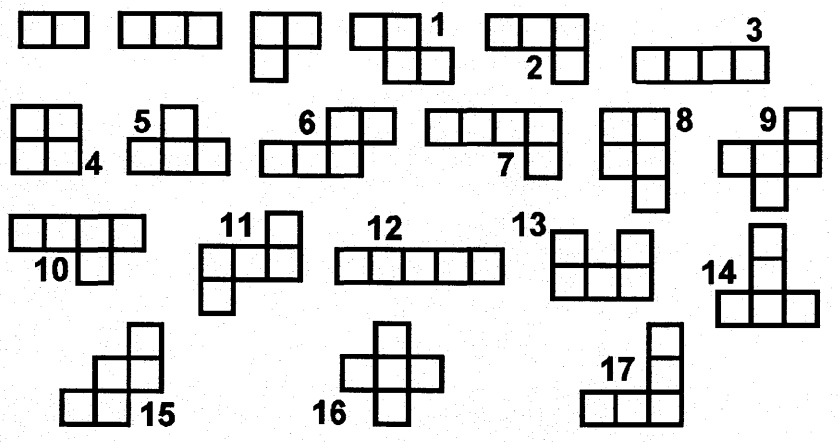


Figure II.2.1 Two-dimensional lattice animals consisting of two to five cells. The numbered animals are labelled for discussion in Table III.5.1.

II.2.4 Shape Similarity Measures Based on Lattice Animals and Polycubes

The similarities of Jordan surfaces in an n-dimensional space can be quantified by using inscribed polycubes.[45] A Jordan surface (or curve in 2-dimensional space) is a simple closed surface that divides a space into two parts. The first part is the bounded domain, which contains all points of the surface, and those "inside" the surface. All other points in the space are unbounded and are considered to be "outside" the surface. Most complicated surfaces can be described as a union of two or more Jordan surfaces.

Polycubes can be used to approximate the shape of a Jordan surface through the act of inscribing. If a Jordan surface is placed on an n-cube lattice, then those n-cubes with all points that fall completely inside the surface will define a polycube. This polycube, however, is dependent upon the orientation of the surface relative to the Cartesian directions of the lattice, where a different orientation of the surface will often inscribe two different polycubes, in terms of shape, and possibly, number of cells for a fixed sized lattice.

Additionally, if a smaller unit cell edge length is used to define the lattice, the "resolution" of the inscribed polycube is said to be "higher." In this case, then, the polycube is most likely going to consist of more cells than a polycube inscribed to the surface using a lattice of lower resolution. Thus, because the shape of a surface is invariant to scaling, resolution is better defined not by the size of the unit cell, but rather by the number of cells in the inscribed polycube. Therefore, for certain finite changes in the unit cell size, there should be invariance to the maximum number of cells that can be inscribed within the surface.

For a given surface J and unit cell size s, there exists a family of inscribed polycubes F(J,s) for all possible orientations of the surface relative to the lattice. A subset A_i(J,s,m) of this family of polycubes will be those polycubes that have the maximal number of cells, m, for the given surface, regardless of orientation, for the unit cell size. These polycubes are the m-cell interior filling animals of the surface J. Therefore, a polycube is an interior filling animal if and only if there are no polycubes with a greater number of cells that can be inscribed into the curve J at the same unit cell size.

Figure II.2.2 shows a Jordan curve with interior filling animals of one to four cells. First note that the cell edge directions do not necessarily correspond to the same lattice directions in each of the four cases, denoting the orientation of the curve may be changed relative to the lattice to maximize the area occupied by the inscribed lattice animal. Also of note should be the size of the unit cells. There is no fixed increment of size change required to get to the interior filling animal that contains one cell more than the previous animal.

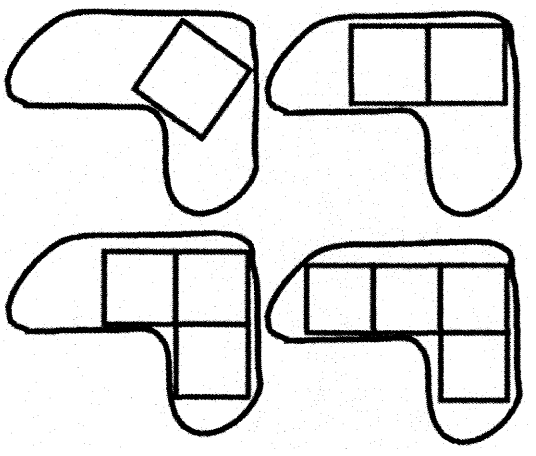


Figure II.2.2 A Jordan curve and the first four interior filling animals.

The Figure shows the generalized property that as the number of cells of the interior filling polycube is increased, a better approximation of the Jordan surface is given by the perimeter of the inscribed polycube. Therefore, based on the resolution at which two surfaces have differing interior filling polycubes, a measure of the similarity of two surfaces can be made.

The similarity index of two surfaces J_1 and J_2 can be defined as the smallest number of cells m_c for which the interior filling polycubes of the surfaces are different for that number of cells and all numbers of cells greater than m_c,

$$i_o(J_1,J_2)=\begin{cases}\min\{m_c : F(J_1,J_2,m) \text{ is empty if } m \geq m_c\} \\ \infty \text{ otherwise}\end{cases}. \qquad (II.2.1)$$

If the Jordan surfaces are the same, then no interior filling polycube exists that is different between the two surfaces and the similarity index is infinite. If the surfaces are different, there must exist some number of cells for which that difference becomes apparent in the inscribed polycubes.

The degree of dissimilarity between two surfaces can then be defined as

$$d(J_1,J_2)=\frac{1}{i_o(J_1,J_2)-(q-1)}. \qquad (II.2.2)$$

The degree of dissimilarity is dependent on the similarity index, as well as a dimension dependent factor q, which is the minimum number of cells a polycube must have in an n-dimensional space before there are two or more distinct polycubes with that number of cells. As can be seen in Figure II.2.1, q is three for two-dimensional space. Additionally, q has a value of three in three-dimensional space.

The degree of dissimilarity can take values between zero and one. In the case of two identical surfaces, the similarity index is infinite, and so the degree of dissimilarity is zero. Two completely different surfaces would have a similarity index equal to q, and thus the degree of dissimilarity would be one.

Finally, the degree of similarity can be defined as

$$s(J_1, J_2) = 1 - d(J_1, J_2). \qquad (II.2.3)$$

The degree of similarity, s (not to be confused with unit cell size), reverses the order of the degree of dissimilarity, such that two identical surfaces have a degree of similarity of one, while two absolutely different surfaces would have a value of zero.

II.2.5 Chirality Measures Based on Inscribed Polycubes

In a process very much similar to that of determining the similarity of Jordan surfaces with inscribed polycubes, one can quantify the chirality of a given surface.[70]

The general conjecture is that most surfaces in a space are chiral, in that the mirror image of a surface is not exactly superimposible point by point upon the original surface by translation and rotation. Additionally, it is assumed that the ratio of achiral polycubes of a given number of cells to the family of all polycubes of that given number of cells approaches zero as the number of cells approaches infinity.

Much like the similarity index, the chirality index of a Jordan surface can be defined as the number of cells at which all interior filling animals of that number of cells and greater are chiral,

$$c_o(J)=\begin{cases}\min\{m_c : A_i(J,m) \text{ is chiral if } m \geq m_c\} \\ \infty \text{ otherwise}\end{cases}. \qquad (II.2.4)$$

As the number of cells increases, the probability of an animal with that number of cells also being chiral increases. Therefore, there is an inscribed animal with some finite number of cells for which the chirality index can be defined for a chiral surface. An achiral surface will inscribe only achiral polycubes at all values of m, and the chirality index of these surfaces will be infinite.

Additionally, the degree of chirality can be defined much like the degree of dissimilarity,

$$\mathrm{X}(J)=\{c_o(J)-(\chi-1)\}^{-1}. \qquad (II.2.5)$$

In this case, the dimensionally dependent factor χ performs the same role as q in the degree of dissimilarity. This factor is the number of cells a polycube must have in an n-dimensional space before it is possible to have a chiral animal. In two-dimensional space, this number is four. This can be verified in Figure II.2.1 as all two- and three-celled lattice animals are achiral. There, however, are two differently shaped lattice animals of four cells that cannot be superimposed on their mirror images. These two animals are commonly referred to as 'Elly' (Figure II.2.1, label 2) and 'Tippy' (Figure II.2.1 label 1).[79]

In three-dimensional space, this dimensionally dependent factor value is also four. For higher dimensions, the value is the dimension of the space plus one.

II.3 Molecular Similarity and Chirality Measures

The concept of molecular chirality measures is deeply intertwined with the concept of molecular similarity measures.[49-60] Generally, molecular chirality measures are based upon mirror image compounds being treated as two separate molecules, and therefore the treatment of measuring the chirality will be analogous to that of determining the similarity of two different molecules.

Carbó *et al.* were the originators of the concept of a quantitative index of similarity between the electronic charge distributions of molecules at a point **r** in space, R_{AB},[49-51]

$$R_{AB} = \frac{\int_{\Omega} \rho_A(r)\rho_B(r)dr}{\left(\int_{\Omega} \rho^2{}_A(r)dr\right)^{1/2}\left(\int_{\Omega} \rho^2{}_B(r)dr\right)^{1/2}}. \qquad (II.3.1)$$

The measure is applied to the integration of the electron density representations ρ of the molecules A and B over the entire three-dimensional space Ω. The measure is also normalized relative to the individual representations of the molecules, resulting in a final index value in the range between zero (total dissimilarity) and one (total similarity).

Hodgkin and Richards later developed a measure[52,53] H_{AB} related to the Carbó index. Their measure

$$H_{AB} = \frac{2\int_{\Omega} \rho_A(r)\rho_B(r)dr}{\left(\int_{\Omega} \rho^2{}_A(r)dr\right)+\left(\int_{\Omega} \rho^2{}_B(r)dr\right)} \qquad (II.3.2)$$

was believed to better account for differences in molecular size that possibly could skew similarity measure results. However, their chirality measure has not been

favoured as much as the Carbó index, to the point that they have done continuing work with the Carbó index.[80]

In either case, the measure is a "quantum" one, in that it measures similarity directly based upon the electronic distribution, and not the nuclear configuration. However, measures based upon nuclear configuration, or more correctly, upon the configuration of points in space have also been developed. Such "structural" measures have been largely based upon the similarity work of Avnir.[81-85]

II.3.1 Quantum Chirality Measures

Quantum chirality measures are generally made by comparing the molecular electron density or electrostatic potentials of two mirror image compounds relative to each other by utilizing the Carbó, or a similar, index. With such indices, it has been possible to perform analysis of molecules as it relates to drug design,[86,87] solvation,[88] optical rotation,[89] and reaction pathways.[90]

Over time, the Carbó index has been refined, especially for chirality measures, by the definition of what are termed the "quantum molecular similarity measures" (QMSM).[88-90] These measures take the general form

$$Z_{A,B}(\Theta) = \iint \rho_A(r_1)\Theta(r_1,r_2)\rho_B(r_2)dr_1dr_2 \qquad (II.3.3)$$

where $\Theta(r_1,r_2)$ is a positive definite operator, defined to take advantage of a specific feature of the quantum chemical calculations. For instance, if the operator is an inverse distance operator ($1/r_{12}$) the measure evaluates the "Coulomb-like" QMSM. By using the Dirac delta function [$\delta(r_1-r_2)$], the similarity of the overlap of the molecules can be studied. While not stated explicitly in Equation, generally these measures are made in three-dimensional space.

This "overlap-like" QMSM is the most useful for chirality measures. If the overlap of a molecule with itself (the "self-similarity" measure) is defined, then chirality can be measured by the Euclidean distance between this self-similarity measure of the molecule (as the maximum overlap value) as compared to the maximal overlap that can be found between two mirror image electronic distributions. For example, if Z_{RR} defines the overlap-like similarity of the R form of a molecule, the overlap-like similarity Z_{SS} of the S form will take on the same value. However, Z_{RS}, the maximal overlap-like symmetry of the R and S molecular forms relative to each other will take on some value less than Z_{RR} as the structure of the chiral molecule will result in regions of the molecules that do not overlap with each other. This is obvious from the concept of chirality: the inability of superimposition of a molecule upon its mirror image.

With both the self-similarity values and the mirror image overlap value, the Euclidean distance

$$D_{RS} = \sqrt{Z_{RR} + Z_{SS} - 2Z_{RS}} \qquad (II.3.4)$$

gives a measure of the chirality of the molecule. The larger the Euclidean distance is, the "more" chiral the molecule can be considered and vice versa. As can be seen in the equation, in the case of an achiral molecule where $Z_{RR}=Z_{SS}=Z_{RS}$, the Euclidean distance will be zero.

The advantage of quantum chirality measures is that they actually represent a measure of chirality of physical representations of molecules that are more relevant to specific properties of interest. The idea of chirality measures that differ based on the property of interest was first stated by Gilat.[91] For instance, an achiral tetrahedron of points becomes chiral if each of the points is assigned a different

mass, leading to mass-based chirality. One can also consider types of chirality, that have differing values for the same molecules, based on nuclear potentials, electrostatic potentials, polarizabilities, or any other molecular physical property. Therefore, quantum chirality measures can be seen as being more relevant to specific studies because they are designed to measure chirality of the specific property that would correlate to experimental results of interest. An example is optical rotation, which would be better represented by an electron density chirality measure, rather than a nuclear mass chirality measure.

The disadvantages of quantum chirality measures are mostly due to the need to find the maximal overlap of two molecules. While efficient algorithms for this step are available, the growing pains associated with their development made the study of the chirality measures without access to the associated software difficult at first. This is not of great concern anymore.

Another disadvantage of the quantum chirality measures is that they are based upon the compared nature of the molecule and its mirror image, leading to an always positive Euclidean difference. There is no specific inherent ability to differentiate between mirror image pairs, but rather, each pair leads to a statement that the members of the pair "have this amount of chirality." Arbitrary rules to assign a differentiation of mirror image pairs, such as positive or negative signs, would not be consistent over all molecules, much like there is no full correlation between molecular L or D designations (which are based on glyceraldehyde as a reference) as opposed to the Cahn-Ingold-Prelog group-priority-based R or S designations.[92]

II.3.2 Structural Chirality Measures

Structural chirality measures[82,83,93-99] are based upon the concept of taking a structural collection of points, and assigning chirality based upon the "distance" that the arrangement of points has from an equivalent achiral arrangement of points. Early attempts at such measures[93,94] usually placed the molecular "centre" at the origin of the space, and then measured the distance (and possibly relative angle) change to the molecular centre after the molecule has been made achiral either by moving atoms, or replacing atoms to create the "nearest achiral" comparison molecule. How the molecular centre was defined was arbitrary, however, and generally was chosen based on the physical property being studied. For instance, the centre could be the centre of mass, nuclear charge, partial charge, polarizability, or other property. Generally, the coordinates of the centre q_c are given by

$$q_c = \sum_i (W_i q_i) \Big/ \sum_i W_i \qquad (II.3.5)$$

where the coordinates of each atom are weighted by the atomic parameter W, which represents the molecular property of interest.

A specific chirality measure developed by Avnir[93,94] utilized the changes in the distance l_d and angle θ of the molecular centre (as defined by atomic volumes) from the stereocentre of the molecule of interest as compared to that for the equivalent centre in the "nearest" achiral molecule. The resultant measure χ was a scaled variant of the changes, such that

$$\chi = 100\ l_d\theta\,. \qquad (II.3.6)$$

With this chirality measure, some success was noted in correlating the chirality of halogenated alkanes to the molecular rotations of plane polarized light.[94]

Further work lead to the concept of "continuous" symmetry measures (CSM) that could be used to measure chirality.[81-84,99] In such measures, the distance of a collection of points to the "nearest" collection of points with a given symmetry group G is given by

$$S'(G) = \frac{1}{n}\sum_{i=1}^{n}\left\|P_i - \hat{P}_i\right\|^2, \qquad (II.3.7)$$

where the sum of the distances of the n points P_i from their corresponding points $\hat{P}_i$ in the collection of points with symmetry group G define the measure. As before in Equation II.3.6, S'(G) is often multiplied by 100 for ease of use.

The "nearest" collection of points of a given symmetry group is found utilizing the *folding/unfolding* algorithm, which can be seen in depth in the original references.[81,82,99] The algorithm is based upon three principles that separate CSMs from earlier attempts at structural chirality measures, namely:

(1) Asymmetric shapes should not be treated as a perturbation of an ideal symmetric shape. All shapes can be described by their position on some single continuous scale.

(2) Assessing symmetry should not be reliant on reference objects.

(3) Evaluation of the symmetry of an object, relative to any symmetry group should be possible.

Once the individual symmetry elements of the symmetry group of interest are determined, the algorithm can be used to find the "nearest" collection of points of a given symmetry. The first step is to place the centroid of the collection of points at the origin of the space, and scale the object so the maximal distance for a point P_i of

the object from the origin is one. Once this is done, the symmetry group elements should be translated so they apply to the origin of this space.

The symmetry elements of the symmetry group need to be applied in some specific order, based upon the connectivity of the points in the object. For example, in 2-dimensional space, there are two possible orders of points for the six vertices of a simply-connected hexagon; the ordering that goes clockwise through the points, or counterclockwise.

The vertices of the object are then *folded* by applying the appropriate symmetry operation upon the appropriate vertex, based upon the previously established ordering. This results in a collection of points that may, or more likely, may not be the same as the original object. The *average point* is the centroid of this new collection of points.

A new collection of points with the desired symmetry is then found by applying the symmetry operations in reverse order upon the average point. This *unfolding* operation defines the coordinates of the new points $\hat{P}_i$. With the defined coordinates of each P_i and $\hat{P}_i$, Equation II.3.7 can be used to calculate the continuous symmetry measure S'(G). However, this measure must be minimized by repeating the algorithm for all possible orientations and orderings of the group symmetry elements.

Structural chirality measures, like the quantum measures, also suffer from the disadvantage of not allowing for the differentiation of mirror images in terms of the chirality measure, but rather, state how "chiral" the molecule is relative to the symmetry element being considered. This is noted by Grimme,[78] who, in the

development of what can be best described as a "quantum-hybrid" CSM notes that, much like the Gilat statement, CSM values are dependent on the symmetry element chosen to represent the "nearest" collection of points. This element could be a mirror plane, point of inversion, or improper axis of rotation, any of which would result in an achiral object.

Other structural chirality measures, such as Hausdorff distance based measures,[82,95] molecular topological based measures,[96] and binary molecular perimeter encoding measures,[97] have also been proposed.

II.3.3 Hybrid Chirality Measures

Recently, Avnir's Continuous Symmetry Measures have been used in the creation of what can be termed "hybrid" chirality measures.[78,99,100] Such hybrid measures apply the CSM method to quantum representations of molecules. The simplest such measure is a direct application of the folding/unfolding algorithm to a collection of connected points that make up a wavefunction isocontour representation of a molecule.[99] Lipkowitz *et al.* were able to take a similar approach to analyze the chirality of several organic molecules.[100]

Grimme[78] proposed a measure based upon the quantum mechanical expectation value of a geometric transformation operator acting upon a wavefunction

$$\langle S \rangle = \langle \Psi | \hat{S} | \Psi \rangle . \qquad (II.3.8)$$

If the wavefunction, as is often the case in most quantum chemistry software packages, is expressed as a function of one electron functions (orbitals), ψ, then the expectation value <S> is the sum of the expectation values of the individual orbitals with occupation numbers η_i. This is important, because these expectation values of

the individual orbitals (χ_i) have an absolute value of one only if the symmetry operation perfectly overlaps the transformed orbital with the original orbital. If the orbital does not perfectly overlap (the symmetry element is not present), then the expectation value takes an absolute value less than one. Therefore, a molecular CSM for a particular symmetry operation can be defined as

$$\mathrm{CSM}\left(\hat{S}\right) = 100\frac{\sum_i \eta_i \left(1 - \left|\chi_i\left(\hat{S}\right)\right|\right)}{\sum_i \eta_i}, \qquad (II.3.9)$$

where the CSM takes a value between zero (the molecule possesses the symmetry element) to a maximal deviation value of one hundred.

However, it was this study, as noted previously, in which the concern was raised about differing chirality measures based upon differing symmetry elements having different measured values for chirality. For instance, the chirality measure of a molecule based on a mirror plane σ might be different from that based on an improper rotation S_n, even though either operation "destroys" chirality. To combat this, Grimme chooses the measure that gives the minimal value.

Because of the Continuous Symmetry Measure basis of the hybrid measures, they have many, if not all, of the same advantages and disadvantages of the CSMs.

II.4 Fuzzy Fragmentation Techniques

Fuzzy fragmentation techniques for creating electron density representations of small parts of molecules have generally been created for one of two purposes. The first purpose is the creation of *ab initio* quality electron density representations of molecules that would be too large to calculate using conventional *ab initio*

techniques. This Molecular Electron Density Lego Assembler (MEDLA) technique[24,27,101-105] has proven quite successful at creating electron densities for large molecules that are "built" from smaller electron density fragments. Such molecular electron densities have been shown to be "very similar" (of the same quality) to electron densities calculated directly at the same level of theory for the whole molecule using *ab initio* techniques.[24,27,101-105]

The second, and more relevant in this instance, purpose of such fuzzy fragmentation is the Shape Group method analysis of "local" shape features of molecules. These features often are relevant in Quantitative Shape-Activity Relationship studies, along with the whole molecule shape features.[27]

Initial work in fragmentation of molecules focused on "crisp" fragmentation schemes such as the "atoms-in-molecules" technique of Bader *et al.*[106-110] In this technique, electron density representations are "carved" out of larger molecular electron density representations along the gradients of the electron density isocontours. Such fragments have been shown to create larger molecules that have properties favourable in comparison to *ab initio* whole molecule properties.[108-110] However, the main drawback of such crisply bounded fragments is that the low isodensity regions of created larger molecules tend to be disjointed, as the crisp boundaries can lead to volumes of "density doubling" where two fragments contribute electron density to the same region of space, or "density bubbling", where a region of space has effectively no electron density because no contributing fragment applies to the specified region.

Another crisp fragmentation technique, where fragmentation is achieved through the use of specified distances from the plane of electron density cross sections can also be used in modelling the electron densities of regions of larger molecules.[111,112] However, in addition to the drawback of crisp fragments (doubling and bubbling of electron density), this fragmentation technique is only useful for creating the electron density cross section of interest. If other analyses need to be performed, the calculations and fragmentation must be repeated.

Fuzzy fragmentation does not suffer the drawback of crisp fragmentation schemes.[101] This is due to the fact that the fragments are created in a similar manner to that of how whole molecular electron density representations are created.

Standard *ab initio* techniques create electron density representations through the Linear Combination of Atomic Orbitals (LCAO) method. In the method, atomic orbitals (φ), in which the electrons of the atom are found, are used in linear combinations to approximate the orbitals of a molecule created from the individual atoms. The electrons originally associated with the atoms (in the atomic orbitals) "populate" these molecular orbitals. How this population is distributed amongst the orbitals can be calculated, meaning that the relative "amount of electrons" that can be associated with a specific molecular orbital (and therefore, atomic orbital) can be expressed in a *density matrix*.

In this method, the electron density ρ of the molecule at a given vector position **r** is

$$\rho(r) = \sum_{i=1}^{N} \sum_{j=1}^{N} P_{ij} \varphi_i(r) \varphi_j(r), \qquad (II.4.1)$$

where P_{ij} is the *ij*-component of the NxN *density matrix* $\underline{\underline{P}}$, which applies to the overlap of the *i*th and *j*th atomic orbitals $\varphi(\mathbf{r})$.

Because the "whole" molecular electron density is a fuzzy object, any fragmentation that mimics the creation of the whole electron density will also result in a fuzzy object. Therefore, a fragmentation scheme based upon Equation II.4.1 results in fuzzy fragments.

If a proposed fragment k of a molecule is defined by the atomic orbitals centred upon certain nuclei of the whole molecule, then the fragment can be created in exactly the same manner as in Equation II.4.1, with the exception that a *fragment density matrix* $\underline{\underline{P}}^k$ is used. The fragment density matrix is based upon the original density matrix $\underline{\underline{P}}$, but changes in the *ij*-components reflect the membership of the atomic orbitals to the sub-group of nuclei used to define the fragment. If both atomic orbitals represented by the *ij*-component of $\underline{\underline{P}}$ are centred on atoms in the fragment, the *ij*-component of $\underline{\underline{P}}^k$ will equal the same component from the whole molecule density matrix. If neither atomic orbital is centred on an atom in the fragment, the *ij*-component of the fragment density matrix is zero. Finally, if one of the two atomic orbitals is centred on an atom in the fragment, the fragment density matrix *ij*-component will be $w(k,i,j)P_{ij}$, where w(k,i,j) is a *weighting factor* with a value between zero and one.

The weighting factor must be a sign-preserving scalar property that is assignable to a specific atomic orbital and scaleable relative to a different atomic orbital. For instance, with a relative non-zero scalar A(i) between the *i*th and *j*th atomic orbitals, the weighting factor could take the form

$$w(k,i,j) = \frac{A(i)}{[A(i) + A(j)]}. \qquad (II.4.2)$$

Based on this definition, if the *co-fragment* (k') of fragment k is defined as the fragment consisting of all nuclei of the molecule not present in fragment k, the weighting scheme results in

$$w(k,i,j) + w(k',i,j) = \frac{A(i)}{[A(i) + A(j)]} + \frac{A(j)}{[A(j) + A(i)]} = 1. \qquad (II.4.3)$$

Scalar properties that can be used as weighting factors include formal charges, or scaled electronegativity values.[41] However, the most commonly used weighting factor is one that arbitrarily takes half of the *ij*-component of the molecular density matrix to become the *ij*-component of the fragment density matrix. Essentially, the electron density arising from the specific atomic orbitals interaction is divided equally between the orbitals.

Formally, then, the electron density of a fragment ρ^k is

$$\rho^k(r) = \sum_{i=1}^{N}\sum_{j=1}^{N} P^k{}_{ij}\varphi_i(r)\varphi_j(r). \qquad (II.4.4)$$

Of special note are the properties of fuzzy fragments. If m fragments are created such that each nucleus is a member of one and only one fragment, then the sum of all the fragment density matrices will return the density matrix of the original molecule, and the sum of the electron densities of all the fragments will equal the electron density of the original molecule.

II.5 Optical Rotation

The theory of optical rotation (the rotation of the plane of plane-polarized light as it propagates through a medium) has been well studied and is quite detailed.

While many texts and books deal with the subject, this discussion is based on that of Mason[48] and is general in nature.

II.5.1 Classical Optical Rotation

While many scientists had previously worked on the phenomenon of optical rotation, it is Biot who is credited with first establishing the approximate law between the angle of optical rotation and the wavelength of light. In this relationship the optical rotation α is inversely proportional to the square of the wavelength λ, and directly proportional to some specific property k of the material the light passes through. Formally, the relationship is

$$\alpha = k/\lambda^2 . \qquad (II.5.1)$$

To formally measure the optical activity at a given wavelength and temperature T, Biot proposed the definition of *specific rotation* based upon the observed rotation, the concentration c of the active substance in grams per millilitre and the path length l the light travels through the substance or solution in decimetres,

$$[\alpha]_\lambda^T = \alpha/lc . \qquad (II.5.2)$$

Additionally, the molar rotation reflects the specific rotation, as it occurs relative to the molar mass M of the substance,

$$[M]_\lambda^T = \left(M/100\right)\ [\alpha]_\lambda^T . \qquad (II.5.3)$$

Fresnel, applying the transverse wave theory of light to the phenomenon of optical activity, decided that the activity was due to *circularly polarized* light. Plane polarized light can be described as the resultant wave of the addition of *left-* and *right-circularly polarized* components of light. In each of these components, the

electric (or magnetic) field of the photon is "screw-like," and so the electric field vector "turns" as the light propagates through the medium. The difference between the two components is the direction of the "turning."

The amplitudes at a given time t of each of these components of the incident light propagated in the z-axis direction through the substance of interest depend on the angular frequency of light ω, the refractive index of the medium n, and the speed of light in vacuum c, and are described as

$$a = a_0\left\{i\cos\left[\omega\left(t - nz/c\right)\right] \mp j\sin\left[\omega\left(t - nz/c\right)\right]\right\}. \qquad (II.5.4)$$

Consider a unit circle in a plane. At time zero, the left and right circularly polarized component vectors start at the origin and end at some point of the circle, and their resultant vector is the plane polarized light vector. As time progresses, as the light passes through a non-optically active medium (the plane is translated in the propagation direction), one component vector rotates clockwise around the origin of the plane, and the other rotates counterclockwise. The resultant vector does not change direction, but the magnitude of the vector oscillates.

Optical activity, then, arises through the difference of refractive index (n_R and n_L) for the right (R) and left (L) circularly polarized components of the plane polarized light wave. This difference comes as a result of the "screw" defined by the electric and magnetic dipoles of the molecule the light is interacting with. If the "turn" of the light is complementary to the "screw" of the molecule, the interaction does not impede the propagation of the light as much as would occur in the case of the opposite circularly polarized component. This impediment means one component of

the light will be slowed as it passes through the medium. This velocity change is observed as a difference in refractive index.

Since the velocity of one circularly polarized component through the medium is greater than the other component, the slower component, at time t, will "lag behind" the other component in traversing the unit circle. Therefore, the resultant vector (the plane of polarization) will rotate along the unit circle.

If $n_L > n_R$, x-direction plane polarized light will rotate towards the y-axis. This is reflected in the amplitudes of the circularly polarized components a_x and a_y. At time zero, the amplitude of the y-axis component is zero. As the light propagates through a path length of the medium, the ratio of the amplitudes will follow the relation

$$a_y/a_x = \tan\phi \quad where \quad \phi = \omega l(n_L - n_R)/2c. \qquad (II.5.5)$$

The wavelength can be expressed as the ratio of the speed of light and the frequency ($\lambda=2\pi c/\omega$), the angle of rotation can be expressed with Fresnel's equation,

$$\phi = (n_L - n_R)l\pi/\lambda, \qquad (II.5.6)$$

where the path length l is measured in the same units as the wavelength.

As time progressed, Boltzmann refined Biot's description of the wavelength dependence of optical activity utilizing a virial equation of the form

$$\left[\left(B/\lambda^2\right)+\left(C/\lambda^4\right)+\left(D/\lambda^6\right)+\ldots\right]. \qquad (II.5.7)$$

Drude subsequently refined this expression to give a general optical rotary dispersion equation,

$$[M]_\lambda = \sum_m K_m\left[\lambda^2 - \lambda_m^2\right]^{-1}, \qquad (II.5.8)$$

where λ_m is the resonant wavelength of light interacting with a group of charged particles. Therefore the molar rotation is the sum of all such interactions, each oscillating at a characteristic frequency, and each multiplied by K_m, the constant that applies to the specific vibration.

II.5.2 Quantum Optical Rotation

The quantum mechanical form of the Drude Equation (II.5.8), formulated by Rosenfeld, gives the angle of rotation ϕ in radians per centimetre of pathlength, where

$$\phi = \left[16\pi^2 N/3hc\right] \sum_m \nu^2 R_{0m} \left[\nu_{0m}^2 - \nu^2\right]^{-1}. \qquad (II.5.9)$$

Here, the number of molecules per cubic centimetre N reflects the concentration of the active portion of the medium, while R_{0m} is the rotational strength (analogue of K_m) of the one-electron transition from the ground state ψ_0 to the excited state ψ_m, with a transition energy $h\nu_{om}$.

Furthermore, the rotational strength is the imaginary part of the scalar product of the electric dipole transition moment μ_{0m} and the magnetic dipole transition moment $\mathbf{m}_{mo}$,

$$R_{0m} = \mathrm{Im}\left\{\left\langle \psi_0 \middle| \hat{\mu} \middle| \psi_m \right\rangle \bullet \left\langle \psi_m \middle| \hat{m} \middle| \psi_0 \right\rangle\right\}. \qquad (II.5.10)$$

II.5.3 Attempts at Theoretical Calculation of Optical Rotations

Many attempts at defining and utilizing theoretical models of optical rotation have been made since the study of the phenomenon began.[113-141]

Mason, in his text,[48] describes four of the early fundamental models. In the independent-systems model, on which Boltzmann developed Equation II.5.7, optical

activity arises from light interacting simultaneously with different, independent parts of a chiral molecule. The first of these parts is the chromophore, which absorbs the light, while subsequent parts form a dissymmetric environment for the chromophore.

In Kirkwood's two-group electric-dipole mechanism,[142] the magnetic dipole moment transitions, as mentioned in the previous section, are considered to be relatively minor contributors to optical activity. In the model, the magnetic moment is actually a result of two separate linear electric dipole moments, whose arrangement relative to each other leads to a helical charge displacement.

The one-electron, or static field model,[143] emphasizes the allowed magnetic dipole transitions compared to the two-group model. In the model, the chromophore is distinct from the other parts (substituents or ligands), and the chromophore alone is involved in the light-absorption process. Since the ligand charge distribution does not change in the light-absorbing process, the field around the chromophore stays unchanged, but is used in the mixing of the magnetic and electric dipole transitions of the chromophore to give non-zero rotational strengths.

In the dynamic polarization method, each allowed magnetic dipole transition of the chromophore, has an accompanying even electric-multipole transition that induces an electric dipole moment in each of the substituents. The magnitude of this induced dipole depends on the substituents' polarizabilities at the incident radiation frequency. Such induced moments enhance the absorption of light, and in chiral molecules, a component of the induced dipole moment works with the magnetic dipole of the chromophore to create non-zero rotational strengths.

Much of the modern theoretical approach to calculation of optical activity properties[119-141] has focused upon the static field[119-132] or dynamic polarization models[130-138,140] of optical activity. In either case, the calculated quantity for the purpose is related to the electric dipole – magnetic dipole polarizability tensor $G'_{\alpha\beta}$,

$$G'_{\alpha\beta} = \frac{-4\pi}{h}\sum_{n\neq s}\frac{\omega}{\omega_{ns}^2-\omega^2}\times \mathrm{Im}\left\{\left\langle \psi_s^0\left|\hat{\mu}_\alpha\right|\psi_n^0\right\rangle \bullet \left\langle \psi_n^0\left|\hat{m}_\beta\right|\psi_n^0\right\rangle\right\} \qquad (II.5.11)$$

The Equation is related to Equation II.5.9 in the dependence of the tensor on the rotational strength and rotational frequency of light ω_{ns} for the transition from the ground ψ_s to the excited state ψ_n wavefunctions relative to that of the incident radiation frequency ω. The transition frequency can be determined from the unperturbed energies of the ground and excited states, such that

$$\omega_{ns} = 2\pi\left(E_n^0 - E_s^0\right)/h \,. \qquad (II.5.12)$$

The optical rotation ϕ in radians per centimetre is then given by

$$\phi = 4\pi N\beta\omega^2\left(n^2+2\right)/3c^2 \quad \text{where} \quad \beta = -(3\omega)^{-1}\left(G'_{xx}+G'_{yy}+G'_{zz}\right). \qquad (II.5.12)$$

As noted earlier, N is the number of molecules per unit volume, and n is the refractive index of the medium.

In the static field approximation,[119] it is assumed that the rotational frequency of the transition is much greater than that of the exciting radiation ($\omega_{ns} \gg \omega$). In this case, Equation II.5.11 can be simplified and expressed in terms of the partial derivatives of the ground state wavefunction with respect to the static electric F_α and magnetic B_β fields, reflecting the perturbation of the ground state by the fields,

$$\omega^{-1}G'_{\alpha\beta} = -\frac{h}{\pi}\mathrm{Im}\left\{\left\langle \frac{\partial\psi_s}{\partial F_\alpha}\middle|\frac{\partial\psi_s}{\partial B_\beta}\right\rangle\right\}. \qquad (II.5.13)$$

Dynamically,[133-135] if the static field approximation is not used, the tensor is given by

$$G'_{\alpha\beta} = \frac{2\pi}{h}\sum_{n\neq s}\frac{1}{\omega-\omega_{ns}}\times \mathrm{Im}\left\{\left\langle \psi_s^0\left|\hat{\mu}_\alpha\right|\psi_n^0\right\rangle \bullet \left\langle \psi_n^0\left|\hat{m}_\beta\right|\psi_n^0\right\rangle\right\} - \frac{1}{\omega_{ns}+\omega}\times \mathrm{Im}\left\{\left\langle \psi_n^0\left|\hat{m}_\beta\right|\psi_n^0\right\rangle \bullet \left\langle \psi_s^0\left|\hat{\mu}_\alpha\right|\psi_n^0\right\rangle\right\} \qquad (II.5.14)$$

A comparison of both methods has been performed[131] and has found, for the most part, that both methods are quite adept at predicting optical rotations of molecules.

The obvious reason for most of this research is developing the ability to predict the optical activity of specific molecules, which would be useful in the assignment of absolute configurations and in structural analysis. Success has been seen in this direction.[120-124,129,130,137-139]

However, of greater relevance to this thesis is the concept of assigning contributions to optical activity from different portions of the molecule of interest.[125,132,137,138,141] In these works, a fuzzy fragmentation scheme is used, as shown in Section II.4, to calculate the perturbation of the wavefunction by the incident radiation at each atom. From this, the contribution of each atom to the overall optical activity is determined.

III CHIRALITY MEASURES VIA ARRAY POINT HANDEDNESS ASSIGNMENT

A viable chirality measure for the electron density representations of fragments or molecules should be equally useful for measuring the chirality of lattice animals. This arises from the array nature of storing calculated electron densities in a computed file. A specific isodensity contour can be considered a Jordan surface within a 3-dimensional space. Within the isodensity contour are array "points" of finite volume that can be considered to be the cells of a lattice animal, or, in the case of toroids or disconnected pieces, as groups of lattice animals. Therefore, a measure that can be applied to lattice animals can conceivably be applied to a volume of space within an isodensity contour.

The development of an algorithm for the measurement of the chirality of an electron density representation is given in stages. First, specialized Cartesian spaces are considered, in the effort to provide a necessary, if mathematically non-rigorous basis, for the continued discussion. From that point, a method of measuring the chirality of lattice animals is given based upon the assigning of a handedness value to each cell of the lattice animal. Several chirality measures can be defined once all cells are assigned a handedness value.

Once the lattice animal chirality measure is established, a discussion of the problems associated with transition from the lattice animal measure to an equivalent electron density measure is necessitated. This transition requires a discussion on the chirality of appropriately face labelled n-cubes, and the concept of chirality based

on differences between the labels of opposing faces of face labelled n-cubes. Such face labelled n-cubes can be used to assign a handedness to each array cell falling within an isodensity contour. Once handedness is assigned to all points, chirality measures equivalent to those for lattice animals can be proposed.

Finally, the algorithm for a computer program developed for assigning handedness to electron density array cells and the subsequent chirality measure generated is discussed. This is followed by a discussion of the testing and application of the algorithm.

III.1 Properties of Specialized n-Dimensional Cartesian Spaces

The properties of n-dimensional Cartesian spaces are well known.[144] Scaling transformations, translations, reflections and rotations in these spaces are accomplished systematically with few constraints. However, the defining of spaces with constraints on reflection, translation and rotation can often be useful in describing some phenomena that would be made more complex if these constraints were not in place. The space of lattice animals is one such space, and is considered here. Additionally, a second space, the *uniform gradient sub-lattice space*, will be introduced. Because these spaces have many of the same properties as Cartesian spaces, the discussion will not be mathematically rigorous.

III.1.1 n-Dimensional Lattice Animal Spaces (R^n_L)

Consider an n-dimensional Cartesian space of a particular handedness defined by n mutually orthogonal unit vectors of non-zero magnitude, and designate this space as $\mathbf{R}^n_L$. An n-cubic lattice may be defined within this space. Within this space

only points with integer coordinates are capable of acting as n-cube *cell centres*. Figure III.1.1 shows the 2-dimensional lattice animal space $\mathbf{R}^2_L$. In the Figure, the grey grid lines are provided for visualization of where cell edges could lie, and are not real objects in the space. The centres of each of the squares would be points allowed for a cell centre.

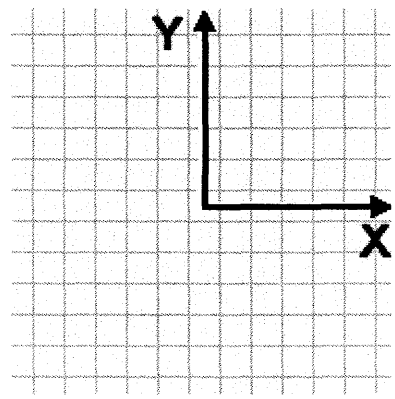


Figure III.1.1 The 2-dimensional lattice animal space $\mathbf{R}^2_L$.

Within this space, the structure of a lattice animal can be enforced if constraints are placed upon reflection, rotation, and translation of the animal:

(1) Scaling is forbidden. Scaling fails to map cell centres onto allowed cell centre points in all cases.

(2) Reflections may only be performed through an n-plane of reflection defined by n-1 axes of the space or by a combination ("diagonal" reflection) of such reflections. The n-planes used must contain the origin of the lattice space. While the resultant animal of such a reflection is a mirror image of the original animal, the cell centres will still fall on allowed points.

(3) Rotation relative to the origin may only be performed within a plane defined by any two of the axes, and only rotations of integer multiples of 90° are

allowed. Such rotations or combinations of such rotations will map all cell centres onto points that are allowed for cell centres.

(4) Translation may only be done by integer multiples of the unit vector along the directions of the coordinate axes. Such translations ensure each cell centre will be mapped to a point allowed to be a cell centre.

Within the lattice animal space, any lattice animal can be described wholly by the coordinates of the cell centres of all the n-cubes that make up the cells of the animal. For the purposes of all subsequent discussion, this animal description should include the coordinates of the primary (centred on the origin) cell from which the other cells have "grown" to create the animal.

A point, known as the *coordinate centre*, can be defined as having the average coordinates of all of the cell centres that define the lattice animal. To obtain the coordinate centre requires the determination of each of the n average directional coordinates of the point. The average *n*th directional coordinate x^n_{avg} of a lattice animal comprised of m cells is defined as the sum of the *n*th directional coordinates of all cells divided by the number of cells:

$$x_n^{avg} = \frac{\sum_{n=1}^{m} x_n}{m} \quad . \qquad (III.1.1)$$

Within the lattice animal space, the distance between the coordinate centre and any given cell centre belonging to a lattice animal is invariant to rotation, translation or reflection. This point may lie within the "body" of the animal by being enclosed within a single n-cube or in a common face between two n-cubes. It also may lie on the "skin" of the animal, which would be a face of an n-cube not shared with another n-cube, or the point could lie "outside" of the animal. A physical

analogy to the coordinate centre would be the "centre of mass" of the animal, assuming that the density of material within the animal is constant for all points that comprise the animal.

Figure III.1.2 shows the 4-celled 2-dimensional lattice animal "Elly", designated **L**, and its mirror image **L***. Based on lattice animal growth principles, each animal has been formed from cell growth starting at an initial cell centred at the origin of the lattice animal space. The coordinates of the cell centres are given for each cell, as well as the coordinate centre, which has been indicated in each animal of the Figure by the letter x.

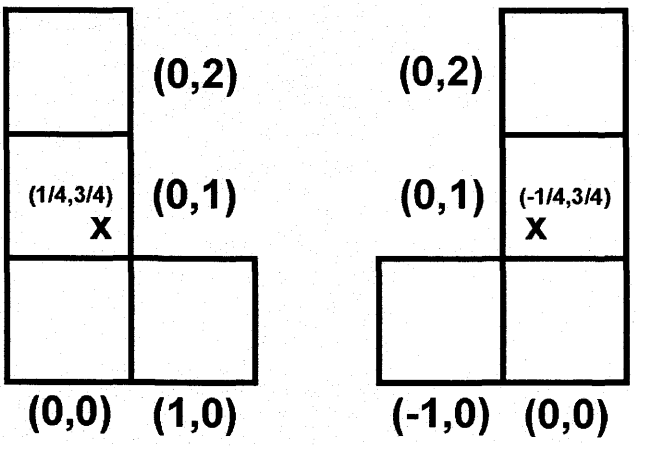


Figure III.1.2 The cell coordinates and coordinate centres (**x**) of the lattice animal "Elly" (**L**), and its mirror image (**L***).

III.1.2 Uniform Gradient Sub-lattice Spaces (R^n_U)

Consider an n-dimensional Cartesian space of a handedness equivalent to the lattice animal space defined by n mutually orthogonal unit vectors of non-zero magnitude, and designate this space as $\mathbf{R}^n_U$. This space can be thought of as an n-dimensional representation of an (n+1)-dimensional space where the first n dimensions are spatial (equivalent to the general Cartesian space $\mathbf{R}^n$), while the

(*n+1*)th dimension is the "dimension of gradient." In the n-dimensional space each point of $\mathbf{R}^n_U$ is assigned a *gradient value* subject to the following constraints:

(1) The gradient has the maximal value for the entire space at the origin. No other point may have this maximal gradient value.

(2) Each point in the space at a specified radial distance from the origin is assigned an equivalent gradient value such that all points with a smaller radial distance from the origin will have a larger gradient value while all those at a larger radial distance will have a smaller gradient value. The function describing this assigning of gradient values must be differentiable at all points. A Gaussian gradient distribution and an inverse squared distance gradient are two of an infinite number of possible gradients.

If the space has been defined correctly, the n partial second derivatives with respect to each of the n spatial coordinates at any given point should define a vector that points towards the origin of the space for all points in the space except the origin. Figure III.1.3 shows an example of a 2-dimensional uniform gradient sub-lattice space.

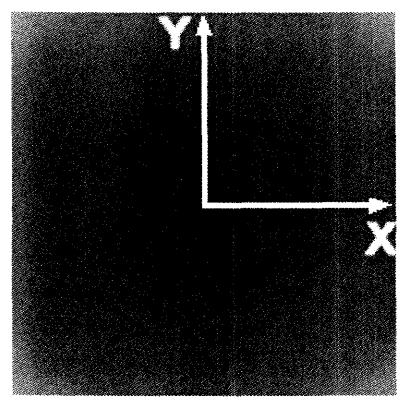


Figure III.1.3 A 2-dimensional uniform gradient sub-lattice space $\mathbf{R}^2_U$.

III.2 Embedding R^n_L Within R^n_U

For the purposes of proposing a chirality measure for lattice animals and later, for calculated electron density representations of molecules and fragments it is useful to consider the "embedding" of the lattice animal space containing the lattice animal of interest into the sub-lattice space. The act of embedding creates a distinct new space with the combined properties and constraints of both spaces, with a few exceptions, based on the following criteria:

(1) The coordinate centre of the lattice animal of interest within the lattice animal space is placed at the origin of the gradient space.

(2) Each of the directional axes of the lattice animal space must be aligned parallel to its counterpart within the sub-lattice space.

Within the embedded space only rotations need be considered and so translation and reflection within the space are prohibited. As specified in the lattice animal space, only integer multiples of 90° rotations are allowed within a plane defined by any two of the axes. However, in the embedded space this must occur about the origin of the uniform gradient sub-lattice space. Figure III.2.1 shows an example of a 2-dimensional embedded space created using the coordinate centre of the 4-celled lattice animal **L** ("Elly") as shown in Figure III.1.2. In the Figure, the black coordinate axes are those of the lattice animal space before embedding, while the white coordinate axes are from the sub-lattice space. Once again the provided visual lattice defines potential cell locations, but the visual lattice is not a real object within the embedded space. The lattice animal **L** would have the same coordinates in the black coordinate frame as it does in Figure III.1.2.

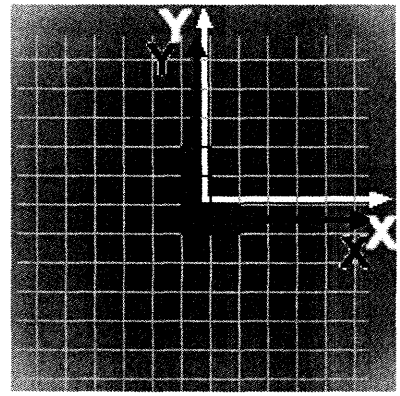


Figure III.2.1 The 2-dimensional embedded space for the lattice animal "Elly".

If an allowed rotation of the lattice animal is performed within the embedded space, the cell centres will not necessarily be rotated from an allowed cell centre position to a new allowed cell centre position within the lattice. However, since the distance from any original cell centre to the coordinate centre of the lattice animal space (the origin of the embedded space) is invariant, the cell centres will be assigned to points with the same gradient value regardless of what rotation is performed. This provides an insight into the nature of array-based chirality measures that will become more apparent as discussion continues; the chirality measure must have suitable constraints on orientation to be effective.

III.3 Rule Set for Determining the Handedness of a Lattice Animal Cell

A chiral lattice animal must have some property which can distinguish itself from its mirror image as is definable within the properties of the space in which it exists. Ordinarily, for any chiral object, this property is that no combination of translation or rotation within the space will allow the superimposition of the object onto its mirror image. More correctly, the object may not contain an improper axis of rotation S_n, where rotation by $\left(\frac{360}{n}\right)^\circ$ and a subsequent reflection in a plane

perpendicular to the rotation axis result in the same object.[145] Furthermore, this property must in some way be applicable to each individual cell of the animal, such that any cell within an animal must either behave identically to the mirror image equivalent cell, or, more likely, it must have some property that is "handed." An analogy relating how a smaller part of a chiral animal may itself be handed is that of a human being and gloves. A person (who in three dimensions, literally, has chiral hands) may have a pair of winter gloves that are designed to fit the left and right hands. The chirality of the hands can be inferred by the chirality of the gloves, and can in some small part be measured by the difference in the gloves. On the contrary, the person may have fingerless mittens that are superimposible on each other in three dimensions. The lack of chirality information provided by the mittens can only lead to ambiguous statements about the chirality of the person's hands.

Within the embedded space, the gradient vector at a given cell centre points towards the origin of the space because of how the *gradient value* at a given point was defined. A line segment coincident to the gradient vector can then be defined between the cell centre and a point on the boundary of the cell (the cell walls). This point of intersection of the cell boundary and the line segment will either coincide with a vertex of the n-cube that partially defines the boundary of the cell, or a point within one of the edges that partially defines the boundary of the cell, or with a point within one of the faces that partially define the boundary of the cell. The vector defined by the cell centre and the point of intersection with the *cell wall* can be used to define the handedness of a cell based on an internal coordinate frame. This vector will be some scalar multiple of the original gradient vector.

A Cartesian coordinate frame can be assigned a handedness based on how it is created. For example, the commonly used 2-dimensional coordinate frame places the positive y-axis perpendicular to the positive x-axis such that a 90° counterclockwise rotation will map a point lying on the x-axis onto the y-axis. However, a coordinate frame created in the same manner based on a clockwise rotation mapping is also valid. If the origins and x-axes of the two spaces were superimposed upon each other, the positive y-axis of the first space would map onto the negative y-axis of the other and vice versa. Any collection of points assigned coordinates in one frame would be the mirror image of the same collection of coordinate assigned points if the object was transplanted into the other space. This property of frame definition allows for the assigning of the handedness to a lattice animal cell.

Within the coordinate frame of the uniform gradient sub-lattice space, the n-component vector between a cell centre and the point of intersection with the cell wall is defined as the *sub-lattice space vector*. The components of this vector can be used to create a new coordinate frame internal to the specific cell with the cell centre acting as the origin of the frame. To accomplish this, a rule set can be established:

(1) The sub-lattice space vector consists of n components

$$^{U}x_i \quad for \quad i = 1,2,3,\ldots,n. \qquad (III.3.1)$$

(2) $^{C}x_1$, the internal cell coordinate frame primary positive axis is chosen to coincide with the direction of the sub-lattice space component vector with the largest magnitude. The particular sub-lattice space vector component used is then designated as "claimed".

(3) Each subsequent positive axis direction ${}^{c}x_i$ for i=2,3,...,n-1 is chosen to coincide with the direction of the largest remaining unclaimed component vector, which is then designated as claimed.

(4) The handedness value of the internal frame of the cell is assigned based on whether the ${}^{c}x_n$ direction coincides with the positive ${}^{R}x_n$ direction of a previously defined handed reference frame of the same dimension (value of +1) or with the negative ${}^{R}x_n$ direction (value of -1). The original handedness of the reference frame is irrelevant as long as the same reference is used in all comparative measures.

(5) If two or more ${}^{U}x_i$ vector components have the same absolute magnitude, the cell is denoted as ambiguously handed and is assigned a handedness value of zero.

In a 2-dimensional space, the commonly used Cartesian frame with the positive y-axis counterclockwise from the positive x-axis serves as the reference frame. Figure III.3.1 shows a square in the Cartesian frame of the embedded space that could represent a single cell of a lattice animal in a 2-dimensional space. Note, first, that the edges of the square are parallel to the axis of the embedded space as a result of the lattice animal space contribution. In the Figure, black lines are drawn from the cell centre of the square to the vertices of the square. Additionally, black lines are drawn from the cell centre to intersect each cell face perpendicular to the face. If the sub-lattice space vector of the considered cell coincides with one of these lines, the handedness of the cell is ambiguous, and is assigned a handedness value of zero. Otherwise, the vector will be contained in a region of the space either

denoted by the grey areas or the white areas. In the Figure, a vector that lies in a grey area of the space has a handedness that agrees with that of the reference space, and the cell would be assigned a handedness value of one. If the vector lies in a white area, the cell handedness opposes that of the reference space and the cell is assigned a handedness of minus one.

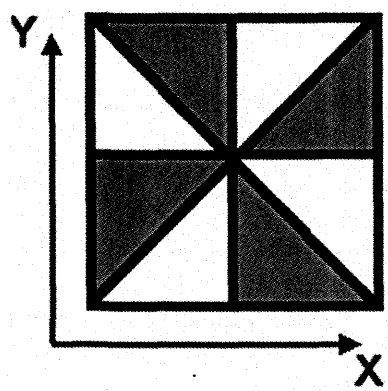


Figure III.3.1 Generic representation of a 2-dimensional lattice animal cell for handedness assignment.

Of special note in Figure III.3.1 is that an allowed rotation in the lattice animal space will map any given grey area exactly onto an existing grey area. Therefore the assigned handedness is invariant to allowed rotations. Figure III.3.2 shows the equivalent 3-dimensional example in a right-handed Cartesian frame.

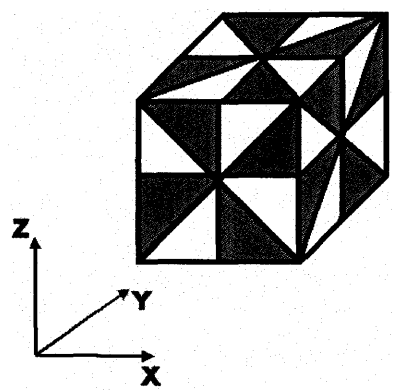


Figure III.3.2 Generic representation of a 3-dimensional lattice animal cell for handedness assignment.

III.4 Cell Handedness Chirality Measure for Lattice Animals

Based on the assignment of the handedness of the cells of a lattice animal, it should be possible to assign a handedness to the animal itself. Figure III.4.1 shows the 2-dimensional lattice animal **L** and its mirror image **L*** in an embedded space that is handed-defined based on the commonly used Cartesian frame. Each animal actually exists in it's own embedded space with the origin coinciding with the coordinate centre of the animal. The dotted lines within each animal denote the lines of reflection that would be allowed in a 2-dimensional lattice animal space, and that subsequently apply to the embedded space. In the case of the lattice animal **L**, the sub-lattice space vectors for each cell are shown. If the spatial coordinates of the cells for the original lattice animals as shown in Figure III.1.2 are used, it is seen that both the cells at (1,0) and (0,1) are ambiguously handed by the rule set defined earlier. By using Figure III.3.1 as a guide, it is also seen that the cell centred on (0,0) has a handedness value of minus one, while the cell at (0,2) has a handedness value of one. The vectors of the equivalent cells of **L*** would also be zero handed for (1,0) and (0,1), but the (0,2) cell would be negatively handed, while the (-1,0) cell would be positively handed.

Unfortunately, a simple chirality measure such as the summing of the handedness values of all cells would determine that both **L** and **L*** are achiral, which is obviously not the case. To counteract this problem, two more issues must be addressed.

The first issue can be resolved by the introduction of a sixth rule to the handedness assignment rule set. Essentially, if any point that belongs to the lattice

animal space allowed reflection surfaces (the dotted lines in Figure III.4.1) coincides with an internal cell (**not** cell wall) point of a cell, the act of reflection would map one or more points within the cell to points already contained within the same cell. The result would be equivalent to the distorted reflection one would see of a person's face if one part of their face were to be placed on one side of a mirror while the remainder was placed on the other side. To resolve this problem, the sixth rule of the rule set states that regardless of the assigned handedness value, the handedness value of a cell that contains one or more points of the allowed reflection surfaces should be reassigned as zero.

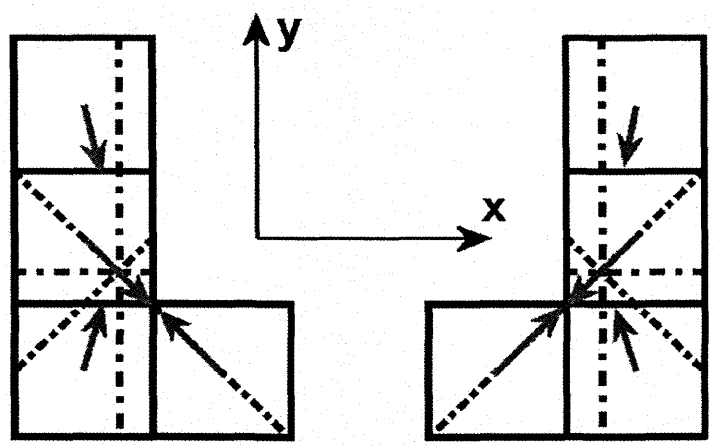


Figure III.4.1 The sub-lattice vectors for the cells of the lattice animal **L** and its mirror image **L***. The dotted lines denote the allowed reflection lines.

However, for a small-number-celled animal, this often will mean that all cells will be assigned a handedness of zero, and no valid chirality determination can be made. To combat this, the concept of **resolution** can be used.

Resolution, in this case, does not have the exact same meaning as is applied to the use of lattice animals in chirality measures of Jordan curves, however, the application is very similar. The perimeter (the union of all edges that belong to one and only one cell – the "skin" of the animal) of the lattice animal can define a lattice-

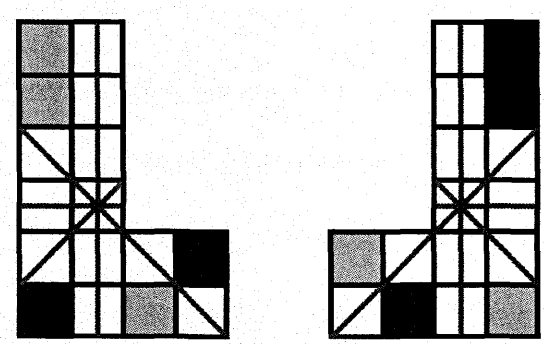
based Jordan curve to which a lattice of finer (smaller cells are used) resolution can be used to define a new lattice animal with the same perimeter as the original animal, but consisting of a greater number of cells. However, unlike the case of the Jordan curve lattice animal chirality measure, where the orientation of the Jordan curve relative to the lattice axes can be changed as the resolution is improved, the application of a finer resolution in the embedded space must follow the constraints of the axes of the space.

A resolution factor f, where f may be a positive integer greater than one, can be defined within the embedded space such that the finer lattice can have edges only of lengths that are 1/f multiples of the unit vector length of the embedded space. As stated above, this new lattice must follow the directional constraints defined by the axes of the embedded space.

Figure III.4.2 shows the lattice animals **L** and **L*** after a resolution factor of two has been applied. In the Figure, white cells are assigned a handedness value of zero, light grey cells are given a value of minus one, and dark grey cells have a handedness of one. A chirality measure has been applied to **L** and **L***. The measure is based on the ratio of one-handed cells to the total number of handed cells.

Based upon the definition of the chirality measure in this specific example, an achiral molecule should give a measure of one-half (0.500) because the number of light grey cells should equal the number of dark grey cells by reflection symmetry. The chirality measure therefore distinguishes between the mirror image lattice animals by assigning each lattice animal a different chirality measure between zero

and one. Additionally, the sum of the chirality measures of a lattice animal and its mirror image would sum up to the total of one.



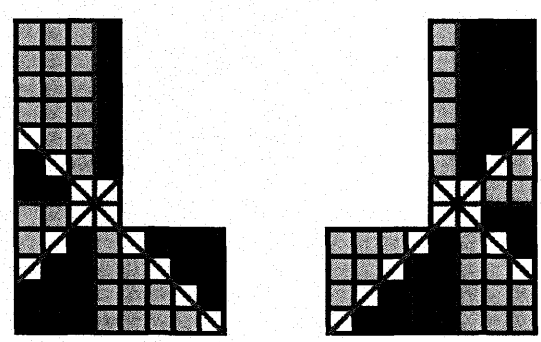
0.600 = 3/5 **0.400 = 2/5**

chirality measure = light gray / total gray

Figure III.4.2 The measurement of chirality for the four-celled lattice animal **L** and its mirror image **L*** at a resolution factor of two.

In the case seen in the Figure, the chirality measure is not necessarily reliable due to the large number of cells of zero handedness. The solution to this problem is an increase in resolution.

Figure III.4.3 shows **L** and **L*** at a resolution factor of four. The chirality measure now is better representative of the chirality of the animals because there are many more handed cells than zero handed cells.



0.538 ~ 28/52 **0.462 ~ 24/52**

chirality measure = light gray / total gray

Figure III.4.3 **L** and **L*** chirality measures based on a resolution factor of 4.

It should be obvious that an increase in resolution factor quickly renders the determination of the chirality measure into an area determination problem for the light and dark regions of the animal. Figure III.4.4 shows the case of an infinite resolution factor. This shows the chirality measure will approach some limiting value as the resolution factor approaches infinity. The value is intrinsic to the animal shape itself, and is invariant to scaling of the cell edge size of the animal.

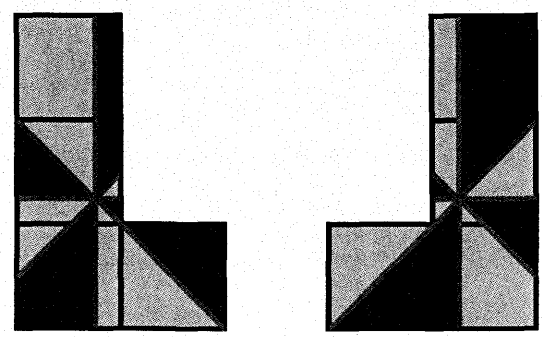


Figure III.4.4 **L** and **L*** chirality measures based on an infinite resolution factor.

III.5 Chirality Measures of Four- and Five-Celled Lattice Animals

The use of lattice animals in the determination of the chirality of Jordan curves (Section II.2.5) does not make the distinction of "amount of chirality" between animals of the same number of cells; an animal is either chiral or achiral. If the Jordan curve only encloses a chiral animal for a given number of cells, and for every number of cells greater than that, then that initial number of cells is used in the definition of the chirality index of the curve. The chirality measure proposed above, however, provides a potential distinction between the chirality of animals with equivalent numbers of cells. To test this statement, a study of the chirality measures of four- and five-celled lattice animals was performed.

In a 2-dimensional space, there are five distinctly shaped lattice animals of four cells, and twelve consisting of five cells. Table III.5.1 lists the animals as numbered in Figure II.2.1, whether the animal is achiral or chiral, and the subsequent chirality measures determined as described above for an infinite resolution factor for the animal and its mirror image.

Animal (**A**)	Chiral / Achiral	**A** : **A*** Chirality Measure
1	Chiral	0.663 : 0.337
2	Chiral	0.531 : 0.469
3	Achiral	0.500 : 0.500
4	Achiral	0.500 : 0.500
5	Achiral	0.500 : 0.500
6	Chiral	0.656 : 0.344
7	Chiral	0.588 : 0.412
8	Chiral	0.575 : 0.425
9	Chiral	0.572 : 0.428
10	Chiral	0.536 : 0.464
11	Chiral	0.500 : 0.500
12	Achiral	0.500 : 0.500
13	Achiral	0.500 : 0.500
14	Achiral	0.500 : 0.500
15	Achiral	0.500 : 0.500
16	Achiral	0.500 : 0.500
17	Achiral	0.500 : 0.500

Table III.5.1 Chirality measures for the seventeen four- and five- celled lattice animals numbered in Figure II.2.1.

As is expected, each of the achiral animals generates a chirality value of 0.500, indicating that the animals are achiral. Each of the chiral animal pairs shows a distinct set of chirality measures that differentiate amongst animals of equivalent numbers of cells. However, one exception is noted in the lattice animal numbered eleven. This five-celled animal, which can be named "Whirly" for purposes of

discussion, is a chiral lattice animal, yet it provides a chirality measure of 0.500 for itself and it's mirror image pair, implying that it is achiral as it concerns the chirality measure. This apparent contradiction warrants closer inspection.

III.5.1 Whirly: The Lattice Animal Equivalent of Centrally-Symmetric Chirality

Figure III.5.1 shows the infinite resolution case of the chirality measure determination of one of the mirror image versions of Whirly. As should be apparent from the Figure, each cell is subdivided into handedness domains of equal area, and therefore the total area of each handedness value within the lattice animal should be equal, leading to a chirality measure of 0.500. Upon closer inspection, the important feature of the Figure is that the coordinate centre of the animal, and therefore, the origin of the allowed mirror lines, occupies the same point as a cell centre of one of the cells.

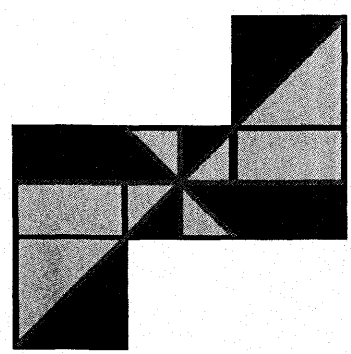


Figure III.5.1 The chirality measure determination at an infinite resolution factor for one mirror image form of the lattice animal "Whirly". Light and dark grey areas indicate points of differing array point handedness assignment. Mid-level grey indicates the allowed mirror lines.

It is the contention of this discussion that such a case where the coordinate centre of a lattice animal coincides with a cell centre is the lattice animal equivalent of a chiral molecule that has no asymmetric atoms (stereocentres). Some examples of such molecules are metal coordinate complexes in which one of three ligands

binds at two sites on an octahedrally coordinated metal ion, resulting in a "propeller" molecule, such as $[Cr(O_2CCO_2)]^{2-}$.[145] An example involving organic molecules is that of 1,1'-spirobiindane,[118,146] for which one of the mirror image structures is given in Figure III.5.2.

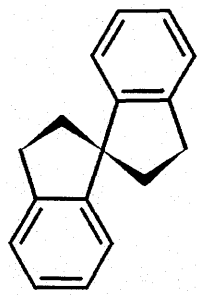


Figure III.5.2 The structure of one mirror image form of 1,1'-spirobiindane.

This molecule can be considered "screw-like" in nature and the "twist" direction would be opposite in the mirror image version. Because there is no improper axis of rotation, this molecule is chiral, yet no carbon atom within the molecule is a stereocentre. The molecule is excellent for the comparison to the case of the lattice animal Whirly, first, because a visual similarity exists between the 2-dimensional structure of 1,1'-spirobiindane and Whirly, but more importantly, because the *electron-averaged coordinate centre* of the space lies on the same point as the central carbon nucleus.

In this case, the electron-averaged coordinate centre of the molecule is defined as the weighted average of each assigned nuclear position, where the weighting is accomplished based on the number of electrons the atomic element possesses. For a chiral molecule with a stereocentre, it is very unlikely that the electron-averaged coordinate centre will coincide with the position of the stereocentre. It is only in the

case of such non-stereocentred chiral molecules that the electron-averaged coordinate centre can reasonably coincide with some "central" nuclear position.

The relationship between the position of an electron-averaged coordinate centre in a 3-dimensional non-stereocentred chiral molecule is analogous to the case in a lattice animal of any dimension where the coordinate centre of the lattice animal coincides with a cell centre, and therefore, a lattice animal cell that contains the coordinate centre within its walls could be conceivably termed a lattice animal "stereocentre" if the coordinate centre point does not coincide with the cell centre.

The presence of non-stereocentred chiral molecules presents a problem, as the lattice animal example of Whirly indicates. The purpose of the proposed method is to differentiate between chiral molecules and their mirror images, for use in Shape Group studies. Obviously, any chiral object that is assigned a chirality value equal to that of an achiral object is not differentiable from its mirror image, and could not be studied effectively in a Shape-Activity Relationship study of chiral molecules. As such, the chirality of the electron density of non-stereocentred chiral molecules will probably not be able to be distinguished using an internal frame handedness based chirality measure, unless the conformation of the molecule renders the central atom "environmentally chiral". In one of the studies presented later (Section V.2), a case of such "environmental chirality" is included, where chirality exists even though the central atom does not have the four different substituents that are usually associated with a carbon stereocentre.

While the case of a non-stereocentred chiral molecule might create problems, the application of internal cell handedness to the measurement of stereocentred

chiral molecules has inherent problems of its own. These problems must be discussed before an algorithm can be developed for the measurement of the chirality of electron density representations.

III.6 Problems of Applying Array Point Handedness Determination to an Electron Density File

III.6.1 Molecular Orientation and Origin Placement in the Array Frame

One of the largest problems that has plagued computational chemists is that of choosing a molecular orientation within a fixed coordinate frame. An infinite number of orientations can be considered. This is doubly problematic for array electron density representations of molecules stored in a computer file, because calculated electron density values at a given point can vary widely for small changes in orientation, especially near nuclear positions. In the lattice animal space, enforcing the constraints for rotation, translation, and reflection solves this problem. Therefore, a potential solution for the problem is enforcing the orientation of the molecule in the array frame by some means. This standard orientation should be applied for all molecules to be studied using the chirality determination method.

The standard orientation scheme chosen was that of a principal moments of charge orientation scheme, with the origin of the space as the centre of the nuclear charge. This particular orientation scheme is easily achieved in the molecular modeling package *Gaussian98*,[147] as it is the default orientation scheme. Once a full electron population analysis is performed in this standard orientation, the electron density representation of the molecule can then be calculated using the in-house software *Rhocalc04*.[148]

However, because the original *Rhocalc* program was developed for use in Shape Group studies, where minor spatial orientation differences are not important due to the curvature domain basis of the method, the program would not necessarily place the origin of the space exactly at the centre of one of the array cells. More specifically, the program chose the centre of nuclear charge to be the "first" cell centre to which the position of all other cell centres was defined. However, if the coordinates of the nuclear charge centre were not integer multiples of the resolution of the array, then the cell that contained the origin of the array space would not place the origin at a cell centre. This is problematic because an exactly calculated mirror image of a molecule would not be created using the program, as origin points of the molecule and its mirror image would not coincide with each other if the array grids were aligned with each other. This would then result in chirality measures that would not be perfectly related to each other.

To combat this problem, *Rhocalc04* chooses an array centred on the origin of the space, and defines the total box size based on the desired resolution such that the array will have an odd number of cells in any of the frame axis directions. This ensures the origin of the space always occurs at an array cell centre, and will result in perfect mirror image calculations.

III.6.2 Non-uniform Electron Density Gradient

One of the constraints of the embedded space used for the creation of a chirality measure for lattice animals is the use of a uniform gradient. Each cell centre of the lattice animal can be assigned a gradient vector that points towards the origin of the space, and the subsequent sub-lattice space vector, and the

handedness rule set assign the handedness value of a cell. Obviously, the electron density gradient of a molecule is not uniform, and thus assigning the chirality of a molecule based on an infinite resolution factor volume is not possible. Figure III.6.1 provides a general 2-dimensional case where the gradient is not uniform over all space. Each ellipse denotes an isoproperty contour much like that of isodensity in three dimensions. The inset shows a specific array cell and the gradient vector for this cell. This vector does not necessarily point towards the origin of the space.

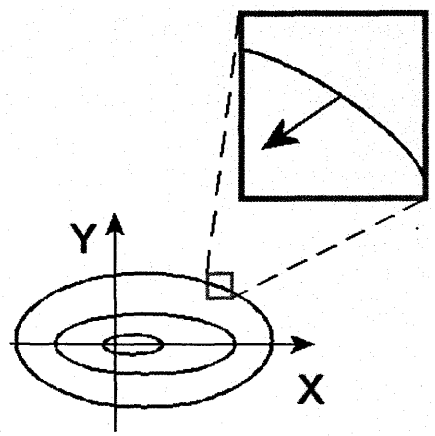


Figure III.6.1 A 2-dimensional example of a non-uniform gradient.

In general, for a molecule, the electron density gradient at a specific point is going to be dominated by the nucleus nearest to that point, and thus the gradient is not uniform unless only one atom is considered. However, for mirror image molecules with stereocentres, the effect of the non-uniform gradient should also be mirrored, and thus a chirality measure based on array cell handedness assignment should still differentiate between the mirror image pairs. As such, assigning the handedness of the cell according to the rule set should still be applicable.

In a fashion, though, Figure III.6.1 is misleading. The actual gradient vector for a point in an array space is dependent on the grid, because it is not calculated

analytically. Therefore, the array grid nature of a stored computer file makes the concept of an isosurface more complex based on what method of approximation is used to define the vector. Figure III.6.2 illustrates this concept. In the Figure, the cell from the inset of Figure III.6.1 is shown as the central cell, with the cells nearest to it also shown.

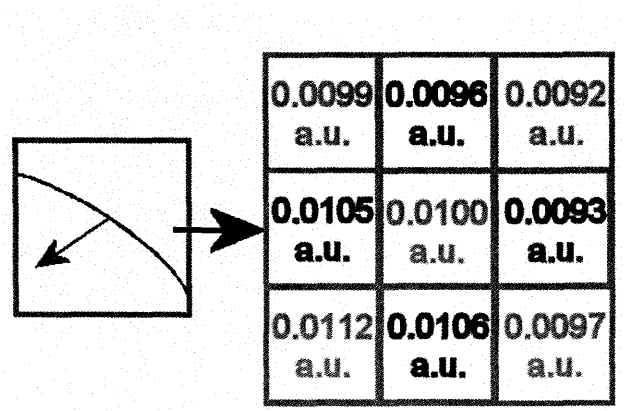


Figure III.6.2 The local electron density array values surrounding a point on an isosurface. The left box shows a portion of an electron density isocontour. The region around the point on the isosurface where the gradient vector starts is expanded on the right.

The Figure shows, that in this case, the isosurface has been chosen at 0.0100 a.u. (arbitrary units). The cell of interest has an assigned property value of 0.0100 a.u., and so it lies on the isosurface. However, none of the surrounding cells has the same property value, and they must be chosen to lie inside or outside of the surface. Therefore an isosurface in an array is actually a construct of some cells with the appropriate property value, and some cell faces that are treated as the boundary between the "inside" and the "outside" of the surface. Computationally, the inside of an isosurface could consist of all points with property values greater to that of the isosurface, or greater than or equal to the property value desired. Such choice in the nature of isosurface would affect how the gradient vector is determined, and

would result in ambiguity in a chirality measure. To combat this problem, the application of *face labelled n-cubes* was chosen as the gradient vector technique.

In an n-dimensional space, an n-cube as defined in the lattice animal theory section, will be an achiral object. Consider, for example, that a cube in 3-dimensional space is achiral. However, the cube can be made into a chiral object if each of its six faces is assigned some labelling property where all the labels are different. One of the most commonly known of these types of face-labelled cubes is a gaming die, where each of the faces is labelled with a number, often, one to six. Figure III.6.3 shows one such face-labelled cube where the faces are transparent.

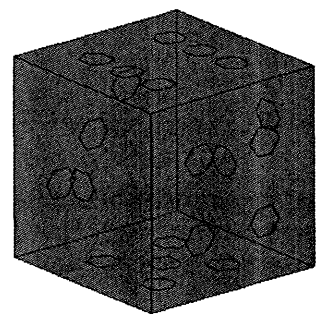


Figure III.6.3 An example of a face-labelled cube.

In the event that the Figure is not clear, the cube shown has the label "six" on the face at the top, "four' on the front right, and "one" on the front left. Opposite of "six" is "five", while "three" opposes "four" and "one" opposes "two". The more commonly seen gaming die place the faces such that the opposing face pairs are six-one, five-two, and four-three.

A rule set to determine the handedness of the face-labelled cube could be created much like that for the sub-lattice space vector. For example, the largest valued label could be assigned the positive x-axis direction, the next largest label the

positive y-axis direction, and the third largest the positive *z*-axis direction. However, the cube of Figure III.6.3 would not be handed-assignable under this scheme, because the two highest label values are collinear. In this case, handedness might be assigned more suitably based on a difference of face labels.

A face-labelled cube can be assigned a handedness based upon the difference in opposing label values. In the case of Figure III.6.3, each of the face-label differences is one, and so this cube could be defined as achiral. For the normal gaming die where six-one, five-two and four-three are the opposing face labels, the label differences are five, three and one. The opposing faces with the largest difference can be placed along the x-axis with the larger-valued face being placed on the positive x-axis. The next largest face difference pair is placed on the y-axis with the larger-valued face on the positive y-axis. The handedness of the cube can then be assigned against a reference frame based on whether the larger-valued face of the smallest face difference pair lies on the positive or negative z-axis in the reference frame.

Face differences can easily be used to determine if a cell is achiral. For example, if any face-pair has a difference of zero, then the faces must be labelled identically, and the labelled cube is achiral because its 2n faces do not all have different labels. More specifically, a mirror plane exists parallel to the two faces in question. Additionally, a cube where two of the face-labelled difference values are equal but non-zero can also be defined as achiral. This would be the same as having a diagonal mirror plane of reflection within the cube.

As an example of the usefulness of face-labelled cubes in handedness determination, consider the array point and its surroundings in Figure III.6.2. The property value inside the cell is irrelevant in assigning the handedness of the cell. Each of the four nearest face-connected neighbour cells have some property value associated with them. This value can be mapped onto the common face of the cell of interest and the neighbour cell. In this way, the cell of interest has been made into a face-labelled 2-cube.

The face-difference pairs of the Figure are (0.0105 a.u. - 0.0093 a.u.) 0.0012 units and (0.0106 a.u. - 0.0096 a.u.) 0.0010 units. Because the face differences are non-zero and not equal, the handedness of the cell is assignable. The larger difference would define the x-axis of the internal frame, and the value 0.0105 a.u. defines the positive x-axis of the internal frame. Because this is a two-dimensional case, the second face-label difference is not needed, as it will lie on the last assignable axis of the internal frame. From visual inspection, it is seen that the larger valued face of this smallest face-pair difference is a 90° counterclockwise rotation from the positive x-axis of the internal frame. Because this agrees with the commonly used two-dimensional Cartesian axis frame, the cell would be assigned a handedness of one.

Utilizing an equivalent technique, an algorithm was developed for assigning the handedness of electron density array cells for subsequent use in a chirality measure of the electron density representations.

III.7 Algorithm for Assigning Internal Handedness of Cells

The algorithm for assigning the internal handedness of cells of an electron density array representation has been used in two programs. The first program, *Chiralizer*, is used for visualization purposes.[149] This program creates an output array file of the same dimensions as the original electron density array where each cell contains the handedness measure of the equivalent cell of the electron density array. The Fortran source code of this program is included in Appendix A. The second program, *Chirality*,[150] which is very similar to the first, calculates cell handedness "on-the-fly" and outputs a very small summary file of the number of points within an electron density isocontour, the number of handed points of positive, negative, or zero handedness, and a resultant chirality measure based on the difference of positively handed points and negatively handed points divided by the total number of handed points. This measure should result in a value of zero for achiral electron density representations, while having a positive value for positive-handed-dominant representations and a negative value for negative-handed-dominant representations.

The discussion of the algorithm will be centred upon the source code in Appendix A.

III.7.1 Inputs for the Programs

Both of the programs accept input of the electron density array information based on the formatted *field file* created by the *Rhocalc04* program. The field file contains information of the dimensional size of the array grid in three dimensions, the resolution of the array grid, which in this instance is the size of a unit cell edge in

bohr, and the coordinates of one of the corners of the "array box" defining the volume of space enclosed. The field file also refers to the named *data file* where the electron density value array is stored. This electron density file is created with *Rhocalc04* by taking the molecular structure information (in the standard orientation) and full electron population analysis output of *Gaussian98*.[147]

The *Chirality* program originally required input of the electron density thresholds in bohr. All points with an electron density value greater than or equal to the lower threshold value and less than or equal to the higher threshold value are assigned handedness values, and are used in the chirality measure determination. All other points are ignored. For the purposes of chirality measures to complement the Shape Group method, the threshold values are chosen to match those used by the Shape Group method programs (10^{-3} e^-/bohr^3 $\leq a \leq$ 10^{-1} e^-/bohr^3).

Finally, both programs have an automatic naming of the output files based on the name of the input field file name. Additional output file name information can also be appended to the output file name.

III.7.2 Outputs of the Programs

Each of the programs has it's own set of output files. *Chirality* creates a small *summary file* consisting of point handedness information and the chirality measure. *Chiralizer* creates two output files. The first of these output files is the handedness array with the same dimensional size as the electron density array. The second is the *chirality field file*, which contains the original information of the field file, with the added reference to the handedness array file. This second file is used by the

visualization program as an input file. Visualizations are performed utilizing the MOLCAD II module[151-153] of the SYBYL molecular modeling package.[154]

III.7.3 The Algorithm

The first step performed by each of the programs after the inputs are made is the reading of the electron density array values. Using a series of nested loops based on the array dimensions read from the field file, the electron density array information is read in a manner equivalent to the way the information was written by the *Rhocalc04* program.

Once the array information is read, the handedness assignment of the cells is done on a cell-by-cell basis using a series of nested loops. These loops have the same dimension sizes as the array grid. However, certain points of the array are considered to be unusable. Each of the cells on the outside of the array box does not have six nearest neighbour cells, and therefore does not have a value that can be mapped on to each of the cell faces. It was decided that the handedness of such points would be non-assignable. In the program each such point is given an arbitrary handedness value of ninety-nine to reflect that the point is not to be considered in the calculation of the chirality measure.

For every handedness-assignable point within the array, the procedure for assigning the handedness is accomplished exclusively through face differences. For each point the electron density values of the nearest neighbour cells are used to define the face differences in each of the three axis directions. These electron density values are assigned the variable name *rho(x,y,z)* based on their position in the array, and the directional face differences are denoted *diffx*, *diffy* and *diffz*. For

the purposes of diagonal reflection achirality determination, three more variables are calculated based on the differences of the face differences in any two of the three axes directions: *diffxy*, *diffxz*, and *diffyz*. Finally, the magnitudes of each of these variables is determined by taking the absolute value of each of the *diff** variables, where * is used to denote an arbitrary string.

The next step determines the *dir** variables. If a *diff** variable is greater than zero, then the direction of the face difference coincides with that of the equivalent direction of the reference space of the array, and the directional variable for that axis is positive one. Additionally, the larger of the two face labels is assigned as *big**, while the other is assigned as *small**. In the case where *diff** is less than or equal to zero, the *dir** variable is negative one and the *big** and *small** assignments are made. Finally, *sumdir* is the sum of the three *dir** values.

The *dir** values can be used to create a new coordinate frame for the array cell which starts as an unlabelled frame. The labelling of the frame will be accomplished based on the ordering of the magnitudes of the face differences.

In the programs, the variable *mircode* is used in various stages. Every point is initially assigned a *mircode* value of zero, which means the point is considered chiral. This *mircode* value is changed based on various checks to a value of one (achiral) or a value of greater than one (potential diagonal reflection). In the case where any particular face difference (a *diff** variable) is zero, then the cell is zero-handed, and the *mircode* is set to a value of one. In a case where a pair of face differences (*diffxy*, for example) has a resultant difference between them of zero, then a diagonal mirror plane may exist and a temporary code of greater than one is

assigned. If the face labels of the two face differences in question coincide, then the point is zero-handed and the *mircode* is set to one, else it is handed and the *mircode* is set to zero.

In those cases where the *mircode* of a point is zero, and handedness can be assigned, the face differences must be ordered from largest to smallest in terms of the directions they occur in the reference array space. Only the magnitude of the face differences is required for this step, and thus the variables *sl** are defined which are the absolute values of the *diff** variables. Because there are three axis directions within the array space, there are six possible ways to sort the directions from largest to smallest. Each of these six possibilities is assigned an *ocode* value within the programs.

Utilizing the four possible values of *sumdir* (3, 1, -1, and -3) and the six possible *ocode* values based on sorting, there are forty-eight possible arrangements of coordinate frames with new axis labels. However, each of these arrangements either corresponds with that of the reference frame, or that of the mirror image of the reference frame, and thus each of the arrangements can be assigned a handedness of positive one or negative one. Table III.7.1 summarizes the possibilities.

To provide a perspective of the basics of the algorithm, Figure III.7.1 shows the handedness assignment of a specific case pictorially. In this case, the face labels of the array points have face differences such that the largest face difference occurs in the y-direction of the reference frame, the next largest is in the x-direction, and the smallest is in the z-direction. Additionally, the face labels are such that the x-axis

face difference has the larger face label coming "before" (lower coordinate value) the smaller value, giving it a *dirx* value of minus one. Similarly, *diry* = +1 and *dirz* = -1.

		ocode	1	2	3	4	5	6
	Large face difference		x	x	y	y	z	z
	Medium face difference		y	z	x	z	x	y
	Small face difference		z	y	z	x	y	x
dirx	**diry**	**dirz**	**sumdir**	**Handedness Value**					
1	1	1	3	1	-1	-1	1	1	-1
1	1	-1	1	-1	1	1	-1	-1	1
1	-1	1	1	-1	1	1	-1	-1	1
-1	1	1	1	-1	1	1	-1	-1	1
1	-1	-1	-1	1	-1	-1	1	1	-1
-1	1	-1	-1	1	-1	-1	1	1	-1
-1	-1	1	-1	1	-1	-1	1	1	-1
-1	-1	-1	-3	-1	1	1	-1	-1	1
Handedness 1 if sumdir is:				3 or -1	1 or -3	1 or -3	3 or -1	3 or -1	1 or -3
Handedness -1 if sumdir is:				1 or -3	3 or -1	3 or -1	1 or -3	1 or -3	3 or -1

Table III.7.1 Forty-eight possible arrangements of labelled axes and the subsequent handedness assignment.

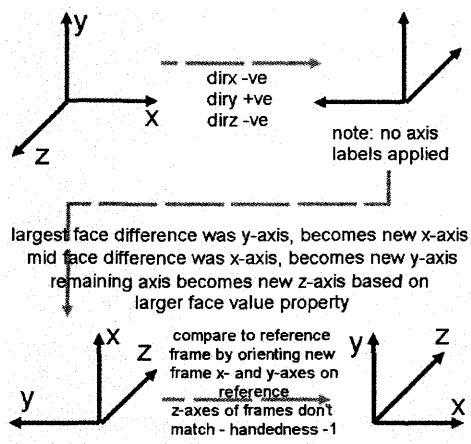


Figure III.7.1 A pictorial example of the application of the algorithm.

The algorithm creates a new unlabelled set of axes at a common origin based on the reference array space axes, such that a *dir** value of positive one means the new axis is directed in the same orientation as the particular labelled reference axis, while a negative one value will orient the new axis along the particular negative axis of the reference frame. Labels are then applied to the axes based on the magnitude of the face differences.

In the case of the Figure, the largest magnitude face difference occurred in the y-axis direction, so the new unlabelled frame receives the x-axis label on the axis that corresponds to the y-axis of the reference frame. Subsequently, the second largest face difference maps the y-axis label on what was the x-axis of the reference, and the z-axis is mapped onto the z-axis of the reference frame.

The newly labelled coordinate frame either is of the same handedness as the reference frame, or is of the opposite handedness. By superimposing the positive x- and y-axes of the new frame upon the x- and y-axes of the old frame through rotations, it is seen that the positive z-axis of the new frame lies on the negative z-axis of the reference frame, and therefore the frame (and the array point to which it applies) are assigned a handedness of minus one. In terms of Table III.7.1, the example of the Figure represents that of an *ocode* of three and a *sumdir* value of minus one, which is assigned a handedness of minus one.

III.8 Testing of the Algorithm with Achiral Molecules

To check the chirality assessment programs, initial testing was performed on the electron density representations, calculated using the restricted Hartree-Fock (RHF) approach, of a series of small achiral atoms and molecules, consisting of: atomic helium, atomic and molecular bromine, molecular fluorine, hydrogen cyanide, ammonia, water, methane, ethane, ethene, ethyne, and benzene. In each electron density representation, each array point should have at least one corresponding mirror image point in a position determined by the spatial position of the array point relative to existing mirror planes of symmetry. Each such pair of points should be

assigned opposing handedness values, leading to an equal number of each type of handed points.

To ensure that the molecules contained planes of symmetry, the highest point group symmetry was enforced during the geometry optimization phase of calculations, which was performed with the 6-31G** basis set.[155-157] The lone exception is ethane, for which the staggered conformation was calculated.

Table III.8.1 shows the array point handedness data for the representations, and includes the number of points assigned handedness values of one, minus one and zero, as well as the total number of points. The data is taken from representations calculated with an array point unit cell edge size of 0.200 bohr.

		Handedness Assignment		
Molecule	Total Points	0	1	-1
Helium	8080	2896	2592	2592
Atomic Bromine	34154	0	17077	17077
Molecular Bromine	55232	4200	25516	25516
Molecular Fluorine	18778	2926	7976	7976
Hydrogen Cyanide	30835	3403	13716	13716
Ammonia	25503	879	12312	12312
Water	19285	1433	8926	8926
Methane	32476	4228	14124	14124
Ethane	49620	1436	24092	24092
Ethene	43442	1398	21022	21022
Ethyne	34838	4534	15152	15152
Benzene	65065	4421	30322	30322

Table III.8.1 Handedness assignments for the array points of the electron density representations of several achiral atoms and molecules.

As is seen in the Table, all of the achiral molecules have the expected equal number of positively and negatively handed points, as well as a significant number of zero-handed mirror points. The lone exception is atomic bromine, which does not

have a single mirror point. This is an artefact of how the electron density calculation program *Rhocalc04* deals with the creation of a Cartesian frame for a single atom with *d*-orbitals. Molecular bromine, for instance, shows no sign of a problem, and, therefore, this is not an issue in any further discussion.

A few remarks can be made based upon the data in the Table. Generally, as larger molecules are considered, the number of mirror points becomes a smaller percentage of the total number of points. This reflects the concept of mirror *planes*, where only a small number of points belong to a plane in the array, regardless of the size of the molecule. Additionally, those highly symmetrical molecules with several mirror planes at 45° or 90° angles to each other will have larger relative numbers of mirror points. As the angle between mirror planes deviates from these values, the planes are less likely to align well with the natural planes of a cubic lattice array, and the number of mirror points will be lowered.

However, it is upon visual inspection where some more interesting aspects of the method become apparent. Figure III.8.1 shows electron density representations of seven of the atoms and molecules out of the twelve listed in Table III.8.1. In the Figure, each representation is shown at an isodensity contour surface of 0.010 e^- /bohr^3 and each point of a surface is assigned a colour based upon the handedness value assigned to it by the program *Chiralizer*.

The most important aspects of the visualization show that in a molecule that only has sigma bonding, the local handedness assignments are dominated by the nearest nucleus. This is not surprising, because the electron density gradients for many parts of a molecule are locally dominated by the nearest nucleus. However,

as is best shown with the example of ethane, the handedness assignment in the bonding regions of the space becomes much more complex. Of greater interest, however, is the effect of pi bonding upon handedness assignment. Ethene and especially, benzene show the effect of the "fattening" of electron density that is a hallmark of pi bonding upon the array point handedness assignment. This is shown best by the central area of benzene, where the differentiation of mirror planes that align well with the array grid are evident, but this differentiation does not extend to the pi bonding regions between the carbon atoms. In these regions, the fat pi bonds have changed the direction of electron density gradient sufficiently that the handedness assignment is more locally complex. Such pi bond effects on an overall chirality measure would not be accounted for by a purely structurally based chirality measure.

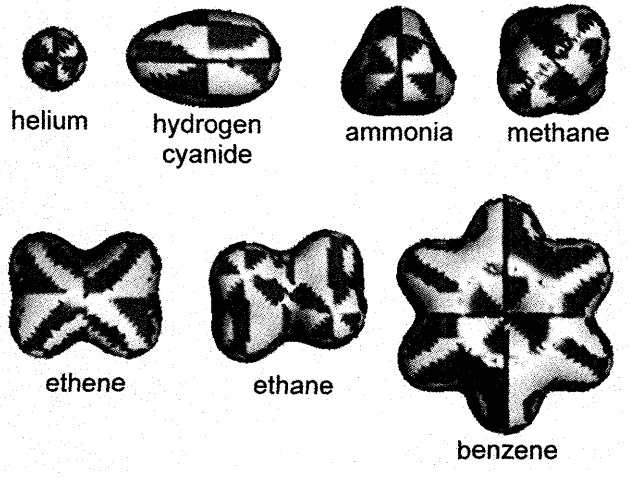


Figure III.8.1 Array point handedness assignments for the electron density representations of seven achiral atoms and molecules. Each representation is shown at a density isocontour of 0.010 e^-/bohr^3 and each array point has a unit cell length of 0.200 bohr.

III.9 Whole Molecule Chirality Measures for Electron Densities

As an initial test of the chirality measure for electron density representations of molecules, the electron density representations of three amino acids were created in

both R and S forms according to the Cahn-Ingold-Prelog (CIP) convention. The geometries were optimized utilizing the 6-31G** basis set, starting from the 6-31G* geometries used by Mezey *et al* in their study of amino acid optical rotation using quantum chirality measures.[89] While these geometries do not necessarily represent the calculated lowest energy conformers,[110,158-162] they do show commonality in the orientation of the carboxyl and amino groups, which is an important factor in Shape Group method studies, as local conformational equivalency tends to reduce the error in the shape similarity measures.

Table III.9.1 shows the array point handedness and chirality measure data calculated for electron density representations of alanine, aspartic acid, and methionine in both the R and S forms. The resolution of the array grid was 0.200 bohr. The chirality measure in this example is defined as

$$\frac{(\#\text{ positively handed points} - \#\text{ negatively handed points})}{\text{total points}}.$$

		Handedness			
Molecule	Total Points	0	1	-1	Chirality Measure
R-Alanine	82504	0	42226	40278	0.023611
S-Alanine	82504	0	40278	42226	-0.023611
R-Aspartic Acid	104248	0	50962	53286	-0.022293
S-Aspartic Acid	104248	0	53286	50962	0.022293
R-Methionine	152585	0	75497	77088	-0.010427
S-Methionine	152585	0	77088	75497	0.010427

Table III.9.1 Handedness assignments for the array points and chirality measures of the electron density representations of three α-amino acids in their *R* and *S* forms.

The data in the Table shows how the nature of the method reveals chirality information. For a mirror image pair of compounds the number of positively and negatively handed points are interchanged between the two, and therefore, the

specified chirality measure will provide one of the pair with a positive value, while the other is given a negative value of the exact same magnitude. However, once the method and the standard orientation are defined, the value is an inherent and singular property of the electron density representation. This method of obtaining a measure of chirality fulfills the desired requirements for the study; it differentiates between the mirror image forms of molecules, the chirality values lie within a continuous range, and the measures are obtained from files that can be directly used for the Shape Group method.

Also of note is that an absolute correlation between the CIP designation of the molecular chirality and the chirality measure does not exist. Because the measure is a function of the whole electron density representation in the molecular standard orientation, while the CIP designation is based upon the arbitrarily assigned priority of groups attached to the stereocentre, this is not surprising.

III.10 Whole Molecule Chirality Measures of Nuclear Potentials

The main feature of using array point handedness to measure chirality is the use of property gradients in labelling the faces of the cells. Because of this feature, it is possible to measure chirality based on any molecular property representation that has a measurable gradient.

To show that the array point handedness method is useful for assigning chirality measures to representations of molecules that are property specific, a small study was conducted on the chirality measures of nuclear potential representations of molecules. A computer program was created to calculate the potential due to nuclear charges for the molecules R-alanine, S-alanine, R-aspartic acid, R-

methionine, and benzene based upon the geometries for the molecules used in the previous electron density studies, at a grid size of 0.200 bohr. The program *Chirality* required only a minor modification to handle these nuclear potential files, in that the upper and lower threshold boundaries were changed to 700 volts and 300 volts respectively, based upon visualization. Table III.10.1 summarizes the nuclear potential chirality measure data, where the chirality measure is calculated as before for the electron density representation study.

		Handedness Assignment			
Molecule	Total Points	0	1	-1	Chirality Measure
R-Alanine	72489	0	37873	34616	0.044931
S-Alanine	72489	0	34616	37873	-0.044931
R-Aspartic Acid	224980	2	114552	110426	0.018340
R-Methionine	330981	1	163619	167361	-0.011306
Benzene	47945	4665	21640	21640	0.000000

Table III.10.1 Handedness assignments for the array points and chirality measures of the nuclear potential representations of previously studied molecules.

Within the Table, many of the same features are seen with nuclear potential chirality measures based upon array point handedness as are seen with those from electron density files. First, mirror image compounds will have measures of equal magnitude, but of opposing sign, based upon an exchange of the number of points of opposing handedness. Secondly, an achiral molecule such as benzene will have a relatively small number of zero-handed points, but the net chirality measure will be zero. However, the most interesting molecule of this Table is R-aspartic acid, which has a negative chirality measure when based upon the nuclear potential, but a positive chirality measure when based upon the electron density (Table III.9.1). This

reinforces the notion proposed by Gilat of differing chirality types, based upon property.

III.11 Effect of Array Grid Size on the Chirality Measure

The most significant source of error in the Shape Group method arises from the grid-based nature of the computer files that contain the electron density representations of the molecule. Two main factors are responsible for this. The first arises from how isoproperty contours are defined in the grid. Since the array cell can be considered to be a property-labelled volume of space, the label value applies to all points within the cell, even though in the programs used, the property value is calculated only at the cell centre. Therefore, isocontours are defined as all cell walls that lie between a cell centre with an property value below the threshold value, and a cell centre with a value above the threshold value. The net result is that isocontours fluctuate as small changes in grid size are made.

The second factor is the placement of nuclei within the grid based upon their coordinates relative to the chosen origin of the space. As the grid size changes, the positions of the lattice walls move relative to the nuclear positions. At some specific grid size, the nuclear position relative to cell centres and walls will move, possibly altering the calculated gradient vector used in finding the Hessian matrix and its eigenvalues, which are used in assigning the cell to a curvature domain.

The array point handedness based chirality measure will also be subject to these errors. Consider the simple examples of a 1-dimensional space with one embedded nucleus (not placed at the origin) at several grid sizes, as presented in Figure III.11.1. In the Figure, a subset of the space is represented by a number line

that starts at the origin, and continues in the direction of the positive numbers (from left to right). The nuclear position is at 0.4875 units upon the line (the central vertical grey arrow). Also, it is assumed the property of interest will decrease radially from the nucleus (like electron density or nuclear potential, for example). Therefore, if the property were analytically calculated, the outer thresholds for the property would lie at a radial distance of 0.3625 units from the nucleus (grey arrows at 0.1250 and 0.8500 units respectively). Finally, the cell walls of the grid are labelled below each line, while the grid size is given at the upper left of each number line. Since the origin is chosen to always be at the centre of a cell, the first cell wall will be at a position that is half the grid size, while subsequent cell walls will exist at a position defined by the position of the previous cell wall plus the grid size.

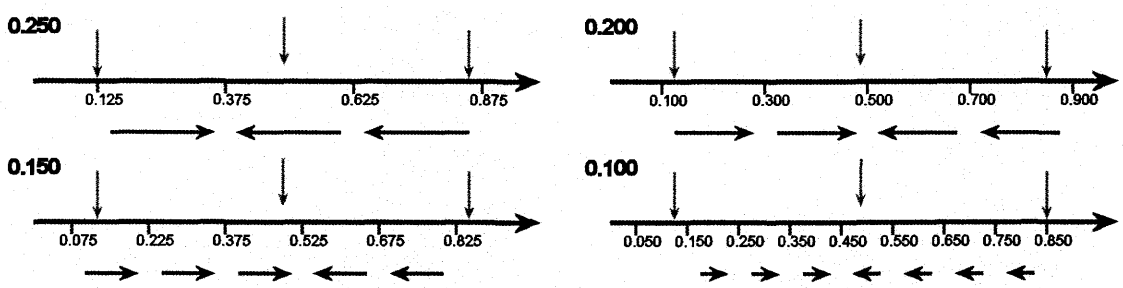


Figure III.11.1 Examples of a 1-dimensional spatial lattice at several different grid sizes. A nucleus placed within the space, but not at the origin, will give rise to chirality measures subject to the errors associated with grid-based techniques. Such errors also occur in the Shape Group method.

In each of the examples of the Figure, a set of left- and right-facing arrows (-1 and +1 handedness assignment, respectively) are given below the number lines to denote the handedness of each cell. These arrows are assigned only to cells within the threshold values. Incidentally, a cell can only be considered to be within the threshold values if the centre of the cell is between the nucleus and the analytical threshold value. This is because the cell property value is calculated at the

coordinates of the cell centre, and therefore if the analytical threshold value is between the cell centre and the nucleus, the actual property value calculated for the cell will be below the threshold value.

In the examples, the handedness of the cells is assigned based upon the face-label differences scheme presented previously. Since, based upon the radial decrease of the property, the cell with the larger value will always be the cell with a cell wall closer to the nucleus, the handedness arrow always points from the cell wall farthest from the nucleus, towards the nucleus.

In the example of grid size 0.250 units, one of the threshold markers (0.1250 units) lies on a cell wall, while the other does not. The isocontour calculated for this property will be correct on one side of the grid, but not at the other. In fact, instead of the analytical distance of 0.7250 units between the threshold ends, the actual distance is 0.7500 units. Then, as the grid size is decreased from 0.200 units to 0.100 units in the other examples, this distance fluctuates from 0.8000 units, to 0.7500 units, to 0.7000 units. During these changes in grid size, the number of cells with assigned handedness goes from three, to four, to five, to seven. Essentially, as the distance between the threshold ends passes through the actual analytic value (between grid sizes 0.150 and 0.100 units), a discrete jump in the number of cells occurs. This is the first factor in grid-based errors; threshold fluctuation.

The second grid-based error is that of the nuclear coordinates relative to important cell features. In the first grid of 0.250 units, the nucleus lies just to the left of the cell centre of the cell it's in, and so the handedness of that cell is –1. At the next grid size, the nucleus is now to the right of the cell centre of the cell it is

contained within, so the handedness is +1. This occurs again for the 0.150 unit grid. Finally, for the 0.100 unit grid, the nucleus is again on the left side of the cell centre, so the nuclear cell handedness changes again. Effectively, if a continuous change in grid size was made between any two grid sizes presented in the Figure, the handedness of the cell containing the nucleus would be seen to change every time the nuclear coordinates moved across a cell centre or a cell wall.

It is these two factors combined together that give rise to the error in the Shape Group method, and the array-based chirality measure. For example, since the Figure presents the case of a single nucleus in the space, the chirality measure should be zero in all the examples. However, if the chirality measure is defined as (left-handed arrows minus right-handed arrows) divided by (total number of arrows), the chirality measure for the example of grid size 0.250 units is 0.333 (1/3). As the grid sizes get smaller, the chirality measure would then take the values zero, -0.200 (-1/5), and 0.143 (1/7).

While the grid size of 0.200 units gives the correct chirality measure, it does so only through a cancellation of errors, which occurs mostly because the nucleus and the analytical threshold values lie close to cell walls. Overall the chirality measure would be seen to fluctuate around the actual value, which would only be exact in the limiting case of an infinitely small grid size. However, the fluctuations should be asymptotic as the limit is approached, and therefore a sufficiently small grid size would render the error acceptable.

Figure III.11.2 shows the effect of changing grid size on the array-based chirality measure calculated for the electron density representations of R-alanine

(dashed line) and R-aspartic acid (solid line) based on the conditions used in the previous electron density study discussed for these molecules. The values of the array-based chirality measure defined earlier have been calculated for electron density representations calculated at grid sizes between 0.30 bohr (coarse grid) and 0.10 bohr (fine grid) at intervals of 0.01 bohr.

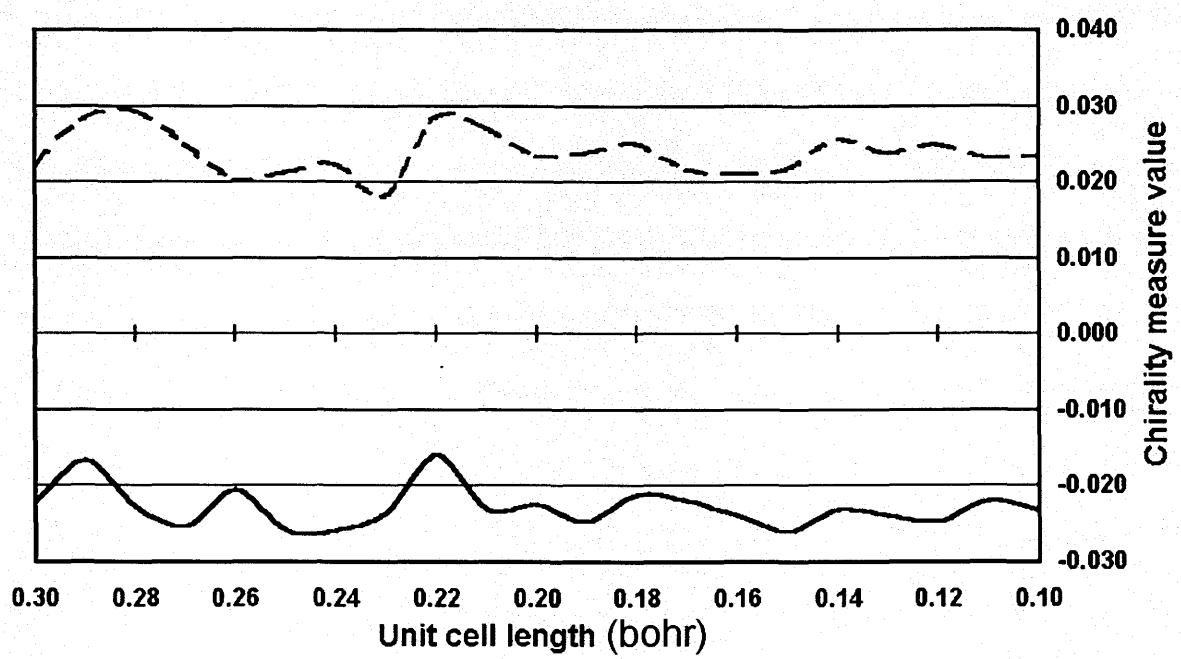


Figure III.11.2 The change in chirality measure as a function of array point unit cell edge size for two amino acids in the R form. The solid line is for aspartic acid and the dashed line is for alanine.

In the Figure, the complex fluctuations seen in the value of the chirality measure arise from the two factors previously discussed. The situation is complicated relative to that presented in Figure III.11.1 because the molecular representations are in a 3-dimensional space. Therefore, the independent errors of each spatial dimension will tend to cancel or augment each other in an unpredictable manner. However, it should be noted from the Figure that the chirality measure takes a consistent value (within reasonable error) with a grid size equal to or less than approximately 0.20 bohr.

Figure III.11.3 shows a visual comparison of the handedness assignments of the points on an electron isodensity contour of R-alanine (0.010 $e^-/bohr^3$) at three different array grid sizes (0.26, 0.20 and 0.16 bohr).

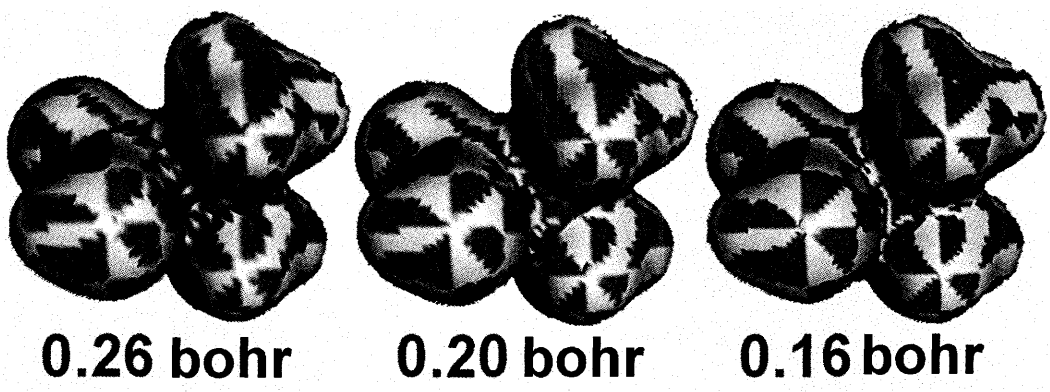


Figure III.11.3 Array point handedness assignments for the electron density representations of alanine in the R form for three different unit cell lengths. Each representation is shown at an isodensity isocontour of 0.010 $e^-/bohr^3$.

In the Figure, it can be seen that the errors associated with changing grid size are a result of fluctuations that occur in the transition regions from one domain of points of common handedness to the other domain. Generally, while the domains themselves remain distinct, it is the border regions that become more defined as the grid size decreases. This better definition of the handedness transition regions of the molecule results in smaller errors in the chirality measure as the grid size decreases.

In the Shape Group method, experience has shown that a grid size of 0.200 bohr gives rise to shape similarity values that have no more than five percent error. Since the same type of errors are associated with the array-based nature of the chirality measure, the error in the measure should also be reasonable in cases where the grid size is 0.200 bohr or better.

IV CHIRALITY MEASURES VIA THE DIFFERENCE DENSITY SHAPE GROUP METHOD

The difference density Shape Group method is an offshoot of the absolute difference density method. Instead of a measure of chirality and shape analysis based upon a single electron density representation of interest, the difference density methods rely on further manipulation of the electron density representations to allow for a new, secondary-shape analysis to provide chirality information.

IV.1 The "Letter" Fragment Analogy of the Shape Group Method

In the early development of the absolute difference density method, a simple analogy was created to help present some of the details of the method without relying heavily on mathematics. This analogy, presented here, simply describes how the absolute difference density method differentiates between totally similar or very similar shapes.

Figure IV.1.1 shows four "molecules" created with structures defined by four substituents bonded to a central "letter" fragment. In a simple sans serif font, such as Arial (the font this thesis is written in), the lower case letters p, q, d, and b can be used to represent these "letter" fragments. Of note, the letters used in the molecules of the Figure have the exact same shapes, as is evidenced visually. Additionally, mirror image pairs exist in the set of letters. For instance, if rotation is not allowed

after reflection, and if only vertical reflection is allowed, then b is the mirror of d, and q is the mirror of p.

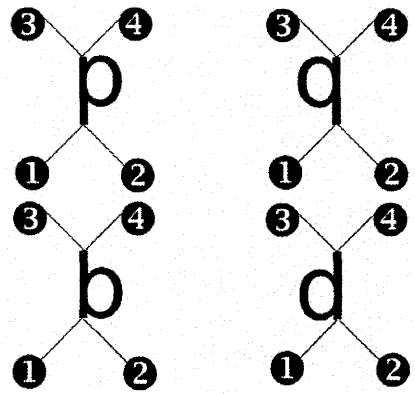


Figure IV.1.1 The "letter" fragment analogy to aid in the description of the absolute difference density method. Four "molecules" are created with different central "letter" fragments. These fragments all have the exact same shape, yet the "molecules" may have different properties.

If the letters of the molecules presented in the Figure were analyzed as a local shape feature of the molecules, the Shape Group method would determine that these fragments have the same shape, and they would, at first glance, have the same properties. This would not necessarily be true, as experience of chirality in chemistry has shown. Therefore a method of further differentiation amongst the letters would be required.

The critical analysis of the shapes of very similar electron density representations (or in this case, letters) requires that the shapes of the objects be altered in some consistent and reproducible manner such that this altered version of the representation can act as a secondary source of information. The absolute difference density method is one such way of altering shapes.

For the letter fragment analogy, the absolute difference density method requires two features. First is the choice of a reference letter. Such a choice in real QShAR studies is usually made as the molecule or fragment of a molecule with the

greatest property of interest. For the letter example, the letter "d" is arbitrarily chosen as the reference. Secondly, the calculation of absolute difference densities requires alignment of the fragments so the absolute difference densities can be calculated point-by-point using simple arithmetic. For letters, the natural alignment we associate with the written English language will serve.

Figure IV.1.2 gives the absolute difference density calculations of the "electron density" of the letter d being subtracted from the "density" of the letters p, q, b, and d. After the subtraction of the density values point-by-point spatially, the absolute value of each point is then taken. For the purposes of the example, the "density" at a given point in the letter is either "zero" (white), or "one" (black). For points where both letters have a density of one or zero, the absolute difference density at that point will be zero. For points where one letter has a zero density and the other has a density of one, the calculation results in an absolute difference density value of one. Finally, the absolute value symbols are not included in the Figure to reduce visual complexity.

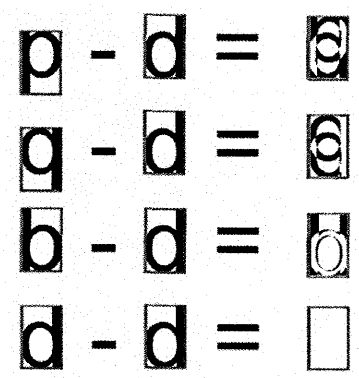


Figure IV.1.2 The absolute difference density "letter" fragment analogy. The letter d is used as the reference letter, and the natural alignment of the letters is used and symbolized by the grey boxes.

As is seen in the Figure, a secondary set of shapes that can be analyzed by the Shape Group method is the result of the absolute difference density method, as applied to the letter fragment example. The shapes of the four letter fragments can now be differentiated from each other.

The letter example does highlight some of the potential difficulties of difference density methods, however. The first is the requirement for alignment. The second is that the reference object will be defined as empty space. These, and other drawbacks are discussed later.

IV.2 The Absolute Difference Density Shape Group Method

IV.2.1 Theory of the Absolute Difference Density Shape Group Method

The absolute difference density method [61] was developed as the initial Shape Group method complementary chirality measure. In this method, certain small fragments of molecules were used to create absolute difference electron densities, which were then analyzed utilizing the Shape Group method as presented in Section II.1.

The concept behind the Shape Group method, when used in QShAR studies, is one of identifying both whole molecule and molecular fragment electron density shape features that can be simply correlated to some specific activity of the molecules. However, as has been discussed previously, when mirror image electron densities are compared, they are considered to be the same in shape by the Shape Group method.

In the absolute difference density method, a comparison of the chirality of the molecule compared to the chirality of the reference is made by choosing some small

common electron density fragment for the molecules (preferably one that will also be used as a "local shape feature" fragment in the study) and analysing the resultant absolute difference electron densities created by taking the absolute values of the result of subtracting point-by-point the electron density of the fragment of the reference molecule from the other molecules of interest. This absolute value requirement originally arose from the needs of the Shape Group method and the software developed for its application. Specifically, the isodensity contour must take on a positive value.

The particular case for which the absolute difference density method was developed was a QShAR study to correlate the shape features of several olfactophore molecules to their concentration in solution before the odour was detected. Such molecules were of interest because many derivatives of a base molecule could give the desired scent, while others might not. Additionally, many of these derivatives had several stereocentres, and often mirror image analogues of heavily scented molecules would have little scent.

Figure IV.2.1 shows the structures of a small subset of the molecules included in the study, labelled by the codes assigned for the study. In the Figure there are two mirror image pairs of molecules (2a-2c and 5b-5e). Additionally, molecule 2a was described as being the strongest in scent (detected at very low concentration in solution) of the entire set.

In the Figure, the molecules on the left are both considered "active" in terms of scent (small concentration in solution before detected by a human test subject), while the two on the right were considered to be "weak" in scent. Because the

molecules are mirror image pairs, the Shape Group method failed to differentiate between the two members of each pair in terms of the various shape features considered. To combat this, the absolute difference density method was applied.

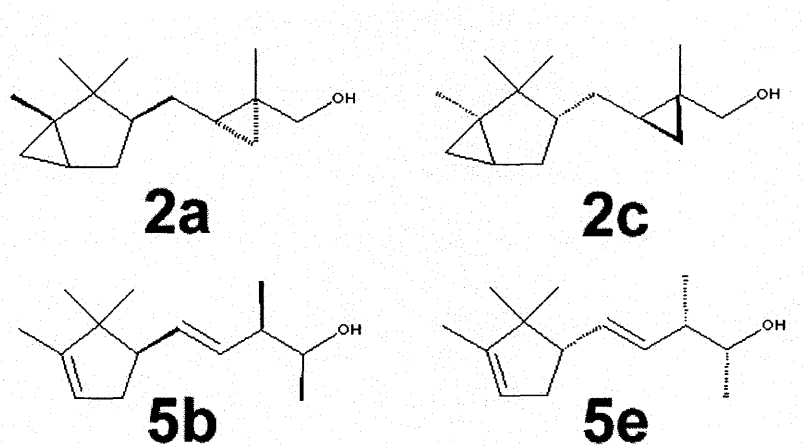


Figure IV.2.1 Four of the molecules used in the olfactophore study.

To apply the absolute difference density method on the electron density fragments of the COH groups present in each molecule required special handling of the data for the molecules. The first step, as also encountered in the letter example, entailed the alignment of the COH groups of the different molecules according to some specific rule set. In the study, the chosen rule set placed the oxygen atom at the origin of the space, while the carbon atom lay on the x-axis. Finally, the y-axis was defined so as to place the hydrogen atom in the xy-plane, with a positive y-coordinate value.

This alignment procedure accomplished several things. First, the placement of the origin at the position of the oxygen atom allowed the atom with the largest associated electron density to be compared most favourably through the absolute difference density method. Second, the alignment of other features was then required so that the point-by-point arithmetic required for the method could be simply

performed while minimizing the error associated between structural differences between the COH groups of the different molecules.

After the alignment procedure, the full population analysis of the molecules was performed using *Gaussian98* utilizing the *nosym* keyword to maintain alignment. Electron density representations were calculated using *Rhocalc*. Because of the placement of the origin and the alignment of the group the creation of absolute difference electron density representations was accomplished through a simple Fortran program that subtracted point-by-point the electron density of the COH group of the reference molecule (2a) from that for the molecule of interest. The absolute value of each point was then taken to give the absolute difference density.

Figure IV.2.2 shows the COH fragmentary electron densities and absolute difference densities of the molecules presented in Figure IV.1.2 at an isodensity contour of 0.001 $e^-/bohr^3$.

In the Figure, the x-axis lies parallel to the top of the page, while the positive y-axis points towards the top of the page, rendering the page as a description of the xy-plane. Because molecule 2a was the reference molecule, its absolute difference density is the subtraction of the electron density of the COH group of 2a from itself, leading to a value of zero at every point.

Visually, from the perspective given, there is very little to distinguish the fragmentary electron densities of the COH groups of each mirror image pair from the other of the pair, except in the carbon atom region, where some lighter shading at the upper right of each COH fragment denotes a "bump" in both the 2c and 5e fragments that would occur on the hidden sides of 2a and 5b. However, differences

do become more pronounced when the absolute difference densities are considered. The difference density of 2c is obviously different from that of 2a, while those of 5b and 5e seem to visually be very similar to each other, with the exception of some small features.

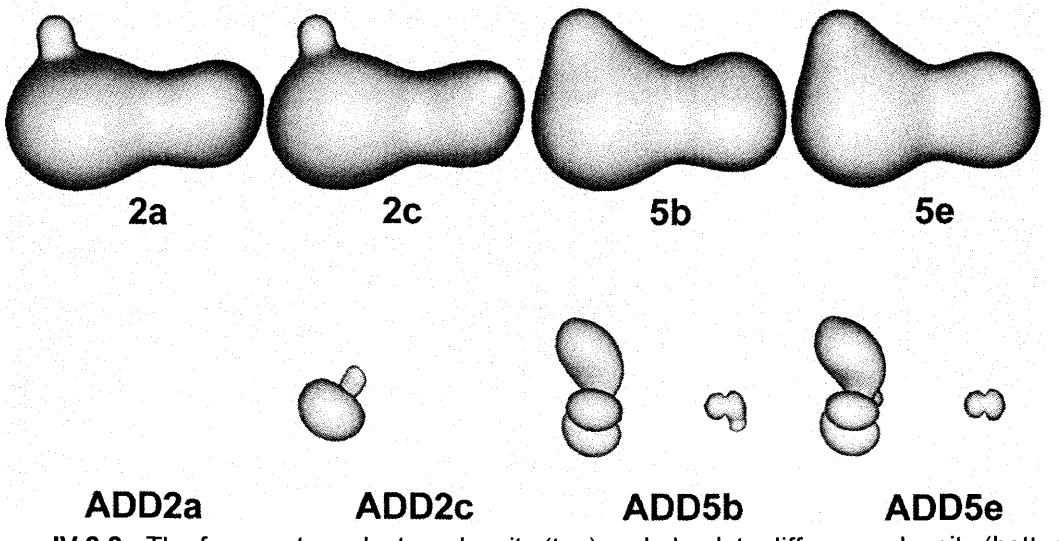


Figure IV.2.2 The fragmentary electron density (top) and absolute difference density (bottom) isocontours of the COH groups of the molecules presented in the previous Figure. All density isocontours are given at 0.001 $e^-/bohr^3$.

Application of the Shape Group method to the absolute difference electron density files is performed in the same manner as for any other electron density file.

Table IV.2.1 presents the shape similarity information of various electron density fragments of the four molecules presented in Figure IV.2.1, relative to the shape of the fragments obtained from molecule 2a.

Molecule	Whole Shape	Head Shape	COH Shape	ADD Shape
2a	1.0000	1.0000	1.0000	1.0000
2c	1.0000	1.0000	1.0000	0.3735
5b	0.6480	0.5439	0.5977	0.1709
5e	0.6480	0.5439	0.5977	0.1636

Table IV.2.1 The shape similarity information for several electron density representations created from the molecules presented in Figure IV.1.1. Whole, Head, COH and ADD denote the whole electron density, head group electron density, COH group electron density, and COH absolute difference electron density representations respectively.

In the Table, the various shape similarity comparisons of the four different electron density representations calculated for each molecule are given. The whole shape refers to the electron density calculated for the whole molecule, while the other comparisons are for fragmentary electron densities consisting of the head group (the substituted ring in the molecules), the COH group, and the absolute difference densities created by aligning and calculating absolute difference density values for the COH groups.

As can be seen in the data, the mirror image pair compounds do give the same shape similarity values relative to the electron density shape features of molecule 2a, as would be expected from the theory of the Shape Group method. However, the shape similarity measures of the absolute difference density representations do show differences that can be used to distinguish the mirror image pairs from each other. It was with these four separate shape features of the molecules that a successful correlation of shape features to activity (scent) was made.[61]

IV.2.2 Disadvantages of the Absolute Difference Density Shape Group Method

As mentioned in the Introduction, several disadvantages were noted with the absolute difference density method that made it awkward to use.

The first of these disadvantages is the alignment procedure required to allow for the point-by-point calculation of the absolute difference density. Such an alignment requires a specific rule set to be defined for the alignment. This is not a great disadvantage, as a similar requirement exists for the array-point handedness method. However, small differences would probably arise through the choice of a different alignment rule set.

A secondary problem associated with the alignment procedure is that it limits fragment sizes that can be compared. For instance, in the olfactophore study, the COH fragment was a common fragment for all the molecules. Because the fragment consisted of three atoms, a coplanar (and therefore, easily aligned) arrangement existed. Such alignment rule sets could easily be defined for up to five atoms, consisting of a stereocentre, and the atoms bonded to it, following some priority rule set, such as the CIP convention. For a larger number of atoms, though, a simple rule set cannot be created. However, because of the proof of the holographic electron density theory,[89,163-165] this need not be a great concern. Simply, the theory states that a non-zero volume of electron density must contain all of the information of the whole electron density representation from which it is obtained. Essentially, since the electron density is a direct representation of the square of the molecular wavefunction (which must contain all information about the molecule), the whole molecule electron density must also contain the information of the molecule. Any piece of the electron density then reflects this whole-molecule electron density (and the information contained within it) based on the electrons and their interaction with each other. Overall, this means the whole molecule chirality information should be present in a fragmentary electron density representation obtained from the whole molecule.

A second disadvantage of the method is the need for a common fragment in all of the molecules of interest. If such a common fragment is not present, a single atom must often be chosen to be the common fragment. Such small fragments do present challenges, the first of which is alignment. Without commonalities,

alignment becomes very arbitrary. The second challenge is that of error. A smaller defined fragment necessarily consists of a smaller number of meaningful points in the electron density file. Therefore any errors associated with the method are magnified. To combat this often requires a finer resolution array to be used, which in turn increases computational, memory, and storage costs of the method. Therefore, while the holographic electron density demands that all of the chirality information of a molecule must be contained in a non-zero volume electron density fragment derived from the molecule, larger fragments and finer resolution arrays are often required for this information to become apparent.

The third disadvantage of the method is the lack of fixed differentiation of achiral molecules from chiral molecules. Because of the theory of the Shape Group method, if an achiral molecule is chosen as the source of the reference electron density used to create the absolute difference densities, then no differentiation of mirror image forms can be made, because the resultant absolute difference densities of mirror image pairs relative to the reference will also be mirror images of one another, and will have the same shape. Essentially, then, the reference molecule must be chiral. Additionally, with a chiral reference molecule, a set of molecules with both chiral and achiral members has no guarantee of assigning a common similarity value to all of the achiral members of the molecule set, indicating their common achirality. Therefore, unlike the array point handedness based chirality measures, there is no continuous scale to provide the chirality information, and no transition from negative numbers to positive numbers, with zero as the indicator of achirality.

The fourth disadvantage is the reference-based nature of the method. In the Shape Group method, the reference shape comes from the reference molecule. In the absolute difference density method, the reference is the electron density representation from the standard molecule subtracted from itself. Thus, the reference shape is empty space, which has no shape features. Therefore, the commonalities $\Delta(X,Y,a,b)$ in the (a,b)-maps used in the shape similarity comparisons given by Equation II.1.9 only exist in cases where there is no difference density isocontour at a given threshold value a, or in cases where the isodensity contours are at every point convex compared to the reference curvature, and are therefore completely truncated before shape analysis begins. This is what leads to very small shape similarity numbers for absolute difference density shape comparisons, as evidenced by the data in the ADD Shape column of Table IV.2.1. Such small similarity numbers possibly do not have the robustness of range that would identify more than the coarsest of shape information.

Additionally, the very complex shapes made through the absolute difference density method exacerbate the situation, where contours embedded within contours can be found due to the absolute value of the density difference being taken. Difference densities can also result in cusp regions for which calculating the gradient vector and the Hessian matrix for curvature determination are difficult and computationally slow, as well as error-prone. As a result, the shape analysis of an absolute difference density file can take up to five times as long as the shape analysis of one of the electron density representations from which the difference density is calculated. On fast computers with small fragments of middling resolution,

this is not a great problem, but for larger fragments, or very fine-grid electron density representations, this could become prohibitive.

But specifically, since it is the absolute value arithmetic of the absolute difference density method that results in this complex shape difficulty, a computational time and reduced error benefit should be obtained if difference densities are used in shape analysis, as opposed to absolute difference densities.

IV.3 The Difference Density Shape Group Method

The difference density Shape Group method suffers from all of the deficiencies mentioned for the absolute difference density method, with the exception of some of the complex shapes, and the contour-in-contour problems. As this is the major flaw of the absolute difference density method, resolving this problem should result in a more robust Shape Group method complementary chirality measure.

The reason the absolute difference density method was the first chirality measure considered was that the software developed for shape analysis relies on electron density information to take on positive values. The simplest solution was to manipulate the data to allow for the application of the method, rather than to manipulate the method to fit the most convenient type of data. However, after a detailed analysis of the Shape Group method programs was undertaken, the adjustment of the Shape Group method through programming change to allow for analysis of difference densities became feasible.

In the absolute difference density method, alignment of the molecules is required for the application of the method. This is also the case for the difference density method. Such alignment allows for the point-by-point calculation of

difference density properties. Consider Figure IV.3.1, where two amino acids, R-phenylalanine and R-serine are presented.

In the Figure, phenylalanine and serine are aligned to allow for the calculation of the difference density representations. In the alignment, the stereocentre is placed at the origin of the space. Then, by following the CIP convention of assigning priority, the positive x-axis of the space is defined so as to contain the bond between the stereocentre and the amino nitrogen. In terms of the Figure, this x-axis is perpendicular to the page. The xy-plane is defined by this x-axis and the bond between the stereocentre and the carboxylic carbon. Finally, the positive z-axis is determined by the placement of the remaining stereocentre-β-carbon bond such that the β-carbon has a positive z-coordinate.

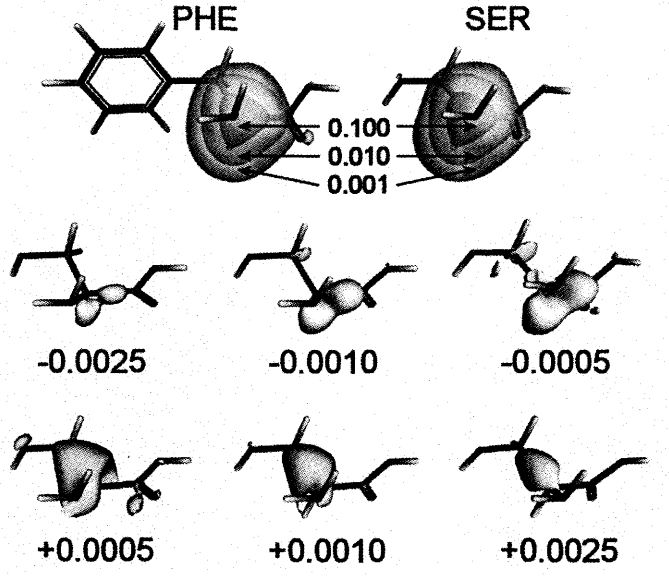


Figure IV.3.1 An example of difference density isocontours of the stereocentre of phenylalanine subtracted from that of serine. Various stereocentre fragmentary electron density contours are given at the top, while the various difference density contours are given at the bottom. All numbers give isocontour values in $e^-/bohr^3$.

With the alignment done, the stereocentre fragmentary electron density can be calculated using the *Gaussian98* and *Rhocalc* programs. Three electron density

isocontours are shown in the Figure for each of the amino acids. It can be noted visually that the isocontours show some differences in shape, reflecting their relative chirality differences.

By utilizing a simple Fortran program, it is possible to subtract point-by-point, the electron density values of phenylalanine from those of serine, to create the difference density file for which sample isocontours are given in the lower part of the Figure.

Some general notes can be made. First, as the magnitude of the difference density contour value is decreased, the general trend is for the isocontours to become larger. This intuitively makes sense, as minor differences between the fragmentary electron densities are more likely to occur away from the nuclei, where both nuclear charge effects and calculated electron density values are lower. Second, the absolute difference density method would create isocontour pictures that are the additive pictures of the isocontours with the same magnitude. In the case of the isocontours of magnitude 0.005 $e^-/bohr^3$ in the Figure, this can lead to complex contour-in-contour shapes located around the β-carbon of the amino acid. The difference density method avoids this problem, and should create shapes that are not as complex as those created by the absolute difference density method. A third note is that if the difference density file was instead created by subtracting the electron density of the serine stereocentre from that of the phenylalanine stereocentre, the isocontours would remain the same, both visually and in terms of shape, but the signs of the isocontours would be opposite to those indicated in the Figure.

Once the difference densities are created, the gradient vectors and the eigenvalues of the Hessian matrix can be calculated using the existing in-house software created for this purpose for the Shape Group method. The only change in applying the Shape Group method to difference densities as opposed to applying it for the absolute difference density method is during the stage of the creation of the (a,b)-maps.

In normal shape group work, as discussed in Section II.1, an electron density representation is analyzed at forty-one separate isodensity contours G(a), each of which are compared at all points on the surface by one of twenty-one reference curvatures b. This results in a 861-member array of shape codes called the (a,b)-map of the electron density representation. The (a,b)-maps of different representations can then be compared point-by-point using Equations II.1.8 and II.1.9 to give a single number measuring the similarity of the shapes.

In the difference density method, the isocontours can take on both positive and negative numbers, and so, the creation of the (a,b) maps must take this into account. To accomplish this, the in-house software was changed to create (a,b)-maps for difference densities that represent the shape codes of eighty-two separate (half positive and half negative) isodensity contour values G(a), where 10^{-3} e^-/bohr^3 $\leq$ $|a| \leq 10^{-1}$ e^-/bohr^3. Each of these isocontours was then analyzed compared to one of the twenty-one reference curvatures, to give a 2 x 41 x 21 member array of 1722 distinct shape codes in the (a,b)-map. These (a,b)-maps can then be compared through an altered version of equation II.1.9,

$$S(X,Y) = \sum_{j=1}^{2}\sum_{k=1}^{N_a}\sum_{l=1}^{N_b}\Delta(X,Y,a_{jk},b_l) \Big/ 2N_aN_b \qquad (IV.3.1)$$

resulting in a single shape similarity value comparing any two given difference density representations.

The use of Equation IV.3.1 resolves some of the difficulties encountered by the absolute difference density method. Because there are a larger number of array members considered in the comparisons, as well as generally less complex shapes, the shape similarity numbers of difference density representations tend to have a greater range than those of absolute difference density comparisons, and are better able to represent chirality information. Additionally, the computational effort required is roughly only double that as compared to normal shape group work, as opposed to five times or more.

Table IV.3.1 shows some shape similarity and difference density shape similarity comparisons for the electron density representations of the stereocentres of some amino acids in their R forms (except for glycine) relative to R-histidine. In the Table the stereocentre electron density shape comparisons are given as "Shape", while "DDS" refers to the difference density shape comparisons. The amino acids compared are histidine, alanine, aspartic acid, glutamic acid, glycine, leucine, lysine, methionine, phenylalanine, tryptophan, serine, and valine. Additionally, the electron density representations were aligned as shown in Figure IV.3.1 and calculated for an array point grid size (resolution) of 0.200 bohr.

When the data of Table IV.3.1 are compared with the data of Table IV.2.1, the ability of the difference density method to provide a greater range of comparison of shapes than the absolute difference density method becomes more apparent. While some of this range is most likely attributable to the fact that the original electron

density shapes in Table IV.3.1 are more similar to the reference shape to begin with, the method itself likely also gives more meaning to the potential range of the difference density shape similarity values.

Name	Shape	DDS	Name	Shape	DDS	Name	Shape	DDS
HIS	1.0000	1.0000	GLY	0.6769	0.6480	PHE	0.9160	0.9112
ALA	0.7722	0.8164	LEU	0.8485	0.8308	TRP	0.9306	0.8644
ASP	0.7931	0.8372	LYS	0.8357	0.8386	SER	0.8669	0.7699
GLU	0.8224	0.8105	MET	0.8261	0.8600	VAL	0.8407	0.7714

Table IV.3.1 Shape similarity and difference density shape similarity comparisons for the stereocentres (α-carbon) of twelve amino acids. The reference stereocentre shape is that obtained from R-histidine. Electron density representations were calculated for a grid size of 0.200 bohr.

Table IV.3.1 also highlights some of the potential deficiencies of difference density shape methods. For instance, glycine is the amino acid least similar in shape to histidine. While there are several factors responsible for this, one of the more important is the chirality of histidine as opposed to the achirality of glycine. Another important factor is that glycine is the only amino acid without a β-carbon, which contributes significantly to the stereocentre fragmentary electron density, regardless of its achirality. Since this reasoning is justifiable, the expectation is that glycine should be the least similar in shape and difference density shape.

The molecule most likely to be nearest to glycine (or second farthest from histidine) in terms of shape and difference density shape is alanine. Because of the very pseudo-symmetric environment of the β-carbon in alanine, and the lack of any major electronegativity differences between the stereocentre and the β-carbon, the expectation should be that alanine has the next lowest shape and difference density shape similarity numbers. While this does turn out to be true for the shape similarity number, it is not true for the difference density value, as the comparisons in the

Table show that the alanine difference density shape is more similar to the reference shape (empty space) than are those for serine or valine, even though the original electron density representations of those amino acids are more similar to histidine. This would imply that these amino acids are less chiral than alanine, which would appear not to be true based on logical reasoning. For instance, alanine is only different from glycine in that a hydrogen atom has been replaced by a methyl group. Because the methyl group can be considered to be closer in size, and more symmetric compared to hydrogen than the -CH_2OH group of serine or the *sec*-butyl group of valine, it can be reasonably expected that the chirality of alanine can be described as "closer" to that of glycine than either serine or valine would be.

Difference density based chirality measures are therefore considered to be complementary to the Shape Group method, and are not suitable to be used as an indicator of chirality because they only tend to deconvolute and enhance the relative chirality information between the original shapes, as opposed to giving a full description of the chirality.

V APPLICATION AND COMPARISON OF THE CHIRALITY MEASURES COMPLEMENTARY TO THE SHAPE GROUP METHOD

In this Section, the Shape Group method and the Shape Group complementary chirality measures are used in studies to correlate optical activity data of differing sorts for three classes of molecules. Through the shape and chirality measures of the fragmentary electron density of the stereocentre in the molecules, simple correlation equations can be shown between the features of the electron density and the magnitude of specific optical rotation for amino acids, the magnitude of molar optical rotations for trisubstituted 2,2'-spirobiindanes, and, as well, the magnitudes of the rotational strengths of methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives.

For each class of molecule, the study is presented in several parts. The first part will be a short discussion of the methods and results of a "motivating study." This is a work by another group in which correlations between experimental data and molecular features are made. Such works offer collected data from various sources, and justify the expectation that such correlations exist, because others have offered results that support the idea.

Results of attempted correlation of the shape features and each type of chirality measure of the stereocentre fragmentary electron density to the experimental data will then be separately discussed, followed by a discussion of the results of the differing methods compared to each other.

V.1 Specific Optical Rotations of Amino Acids

V.1.1 Carbó's Quantum Similarity Measures

In the latter part of the 1990's the Carbó-Dorca group, along with visiting scientist, P.G. Mezey, studied the quantum chirality measures of amino acids in an attempt to correlate the QSM data of mirror image compounds with specific rotation data for amino acids,[89] partially in the hope of gaining some insight on the applicability of the holographic electron density theorem[163-165] to a real-world example that was calculable at the time.

By utilizing fuzzy fragmentation techniques similar to those presented in Section II.4, and quantum self-similarity measures (Equation II.3.3) based on the Dirac delta function [$\delta(r_1-r_2)$], the similarity of the overlap of the mirror image forms of fragments of the amino acids were studied.

Eight amino acid molecules were involved in the study: lysine, glutamic acid, leucine, methionine, serine, valine, aspartic acid, and alanine. These molecules were geometry optimized, in their non-ionic forms, to a local energy minimum with the 6-31G* basis set using an earlier version of *Gaussian*.[147] The geometry optimization results can be easily shown not to be at the global energy minimum, because, if the optimizations are performed with the same basis set from starting geometries considered in other papers[110,158-162] the final geometry often has a lower energy than the geometry used by the Carbó group. However, more important in this case was the concept of conformational equivalency of common substituents. In each of the molecules in the study, the amino group and the carboxylic group, as well as the α-carbon hydrogen atom, have the same local conformational profile,

though they are not actually constrained to meet this requirement. This conformational equivalency provided two distinct advantages. First, any conformational variability came only from the final substituent on the α-carbon. In terms of chirality, this meant the contributions to the molecular chirality of all but this final substituent should be nearly identical in all the molecules due to the nearly identical conformations. Secondly, in the specific molecular conformations considered, there was no possibility of the final substituent interacting with the amino or carboxylic groups via intramolecular hydrogen bonds, which in turn reduced potential conformation and energy problems introduced by this factor.

Figure V.1.1 shows the structures and the conformations of the eight amino acids considered.

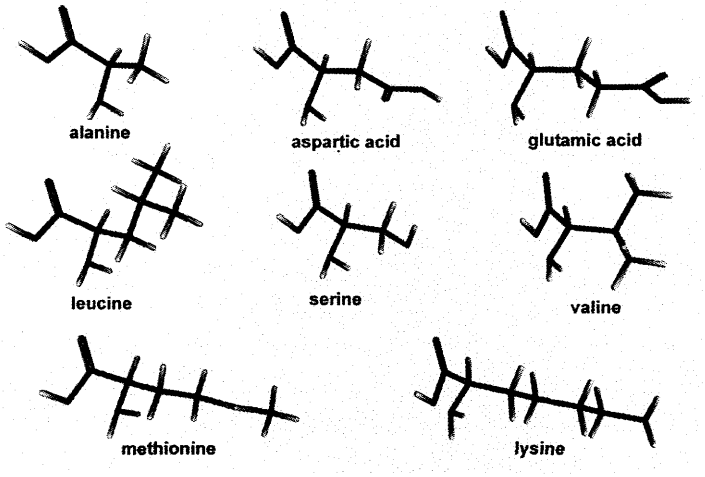


Figure V.1.1 The eight amino acids of the Carbó Quantum Similarity study.

In the Figure, the carboxyl group of each amino acid is on the left, while the hydrogen and amino groups attached to the α-carbon lie above and below the stereocentre. This leaves the final substituent to trail off to the right of the stereocentre.

In the study, two fragments derived from the molecules were considered, namely the stereocentre, and the three-atom fragment consisting of the stereocentre, the amino nitrogen, and the carbonyl carbon (called the NCC fragment). Since, in both cases, there were a small number of atoms in the fragments, superimposition of the atoms in mirror image forms could be done exactly. Once this superimposition of the mirror image fragment forms was performed, the quantum self-similarity measures (Z_{RR} and Z_{SS}), and the overlap measure Z_{RS} were calculated. The chirality of the fragment was then described via the Euclidean distance D_{RS} as defined by Equation II.3.4. Attempts were then made to correlate the chirality information of the amino acids with the magnitudes of the specific optical rotation $[\alpha]_D$ in aqueous solution at the sodium D line (589.3 nm), as found in the 63rd edition of the CRC Handbook of Chemistry and Physics.[71]

In the study, the group was able to establish correlations between the Euclidean distances quantifying the chirality of the electron density fragments and the magnitude of the specific rotations for the amino acids. The stereocentre fragments provided a correlation with an R^2 value of 0.781, while the NCC fragments provided a correlation with an R^2 of 0.832. In the case of the carbon fragment, the chirality of the molecule is modelled through the chirality of the stereocentre, providing a description of the chiral environment to which the chromophore is subjected. The NCC fragment, being larger and more able to represent the complexities of the whole molecule, might be expected to provide a better correlation. In either case, the holographic electron density theorem rationalized that the chirality of the fragments must contain the chirality information for the whole

molecule as well, and therefore such fragmentary correlations with a molecular property are to be reasonably expected.

Additionally, the study particularly mentioned the fact that deviations from a perfect correlation of the specific rotations and the quantum similarity measures could be rationally explained. The example mentioned was serine, whose ability to develop potential intramolecular hydrogen bonds due to the amino group and the β-carbon hydroxyl group tended to lead to a deviation from the correlation equation to a larger extent than the other molecules of the study.

However, a second interpretation not considered in the Carbó study, relevant to this work, is also possible. This interpretation is based on the study-specific conformations used. Of the eight amino acid molecules of the study, the conformations could be described as falling into one of three categories: staggered, angled, and local-symmetry maximized. Figure V.1.2 shows three of the amino acids of the study (lysine, valine, and serine), and the category into which each of these molecules falls.

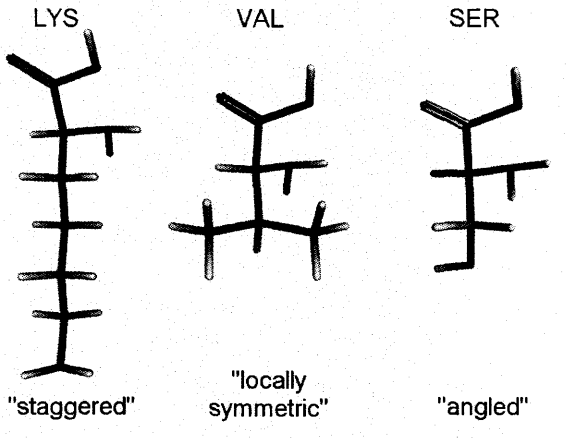


Figure V.1.2 Three of the molecules of the Carbó study defined by conformational type.

In each case the carboxylic group is found at the top, coming out of the page from the α-carbon. This leaves the amino group and hydrogen attached to the stereocentre to be "below" the page. The final substituents are seen at the bottom, starting with the β-carbon, which is above the page.

In most of the molecules of the Carbó study, the "local symmetry" around the β-carbon is maximized. This means that the groups attached to the β-carbon are conformationally arranged such that, if the α-carbon and its attached groups are considered to be a featureless sphere, then a pseudo-mirror plane containing the α-carbon-β-carbon bond and the α-carbon-carbonyl-carbon bond exists. This pseudo-mirror symmetry through the β-carbon results in valine falling into a category by itself. While all of the other molecules of the study have two hydrogen atoms bonded to the β-carbon, the valine conformation that maximizes local symmetry must place the two β-carbon methyl groups of valine as symmetrically as possible around the carbon atom to make the pseudo-mirror, leading to the "locally symmetric" category.

The difference between the "staggered" category and the "angled" category then arises from the remaining group bonded to the β-carbon. In lysine (and all the other molecules of the study, besides valine, serine, and leucine) the staggered conformation of the chain of these final substituents tends to maximize the local symmetry of the whole chain at each carbon atom, except possibly at the end farthest from the stereocentre. Serine, however, has the hydroxyl group in an energetically favoured "angled" conformation rather than placing the group in the defined pseudo-mirror plane. Therefore, because this group is directly attached to

the β-carbon, the "local symmetry" of the β-carbon in serine is reduced relative to the β-carbons of the other molecules. Finally, since this occurs relatively close spatially to the α-carbon, the α-carbon chirality has a larger "desymmetrizing" contribution from the β-carbon in serine than it does in any other of the molecules, leading to the larger deviation from the correlation that was noted. For the "staggered" molecules, such desymmetrizing effects, if they occur, occur spatially much farther away from the stereocentre, and should therefore be less noticeable.

In the one case where the local pseudo-symmetry is not maximized (leucine), the desymmetrization of the β-carbon environment has the potential to radically affect the desymmetrization of the stereocentre electron density as well.

V.1.2 Array Point Handedness Chirality Measures and Amino Acids

V.1.2.1 Carbó's Test Set

The first step in attempting to match a correlation between the magnitude of optical rotation of amino acids and the shape and chirality features as defined by the array point handedness measures was that of applying the techniques directly to the molecules as studied by Carbó. As correlations had been found with the quantum similarity measurements of the stereocentres and the optical rotation, these stereocentre fragments were chosen as the starting point.

Upon obtaining conformational information for the eight amino acids from a website cited in the paper,[89] each of the R-form amino acids was geometrically reoptimized utilizing the RHF/6-31G** basis set as opposed to the 6-31G* basis set with which the molecules had initially been optimized. This was done because it was felt that the stereocentre electron density fragment would benefit from the

added hydrogen atom polarization contribution, placing a more equal emphasis on all of the atoms that contribute orbitals to the description of the stereocentre. This reoptimization did not eliminate the conformational equivalency that had been established for the amino and carboxylic groups.

Upon the completion of the geometry reoptimization, the full electron population calculations were performed using *Gaussian98*. This was followed by the creation of the fragmentary electron density files using *Rhocalc04* at an array grid size of 0.200 bohr. These files were then analyzed with the Shape Group method software, as well as the program *Chirality*, to determine the (a,b)-map shape similarity values and the absolute array point handedness chirality measures for the stereocentre electron density fragments. Only the absolute values of the array point handedness chirality measures were required because differentiation of the mirror image forms of the same molecule was not required, only a quantification of the complementary chirality information.

Table V.1.1 shows the relevant data for this initial study.

In the Table the stereocentre shape similarity comparisons are made using the stereocentre electron density fragment from lysine as the reference, as lysine has the largest magnitude of specific rotation of the amino acid molecules considered. These shape similarity numbers themselves tell an interesting story. The three stereocentre fragments with lowest shape similarity compared to the lysine fragment come from leucine, aspartic acid, and serine. Two factors contribute to these lowered similarity numbers. The first factor is the conformation category into which each molecule falls. Serine, as noted earlier, is "angled" compared to the lysine

"staggered" conformation, and can be expected to have a desymmetrizing effect on the stereocentre electron density that is evident in the shape similarity values. The leucine stereocentre also has such desymmetrization effects because the molecular conformation used in the study is the only one in which the local symmetry around the β-carbon is not maximized by placing both hydrogen atoms in positions where the pseudo-mirror can be created.

Amino Acid	Experimental $\lvert[\alpha]_D\rvert$ (deg)	Stereocentre Shape	Absolute Chirality	Calculated $\lvert[\alpha]_D\rvert$ (Eqn. 1)	Calculated $\lvert[\alpha]_D\rvert$ (Eqn. 2)
LYS	14.6	1.0000	0.015017	14.77	12.02
GLU	11.5	0.9267	0.013505	11.43	9.74
LEU	10.8	0.8542	0.058655	9.97	7.09
MET	8.11	0.9462	0.004995	7.68	10.38
SER	6.83	0.7934	0.036054	6.10	4.49
VAL	6.42	0.8712	0.009157	7.48	7.75
ASP	4.7	0.8345	0.010374	6.00	6.29
ALA	2.7	0.8754	0.003855	2.21	7.91
A =	55.929	Equation 1 = A + B/Shape + C/(Abs. Chirality)			
B =	-38.730	R^2 = 0.958			
C =	-0.037				
A =	40.928	Equation 2 = A + B/Shape			
B =	-28.907	R^2 = 0.383			

Table V.1.1 Shape similarity values and absolute array point handedness chirality measures for the stereocentre electron density fragments obtained from the eight amino acids used in the Carbó study. Two different correlations were attempted. The first used both the shape and chirality measures, while the second used shape similarity values alone. Experimental optical rotation data are presented for the molecules in aqueous solution at approximately twenty-five degrees Celsius (+/- four degrees).

Aspartic acid, and its second carboxyl group in close proximity to the stereocentre, provides a view of the second factor of lowered shape similarity numbers in this study. In cases where electronegative elements are close to the fragment of interest, the expectation is that the fragment electron density will be drawn towards those elements, leading to differences in shape features. Therefore,

serine, which is affected by both factors mentioned, can reasonably be expected to give the smallest shape similarity to a fragment obtained from lysine, where these two factors do not exist.

The absolute array point handedness data also have interesting features. Essentially, due to the nature of the method, the expectation is that the chirality measure should be zero if a mirror plane exists, and the measure should increase as deviations from achirality occur. For single atom fragmentary electron densities, these deviations from chirality can be expressed as a combination of different factors.

The first of these factors is the notion of chirality inherent in the shape of the fragment. In an achiral electron density fragment with a mirror plane, each array point has a mirror image analogue of opposite handedness. Since each point has a mirror image pair, the net chirality measure would be zero. In a chiral fragment, such a one-to-one cancellation of points is not seen, as one portion of the electron density will give a set of points of one handedness type, while another portion might give a set of points of the other handedness. However, the number of points in each set is unlikely to be equal, leading to an abundance of one type of point, which leads to a non-zero component to the chirality measure. This is the "shape-based chirality" component of the measure.

The second of these factors, as also seen in Avnir's early chirality measures[93,94] (Section II.3.2), are the changes in the distance and relative orientation of the centre of a molecule when compared to its nearest achiral equivalent. This provides a way to categorize some of the deviations from achirality.

The array point handedness measure accomplishes this type of categorization when a standard orientation is defined. An achiral molecule, in its standard orientation, will align itself well within a space, so that the mirror planes within the molecule will coincide with the planes defined by the axes of the space (and, therefore, the array), leading to a chirality measure of zero. As deviations from achirality occur, however, the standard orientation of the molecule places the molecule in the lattice in some other inconsistent manner. For a fragmentary electron density, such as the stereocentres of amino acids, this means the placement of the fragment in the array provides some information of deviation of the centre of the molecule and the angle of change, much like Avnir's original measure. This factor can be called the "whole-molecule chirality" component of the chirality measure.

Consider Figure V.1.3, which shows the amino acids histidine and lysine in their standard orientations in the same space. In the Figure, the whole molecule and the stereocentre electron densities are shown at 0.010 $e^-/bohr^3$ and a grid size of 0.200 bohr. Additionally, points of same handedness value are grouped by colour.

In the Figure, the standard orientation of the lysine molecule aligns itself relative to one set of array planes (those parallel to the page) in such a way that the staggered carbon chain is only slightly rotated relative to the plane of the page as compared to an achiral staggered chain (such as hexane) where one could imagine one of each of the two hydrogen atoms attached to a mid-chain carbon atom lies above the page while the other lies below. The carboxyl and amino groups are the main source of this slight deviation. However, this is balanced by the pseudo-

symmetry at each of the carbon atoms of the staggered chain, resulting in a small deviation from a nearest achiral molecule.

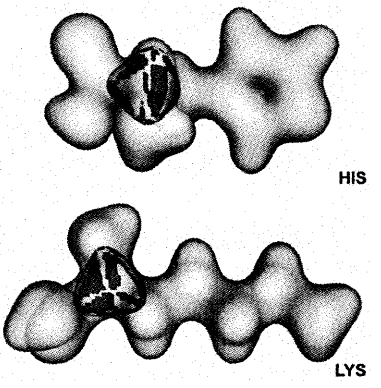


Figure V.1.3 Whole molecule and stereocentre electron densities of histidine and lysine in their standard orientations. The stereocentre electron densities are coloured by differing array point handedness domains.

Histidine in the Figure, however, would have its side chain fall into the "angled" category of conformations. Therefore, while the carboxyl and amino groups have conformational equivalency between the two molecules, it is the side chain that effectively determines the difference in orientations of the molecules in the same space. Since histidine is not as locally pseudo-symmetric, the orientation reflects this to a greater degree. Therefore, when the array point handedness determinations are made for the stereocentre electron densities the expectation is that histidine should show a larger deviation from an achiral orientation. In the Figure, this is seen as a disparity in the relative sizes of the areas of same-handedness as compared to the regions of opposite handedness. In lysine, this disparity is not visually as great.

The data in Table V.1.1 reflect this whole-molecule chirality component. For serine and leucine, the two molecules that can be reasonably described as the least pseudo-symmetric around the β-carbon due to conformational desymmetrizing

effects, the absolute chirality measures have the largest magnitudes, signalling the largest deviation from whole-molecule achirality. Alanine, with a highly pseudo-symmetric β-carbon environment shows the lowest chirality measure. The remaining five "staggered" and "locally symmetric" molecules take values in a relatively small range, with methionine and its standard orientation dominating sulphur atom falling closest to the alanine value, while lysine takes the largest value of the range.

In attempting to correlate the shape and chirality features of the stereocentres to the specific rotation data, it was decided that a simple correlation equation could be used. Other Shape Group method QShAR work[27,61] had shown success with correlating shape features and experimental data by using equations in which the shape features were inversely proportional to the calculated results.

To calculate predictive data, a simple equation (Equation 1) was set up such that the calculated specific rotation magnitude equalled A + B/Shape + C/(Abs. Chirality). The coefficients of the equation were determined by utilizing the Solver function of Microsoft Excel to maximize the R^2 value obtained by doing a linear plot of the calculated magnitudes of the specific rotation versus the experimental data, under the constraint of minimizing the sum of the squares of the differences between the experimental data values and the calculated values. The resultant R^2 value was 0.958, indicating a very favourable comparison between the calculated data and the experimental data. The calculated data using this correlation equation can be found in Table V.1.1.

The regression equation itself can be further analysed. The coefficient A (55.529) can be thought of as the "maximal" specific rotation for this set of amino acid molecules, arising from the conformational equivalence seen in all but the side chain of the molecules, as it relates to the shape of the stereocentre. This value is much like a half-cell potential in that it has no meaning in itself, but can only be utilized with the other "half-cell", the side chain, to provide meaningful results. With this in mind, the coefficient B (-38.730) represents the other "half-cell" to provide an overall picture of the contribution of the side chain to the shape of the stereocentre. This coefficient is modified inversely by the shape of the stereocentre, relative to that of the stereocentre from lysine. Additionally, the modification acts as a "correction" to coefficient A to account for the side chain "half-cell." Since lysine has the largest shape similarity value, it is the least corrected, while serine, with its lowest shape similarity measure, would have the greatest correction. The negative sign of the coefficient shows the correction occurs to bring the calculated data to a lower value from the "maximal half-cell" value.

However, when inverse shape is the only variable used for the correlation equation (Equation 2), the maximal R^2 value is found to be less than 0.4, indicating that the shape of the stereocentre, in this case, is not enough to provide a description of the specific rotation data. This arises from the comparative nature of the Shape Group method. Since the stereocentre of the lysine molecule is used as a reference, all of the shape similarity numbers have a hidden comparison of the chirality of the stereocentre it was obtained from, as it compares to the chirality of the lysine stereocentre. Effectively, in part, the chirality of the reference must be

"corrected for," because the reference is not achiral, and in part, the remaining chirality information must be separated from the shape information that does not directly arise from the chirality.

The coefficient C (-0.037) determines the base of this correction, after it is divided by the absolute chirality measure. Alanine, with its very small chirality measure, will give rise to the largest correction for chirality. Inherently, this makes sense, as we would expect alanine with its highly symmetric β-carbon environment to be less chiral than lysine. Together, this small measure and the coefficient mean that the calculated specific rotation for alanine is further corrected by almost ten degrees from the value of coefficient A, while that for leucine is corrected by about only half a degree.

Another manner in which to think of this chirality correction can be explained with the data of leucine and valine. The shape similarity numbers of the stereocentre electron densities of these molecules are nearly the same (0.8542 and 0.8712), and one could assume that they are therefore very similar in shape to each other. However, this is not necessarily true. It is the nature of the Shape Group method that the reference is assigned the highest possible value, and so this type of differentiation is not straightforward.

However, when the chirality measures of leucine and valine are considered (0.058655 and 0.009157), differentiation of their very similar shape numbers occurs, as the chirality measure hints that valine has a lesser deviation from achirality (smaller value) than does leucine. As stated previously, this larger chirality measure

for leucine can be explained by the fact that the environment around the β-carbon is not locally pseudo-symmetry maximized in the conformation used.

V.1.2.2 Conformational Changes (Pseudo-Symmetry Maximization)

The lack of local pseudo-symmetry in all conformations was perceived as a potential drawback of the Carbó study. Leucine, in the study, does not maintain the idea of conformational equivalency to its greatest extent, as the two hydrogen atoms bonded to the β-carbon are not in the pseudo-symmetric positions seen in the other "staggered" molecules. To check the effects of this lack of equivalency, the molecules of the Carbó study were geometry reoptimized from several different starting geometries. The goal of this geometry reoptimization was to find the lowest energy geometrically equivalent conformations (with similar carboxylic and amino group conformations) of the molecules of the study, under the constraint of maximizing local pseudo-symmetry around the β-carbon.

Only two molecules of the eight considered were not already at the lowest energy conformation that maintained the pseudo-symmetry requirements. These molecules were leucine and lysine. For lysine, a slightly lower energy conformation at the RHF/6-31G** level of theory (0.1 milliHartree ≈ 0.26 kJ/mol) was found by rotating the side chain amino group. For leucine, the pseudo-symmetry requirement led to a conformation that was of higher energy (0.6 milliHartree ≈ 1.56 kJ/mol) than the conformation used in the Carbó study. These energy differences show that the pseudo-symmetry maximized geometries for these two compounds are not energetically different from their Carbó equivalents so as to not be "comparable" to the remaining molecules with unchanged geometries.

Figure V.1.4 shows the Carbó study conformations, as well as the reoptimized conformations of leucine and lysine in their standard orientations. The stereocentre electron density fragments are shown at an isosurface value of 0.050 $e^-/bohr^3$ and the points are coloured by regions of common array point handedness assignment.

The Figure also provides information about the shape similarity of each stereocentre electron density fragment as compared to that obtained from the Carbó study conformation for the given molecule, as well as the absolute chirality measure. It should also be noted that the energy applies to the conformational change for the whole molecule, and not just that for the stereocentre.

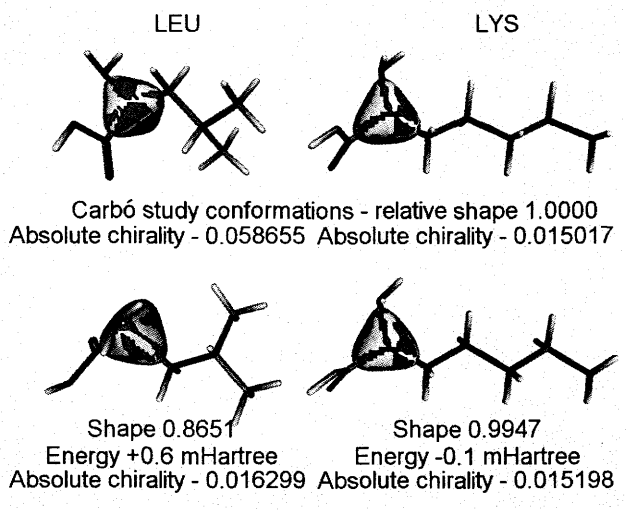


Figure V.1.4 The Carbó study conformations (top), as well as the reoptimized conformations (bottom) of leucine and lysine. Reoptimization was performed to find the lowest energy β-carbon-pseudo-symmetric conformation. The electron density isosurfaces of the stereocentre fragments are shown at 0.050 $e^-/bohr^3$ and are coloured by regions of common array point handedness assignment.

In the Figure, lysine undergoes a relatively small conformational change, as only two hydrogen atoms are effectively moved in the space. As would be expected, this change has little effect on the standard orientation, as it is defined by the principal moments of nuclear charge. Because of this, the shape of the stereocentre electron density does not appreciably change (0.9947) from the one conformation to

the other, nor does the absolute chirality measure. This can be visually verified by the fact that the regions of common array point handedness barely change in the Figure.

Leucine, however, underwent a fundamental conformational change as a result of the requirement of local pseudo-symmetry maximization. The first effect of this change is that the shape of the stereocentre electron density fragment undergoes so radical a change that it has a shape similarity of 0.8651 compared to the Carbó study conformation stereocentre fragment. This shape change is balanced by a large change in the absolute chirality measure. The reoptimized conformation has a stereocentre fragment that should be less chiral because the β-carbon local pseudo-symmetry has been maximized, and this is seen as the value of the chirality measure changes from 0.058655 to 0.016299. The movement towards a lower value indicates the fragment has become less chiral. Of secondary importance is that the reoptimized leucine conformation places the molecule in the "locally symmetric" conformation category, and gives rise to a chirality value in the same range as the other "staggered" and "locally symmetric" molecules.

With the two new conformational representations of leucine and lysine, it was possible to perform the shape similarity – chirality measure analysis again. Table V.1.2 shows the results of that study, presented in the same form as Table V.1.1.

A few points need to be made about the Table. First, even though six of the molecular conformations were not changed as compared to the Carbó study, because they were found to be the lowest energy conformationally equivalent conformations, some of the shape similarity numbers for those molecules have

changed slightly due to the change in the conformation of the reference molecule, lysine. Secondly, however, the absolute chirality measure numbers of those stereocentre electron density fragments for those six molecules did not change. Only the measures for lysine and leucine, as seen in Figure V.1.4, changed.

Amino Acid	Experimental $[\alpha]_D$ (deg)	Stereocentre Shape	Absolute Chirality	Experimental $\lvert[\alpha]_D\rvert$ (deg)	Calculated $\lvert[\alpha]_D\rvert$ (deg)
LYS	14.6	1.0000	0.015198	14.6	13.37
GLU	11.5	0.9267	0.013505	11.5	10.44
LEU	-10.8	0.9353	0.016299	10.8	11.14
MET	-8.11	0.9450	0.004995	8.11	7.57
SER	-6.83	0.7952	0.036054	6.83	5.67
VAL	6.42	0.8722	0.009157	6.42	7.14
ASP	4.7	0.8329	0.010374	4.7	5.66
ALA	2.7	0.8772	0.003855	2.7	3.10
A =	49.301	Calculated $\lvert[\alpha]_D\rvert$ = A + B/Shape + C/(Abs. Chirality)			
B =	-34.066	R^2 = 0.958			
C =	-0.028				

Table V.1.2 Shape similarity values, absolute array point handedness chirality measures, and calculated absolute specific rotation data for the stereocentre electron density fragments obtained from the eight amino acids with maximized local pseudo-symmetry geometries.

In the Table, when an inverse property correlation equation is used, and the R^2 value is maximized, as was done in Table V.1.1, the calculated absolute specific rotation data show an R^2 value of 0.958 relative to the experimental absolute data. While this value is serendipitously the same as that found for the data in Table V.1.1, it arises from a different correlation equation. The correlation equation of Table V.1.2, while it takes the same form as in the previous Table, does have notable changes in the three coefficient values.

The coefficient value of A decreased to 49.301, from a value over 55 in the previous Table. Since the shape similarity numbers and absolute chirality measure

numbers only change marginally except for leucine, this large change can be attributed almost exclusively to the conformational change in leucine. Because the new conformation of leucine can be considered to be more pseudo-symmetric than the Carbó conformation, and therefore the whole set of molecules can be considered to be "less chiral" than the Carbó set, this coefficient value, which can be thought of as the maximal absolute optical rotation for the theoretically "most chiral" conformationally equivalent molecule that could fit in the set, should be lower.

With this reasoning in mind, the expectation would be for the other two coefficients of the correlation equation to be lower as well, as less correction for the shape and chirality differences would be required because the set of molecules are "matched" better conformationally. This is confirmed as both B (-34.066) and C (-0.028) are seen to be of lesser magnitude than the same coefficient values for the Carbó set (-38.730 and -0.037 respectively).

The key to the changes of the Table, of course, is the change in data for leucine. The leucine shape similarity number is much higher than for the previous conformation, because of the maximal pseudo-symmetry requirement for the study. This, combined with the smaller value for the coefficient B, means that all the molecules, but especially leucine, have a much smaller correction to the value of A to account for the shape similarity differences of the molecules. For example, in the Carbó conformation, the "shape half-cell correction" defined by B/(Shape) has a value of approximately –45.3 degrees. In the pseudo-symmetry-maximized conformation, this shape correction value is approximately –36.4 degrees.

The chirality correction would also decrease for all molecules, except leucine, because the lower value for C combined with the unchanged absolute chirality measures results in a lower chirality correction overall. Leucine, however, sees an increase in the magnitude of the C/(Abs. Chirality) correction, from a value of approximately –0.6 degrees for the Carbó conformation, to a value of approximately –1.7 degrees with the new conformation.

Finally, while the data of Table V.1.2 seem to point out that the array point handedness method combined with the Shape Group method can accommodate some conformational change without affecting correlation discovery, there is not enough changed data in the Table to say so definitively.

V.1.2.3 Additional Amino Acids

The relative success of the studies previously mentioned led to the question of whether the method could handle a larger set of molecular specific rotation data. Sources[71,72] have specific rotation data for thirteen single stereocentre amino acids: the eight in the Carbó set, as well as arginine, cysteine, tryptophan, phenylalanine, and histidine.

The structures of the thirteen amino acids in their S-form lowest energy non-zwitterionic pseudo-symmetry-maximized conformations are given in Figure V.1.5.

As in the follow-up to the Carbó conformation study, the geometrical structures of the thirteen amino acids were optimized with *Gaussian98* utilizing the 6-31G** basis set in the RHF method to find the lowest energy conformation that maximized the pseudo-symmetry around the β-carbon (often by placing two hydrogen atoms so they "straddle" the α-carbon-β-carbon bond) while maintaining the conformational

equivalency of the carboxylic and amino groups of the α-carbon. The optimized structures were then subjected to full population analysis, and the electron density fragments of the stereocentres were calculated at a resolution of 0.200 bohr using *Rhocalc04*. Shape similarity analysis and chirality measure analysis were then performed with the relevant programs.

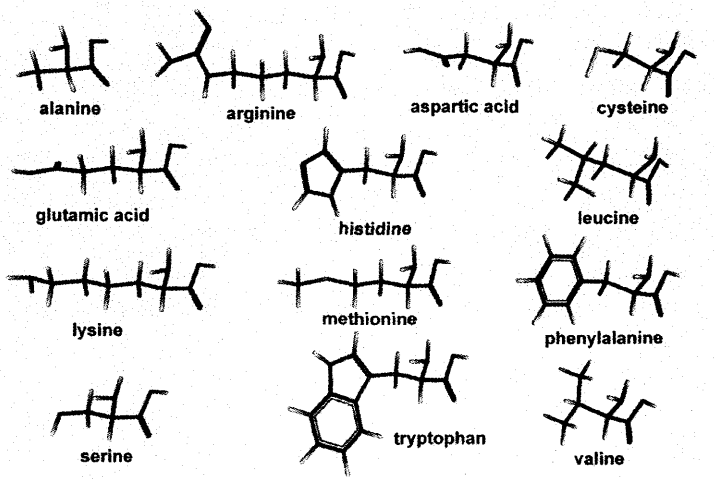


Figure V.1.5 Thirteen amino acids in their S-form lowest energy pseudo-symmetry-maximized conformations for which specific optical rotation data are available.

Table V.1.3 summarizes the data for the thirteen amino acids. The Table includes the experimental absolute specific rotation data. Additionally, the stereocentre shape and absolute chirality data are given. Finally, the calculated absolute specific rotation data, R^2, and regression coefficient data, based upon inverse-property correlation equations, are also given.

The presentation of the Table is slightly different from the previous Tables of this Section. Of the five additional amino acids added to the original Carbó set, four have absolute specific rotation values that are larger than that for lysine, which was the original study reference molecule. Therefore, the Table presents six different studies, where five different reference molecules are used. It must be noted then, that as the reference molecule changes, the shape similarity values for any given

			Realtive Shape Similarity and Calculated Optical Rotation					
			LYS		LYS inc. ARG		CYS	
Amino Acid	Experiment $\lvert[\alpha]_D\rvert$ (degrees)	Absolute Chirality	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)
HIS	39.01	0.012993						
PHE	35.14	0.015431						
TRP	31.5	0.013337						
CYS	16.5	0.026908					1.0000	16.79
LYS	14.6	0.015198	1.0000	13.37	1.0000	13.79	0.8653	10.51
ARG	12.5	0.009480			0.9679	11.44	0.8678	9.81
GLU	11.5	0.013505	0.9267	10.44	0.9267	10.67	0.8494	9.55
LEU	10.8	0.016299	0.9353	11.14	0.9353	11.40	0.8689	10.78
MET	8.11	0.004995	0.9450	7.57	0.9450	7.81	0.8531	7.13
SER	6.83	0.036054	0.7952	5.67	0.7952	5.50	0.8126	8.56
VAL	6.42	0.009157	0.8722	7.14	0.8722	7.20	0.8304	7.84
ASP	4.7	0.010374	0.8329	5.66	0.8329	5.60	0.8778	10.48
ALA	2.7	0.003855	0.8772	3.10	0.8772	3.13	0.8021	3.18
Regression		(constant) A:	49.3012		52.0951		54.0718	
Coefficients		(inverse shape) B:	-34.0660		-36.4180		-36.5148	
		(inverse chirality) C:	-0.0284		-0.0287		-0.0207	
		R^2	0.9582		0.9581		0.6203	
			TRP		PHE		HIS	
Amino Acid	Experiment $\lvert[\alpha]_D\rvert$ (degrees)	Absolute Chirality	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)	Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (degrees)
HIS	39.01	0.012993					1.0000	39.04
PHE	35.14	0.015431			1.0000	33.09	0.9160	25.94
TRP	31.5	0.013337	1.0000	30.59	0.9429	24.65	0.9306	28.53
CYS	16.5	0.026908	0.8646	12.52	0.9004	17.84	0.8571	14.90
LYS	14.6	0.015198	0.8530	10.52	0.8751	13.28	0.8357	11.24
ARG	12.5	0.009480	0.8557	10.73	0.8820	14.38	0.8394	12.72
GLU	11.5	0.013505	0.8439	9.00	0.8586	10.20	0.8224	8.68
LEU	10.8	0.016299	0.8762	14.17	0.8791	14.01	0.8485	13.69
MET	8.11	0.004995	0.8513	9.48	0.8794	13.58	0.8261	11.80
SER	6.83	0.036054	0.8625	12.24	0.8408	6.94	0.8669	16.57
VAL	6.42	0.009157	0.8621	11.72	0.8493	8.31	0.8407	13.05
ASP	4.7	0.010374	0.8122	3.47	0.8357	5.69	0.7931	2.80
ALA	2.7	0.003855	0.8080	1.79	0.8050	-1.21	0.7722	1.06
Regression		(constant) A:	147.7844		172.1445		177.9058	
Coefficients		(inverse shape) B:	-116.7674		-138.8227		-140.3106	
		(inverse chirality) C:	-0.0057		-0.0035		0.0187	
		R^2	0.8143		0.8980		0.8399	

Table V.1.3 Six separate studies involving shape similarity values and absolute array point handedness chirality measures for the stereocentre electron density fragments obtained from thirteen amino acids with energy minimized local pseudo-symmetry maximized geometries. Calculated specific rotation data are obtained with a simple inverse property correlation equation.

molecule can change, but the absolute chirality measures remain the same, regardless of reference molecule. The first study presented in the Table (LYS) is a reprise of the data from Table V.1.2, and therefore is given only for the sake of completeness. The next study incorporates the data from arginine, which has a specific rotation magnitude less than that of lysine, into the lysine referenced study to give the (LYS inc. ARG) study.

Arginine, by its structure, can be seen to be very similar to lysine in many respects. The side chain substituent on the α-carbon in arginine comprises of a four-carbon chain with two nitrogen-based groups on the fourth carbon. Lysine also has a four-carbon chain, but only has the one amino group attached to the fourth carbon. However, since this fourth carbon atom is a relatively far distance spatially from the stereocentre, the expectation would be that the difference between the two should be small in terms of the relevant shape and chirality data, much like when conformational change was performed on lysine in Figure V.1.4.

The data in the Table confirms this reasoning. First, the shape similarity value of the stereocentre in arginine is 0.9679, which is the most similar to lysine of the eight other amino acids of the set. The absolute chirality value of arginine is also in the appropriate range for "staggered" and "locally symmetric" molecules, but is lower than that for lysine. As there are two nitrogen-based groups on the fourth carbon, as opposed to two hydrogen atoms, the local pseudo-symmetry can be described as higher based on the larger contribution of these two groups to the standard orientation. Higher local pseudo-symmetry translates into a lower chirality measure.

The net result is that arginine fits relatively seamlessly into the previously established eight amino acid correlation equation. While each of the correlation coefficients has been increased slightly with the addition of arginine, the R^2 value remains essentially unchanged at 0.958, and all previously calculated specific rotation data only change by 0.5 degree.

Each of the remaining four studies adds one more amino acid to the previously used sets. The added amino acid is the next in the series of increasing specific rotation magnitude, and therefore the added amino acid becomes the new reference molecule for the entire set. As mentioned previously, this changes the shape similarity data for the molecules, but not the absolute chirality measure.

The next amino acid added to the set, bringing the total to ten amino acids, was cysteine.

The cysteine study begins to show the importance of the standard orientation to the array point handedness method. Cysteine, like serine, would fall into the "angled" conformation category. This can be seen in the absolute chirality measure for the stereocentre of cysteine (0.026908), which is seen to be larger than the range of chirality measures for "staggered" and "locally symmetric" molecules (~0.004 – 0.016) in the study. However, the value is also much less than that for serine, the other "angled"-conformation molecule (0.036054) in the ten-molecule set. The culprit in this disparity is the sulphur atom of the cysteine, due to the way it dominates the evaluation of the standard orientation. In the case of methionine, the orientation-dominating sulphur atom (due to its large nuclear charge) tends to give the stereocentre electron density a low absolute chirality measure. Equivalently, the

sulphur atom in cysteine also will have the same effect. This large charge, combined with the placement of the sulphur atom anti to the carboxyl group (and therefore in the pseudo-mirror plane between the two β-carbon hydrogen atoms) means the standard orientation of the molecule aligns itself with the array of the electron density file in a manner that would be best described as between that of "staggered" and other "angled" molecules.

Additionally, because the sulphur atom is spatially less distant from the stereocentre in cysteine than in methionine, it can affect the electron density of the stereocentre of cysteine to a greater extent. This electron density effect, combined with the differences between "angled" and "staggered" conformation molecules, can be seen in the shape similarity numbers of stereocentres relative to that from cysteine. The "most similar" molecule to cysteine, when it is used as a reference, is aspartic acid (0.8778). However, many of the other "staggered" molecules have shape similarity values that are not much different than this value. Most interestingly, serine, the other "angled" molecule, which might be expected to have a high shape similarity, actually has the second lowest shape similarity number, followed only by alanine. This large shape difference directly arises from the electron density effects on the stereocentre that come from the electronegativity differences between the sulphur atom in cysteine as compared to the oxygen in serine.

The net effect of these factors is that the cysteine study results in a much lower R^2 value (0.620) for the correlation equation than has been seen in the previous

studies, even though the correlation equation coefficients generally do not change very much from the lysine based studies.

The interpretation of these similar correlation coefficients between studies (so far) is that the shape effects and chirality effects as defined within the standard orientation show that even though there is a difference between "angled" and "staggered" conformations of the molecules, the standard orientations are still relatively the same between molecules in terms of arrangement of the carboxyl and amino groups of the α-carbon in the space. This is mostly due to the fact that the only reason a molecule is considered to be "angled" is due to the placement of a single hydrogen atom out of the pseudo-mirror plane. Figure V.1.6 visually confirms this. In the Figure, the structures of the amino acids histidine, tryptophan, phenylalanine, cysteine and lysine are presented in their standard orientations in the same space.

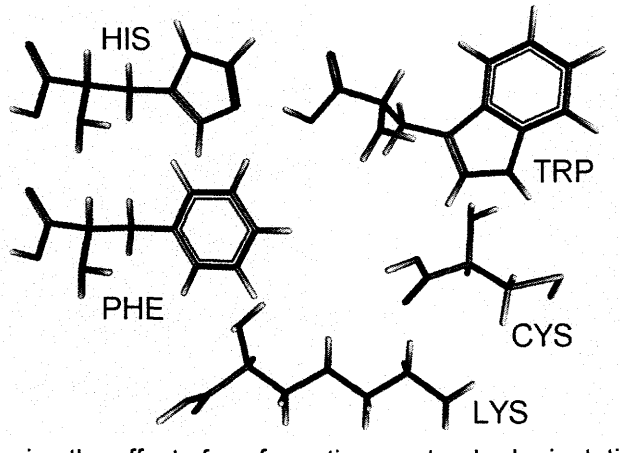


Figure V.1.6 Comparing the effect of conformation on standard orientation of five amino acids in their lowest energy pseudo-symmetry maximized conformations within the same space.

Upon examining the structures of lysine (LYS) and cysteine (CYS) in the Figure, it can be seen that the carboxyl group and amino group of cysteine show a slight rotation relative to a horizontal line through the Figure as compared to lysine.

This arises from the "angled" conformation of cysteine, and is partially responsible for the differences in chirality measure for the stereocentre electron densities of these two molecules. Since the rotation can be described as minor (as compared to the effect seen for histidine, tryptophan, and phenylalanine in terms of the standard orientation) the expectation is that the chirality measures of cysteine and lysine can be expected to be "comparable" with each other, and would result in correlation coefficients that would be little changed by the addition of the molecule cysteine to a set of amino acids. This is indeed what is seen in Table V.1.3. However, if this is true, the major rotations in standard orientations seen in the other three amino acids of the Figure might mean the results might not be as "comparable," nor the correlation coefficients generally the same if one of these amino acids were added to the set. This lack of "comparability" is implied by the experimental data for the absolute specific rotations. The first ten amino acids studied have specific rotations between approximately two and sixteen degrees. The next amino acid in the full set, the first for which this "incompatibility" might be an issue, is tryptophan, which has a specific rotation magnitude of almost thirty-two degrees, which is double that of the established range maximum value. Needless to say, this hints at a fundamental change occurring between the first set of amino acids, and the others yet to be studied.

This is indeed the case when the tryptophan (TRP) study is considered (Table V.1.3). The tryptophan conformation used is "angled", but the angle now arises from a very large double ring structure connected to the β-carbon instead of a hydroxyl group, like in serine, or an –SH group, as in cysteine, where the "angle" comes from

the placement of a lone hydrogen atom in a non-staggered manner. The double ring structure, in fact, dominates the standard orientation of the molecule, such that the carboxyl group and amino group are rotated greatly from their positions within the space relative to those from lysine or cysteine (Figure V.1.6). Such a rotation within the array of the space has the potential to affect the chirality measure based on the whole-molecule chirality component. Additionally, while the double ring structure is "angled" relative to the stereocentre, the angled portion has atoms on both sides of the pseudo-mirror plane defined by the carbonyl carbon, the stereocentre, and the β-carbon. This tends to reduce the desymmetrization effect of the β-carbon.

Since the array point handedness chirality measure is essentially partly a measure of the deviation from achirality of the molecule in its standard orientation based on how the substituents affect the orientation relative to the array space, and partly a measure of the desymmetrization of the stereocentre by the substituents, large changes in either or both factors can lead to counter-intuitive chirality measures.

This is seen in the chirality measure of tryptophan (0.013337) in the Table. This value falls into what would, in studies until this point, be called the "staggered" molecule chirality measure range. From the earlier studies, the expectation for the chirality measure of an "angled" molecule is in the range of (0.025 – 0.035). A fundamental difference has occurred, and the new amino acid, tryptophan, may not fit in well with the previous set of amino acids because, in all the previous amino acids, the dominant features have either been those attached to the α-carbon (which are conformationally equivalent in all molecules), or are "staggered" or slightly

"angled" compared to the pseudo-symmetry plane. In tryptophan, the rotation of common groups in the array space compared to the other molecules in the space is of a much different magnitude. The chirality measure may not be measuring the deviation from achirality in tryptophan compared to a similar reference point used for comparing the previous amino acids. The chirality measures for phenylalanine (0.015431) and histidine (0.012993) in the Table also show this type of radical change.

In the Table, it is seen that the correlation of experimental and calculated data in the TRP study has an R^2 value of 0.8143, and this occurs with a large change in the magnitude (relative to the LYS or CYS studies) of the correlation coefficients for the correlation equation, where coefficients A and B almost triple from the previous study, while the value of C decreases by a factor of four. Therefore, the "half-cell" specific rotation contribution (coefficient A) has been changed from that seen in previous molecules to account for the large change in positions of the common conformationally equivalent portions of the molecule in the array space. This in turn means a much larger shape "correction" (coefficient B) must take place to counteract the larger coefficient value of A. Finally, the much smaller chirality measure correction coefficient C shows the incompatibility of the chirality measures within the set. The tryptophan chirality measure "does not belong" in the set and therefore the overall contribution of chirality measures in the correlation must be reduced. This can be seen with alanine. Alanine, with the smallest chirality measure, would have the largest chirality correction to the specific rotation as defined by C/(chirality measure). In this study, that value would be approximately -1.5 degrees (~ -

0.0057/0.003855), while in many of the previous studies, this value would be more like negative six degrees.

However, the saving grace in the whole study is that the inherent chirality within the reference shape (tryptophan in this case) must have a noticeable effect on the shape similarity numbers of the other molecules, based on their differences in chirality, as shape alone almost accounts for the correlation R^2 value that is seen. This trend is continued when the amino acid phenylalanine is added to the set.

In the phenylalanine referenced study (Table V.1.3), the correlation equation coefficients A and B continue to increase in magnitude compared to the previous study, while the chirality measure correction value C (-0.0035) becomes even less of a factor in the correlation, indicating the mismatch within the set of the chirality measures as they are influenced by the standard orientations of the molecule. However, the shape similarity numbers, and their implied chirality information relative to the reference molecule (phenylalanine) manage to convey enough information to provide a surprisingly good correlation of the features of the stereocentre electron density to the absolute specific rotation data, as the R^2 value for the correlation equation calculated data compared to the experimental data in this set is 0.8980.

Finally, the addition of the last amino acid (histidine) to the set continues the trends seen in the correlation coefficients, with the exception of coefficient C, where the magnitude increases (0.0187) over the phenylalanine-referenced study. However, this magnitude increase is coupled with a change in sign, meaning that the correction now "adds back" some optical rotation magnitude instead of a correction

by "subtracting." In fact, a look at the changes in the value of C from the LYS study to the HIS study show that a consistent change of C towards a more positive value occurs as new amino acid data are added to the set.

This change can be said to occur in a linear fashion, because when a graph is plotted of the values of C for a study versus the reference amino acid absolute specific rotation angle, a linear least squares fit of the data gives an R^2 value of 0.87. This is shown in Figure V.1.7, where the coefficient values of A, B, and C (diamond, triangles, and squares) are plotted versus the reference molecule absolute specific rotation for five of the studies listed in Table V.1.3, with the (LYS) and (LYS inc. ARG) studies being considered to be essentially equivalent. In the Figure, a black line indicates the best least squares fit for the data of each coefficient.

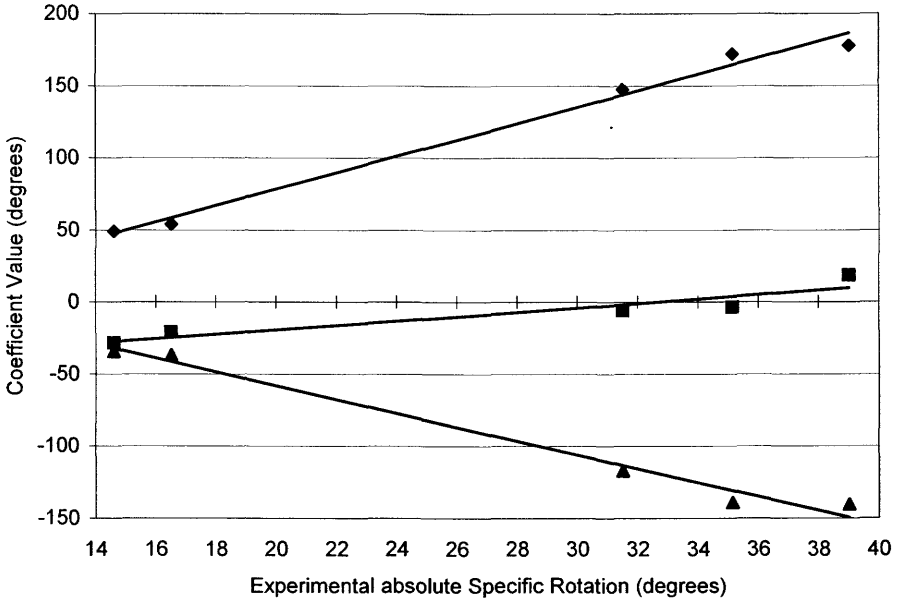


Figure V.1.7 The correlation coefficients A (diamonds), B (triangles), and C (squares – multiplied by 1000 for convenience), plotted versus the reference molecule experimental absolute specific rotation data (in degrees) for five of the studies in Table V.1.3. The black lines in each set indicate the best-fit linear least squares line. The R^2 values for the correlations represented by these lines are 0.99, 0.98, and 0.87 respectively for coefficients A, B, and C.

As is seen in the Figure, the changes of the values of A and B also occur linearly compared to the absolute specific rotation data of the added amino acid, where in both cases the plot of the coefficient of the correlation equations and the specific rotation data have R^2 values greater than 0.98. These very large R^2 values tend to confirm the linear nature of the correlation, despite the relative paucity of data points. The relationship for the coefficient C values is not as well established linearly due to the dominance of the shape similarity numbers in the regression equations of the TRP and PHE studies. This dominance is seen as a near-zero value for C, indicating that the chirality correction is almost irrelevant in those studies.

A way in which to interpret the linear nature of the change of the correction from the chirality measure (coefficient C), and the other coefficients in general, is based on the concept of subsets of data. It is seen that the first ten amino acids comprise a subset of the group of thirteen amino acids. The amino acids of the subset have commonalities based on their "staggered" or "slightly angled" conformations. The relatively small differences between their conformations (and therefore, standard orientations relative to the array space) result in a set of chirality measures that are "comparable," in that the whole-molecule chirality component of the measure should be similar, so that the differences in value directly comment on deviation from achirality (shape-based). This is advantageous, because the chirality differences in the stereocentre electron densities as expressed in the shape similarity numbers alone are probably not differentiated enough to be fully seen.

This was shown in the Carbó conformation study, where shape alone could only account for a correlation equation with an R^2 value below 0.4.

The second subset consists of the remaining three amino acid molecules. When one or two of them are added to the first subset of ten, there is not a significant enough sample size to see how the large change in standard orientations between the two subsets expresses itself in the chirality measure numbers. Therefore the chirality measure numbers have little contribution to the calculated specific rotation data, as the values cannot fully distinguish between molecules from the two subsets. However, the chirality differences that are hidden within the shape similarity numbers become more apparent, so the shape alone becomes able to provide reasonable to surprisingly good correlations of the calculated and experimental data despite different reference molecules. As the subset grows from two to three molecules, the differences in chirality measures begin to become more important again, as differentiation between the molecules of the second subset can only occur again through the chirality measures. The linear trend in the coefficient C therefore suggests, that if data were known for other amino acid molecules that could be considered to fit in the second subset, the need for the chirality measure correction would again become greater to help differentiate amongst the chirality of all the molecules of the subset. Of course, this correction would still be of less importance overall, as the other coefficients change more rapidly than the chirality correction coefficient, as is evidenced by the slopes of the lines in Figure V.1.7.

V.1.3 The Difference Density Shape Group Method and Amino Acids

In Section V.1.2, the Shape Group method was complemented with array point handedness chirality values to simply correlate absolute specific rotation data of amino acids with the shape and chirality of the electron density fragments of the stereocentre of the molecules. In this Section, all the studies were performed with the same conformations of molecules, and therefore, all of the data will be presented in a similar manner, including shape similarity values for the stereocentre electron densities that are exactly the same as in the previous Section. However, in these studies, the comparative chirality of molecules is assessed through the use of the difference density Shape Group method, where the electron density representation of the reference molecule is subtracted from that of each of the other molecules to give difference density representations. The shapes of these difference densities are then compared to that of the reference subtracted from itself, effectively comparing the shapes to empty space.

The main difference between the two Sections, then, is that the difference density Shape Group method requires a second calculation of each electron density fragment in a specifically-oriented form to allow for the point-by-point subtraction of electron densities. This orientation was defined as is shown in, and explained following, Figure IV.3.1, where the positive x-axis was chosen by moving from the α-carbon (the chosen origin of the space) along the α-carbon-amino-nitrogen bond. The xy-plane was then defined by this x-axis, and the line signifying the bond between the α-carbon and the carbonyl carbon atom (the y-coordinate of the

carbonyl atom is positive). Finally, the positive z-axis was chosen dependent on the placement of the third non-hydrogen substituent on the α-carbon.

V.1.3.1 Carbó's Test Set

Table V.1.4 shows the data obtained from the amino acid molecules in conformations as were used in the Carbó study, as well as in the array point handedness study presented in Table V.1.1.

Amino Acid	Experimental $[\alpha]_D$ (deg)	Stereocentre Shape	Diff. Density Shape	Experimental $\lvert[\alpha]_D\rvert$ (deg)	Calculated $\lvert[\alpha]_D\rvert$ (deg)
LYS	14.6	1.0000	1.0000	14.6	13.37
GLU	11.5	0.9267	0.8974	11.5	10.31
LEU	-10.8	0.8542	0.8879	10.8	11.93
MET	-8.11	0.9462	0.8972	8.11	9.78
SER	-6.83	0.7934	0.7468	6.83	4.90
VAL	6.42	0.8712	0.7918	6.42	5.56
ASP	4.7	0.8345	0.7858	4.7	6.34
ALA	2.7	0.8754	0.7635	2.7	3.45
A =	32.682	Calculated $\lvert[\alpha]_D\rvert$ = A + B/Shape + C/(DD Shape)			
B =	23.492	R^2 = 0.862			
C =	-42.809				

Table V.1.4 Shape similarity values, difference density shape values, and calculated absolute specific rotation data for the stereocentre electron density fragments obtained from the eight amino acid conformations seen in the Carbó study.

The data of the Table show some interesting features and some interesting comparisons to the data of Table V.1.1 that arise from the differences between the methods for measuring the chirality, which can be discussed most effectively in terms of four separate groups. The first of these groups consists of lysine by itself. As the reference, it is set apart from the others by the nature of the Shape Group method. In the difference density method, lysine becomes the basis of comparison,

while in the array point handedness method, lysine is itself compared to the achiral ideal represented by a chirality measure of zero.

The second group of amino acids consists of valine and aspartic acid, which have similar difference density values to each other, while their array point handedness values (Table V.1.1) are relatively the same as well. Therefore, for this group, while it is possible that the difference density shape values are serendipitously the same, while the actual chirality of the molecules is different, it is not likely. In this group, then, the difference density method gives similar chirality information to the array point handedness method, despite the difference in reference points of the two methods.

The third group of serine and alanine also have similar difference density shape numbers to each other, but are seen by the array point handedness method to respectively be the second most chiral, and the least chiral molecules, indicating that the similarity in the numbers is an artefact of the Shape Group method, and mutual "chirality" similarity between the two probably does not exist. However, the shape similarity numbers of the two show a significant difference, and so the two are still distinguishable from each other. As long as a difference occurs in one or both of the two types of the shape similarity values, the possiblity of deconvoluting the chirality component information from the shape information still exists.

Problems, however, arise, in the last group, which consists of glutamic acid, leucine, and methionine. These molecules are grouped together because their difference density shape group similarity numbers are effectively the same. This similarity is interesting, though, when contrasted with the shape similarity numbers of

the stereocentres. In the shape similarity numbers, leucine is much less similar to the reference than are the other two amino acids of the group, indicating fundamental differences between the shapes must exist, and thefeore, like serine and valine from the third group, leucine is differentiable from the other two. However, the closeness of both the shape similarity and difference density shape similarity values for glutamic acid and methionine means that differentiation between the two might not take place, unlike for the array point handedness data for the same molecules from Table V.1.1. In that Table, it is seen that glutamic acid and methionine were assigned chirality measures at either end of the range for "staggered" conformation molecules, with lysine itself having a chirality measure closer to that for glutamic acid. The expectation that the difference density similarity values of methionine should be less than that for glutamic acid is not borne out by the actual difference density data and is a drawback of the Shape Group method based measures. Simply stated, similar shape similarity measures may occur that may represent true similarity between non-reference shapes, or the values may be on "different faces of the mountain," but at the "same elevation" compared to the reference "summit." Overall, this occasional lack of differentiation could be a fatal flaw for the method, as it shows the method does not measure differences in chirality of molecules *per se*, but rather it potentially provides a different view to differentiate between very similar shape similarity values.

Fortunately, in this specific study, when a simple inverse property correlation equation is applied, the difference density shape data combine with the original shape data to give an R^2 of 0.862 between the experimental and calculated absolute

specific rotation values. This correlation, while not as good as that for the array point handedness chirality measures based study, still is very much better than that achieved using the shape of the stereocentres alone, which had an R^2 value of less than 0.4 (Table V.1.1). The lack of differentiation by the difference density shape similarity values and the shape similarity values for methionine and glutamic acid, despite the chirality differences known to exist, as shown through the array point handedness method, must be the main culprit for the correlation differences.

In terms of correlation coefficients, a large difference is noted between the data of Tables V.1.1 and Table V.1.4. In the difference density study, the basic "half-cell" coefficient A represents the "minimal" specific rotation contribution from a stereocentre belonging to the set of molecules, as opposed to the "maximal" contribution as seen in Table V.1.1. This assigning of "maximal" and "minimal" is dependent on the sign of the shape similarity correction coefficient B. In the difference density study, this value is positive, and therefore shape corrections are "added" to the base value of A. In the array point handedness study, the coefficient is negative, and so the correction is subtracted. Since the corrections are being "added" in the difference density case, the original "half-cell" coefficient A has a smaller magnitude.

The chirality correction coefficient C shows the largest change between the studies, as the methods of measuring chirality are fundamentally different, and are expressed on differing scales. Therefore, in the difference density study even the reference molecule specific rotation has a correction for chirality of over forty degrees, while others have larger corrections. This is markedly different from the

array point handedness study, where alanine sees the largest correction of about ten degrees. Another manner of thinking of the coefficient C, then, in the difference density case, is less that of a chirality correction, and more of a correction for redundant shape information inherently found in both the shape and difference density shape similarity values. Such duplication of information arises from the use of the same reference for both sets of data.

V.1.3.2 Additional Amino Acids in Pseudo-Symmetry Maximized Forms

Table V.1.5 presents the difference density method equivalent of the six studies given in Table V.1.3. It should be recalled that in these studies, the lowest energy psudo-symmetry maximized conformations of the molecules were used, and each subsequent study added one more molecule to the set. In many cases this new molecule had the largest absolute specific rotation, and therefore the molecule became the new reference to which the shapes of the electron densities calculated for the other molecules was compared.

The LYS study presented in the Table is the difference density study equivalent of the array point handedness study presented in Table V.1.2. In that study, the conformations of lysine changed slightly from the Carbó study, while that for leucine changed drastically, due to the requirement for the lowest energy pseudo-symmetric conformations. All other molecular conformations remained the same. As before, the change in the conformation of the reference molecule contributed to small changes in the shape similarity numbers for most of the molecules, while leucine saw a drastic change in the shape similarity of its stereocentre electron density as compared to that of lysine.

		Realtive Shape Similarity and Calculated Optical Rotation								
		LYS			LYS inc. ARG			CYS		
Amino Acid	Experiment $\lvert[\alpha]_D\rvert$ (degrees)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)
HIS	39.01									
PHE	35.14									
TRP	31.5									
CYS	16.5							1.0000	1.0000	17.36
LYS	14.6	1.0000	1.0000	13.34	1.0000	1.0000	12.92	0.8653	0.8229	10.41
ARG	12.5				0.9679	0.9959	13.41	0.8678	0.8513	9.60
GLU	11.5	0.9267	0.8725	8.81	0.9267	0.8725	8.71	0.8494	0.8223	8.67
LEU	10.8	0.9353	0.9311	11.75	0.9353	0.9311	11.36	0.8689	0.8212	10.87
MET	8.11	0.9450	0.9111	10.45	0.9450	0.9111	10.22	0.8531	0.8589	7.69
SER	6.83	0.7952	0.7542	5.19	0.7952	0.7542	5.12	0.8126	0.7507	7.42
VAL	6.42	0.8722	0.7909	5.23	0.8722	0.7909	5.37	0.8304	0.7513	9.55
ASP	4.7	0.8329	0.7931	6.68	0.8329	0.7931	6.57	0.8778	0.8829	9.54
ALA	2.7	0.8772	0.7787	4.20	0.8772	0.7787	4.47	0.8021	0.8086	3.54
(constant) A:			33.3354			32.7788			72.2926	
(inv. CC shape) B:			24.0236			19.5226			-81.7722	
(inv. DD shape) C:			-44.0134			-39.3768			26.8440	
R^2			0.7602			0.7814			0.6101	
		TRP			PHE			HIS		
Amino Acid	Experiment $\lvert[\alpha]_D\rvert$ (degrees)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)	CC Shape Similarity	DD Shape Similarity	Calculated $\lvert[\alpha]_D\rvert$ (deg.)
HIS	39.01							1.0000	1.0000	42.35
PHE	35.14				1.0000	1.0000	34.35	0.9160	0.9112	27.81
TRP	31.5	1.0000	1.0000	31.84	0.9429	0.9692	26.15	0.9306	0.8644	26.83
CYS	16.5	0.8646	0.8805	13.35	0.9004	0.8952	18.13	0.8571	0.9320	20.89
LYS	14.6	0.8530	0.8459	10.34	0.8751	0.8765	13.64	0.8357	0.8386	11.96
ARG	12.5	0.8557	0.8597	11.33	0.8820	0.9037	15.47	0.8394	0.8493	13.25
GLU	11.5	0.8439	0.8596	9.89	0.8586	0.8312	9.59	0.8224	0.8105	7.91
LEU	10.8	0.8762	0.8520	13.36	0.8791	0.8730	14.19	0.8485	0.8308	13.38
MET	8.11	0.8513	0.8420	9.94	0.8794	0.8679	14.09	0.8261	0.8600	11.88
SER	6.83	0.8625	0.7714	7.47	0.8408	0.7707	4.53	0.8669	0.7699	11.54
VAL	6.42	0.8621	0.8036	9.24	0.8493	0.7907	6.70	0.8407	0.7714	7.78
ASP	4.7	0.8122	0.7983	2.71	0.8357	0.7961	4.51	0.7931	0.8372	4.93
ALA	2.7	0.8080	0.8902	6.68	0.8050	0.8310	-0.05	0.7722	0.8164	-0.20
(constant) A:			154.5540			179.7240			197.9260	
(inv. CC shape) B:			-87.6558			-124.2126			-107.9001	
(inv. DD shape) C:			-35.0630			-21.1646			-47.6729	
R^2			0.8865			0.9069			0.8965	

Table V.1.5 Six separate studies involving shape similarity values (CC) and difference density shape similarity values (DD) for the stereocentre electron density fragments obtained from thirteen amino acids with maximized local pseudo-symmetry geometries based upon the reference molecule that has the largest absolute specific rotation for the set. Calculated specific rotation data are obtained with a simple inverse property correlation equation.

Additionally, since lysine is used as the reference for the difference density Shape Group method similarity numbers as well, changes are seen in these values as compared to the same category of data in Table V.1.4, though again, with the exception of leucine, these changes are not very significant.

The largest change in the difference density shape numbers due to the conformational change in lysine affects the group of glutamic acid, leucine, and methionine. As opposed to the Carbó conformation based study, where these molecules essentially had very similar difference density shape similarity values, the change in the conformations has provided some further differentiation within the group. Because methionine and glutamic acid did not experience conformational change, this differentiation is due to the change of conformation of lysine. The leucine value change is therefore some combination of the effects of its own conformational change and that of lysine.

Because the lack of differentiation in this group was suspected as being the culprit for a lesser correlation for the difference density study as compared to the array point handedness study, the lower R^2 value (0.7602) for the conformational change study compared to the Carbó study is initially puzzling.

A simple explanation for this would be that the error in the Shape Group method reduced the correlation in this specific case, yet this is not a satisfying explanation, as such errors should be just as common in the Carbó version of the study. Additionally, the rather large difference in R^2 values implies something more fundamental.

A possible explanation arises from the Carbó method of maximizing the overlap of mirror image forms for quantum self-similarity studies. If the assumption is made that the Carbó study did not maximize the pseudo-symmetry of the β-carbon in all of the molecular conformations because the maximal overlap was found to be not as great, then the possibility that maximizing the pseudo-symmetry potentially

reduces differentiation of the relative chirality information of the stereocentres must be considered. Conceptually, this seems reasonable, as conformations that are made more pseudo-symmetric (as happened with leucine in this specific case) should result in a lessening of chirality. Forcing leucine through conformational change to fall in the same "chirality range" as many of the other amino acids increases the number of "relatively equal chirality shapes" that need to be differentiated from each other. It is possible that the limitations of the difference density method mean it is not robust enough to accomplish this.

The array point handedness data for leucine backs up this reasoning, as the chirality measure showed a remarkable change from a larger value to a smaller value resulting from the conformational change to maximize pseudo-symmetry. This large change comes not only from the change in shape (a localized phenomenon) due to conformational change, but also due to the change in standard orientation (a molecular phenomenon). Therefore the array point handedness method reflects conformational (and subsequent relative chirality) change through two separate processes. The difference density Shape Group method is potentially less able to pick up on this chirality change because it only looks at local shape changes resulting from conformational differences, and has no way of gathering information on how these changes affected the molecule as a whole.

Arginine was then the next molecule added to the amino acid set (LYS inc. ARG study of Table V.1.5). This led to the same type of situation as occurred when arginine was added to the set in the array point handedness study. In that case, arginine, because of its high structural similarity to lysine, fit in well with the

correlation of the LYS study with little change to the R^2 value, or the correlation coefficients. In the difference density study, this was also the case, as the R^2 value actually improved slightly (up to 0.7814 from 0.7602). And while the changes in the correlation coefficients appear to be larger than the changes seen in the array point handedness version of the study, it must be remembered that the array point handedness and difference density methods work on differing scales of numbers. However, the difference density Shape Group method still gives a lesser correlation compared to its array point handedness equivalent, further enhancing the idea that the difference density data might not be robust enough to pick up relative chirality data for sets of "comparable" molecules as well as the array point handedness measures.

The cysteine (CYS) difference density study also suffers from many of the same changes in features as the array point handedness study (Table V.1.3), resulting in a low correlation value (0.6101) that is very similar to the earlier study value (0.6203). In the array point handedness study, this was attributed to the fact that the cysteine chirality measure number fell between those of "staggered" and "angled" molecules, even though it is clearly angled itself, leading to trouble in differentiation of the molecules through the whole-molecule-based standard orientation contributions to the chirality measures. In that study, the correlation coefficients remained much the same as in the LYS study, because cysteine didn't fit the previously established correlation. The blame for the discrepancy was placed solely on the chirality measure. Since it is now seen that both studies give similar poor correlations, the shape similarity values of the stereocentre electron densities,

as the common feature to both studies, must also be a large contributor to the problem.

If the shape similarity numbers are examined more closely, it is seen that no molecule stereocentre electron density can be described as "very similar" (shape similarity values above 0.9) to that from cysteine. In the lysine-based studies this is not the case, as many of the higher magnitude specific rotation molecules have stereocentre shapes that are similar to the stereocentre from lysine. Due to the notion that shape equivalency results in activity equivalency, these "very similar" shapes are to be expected. However, when cysteine is used as the reference, the shape similarity numbers show a marked drop that is not reflected in cysteine having experimental data that is markedly different from the other molecules of the set. Either the experimental data being used is incorrect, or the electron density representation calculated for cysteine is flawed in some way, possibly due to the inadequacy of the basis set to represent the sulphur atom as well as it represents other atoms in the basis set.

In the difference density CYS study, then, this incompatibility of cysteine with the other molecules of the set is reflected in the change in correlation coefficients as compared to the lysine-based comparison studies. However, this change is better shown through a comparison of all studies, much like was given in Figure V.1.7, and so this will be discussed at a later point.

The next study added tryptophan as the reference molecule. This addition, reflected in the TRP study of Table V.1.5, also continues the trend of having no other molecular stereocentre electron density that is "very similar" to the reference. In this

case, though, the large jump in the experimental data range suggests that this should be what happens, unlike when it occurs in the CYS study. This is also confirmed as the correlation R^2 value is 0.8865, which indicates the correlation equation is calculating data from the shape features of the molecules with relatively good comparison to the experimental data.

What is more interesting in this study, though, is that the difference density method (0.8865) now does a better job than the array point handedness method (0.8143) in contributing to a good correlation, unlike the lysine-based studies, where the array point handedness method gave a better correlation. While the array point handedness method was less able to handle the differences that arose due to large changes in standard orientation within the set, that were subsequently reflected in non-intuitive chirality measure numbers, the difference density method, with its enforced orientation requirement is less affected by these changes. Quite simply, the reference-based nature of the difference density method becomes an advantage as large changes are made, while the array point handedness method is best at dealing with "comparable" molecules of a set where differentiations of small differences are more important. The most suitable manner in which to describe this is based on the commonalities of three of the groups bonded to the stereocentre. Since these common groups also have conformational equivalency, the large source of difference density features comes almost exclusively from the non-common substituent.

This is confirmed by the difference density studies of phenylalanine (PHE) and histidine (HIS) as reference molecules as well. In both these studies, an

improvement in the correlation R^2 value is seen as the difference density method is used, though in the PHE study case, it is a slight improvement.

These three studies, looked at as a group, tend to confirm the concept of "comparable" subsets of data. Within the subset of histidine, phenylalanine, and tryptophan, the "very similar" (shape similarity values over 0.9) shapes of molecules of the subset compared to each other become notable again, indicating their basic difference from the other molecular subset, and commonness to each other. Additionally, the "very similar" shape similarity number for cysteine compared to phenylalanine, as opposed to its "lesser" similarity values compared to tryptophan and histidine, hint at a fundamental uniqueness of cysteine, perhaps placing it in a subset by itself.

Regardless, it appears that the difference density method is better able to complement the Shape Group method for large data ranges with differing subsets of molecules present, while the array point handedness is best at distinguishing between molecules belonging to one subset, which have smaller differences in experimental data.

In the transition from study to study that occurs by adding a new molecule to act as the reference, changes in the correlation coefficients occur, much as they did in the array point handedness studies for the same molecules. In those studies, the changes in coefficients were found to occur essentially linearly when the coefficients were plotted versus the specific rotation magnitude of the added amino acid (Figure V.1.7).

Figure V.1.8 shows the same type of plot of correlation coefficients versus the specific rotation data for the difference density studies of Table V.1.5. In the Figure, the diamonds denote the coefficient A values, the triangles denote the coefficient B (shape correction) values, and the squares denote the C coefficient (chirality correction) values. Additionally, linear best-fit lines are shown for each set of coefficient data.

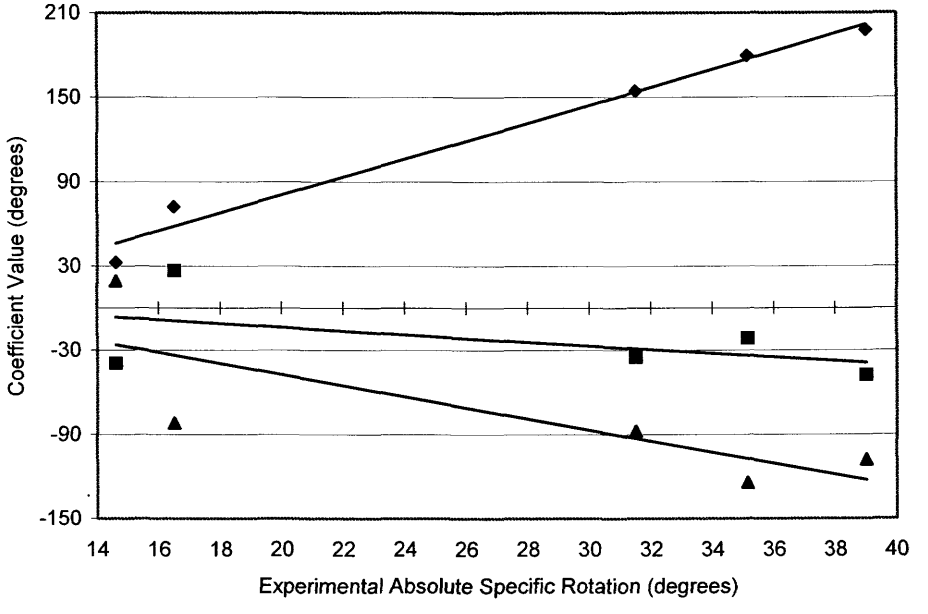


Figure V.1.8 The correlation coefficients A (diamonds), B (triangles), and C (squares), plotted versus the reference molecule experimental absolute specific rotation data (in degrees) for five of the studies in Table V.1.5. The black lines in each set indicate the best-fit linear least squares line. The R^2 values for the correlations represented by these lines are 0.98, 0.61, and 0.25 respectively for coefficients A, B, and C.

In the Figure the best-fit lines for coefficients B and C show a distinct lack of linearity, unlike what was seen in Figure V.1.7. The lack of linearity (indicated by the R^2 values of 0.61 and 0.25 for the coefficients B and C, respectively) is a function of the single-referenced nature of the correlation equation in the difference density method studies, as opposed to the array point handedness studies. Because the reference molecule is used for both shape similarities and difference density shape

similarities, the net result is that the coefficients B and C have a complementary relationship with each other. The two coefficients both tend to correct coefficient A through subtraction (leading to negative slopes in both cases) and therefore correction is done through the data (whether shape or difference density shape) that is best able to accomplish differentiation between the molecules, based on the concept of subsets of molecules. Since the best data to do this might change from study to study, the net result is that the coefficient changes are not linear.

In the array point handedness method, the chirality measure correction (coefficient C) is actually referenced to the value for an achiral molecule (zero) in all cases. Since two separate references are used between the shape and chirality data, the coefficients can work against each other in all cases, acting as two independent corrections to coefficient A (shown by slopes in different directions in Figure V.1.7 for coefficients B and C). Therefore, changes in coefficients can occur in a linear manner, as only the reference of the shape data changes.

The final item that should be noted (though it must not be taken out of context, due to the non-linearity problems in Figure V.1.8) is that the slope of the change in coefficient C has a lower magnitude than the slope for coefficient B. This was also seen in the array point handedness study plot (Figure V.1.7). Overall, this means that either relative chirality determination method becomes less relevant as the data range increases. However, as seen by the good correlation results within either set of studies, this is because the Shape Group method seems quite capable of making the differentiations in the large data ranges by itself. Therefore either chirality

determination method seems best suited for use in cases where further differentiation of shape is required to obtain all relevant information.

V.1.4 Comparison of the Chirality Measures

This Section will summarize some of the comments made in the previous Sections as it relates to comparisons made between the two methods for assessing chirality information. As was seen in those Sections, several reasons for differences between the two methods were apparent.

The first reason for the disparity in the two types of chirality measures is based upon the reference concept. In the difference density method, the reference molecule is chosen, and all shape similarity comparisons are made relative to that molecule. Since the same reference molecule is used for the shape similarity as well as the difference density shape similarity measures, the large differences between molecules will not be important, as large differences will always result in radically different shape similarity numbers. However, the small differences in electron density will be magnified by the dual role of the reference, possibly causing the small differences to be given more weight than should actually be given, resulting in "unexpected" trends, and lower correlations between experimental and calculated data when the data range is small.

The array point handedness method, however, uses the concept of achirality as the reference. The chirality measure indicates deviation from achirality, and therefore does not depend on any of the molecules directly. This detachment of the chirality measure from the reference molecule allows the approach "from both ends" of the information present in the electron density, and therefore differences in

electron density (and consequently shape) tend to be weighted based upon their magnitude.

A second difference between the two methods is based on the idea of localized versus more global chirality. In the difference density method, the chirality measure is based on how the electron density of the stereocentre differs from that of the reference (and its inherent chirality). Since all the information about the rest of the molecule is neglected (due to the orientation restriction), the chirality is only a function of the local electron density of the stereocentre. In the array point handedness method, two factors contribute to the chirality measure. If there are differences in electron density representations, this results in changes in the number of points of each handedness type, which changes the value of the measure. This is the "shape-based" chirality component of the measure. However, the use of the standard orientation places the electron density representation in the array so that the handedness of the points of the calculated electron denisty are also dependent on the neighbour array points. The standard orientation (and therfore the array point handedness method) also uses the whole molecule to help describe the chirality. This results in situations where subsets of data can be encountered in the array point handedness studies where, as long as the molecules belong to the same subset, good comparisons can be made due to the "comparability" of the molecule.

This can be seen in Figure V.1.9, which show the calculated data versus the experimental magnitudes for specific rotation for the LYS studies (lowest energy pseudo-symmetry maximized conformations) using the array point handedness method correlation equations (squares, with dashed best-fit lines) and the difference

density method correlation equations (diamonds, with solid best fit lines). The grey line in the Figure represents the experimental data plotted against itself, giving the line of perfect correlation.

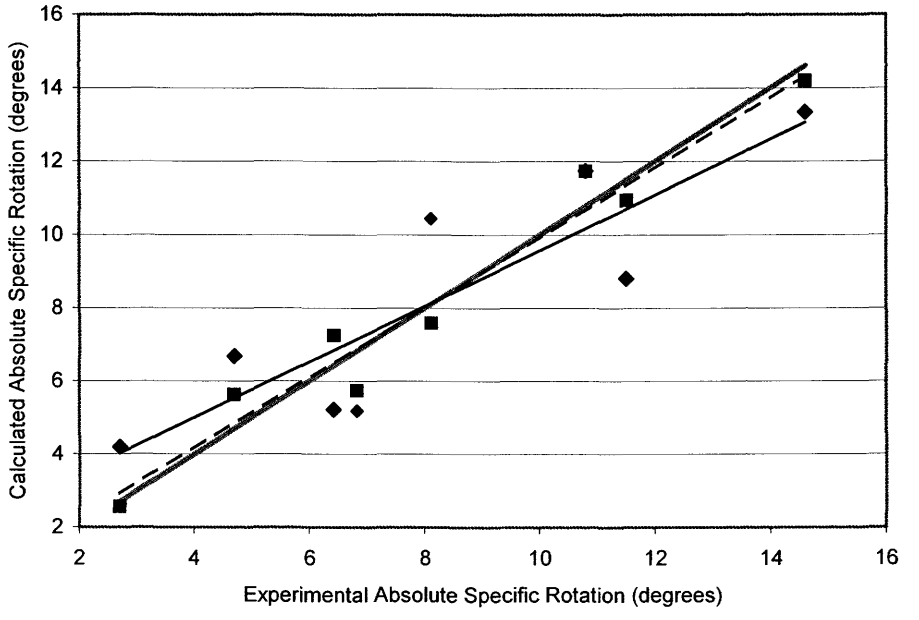


Figure V.1.9 Comparison of the calculated data versus the experimental data for the lowest energy pseudo-symmetry maximized conformation studies using lysine as the reference (LYS studies of Tables V.1.3 and V.1.5). Array point handedness data are represented by squares and the dashed best fit line, while the difference density data are presented as diamonds and the black solid line. The grey line denotes a perfect correlation where the slope equals one.

In the Figure, the best-fit lines show that the array point handedness data were better able to represent the experimental data through the simple inverse property correlation equation than was the difference density data. In this case, the Shape Group method is not enough to differentiate between the molecules because of the small range of the experimental data. Further differentiation of the shapes of the stereocentres of the molecules by either chirality method do result in an improvement of the correlation (R^2 less than 0.4 for shape alone [see Table V.1.1]), but the array point handedness method is better able to differentiate amongst the molecules because the molecules belong to a "comparable" set where the standard

orientation contributions to the chirality measure are similar for all of the molecules. This can be seen by the ability of the chirality measure to differentiate between "staggered" and "angled" conformations within the set (Table V.1.2).

The difference density method is less able to differentiate amongst the molecules because the chirality is only looked at on the local scale due to the fixed-orientation requirement of the method. Because of the lack of differentiation, the correlation is found not to be as good.

Figure V.1.10 shows the opposite case as it is seen in the histidine (HIS) referenced studies presented in Tables V.1.3 and V.1.5. In this case, the differing subsets of molecules (as can be confirmed by the "very similar" shape similarity values over 0.9 for histidine, tryptophan, and phenylalanine, as well as by the large difference in the specific rotation values between subsets) within the whole set tend to make the situation more complex than the array point handedness method can efficiently handle. It is less able to differentiate well between the relative chirality components of the molecule. This occurs because when different subsets of molecules are present, the chirality measure can only differentiate well between molecules in one set, but not necessarily in two different subsets due to differing "whole-molecule" chirality contributions. Since there are two subsets, the "comparability" of the chirality measures is not the same between sets, mostly due to the standard orientation component of the chirality measures not being similar between the subsets.

However, the Shape Group method based similarity values are better able to differentiate between subsets of molecules, because the shape differences tend to

be very large between subsets. In this case, it is found that the difference density method is better able to correlate the shape and chirality features to the experimental data. As the range of experimental data becomes greater, the need for chirality correction becomes less necessary, and so the benefit of using the difference density method over the array point handedness method is not as great as when the opposite situation occurs, as in the lysine-based study (Figure V.1.9).

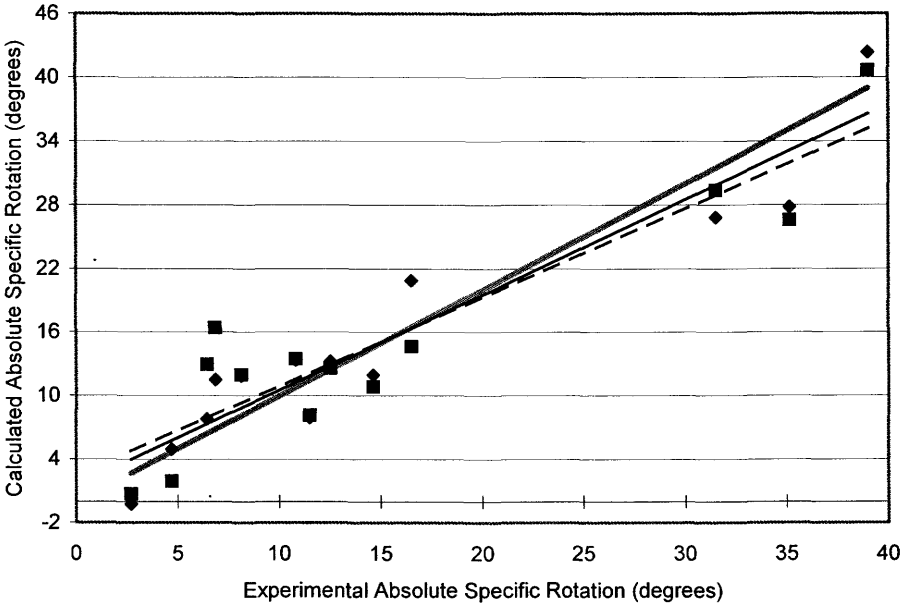


Figure V.1.10 Comparison of the calculated data versus the experimental data for the lowest energy pseudo-symmetry maximized conformation studies using histidine as the reference (HIS studies of Tables V.1.3 and V.1.5). Array point handedness data are represented by squares and the dashed best fit line, while the difference density data are presented as diamonds and the black solid line. The grey line denotes a perfect correlation where the slope equals one.

The final comparison to be made is for the cysteine-based studies. Neither method deals particularly well with the CYS studies (Tables V.1.3 and V.1.5). In fact, Figure V.1.11 shows that each method results in almost the same "best-fit line," despite significant differences in the calculated data values for many molecules. Specifically, in this case, the use of the term "best-fit line" is used for convenience

only, as no actual "linear" relationship can be inferred from the data. At best, similarities between the calculated specific rotation data for the two different methods arise solely from the use of the same shape similarity data in each case. Neither chirality method is able to extract any more useful, relevant information.

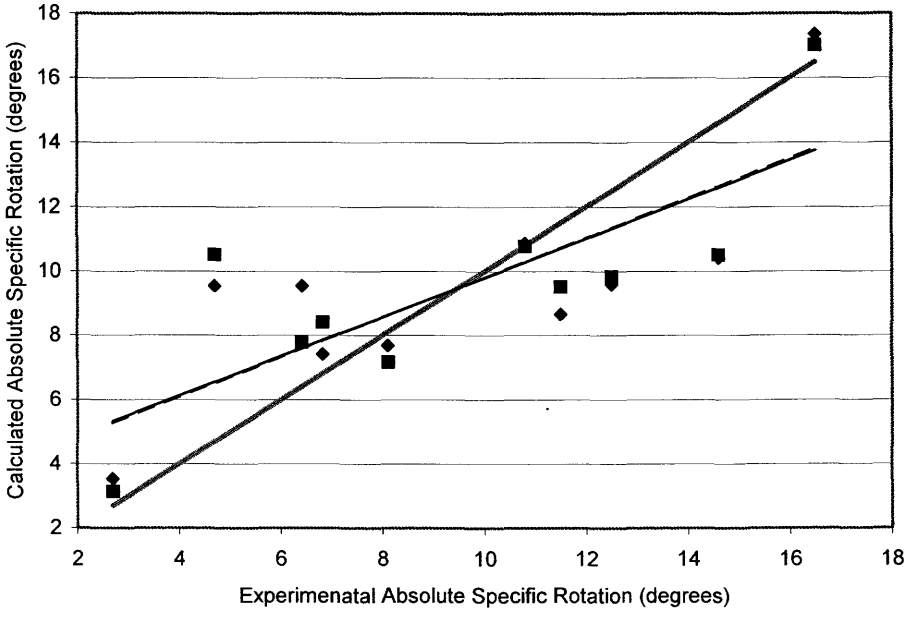


Figure V.1.11 Comparison of the calculated data versus the experimental data for the lowest energy pseudo-symmetry maximized conformation studies using cysteine as the reference (CYS studies of Tables V.1.3 and V.1.5). Array point handedness data are represented by squares and the dashed "best fit line," though no linear relationship exists, while the difference density data are presented as diamonds and the black solid "line." The grey line denotes a perfect correlation where the slope equals one.

If the subsets-of-data model is used, it is possible that cysteine could fall in a subset by itself. In Figure V.1.6, the standard orientation of cysteine (and therefore its "comparability") appears to be closer to that of lysine than it is to any of the molecules in the subset of histidine, phenylalanine, or tryptophan, indicating that cysteine should be in the lysine subset of molecules. However, when the shape similarity data of the CYS study are considered, it seems that none of the lysine-including subset of molecules is "very similar" to cysteine, indicating that it does not

necessarily belong in the LYS subset. This discrepancy either arises from the lack of the ability of the method or basis set to represent well the electron density of cysteine, or the experimental data used for cysteine is not correct in the data sources.[71,72]

V.2 Absolute Molar Optical Rotations of Trisubstituted 2,2'-spirobiindanes

V.2.1 Derflinger's Use of Pairwise Interactions

The Principle of Pairwise Interactions (PPI) can be best described as a variant of chirality function theory.[73] Within the chirality function theory, an achiral molecular skeleton is modified by the addition of ligands. Such modifications can result in a description of the properties of the molecule based upon chirality functions $\varphi(\lambda_1, \ldots, \lambda_n)$ that are dependent on ligand-specific parameters λ_i determined through previous application.

In the PPI method, all interactions arising from more than two parts of a molecule are considered not to contribute a great deal to the molecular property being modelled. As part of this process, the only interactions that are considered are those of the ligands with each other.

Consider Figure V.2.1, which shows the structure of 2,2'-spirobiindane, as well as an electron density isocontour (0.010 $e^-/bohr^3$) for the molecule geometry optimized (3-21G* basis set)[166,167] in an enforced D_{2d} group symmetry, as well as in its lowest energy C_1 symmetry conformation. In the Figure three sites are labelled as positions for ligand substitution. A fourth site, which is not labelled, could also

have a ligand attached, but for the purposes of this discussion it is assumed the fourth ligand is a hydrogen atom.

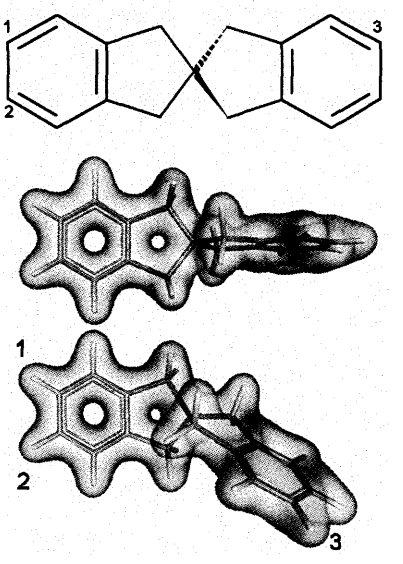


Figure V.2.1 The structure of 2,2'-spirobiindane (top) as well as electron density representations at 0.010 e^-/bohr^3 for the molecule with enforced D_{2d} symmetry (middle), as well as in its lowest energy conformation (C_1 symmetry - bottom). The numbers one to three indicate sites where different non-hydrogen atom substituents can be placed to make the central atom a stereocentre.

Within the enforced D_{2d} symmetry, an interaction between ligand pairs is non-zero by symmetry if it is not invariant under improper rotation. Therefore, if only pairwise interactions (μ_{ij}) are taken into account, the interactions between the ligands at sites one and two, and three and four, are zero. This means the entire chirality function describing the molecule consists of only four pairwise interactions,

$$\varphi = \mu_{13} + \mu_{14} + \mu_{23} + \mu_{24}. \qquad \text{(V.2.1)}$$

Such pairwise interactions can be tabulated for differing types of ligands, as long as a reference ligand is used to define the base. Hydrogen atoms are usually treated as the reference.

Derflinger,[76] using the experimentally determined molar optical rotation data (589 nm) of Neudeck, Schlögl, *et al.*,[74,75] was able to calculate molar rotation data

with a good correlation to the experimental values for many trisubstituted 2,2'-spirobiindane molecules.

Using Derflinger's collected data as a source,[76] modelling the absolute molar rotations using chirality measures to complement the Shape Group method was attempted.

V.2.2 Array Point Handedness Chirality Measures and Trisubstituted 2,2'-spirobiindanes

The Principle of Pairwise Interactions treats molecules based on how the interactions between ligands change under symmetry elements, requiring that the molecular skeleton be constrained into some chirality point group. For 2,2'-spirobiindane, this is D_{2d} symmetry. However, once ligands are attached to the molecule in actual theoretical calculations, this symmetry is lost, and therefore enforcing the skeletal symmetry seemed to be an unnecessary feature of the PPI as compared with the Shape Group method.

Calculation of trisubstituted 2,2'-spirobiindane molecules was undertaken using *Gaussian98* using the RHF/3-21G* level of theory. This "smaller" basis set (as compared to the amino acid studies) was used to compensate for the larger number of atoms in the spirobiindane molecules. Such compensation was required to make calculations possible based upon the computational resources available.

The 2,2'-spirobiindane molecule was first geometry optimized from an enforced D_{2d} symmetry. This geometry was then reoptimized under no constraint to give the lowest energy C_1 symmetry conformation. These conformations, as well as an electron isodensity contour for the whole molecule can be seen in Figure V.2.1.

Eleven molecules were chosen from the data presented by Derflinger, the structures of which are seen in Figure V.2.2. These molecules were chosen for the commonality of the -C_2H_5 group at site one (Figure V.2.1), while allowing for two subsets of ligand substitution at site two (-$COCH_3$ or-CH_3). Such commonalities were expected to be important based on the idea of conformational equivalency, a concept discussed in the amino acid study presented earlier.

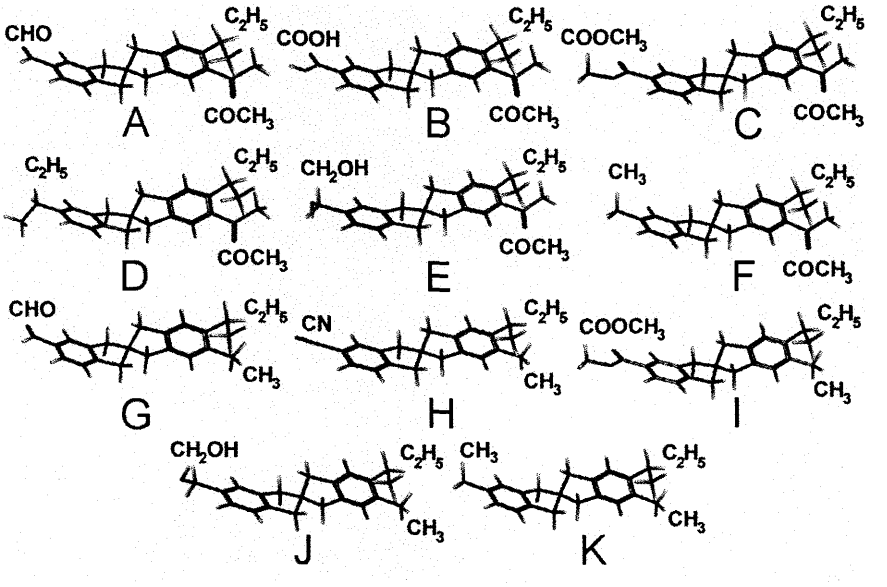


Figure V.2.2 The eleven trisubstituted 2,2'-spirobiindane structures. Each have the common substituent –C_2H_5 at site one (see Figure V.2.1), while two subsets exist where six molecules have a -$COCH_3$ group at site two, and five molecules have a –CH_3 group at site two.

The molecules were created using *Gaussian98* based on the previously calculated 2,2'-spirobiindane molecule and were geometry optimized in steps to ensure the conformational equivalency amongst the molecule set.

The first step in this process was the geometry optimization of the "site one template molecule." The original molecule in its C_1 symmetry had the site one ligand added and was subsequently reoptimized for geometry. This template was then used to create two further templates for the site two substituted ligand subsets.

Individual molecules were then geometry optimized again after the appropriate ligand was attached to site three of the appropriate template molecule.

The net result of the template-based system of creating the molecules was that conformational equivalency was maintained to the highest possible extent between molecules, as can be seen in Figure V.2.2. In the Figure, the nearly-identical conformation of the molecular skeleton is seen to be the main conformational feature of the molecules. The common ligands at sites one and two also show high conformational equivalency. Therefore, the differences in standard orientation for molecules in the same subset should almost entirely arise from the nature of the ligand attached at the third site.

In the molecules the central atom acts as a stereocentre if three different ligands are attached. In the two cases where two common ligands are involved in the same molecule (D and K in Figure V.2.2), conformational chirality through the helicity of the skeleton of the molecule still applies, and so the central atom can be considered to be a pseudo-stereocentre.

Fragmentary electron density calculations were performed using *Rhocalc04* with the 3-21G* basis set. As in the amino acid study, the stereocentres of the molecules were chosen as the fragments of interest. Shape analysis was then performed with the suite of programs established for that purpose.

Due to the nature of the molecules, in that their conformational equivalency was largely determined by the large skeletal base molecule, and the fact that all the ligands are significantly spatially distant from the stereocentre fragment, it was assumed that the Shape Group method might not be able to differentiate well

between the molecules based on the shapes of the stereocentres due to the rather small changes in electron density that were expected. The shape similarity data of Table V.2.1 somewhat bears this out, as all the stereocentre electron density shapes can be said to be "very similar" to that of molecule A, which was chosen as the reference, as it has the largest magnitude of molar optical rotation seen in the set.

	Ligands			Expt.		Calc.
Molecule	Group 1	Group 2	Group 3	$\lvert[M]_D\rvert$	Shape	$\lvert[M]_D\rvert$
A	C_2H_5	$COCH_3$	CHO	32.2	1.0000	30.76
B	C_2H_5	$COCH_3$	CO_2H	28.1	0.9964	29.24
C	C_2H_5	$COCH_3$	CO_2CH_3	27.9	0.9923	27.50
D	C_2H_5	$COCH_3$	C_2H_5	15.3	0.9566	11.66
E	C_2H_5	$COCH_3$	CH_2OH	12.8	0.9654	15.67
F	C_2H_5	$COCH_3$	CH_3	11.6	0.9495	8.37
G	C_2H_5	CH_3	CHO	20.0	0.9691	17.34
H	C_2H_5	CH_3	CN	17.0	0.9619	14.08
I	C_2H_5	CH_3	CO_2CH_3	16.0	0.9727	18.95
J	C_2H_5	CH_3	CH_2OH	7.9	0.9527	9.86
K	C_2H_5	CH_3	CH_3	5.5	0.9549	10.88

Regression coefficients:		Calculated = A + B/(Shape)
A	451.87	
B	-421.10	
R^2	0.87553	

Table V.2.1 The shape similarity measures of the electron densities of the stereocentres of the eleven trisubstituted 2,2'-spirobiindane molecules presented in Figure V.2.2. Using an inverse property relationship, a good correlation of the calculated magnitudes of molar optical rotation values $\lvert[M]_D\rvert$ can be made to experimental values.

However, in terms of differentiation, the Shape Group method seems to differentiate fairly well amongst the molecules of the set. Of greater interest was that the shape similarity numbers seemed to mirror fairly well the molar optical rotation data for the molecules. To check this, a shape-based-only inverse property relationship (as in the amino acid based studies) was attempted in the calculation of the molar rotation data.

As in the amino acid based studies, the regression equation was defined (A + B/Shape) and calculated molar rotation data were calculated based on a starting set of regression coefficients. These calculated data were then plotted versus the experimental data in Microsoft Excel and the best-fit least squares line was determined. The Solver function of the spreadsheet was then used to maximize the R^2 value of the best-fit line by changing the regression coefficients under the constraint of minimizing the sum of the distances between the calculated and experimental data points. The best-fit line regression coefficients and the calculated absolute molar optical rotations are also given in Table V.2.1.

The correlation determined by the shape similarity numbers alone was quite surprising, in that a large R^2 value was noted (0.8755). Because the shape similarities were expected to be "very similar," it was supposed that differentiation of the stereocentre electron densities would be difficult, and that the chirality of the molecules would most likely need to be determined by a complementary method. However, it appears that enough differentiation in the shapes occurs such that the chirality of the stereocentres can be determined quite well by shape alone.

In the case of the amino acid molecules, while conformational equivalency was maintained, it was not enforced through conformational restraints. Therefore, while the common groups of the amino acids were in "equivalent" conformations, this did not mean that there were not some small conformational differences between the common groups brought on by the differences in the side chain conformations. These common groups, as well as the different side chains for the amino acids, were directly bonded to the stereocentre, and therefore any conformational differences

between the common groups of the molecules would show up in the electron density of the stereocentre because the atomic orbitals of the atoms directly bonded to the stereocentre are used in determining the electron density of the fragment (Equation II.4.1). Because the electron density is affected by these changes, the shapes would also be affected in some non-predictable manner.

In the spirobiindane-based molecules, the ligands are far from the stereocentre, and the conformational equivalency was maintained to its highest possible extent by the template-based system in which the geometry optimizations were carried out. Since the same skeletal template was used as a base for all the molecules, the conformation of the molecules are probably much more similar within the set than was possible for the amino acid molecules. Any variations in conformation are likely only to arise at sites one and two (the common ligands), and since these changes are far away from the stereocentre, the shape of the stereocentre is less affected by small conformational differences. However, the overall nature of the ligands themselves must still be felt at the stereocentre to account for the shape similarity differences that are seen.

With a good correlation established between the molar rotation data and the shape of the stereocentre alone, the need for further chirality differentiation was not actually present. However, if the assumption is made that the chirality of the shape is the predominant component of the shape similarity differences seen in this set of molecules then the application of chirality differentiation methods should not really enhance the chirality information in any significant manner.

Table V.2.2 shows the shape similarity data, as well as the absolute array point handedness chirality measures for the electron density representations of the molecular stereocentres in the study. As well, a regression equation was attempted where calculated data were based on the inverse of both the shape and chirality measure data. The R^2 value for this correlation equation is 0.9129, which is a slight improvement over the value of 0.8755 found for the inverse shape-only correlation equation.

	Ligands			Expt.			Calc.
Molecule	Group 1	Group 2	Group 3	$\lvert[M]_D\rvert$	Shape	|Chirality|	$\lvert[M]_D\rvert$
A	C_2H_5	$COCH_3$	CHO	32.2	1.0000	0.001712	31.87
B	C_2H_5	$COCH_3$	CO_2H	28.1	0.9964	0.002343	30.00
C	C_2H_5	$COCH_3$	CO_2CH_3	27.9	0.9923	0.003284	27.93
D	C_2H_5	$COCH_3$	C_2H_5	15.3	0.9566	0.000380	13.10
E	C_2H_5	$COCH_3$	CH_2OH	12.8	0.9654	0.004492	14.75
F	C_2H_5	$COCH_3$	CH_3	11.6	0.9495	0.000190	12.50
G	C_2H_5	CH_3	CHO	20.0	0.9691	0.011829	16.43
H	C_2H_5	CH_3	CN	17.0	0.9619	0.009214	12.86
I	C_2H_5	CH_3	CO_2CH_3	16.0	0.9727	0.004049	18.40
J	C_2H_5	CH_3	CH_2OH	7.9	0.9527	0.000884	9.37
K	C_2H_5	CH_3	CH_3	5.5	0.9549	0.013546	9.27

Regression coefficients:		Calculated = A + B/(Shape) + C/|Chirality|
A	497.05	
B	-465.86	
C	0.00116	
R^2	0.91286	

Table V.2.2 The shape similarity measures and the array point handedness chirality measures of the electron densities of the stereocentres of the eleven trisubstituted 2,2'-spirobiindane molecules presented in Figure V.2.2. A slight improvement over the shape-only based correlation (Table V.2.1) is seen.

The correlation coefficients A, B, and C in the inverse property regression equation can be thought of in exactly the same terms as in the amino acid studies seen earlier. The coefficient A is the "half-cell" optical rotation that can be assigned to a general molecule of the set with the common features (the skeleton and the

ligand at site one), which is then corrected by the other "half-cell" which is defined by both coefficient B and the shape similarity value, as a function of the non-common groups amongst the molecules. This is the "shape correction" to the coefficient A. In the specific regression equation, these two "half-cells" combine to give calculated molar rotation values in the range of 6.41 degrees for molecule F to 31.19 degrees for molecule A. For the most part, this is not much different than the range established for the shape-only based equation, where the molar rotation values for molecules F and A are 8.37 and 30.76 degrees respectively.

While the shape-based range does not change all that much between the two, the "chirality correction" based on coefficient C can be extremely large in some cases, which was an unexpected result based on how well shape accounted for the chirality alone. The largest chirality correction occurs for molecule F, and is over four degrees, and is due to the very small chirality measure. As F was one of the molecules with the largest distance between the calculated and experimental molar rotation value in the shape-only regression equation, the added chirality correction serves to reduce this distance, thus increasing the correlation R^2 value. Molecule D also shows a large chirality correction. However, some of the molecules of the shape-only study that were not well represented in the regression calculated data did not benefit all that much by chirality correction. For instance, molecule K only sees a correction of less than a tenth of a degree, which does not improve its place in the set as compared to the shape-only study.

As stated in the discussion of the amino acid studies, the array point handedness measure combines two aspects of chirality in the measure. The first,

the local chirality of the stereocentre electron density is generally seen somewhat in the shape of the stereocentre, but often needs to be augmented to deconvolute the shape features directly related to the chirality. If this were not so, the amino acid studies would have shown very good shape-only regressions themselves. The other chirality component comes from the alignment of the stereocentre in the molecule (and space) as defined through the standard orientation. This allows information about the whole molecule to be recovered from the electron density representation.

In the trisubstituted 2,2'-spirobiindane molecule set, the shape of the stereocentre is very adept at describing a large component of the chirality on the local scale, leading to the good shape-only correlation. The chirality measure, then, can only enhance the correlation by the whole molecule component, which changes when the standard orientation changes. However, since these molecules have large common groups in the skeleton and at ligand position one, these tend to dominate the standard orientations of each molecule. Any differences in the standard orientation then must come from the identity of the ligands at sites two and three. If these ligands are "small" in terms of their effect on standard orientation determination, then their chirality measures may not be able to reflect the "whole molecule" component of the chirality because the rest of the molecule tends to dominate the standard orientation, and the smaller differences might not be seen.

Arguably, the three molecules with the "smallest" ligands at sites two and three are molecules G, H, and K. In these molecules, site two has a methyl group, and the site three locations are home to a CHO, a CN, and a methyl group respectively. These molecules also have the largest absolute chirality measures, and are

therefore subject to the least amount of chirality correction after the shape contribution to the calculated molar rotation is considered. The resultant effect is that these three molecules are the least well represented in the calculated data as compared to the experimental data, having anywhere from a 3.5 (molecule G) to a 4.2 (molecule H) degree "distance" between the two data types of molar rotation values. If the chirality measure were better able to account for these smaller differences in standard orientation, the expectation would be that the correlation would become better.

Overall, at least for this particular set of molecules, it appears that when the groups that give rise to the chirality of the molecule are spatially distant from the stereocentre, the shape of the electron density of the stereocentre is better able to describe the chirality in terms of a property such as molar optical rotation. The absolute array point handedness chirality measure is then able to augment this information somewhat with its "whole molecule" chirality component. However, as the chirality-inducing groups are found to be more distant from the stereocentre, the ability of the chirality measure to account for the whole molecule chirality diminishes because the small groups might not be able to influence a significant change in the standard orientation of the large base molecule of the set. For those molecules, the chirality correction will be minimal, possibly affecting the overall correlation.

V.2.3 The Difference Density Shape Group Method and Trisubstituted 2,2'-spirobiindanes

As in the array point handedness chirality measure case, the good shape-only correlation seen in Table V.2.1 hinted that the difference density Shape Group

method would probably not improve the correlation by any great deal, as the shapes of the stereocentres were able to reflect the chirality of the molecules.

The difference density method, as was seen in the amino acid molecule studies, requires the defining of a fixed orientation of the molecules to allow for the point-by-point subtraction of electron density files for the creation of the difference density files. In the amino acid case, this was accomplished using the Cahn-Ingold-Prelog notion of priority of groups. This worked well due to the fact that three common groups were present in all of the amino acids.

For the substituted spirobiindane molecules, however, there are only two common groups, and two subsets (each with about one half of the entire set of molecules) of a third common group. The CIP convention, it was decided, was probably not the best choice of priority assignment in this case.

Because the stereocentre is spatially distant from the sites of substitution, describing the orientation requires the assigning of labels to the bonds from the stereocentre to the nearest carbon atoms of the two five-member rings to which the stereocentre belongs. From the stereocentre to each site of substitution, there are four carbon-carbon bonds that must be traversed for the shortest through-bond distance. Since these shortest distance trips all use a different carbon atom as the first point visited after the trip starts from the stereocentre, the labels of sites one to four can be mapped onto these four different carbon atoms to allow for the orientation scheme.

Therefore, the orientation of the molecule was done so that the stereocentre was placed at the origin. As the site with a group common to all molecules, the

mapped analogue carbon for site one was chosen to lie on the x-axis of the space. This x-axis, as well as the site two analogue (the next site with common groups, if the hydrogen atom of site four is ignored) defined the xy-plane in a manner such that the site two analogue had a positive y-coordinate. The positive z-axis was then defined based on the position of the site three analogue.

Once the orientation scheme was applied, a full electron population analysis for each molecule was performed with *Gaussian98* with the *nosym* keyword to maintain orientation. Chiral centre electron density fragments were created from this population analysis with the 3-21G* basis set. Since these files were spatially compatible, the program for creating difference density files could then be used. Shape group analysis was then performed with the suite of programs developed for that purpose.

Table V.2.3, which shows the regression equation based on the shape similarity and difference density shape similarity values for the stereocentre electron densities, confirms the supposition that little improvement over the shape-only based correlation would be seen. However, there are some features of note in the Table that should be discussed.

The first item of note in the Table would be the difference density shape similarity values for molecules A, B, and C. Since A is the reference molecule, the difference density of A subtracted from its own electron density is empty space. The values of 1.0000 for the difference density shapes of B and C seem to indicate that the electron density files of B and C are exactly the same as that for A, but this can not be true, as the shape similarity numbers indicate. These "exactly similar"

difference density shape values are an artefact of the Shape Group method, which uses isodensity contour values that are given by the relation 10^{-3} e$^-$/bohr^3 $\leq |a| \leq 10^{-1}$ e$^-$/bohr^3. Therefore, while the difference densities of molecules A and B are not exactly totally empty space, any non-zero values within the array points of the file are probably of a magnitude less than 10^{-3} e$^-$/bohr^3, and are therefore not accounted for in the difference density Shape Group method. However, the "highly similar" numbers in both the shape similarity and the difference density shape similarity values indicate true "near similarity" between the molecules which is reflected in their similar molar rotations, as well as in their similar array point handedness chirality measures (Table V.2.2).

	Ligands			Expt.		Diff. Dens.	Calc.
Molecule	Group 1	Group 2	Group 3	$\lvert[M]_D\rvert$	Shape	Shape	$\lvert[M]_D\rvert$
A	C_2H_5	$COCH_3$	CHO	32.2	1.0000	1.0000	31.32
B	C_2H_5	$COCH_3$	CO_2H	28.1	0.9964	1.0000	29.41
C	C_2H_5	$COCH_3$	CO_2CH_3	27.9	0.9923	1.0000	27.21
D	C_2H_5	$COCH_3$	C_2H_5	15.3	0.9566	0.9485	12.52
E	C_2H_5	$COCH_3$	CH_2OH	12.8	0.9654	0.9707	15.25
F	C_2H_5	$COCH_3$	CH_3	11.6	0.9495	0.9482	8.41
G	C_2H_5	CH_3	CHO	20.0	0.9691	0.9829	16.11
H	C_2H_5	CH_3	CN	17.0	0.9619	0.9335	17.19
I	C_2H_5	CH_3	CO_2CH_3	16.0	0.9727	0.9890	17.53
J	C_2H_5	CH_3	CH_2OH	7.9	0.9527	0.9672	8.29
K	C_2H_5	CH_3	CH_3	5.5	0.9549	0.9529	11.06

Regression coefficients:	Calculated = A + B/(Shape) + C/(DD Shape)
A 464.654	
B -529.43	
C 96.09	
R^2 0.89842	

Table V.2.3 The shape similarity measures and the difference density shape similarity measures of the electron densities of the stereocentres of the eleven trisubstituted 2,2'-spirobiindane molecules presented in Figure V.2.2. A slight improvement over the shape-only based correlation (Table V.2.1) is seen.

Additionally, because the difference density Shape Group method only reveals the local relative chirality differences between the stereocentre electron densities, the calculated molar rotation values for molecules G, H, and K were expected to not improve compared to the array point handedness based measure study. In Table V.2.3 this is seen to be the case, even though molecule H is actually represented quite well, as the values of the calculated molar rotations for G, and especially K deviate to a greater extent from the experimental values than do the array point handedness-based calculated values. This deviation, combined with larger deviation for molecules D, E, and F, results in a lower correlation R^2 value overall.

Next, the data of the Table (as well as the data of Table V.2.2) generally show the largest relative differences between the calculated data and experimental data for the molecules with the molar rotations of the least magnitude. Partially, this is attributable to the process for obtaining the regression coefficients. The linear R^2 value was maximized under a constraint of minimizing the sum of the distances between the calculated and experimental data points. These distances are not weighted by the experimental data, and are therefore not "relative." Other factors must also be considered, including the discrete nature of the shape similarity measure (small differences between two shapes may not be "significant enough" to trigger a change in the shape similarity value), and the fact that the relative error in the experimental data probably is most likely largest for the molecules with the smallest measured values.

The final feature of the Table to discuss is that of the regression coefficients themselves. When the difference density studies of the amino acids are considered

(Table V.1.5), the studies with the best correlation results (TRP, PHE, and HIS) have a large positive value for A, a lower magnitude negative value for B, and low magnitude negative value for C. This indicated the molecular set "half-cell" component given by A was negatively corrected for by the shape similarity values, and then negatively corrected for by the chirality measures. In the spirobiindane molecule study, this is not the case. Here, the coefficient A has a magnitude that is less than that for B, meaning the "half-cell" shape correction (B/Shape) "takes away" from coefficient A to give a negative value. The chirality correction (C/DD Shape) then "adds back" to give a positive value. Since the shape-only correlation equation of Table V.2.1 gives values of A and B much like those seen in the amino acid studies (in terms of relative magnitude between A and B), the difference density shape numbers must be the root cause of this change in relative coefficient magnitudes.

A possible reason for this change stems from the concept of common groups. In the amino acid studies, the stereocentres of all the molecules have attached three common groups (carboxyl, amino, and hydrogen atom). Only the side chain starting at the β-carbon was different amongst the molecules. Major differences in the shapes of the stereocentres should then have occurred due to the influence and conformation of the side chain, and should occur in a mostly spatially localized area in the "side chain"-"stereocentre" bonding region, while minor differences would be associated with small conformational differences in the common groups and the interaction of the side chain and common groups as occurred "through" the stereocentre electron density (e.g. electron withdrawing and donating effects).

Therefore coefficient A represents the physical chiral property "half-cell" contribution of the common amino acid template (aside from the side chain) in its "most chiral possible form." The differences in the side chain are then accounted for mostly by the shape correction, while the smaller electron density differences and chirality components are accounted for within the chirality measure or difference density shape values. In the trisubstituted 2,2'-spirobiindane molecules, however, there are only two common groups within the molecule (the site one $-CH_2CH_3$ group and the site four hydrogen atom). Therefore the large shape-based differences will occur not at one, but rather at two, spatially different parts of the stereocentre electron density. Since the difference density method should pick up on these two separate regions of large differences (because the regions of difference should result in regions with notable electron density difference values) it becomes a more important indicator of how two electron density representations differ than when there is one non-common group. For instance, if two shapes are "very similar" to the reference shape, as are molecules E and H in Table V.2.3, then if they also have the same (relatively) difference density shape values, it is more likely that they are truly similar in shape to each other than if the difference density numbers differ, as they do for these molecules.

In the correlation equation, this results in a "half-cell" coefficient A that can only approximate the "most chiral possible" molecule of the set due to the lack of a third common group. The shape correction then reduces this value below zero, and then it is up to the difference density based correction to try and "add back" based on how it determines the shape differences arise from the two large regions of electron

density difference that would be expected to exist based on the differences of two non-common groups.

This points to another possible limitation in both methods of chirality determination. If there are less than two common groups attached to the stereocentre, then neither method will probably be able to establish a good correlation with a physical property because the shape similarity values and the chirality assessment value would not be able to fully deconvolute the three or four large regions of shape difference of the stereocentre. To do this would require a third or fourth measure based on the stereocentre electron density.

V.2.4 Comparison of the Chirality Measures

Based on the good shape-only correlation seen in Table V.2.1, and the data of Tables V.2.2 and V.2.3, a comparison of the chirality measures is not largely relevant in the trisubstituted 2,2'-spirobiindane molecules study. However, for the sake of completeness, Figure V.2.3 presents the calculated absolute molar rotation data plotted versus the experimental data, as was done in Figures V.1.9-V.1.11 for a few of the amino acid studies.

In the Figure, the squares and the dashed best-fit line are for the array point handedness study data, while the diamonds and the solid black line represent the difference density study data. The solid grey line gives the line of a perfect correlation.

As the shape-only regression was so successful in relating shape to chirality, it is not surprising that the best-fit lines for both study types are nearly identical, even though some differences between the calculated molar rotation values can be seen

between the methods, particularly for molecules E (experimental value of 12.8 degrees), H (17.0 degrees), and K (5.5 degrees). Two of these three molecules are those mentioned in the array point handedness study discussion as having "small" site two and three ligands, which cannot exert their influence on the standard orientation of the molecule sufficiently to help correct the chirality. This shows that the array point handedness method probably has a spatial distance from the stereocentre limit to its effectiveness in differentiating amongst different non-common groups.

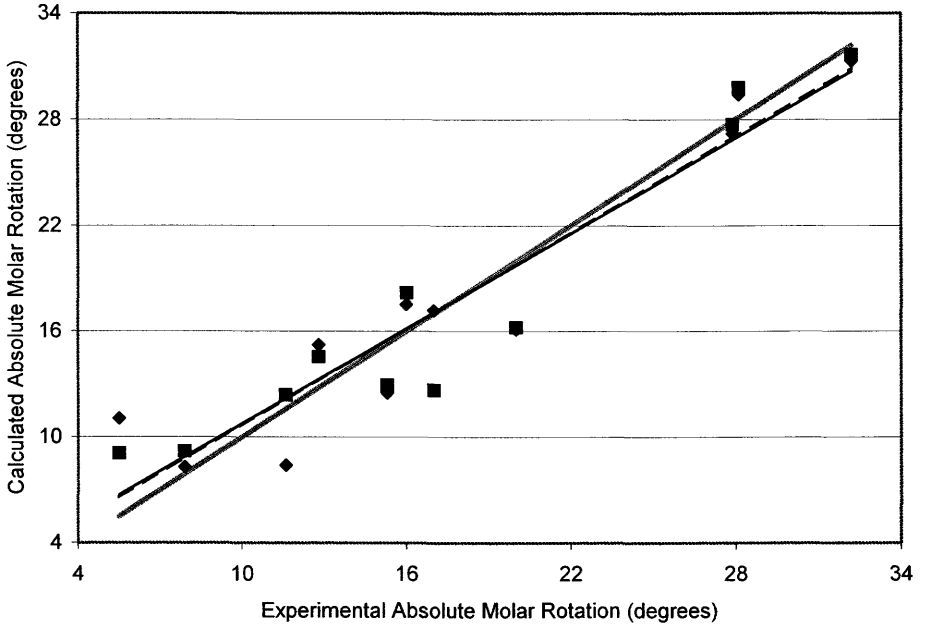


Figure V.2.3 Comparison of the calculated data versus the experimental data for the eleven trisubstituted 2,2'-spirobiindane molecules (Table V.2.2). Array point handedness data are represented by squares and the dashed best fit line, while the difference density data are presented as diamonds and the black solid line. The grey line denotes a perfect correlation where the slope equals one.

The difference density method is better able to account for the molar rotation of molecule H, due to its small difference density shape similarity value, which allows for the largest correction for chirality and regional shape differences that arise because the molecule only shares two common ligands with the reference molecule.

The shape group method is better able to account for the molecule E molar rotation because it does share three common ligands with the reference molecule, and therefore the shape differences tend to be more localized, yet the chirality can be considered from the whole molecule point of view because of the standard orientation component of the chirality measure.

V.3 Absolute Rotatory Strengths of Methyl-Substituted bicyclo[2,2,1]heptan-2-one Derivatives

V.3.1 Grimme's Use of Hybrid Chirality Measures

Bicyclo[2,2,1]heptan-2-one, or norcamphor, as it will be referred to from this point on, is a bicyclic molecule than can be best described as a cyclohexanone-based molecule where cyclohexanone has been modified by the addition of a bridging carbon atom between one of the carbon atoms bonded to the carbonyl atom and the carbon atom furthest from this atom.

Various derivatives of norcamphor (including camphor) are known that are methyl-group-based, where two or three methyl groups are placed at various points on the norcamphor skeleton. The structures of norcamphor and seven of these derivatives are given in Figure V.3.1. Each structure is labelled by its common name, as well as by a letter. The letters will be used in later Tables of data to represent the molecules.

Each of the molecules of the Figure have a non-zero rotatory strength, indicating that optical activity will be seen for the molecules. The magnitudes of these rotatory strengths are presented in a paper by Tokiwa and Kamiya,[77] wherein the rotatory strengths of the eight molecules were calculated based upon an out-of-

plane bending of the carbonyl group. The bending of the carbonyl group required by the desired rotatory strengths in the calculated structures tended to match the bending seen in X-ray diffraction data for the molecules.

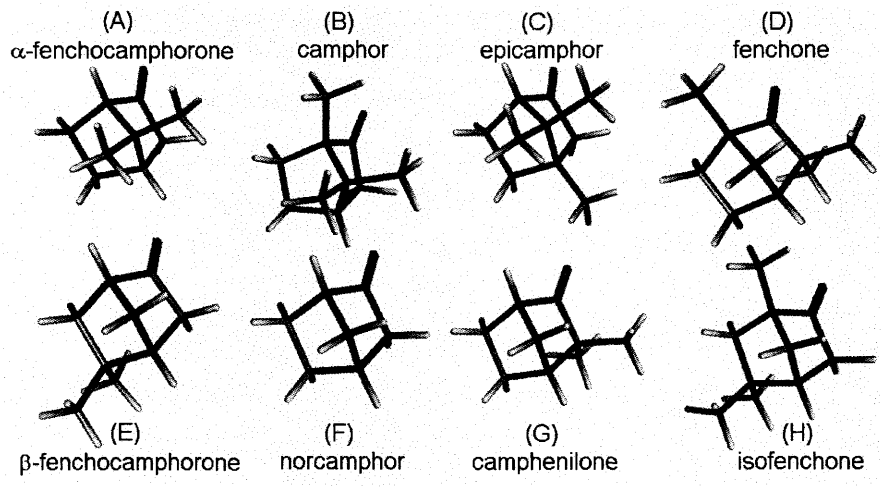


Figure V.3.1 Norcamphor and seven of its methyl-based derivatives.

Grimme, utilizing his hybrid chirality measure (Section II.3.3), has established a correlation between the chirality induced in the carbonyl group by the molecular environment and the absolute rotatory strengths.[78] In his work, the CSM($\hat{C}_2$) local measures of chirality were found for the oxygen lone pair n orbital and the antibonding π^* orbital of the carbonyl group. These are the orbitals involved in the electron transition that leads to the first circular dichroism band for the molecules.

Because these CSM values are maximized (the "most chiral") for an optimized specific C_2 axis for each orbital, Grimme associated larger magnitudes of rotatory strength with larger angles between the optimal axes for the two differing orbitals. Specifically, since his lone pair orbital axis calculated for all molecules essentially coincided with the C-O bond axis, he stated that the optical activity of these compounds was directly related to how the chiral environment of the chromophore perturbed the diffuse π^* orbitals spatially relative to the C-O bond. This perturbation

is noted by the deviation of the optimal C_2 axis for the π^* orbital from the axis defined by the C-O bond.

For seven of the eight molecules studied (camphenilone was not used as it was an outlier), Grimme was able to establish a very good correlation between the angle between the axes and the absolute rotatory strengths. Based on his value for R of 0.961 for the best-fit line of the plotted data, his correlation had an R^2 value of approximately 0.91. However, by using values estimated from the plot of the data presented in the paper,[78] if camphenilone were included in the correlation, it appears that the R^2 value for the correlation would actually fall somewhere between 0.6 and 0.7.

V.3.2 Array Point Handedness Chirality Measures and Methyl-Substituted bicyclo[2,2,1]heptan-2-one Derivatives

Norcamphor and its seven methyl-substituted derivatives have interesting features that are important to the discussion of rotatory strength. First, due to the bridging of the cyclohexanone ring with a carbon atom, the conformational flexibility of the molecules is very small. Such rigidity is part of what makes this family of compounds a favoured choice for optical rotation modelling. Because very few conformations of the molecules are energetically accessible, rotatory strength and its related optical rotation measures can potentially be calculated without worrying about contributions from many different conformations, each statistically weighted based on their relative energies.

A second feature is that the optical properties of the molecules are based upon the first circular dichroism band from the $n\rightarrow\pi^*$ transition seen in the carbonyl

group.[78] This transition occurs specifically from an occupied lone pair orbital of the oxygen atom of the carbonyl group to an antibonding π orbital of the group. Since the carbonyl group (the chromophore) by itself is mirror symmetric, an independent-systems explanation of optical activity (Section II.5.3) describes how the optical activity occurs; the achiral chromophore is subjected to a chiral environment, leading to a non-zero rotatory strength. Alternatively, in the static field model,[78] such a carbonyl group transition is electronically dipole-forbidden, yet magnetically dipole allowed (as in, for example, acetone). Therefore the chiral surroundings of the chromophore must enhance the electronic dipole component of optical activity if rotatory strengths are to be non-zero.

The third, and most interesting, feature in the molecules as it relates to this thesis, is that each molecule contains two stereocentres. If each molecular skeleton can be described as a cyclohexanone ring to which a bridging carbon atom has been added, then each of the two "bridgehead" carbons of the original cyclohexanone ring are stereocentres, as they both have four differing bonded groups (the bridging carbon, the hydrogen atom, and two different paths around the original ring to get to the carbonyl carbon). Figure V.3.2 shows the structure and molecular electron densities (0.010 $e^-/bohr^3$) of two of the norcamphor derivatives. In each molecule the electron density isocontours (0.010 $e^-/bohr^3$) of each stereocentre is also shown, with points of equal array point handedness grouped by colour.

Having two stereocentres in the molecule makes for an interesting counterpoint to the previous amino acid and spirobiindane studies where only one stereocentre

was present. Specifically of interest would be whether the shape and chirality of the first stereocentre (which is bonded directly to the carbonyl carbon, and should model well the chiral environment the chromophore experiences) would be sufficient to correlate well with the rotatory strength, or whether additional information about the chirality of the molecule found in the shape and chirality measures of the second stereocentre would be important.

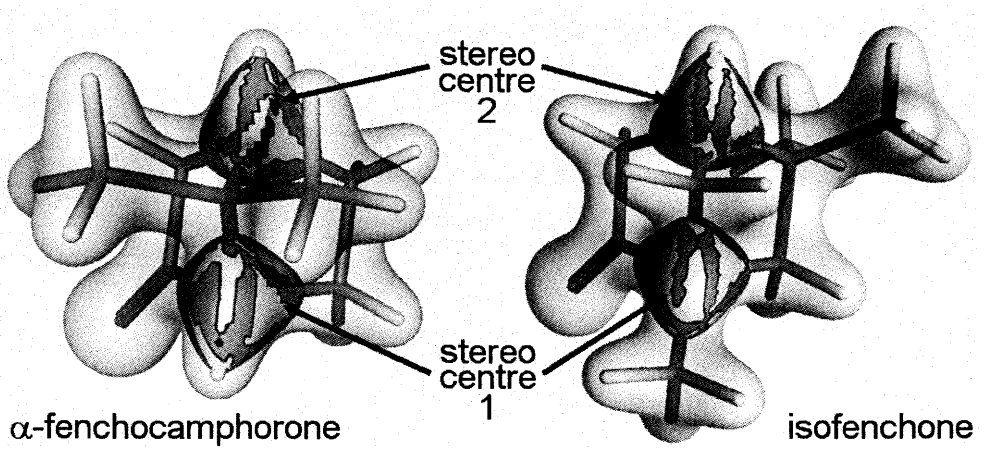


Figure V.3.2 The whole molecule and stereocentre electron density representations (0.010 $e^-/bohr^3$) of two methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives. The stereocentre electron density fragments are coloured by regions of same array point handedness assignment.

Calculations for the molecules were done in the same manner as for the previous amino acid and spirobiindane studies. First, the norcamphor molecule was created in *Gaussian98* and geometry optimized at the RHF/6-31G** level of theory. The resultant geometry was then used as a template for the methyl-substituted molecules. Methyl substitutions were made to the norcamphor template geometry, and geometry optimizations (6-31G**) were performed to find the lowest energy conformations of the molecules. Due to the conformational rigidity, every starting geometry attempted resulted in the geometries for the molecules seen in Figure V.3.1.

Electron densities of the stereocentres were calculated from the full electron population analysis of the molecules calculated with *Gaussian98* on the lowest energy geometries with the 6-31G** basis set using the program *Rhocalc04*. Shape group analysis and array point handedness chirality measures were then determined utilizing the appropriate programs.

Table V.3.1 shows the shape similarity (with α-fencamphorone as the reference molecule) and absolute chirality measure data for the electron densities of the two stereocentres for norcamphor and its seven methyl-based derivatives as represented by the single letter labels assigned to the molecules in Figure V.3.1. Also shown in the Table are the experimental absolute rotatory strengths (|R|) in cgs units as collected in the paper by Tokiwa and Kamiya.[77] Additionally, the Table contains calculated absolute rotatory strength data calculated from several inverse-property correlation equations based upon the shape similarity and chirality measure data of the two stereocentres. The regression coefficients of each of the equations, as well as the R^2 values between the experimental and calculated data are also given.

A look at the shape similarity data in the Table confirms some intuitive ideas about the shape of the electron density representations of the stereocentres. Because the first stereocentre of the reference molecule does not have a methyl group attached to it, the molecules that do have a methyl group in this position (B, D, and H) have the lowest shape similarity values. In fact, in comparison to other shape similarity values seen in the amino acid and spirobiindane studies, these values are significantly lower.

The main reason for this is the conformational rigidity of the molecules. If a methyl group is bonded to the first stereocentre, the other groups around the stereocentre (most notably, the carbonyl group) must shift away from the methyl group to avoid steric problems. Such shifting tends to reduce the shape similarity because the electron density of the stereocentre changes in step with the conformational shifts. Additionally, since the group being replaced is a hydrogen atom, this effect is compounded due to the relatively small contribution the hydrogen atom makes on the shape of the stereocentre in the molecules where it is present, compared to the larger effect a methyl group will have.

Molecule	Expt. |R| (10^{-40} cgs)	Stereocentre 1 Shape	Stereocentre 1 |Chirality|	Stereocentre 2 Shape	Stereocentre 2 |Chirality|	3-Term Calc. |R| (10^{-40} cgs)	4-Term (S) Calc. |R| (10^{-40} cgs)	4-Term (C) Calc. |R| (10^{-40} cgs)	5-Term Calc. |R| (10^{-40} cgs)
A	6.26	1.0000	0.003605	1.0000	0.023400	6.62	6.65	6.44	6.41
B	4.60	0.6290	0.016615	0.8738	0.005964	2.34	3.09	4.45	4.25
C	4.47	0.8725	0.028655	0.8027	0.006156	2.38	2.25	4.25	4.51
D	2.17	0.5797	0.015396	0.7989	0.050647	2.27	1.49	1.36	1.69
E	1.43	0.7617	0.030153	0.8314	0.033968	2.17	2.59	1.32	1.00
F	1.25	0.7872	0.045078	0.7941	0.054658	2.03	1.85	0.92	0.90
G	0.83	0.7701	0.030991	0.7943	0.034826	2.17	1.85	1.30	1.38
H	0.45	0.5863	0.061474	0.8161	0.015038	1.47	1.69	1.43	1.31

Regression Equations		Regression Coefficients			
Basic eqn	BASE = A + B/(S1) + C/(AC1)	A:	2.9998	20.3244	1.2043	-8.5327
3-Term	Calc. |R| = BASE	B:	-1.0562	-1.3598	-0.8529	-0.6643
4-Term (S)	Calc. |R| = BASE + D/(S2)	C:	0.0169	0.0038	0.0187	0.0262
4-Term (C)	Calc. |R| = BASE + D/(AC2)	D:		-13.3660	0.0207	7.3680
5-Term	Calc. |R| = BASE + D/(S2) + E/(AC2)	E:				0.0229
		R^2:	0.5728	0.6215	0.9353	0.9462

Legend | R | is the absolute rotatory strength
S1 is stereocentre 1 shape similarity — AC1 is stereocentre 1 absolute chirality
S2 is stereocentre 2 shape similarity — AC2 is stereocentre 2 absolute chirality

Table V.3.1 The shape similarity and array point handedness absolute chirality measures for the two stereocentres on the methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives. This data is used to create several inverse property regression equations used in the calculation of absolute rotatory strengths.

The other trend in the shape of the first stereocentre can be seen based on the groups attached to the bridging carbon. In molecules A, B, and C, there are two methyl groups attached to this carbon, and so the shape similarity numbers of B, in

comparison to D and H (methyl group on the stereocentre), and C, in comparison to E and F (hydrogen atom on the stereocentre), are larger, because these molecules share more commonalities with the reference molecule.

Similar reasoning can be used for the shape similarity measures of the second stereocentre. Molecules A, B, and C, should be the most similar in shape due to the methyl laden bridging carbon, but it is seen that the shape similarity of C is somewhat lower due to the methyl group bonded directly to the second stereocentre. However, no distinct trend can be seen in the shape of the second stereocentre of the molecules D, E, G, and H, where two methyl groups are placed on a carbon atom of the original cyclohexanone ring next to the stereocentre. This is most likely due to the symmetric change that is seen. While two hydrogen atoms have been replaced by methyl groups, the replacement occurs in such a manner that the spatial effects on the electron density of the stereocentre are balanced, and so little change should be seen. Therefore the second stereocentre electron density shapes of these molecules are not significantly different from that of molecule F.

As for the absolute chirality measures, there appears to be no fixed trend in the values. For instance, molecules B, D, and H have a methyl group on the first stereocentre, as well as two more methyl groups on another carbon atom. For molecules D and H, where these two methyl groups are placed on a carbon atom on either side of the second stereocentre, it might be expected that these two might have similar chirality measures, because on a more local electron density scale, the two could be considered to be "near-mirror" images of one another. However, this is not the case, as B and D have similar values, while H has the largest chirality

measure seen for any stereocentre electron density studied for all classes of molecule so far. This points out the importance of the standard orientation in the chirality measure. The main difference in these molecules is how they align in the array space based on the principal moments of nuclear charge. This factor, combined with the electron density differences seen in the shape similarity numbers, gives the differing chirality measures.

The first attempted correlation equation involved the shape similarity and chirality measure data for the first stereocentre. As the stereocentre that is bonded directly to the carbonyl carbon, the electron density of the stereocentre should model very well the chiral environment the carbonyl group experiences in the molecule. An inverse property equation was set up as in previous studies for amino acids and spirobiindane molecules. This equation, labelled the BASE equation in the Table, had the form A + B/(Shape 1) + C/(Abs. Chirality 1).

Starting from a basic set of regression coefficients, values for the absolute rotatory strengths were calculated using the equation. These values were then plotted against the experimental data, and the linear least-squares best-fit line was then determined, as well as the R^2 value for the correlation. By utilizing the Solver function in Microsoft Excel, this R^2 value was maximized by changing the correlation coefficients under the constraint of minimizing the sum of the distances between each experimental-calculated data point pair.

The BASE equation is not very successful at providing calculated data that correlate well with the experimental data (R^2 of 0.5728). Most notably, the equation does not provide highly differentiated calculated values for molecules B to G. Since

these molecules have rotatory strength values that occupy over sixty percent of the entire range, the correlation is low.

The next regression equation attempted used the inverse shape and chirality data from both stereocentres, leading to a five-term equation with the form BASE + D/(Shape 2) + E/(Abs. Chirality 2). This equation takes the original BASE equation and adds two terms for the data of the second stereocentre. The R^2 value for the calculated data from this equation relative to the experimental data (0.9462) shows a good correlation can be made if the data from both stereocentres is taken into account. However, overall, this correlation could be misleading, as a five-term equation should be able to model a set of eight data points easily without a true association being present.

Therefore, attempts were made with four-term regression equations using the shape and chirality data of the first stereocentre (the BASE equation) combined with either the shape or the absolute chirality data for the second stereocentre.

In the case of the shape of the second stereocentre being used as the third piece of data for the molecule [4-Term (S)] equation, the R^2 value for the correlation equation rises from 0.5728 for the BASE equation to 0.6215 for the four-term equation. However, if the absolute chirality measure data is used from the second stereocentre [4-Term (C)], this value is 0.9353, indicating a strong correlation that is less reliant on many pieces of data.

The relatively small increase in R^2 value as the correlation equation changes from the BASE equation to the [4-Term (S)] equation, coupled with the small decrease in R^2 value as the 5-Term equation is reduced to the [4-Term (C)] equation

indicates the shape of the second stereocentre is relatively unimportant in the calculation of the absolute rotatory strengths of the molecules. By the holographic electron density theorem, this is not necessarily surprising. In the theorem, the electron density of the molecule as a whole must be completely described in either of the stereocentre electron densities, and therefore the shape of one should be all that is required to get this information. However, in a practical sense, the method will be better able to pick this information out from the stereocentre electron density closer to the chromophore, because the optical activity is associated with this part of the molecule, and the relevant information is more notable in this stereocentre as opposed to the other.

However, if the regression coefficients are considered in the [4-Term (C)] equation correlation, an interesting feature is noted. The basic absolute rotatory strength constant A (the "half-cell") for the set of molecules is 1.2043. This constant, much like was seen in the direct optical rotation data based studies, is subsequently corrected by coefficients modified by the shape of the first stereocentre and the chirality measures of both stereocentres. These corrections are proportional to the regression coefficients B, C, and D, and the coefficients are inversely modified by the shape or chirality data. It is shape correction in the equation (B/Shape 1) that is most notable in the equation. The value for B (-0.8529) has a relatively small magnitude relative to the rotatory strength of the reference molecule. This is not what was seen in the optical rotation studies of spirobiindanes and amino acids, where the value of B often had a magnitude that was two to three times greater than the actual reference experimental value for the optical rotation. The large value of B

was required to counteract a larger value of A. While the difference in the type of data must be partially responsible for this, the rotatory strength still is a direct contributor to the optical rotation (Equation 2.5.9), and so the same type of trend in relative coefficient magnitude could reasonably be expected to occur in the norcamphor study case. Yet, this is not so.

Upon closer examination, the shape similarity values fall in a range between 0.5797 and one. When this range is expressed in terms of the (B/Shape 1) correction to A, the range becomes approximately –1.47 to –0.85. If the values of A and the B/Shape correction are subsequently added, the overall range of calculated rotatory strength contribution from the shape of the first stereocentre is approximately –0.27 to 0.35. Between these two values is a potential to differentiate amongst the rotatory strengths of the molecules of about 0.62 units. This, when compared to the reference rotatory strength value of the reference molecule (6.26), means that the shape similarity numbers for the set of molecules only account for about ten percent of the calculated rotatory strength range needed. Since the correlation R^2 value for the overall [4-Term (C)] equation is very good, this means that the shape of the first stereocentre can not be a main contributor to the overall correlation.

At first thought, this seems unreasonable, as the shape of the stereocentre, through the holographic electron density theorem, contains information about the shape of the molecule. It is expected that this shape information would be required to model the rotatory strengths. However, recall that the array point handedness chirality measure assesses chirality based on two different factors. The first of these

factors is the whole-molecule chirality contribution, given by how the electron density fragment is aligned in the array space based on the standard orientation defined by the whole molecule arrangement of nuclear charges. It was this contribution that was most relevant to the spirobiindane study. In that study, the standard orientation of each of the molecules was dominated both by the large molecular skeleton, and by the fact that the non-common ligands were distant from the stereocentre. Therefore, while the chirality measure was more adept at looking at the whole molecule chirality contribution, the relative significance of the chirality measure to calculated optical rotations was low because of the minimal effect ligand change had on standard orientation.

The other factor in the chirality measure is that the electron density "shape" is chiral, and the measure serves to bring this information out. Recall that, when the amino acid molecules were described as being "comparable" based on similar standard orientations in the space, such as in the LYS study, the whole-molecule chirality factor would tend to be very similar in the molecules, and therefore differences in the chirality measure would indicate directly the differences in the chirality of the molecules. Generally, the chirality measure information would not be able to provide a good correlation to the optical rotation by itself though, as the "comparability" of the whole-molecule chirality component is only "general" due to "similar" standard orientations.

In the case of the [4-Term (C)] equation, though, the two stereocentre chirality measures are able to work together to isolate the whole-molecule chirality component due to standard orientation differences. For instance, if an electron

density fragment is placed into a space, and the chirality measure is calculated, the value, while specific for that fragment in that orientation, could be found for the same fragment in a different orientation (integer multiples of ninety degree rotations about each of the spatial axis being a good example of this [Figures III.3.1 and III.3.2]). However, two chirality measures from different fragments within the molecule tend to "solidify" the whole-molecule chirality component, as fewer orientations of the whole molecule will result in the chirality measures for two stereocentre electron densities that are unchanged. Once the whole-molecule chirality component is isolated in this manner, the chirality measures are able to look at the fragment "shape-based" chirality more rigorously, which results in a lesser reliance on the shape similarity numbers for the correlation.

To assess whether this type of reasoning could explain the observed lack of reliance on shape similarity data in the correlation given by the [4-Term (C)] equation, an attempt was made to calculate absolute rotatory strength data for norcamphor and its derivatives based only upon the chirality measure data for the electron density fragments of the two stereocentres.

Table V.3.2 shows the result of this attempt. As before, the regression equation was set up and a starting set of regression coefficients was used. The calculated data were then plotted against the experimental data, and the R^2 value of the best-fit least-squares line was maximized using the Solver function in Microsoft Excel, by varying the regression coefficients under the constraint of minimizing the sum of the distances between the calculated and experimental data point pairs.

As seen by the R^2 value of the best-fit line in the Table, the supposition that the first stereocentre electron density shape contributes little to the overall correlation of the [4-Term (C)] equation (Table V.3.1) is confirmed, as removing the shape similarity data component only reduces the R^2 value from 0.935 to 0.927. However, the most interesting comparison occurs for the regression coefficient that is modified by the second stereocentre chirality measure in the two Tables. In the [4-Term (C)] equation (Table V.3.1), the value for D is 0.0207. In Table V.3.2, the value for C is 0.0208. Essentially, dropping the shape similarity component of the regression equation has no effect on the overall contribution to the calculated rotatory strength by the chirality measure of the second stereocentre. Any change seen between the two correlation equations is seen through the chirality information of the first stereocentre.

Molecule	Stereocentre 1 Abs. Chirality	Stereocentre 2 Abs. Chirality	Experimental |R| (10^{-40} cgs)	Calculated |R| (10^{-40} cgs)
A	0.003605	0.023400	6.26	6.39
B	0.016615	0.005964	4.60	4.63
C	0.028655	0.006156	4.47	4.01
D	0.015396	0.050647	2.17	1.64
E	0.030153	0.033968	1.43	1.21
F	0.045078	0.054658	1.25	0.76
G	0.030991	0.034826	0.83	1.18
H	0.061474	0.015038	0.45	1.64
A =	-0.0713	Calculated |R| = A + B/(Abs. Chirality 1)		
B =	0.0201	+ C/(Abs. Chirality 2)		
C =	0.0208	R^2 = 0.927		

Table V.3.2 The array point handedness absolute chirality measures for the two stereocentres of the methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives. This data is used in the calculation of the absolute rotatory strengths |R| of the molecules.

In terms of the calculated rotatory strengths between the two correlation equations, a simplistic view of the lack of change in the second stereocentre regression coefficient would be that the whole-molecule chirality component of the system is modelled wholly by the second stereocentre chirality measure. Since this measure defines the whole-molecule chirality component, the first stereocentre chirality measure can focus on the shape-based chirality features of the electron density near the chromophore. Because this focus can be absolute, the requirement for shape similarity based information in the correlation is reduced, and in this case, is really not required. However, a more realistic description of the situation is that both chirality measures serve to "nail down" the whole-molecule chirality component, allowing the shape-based features of both chirality measures to come to the forefront, meaning that shape similarity values of either stereocentre electron density are not required.

If this type of chirality determination at a few fragmentary electron density portions of the molecule holds out for several different sets of molecules, then the array point handedness chirality measure will become a very useful tool in QShAR studies. However, it is possible that the deconvolution of chirality components in the measures works well in this set of molecules because of the conformational rigidity imposed on the molecules by the bridging carbon. Without this inflexibility, it is possible that the chirality measure contributions to calculated properties will be like those seen in the "comparable" amino acids, regardless of the number of fragments used, meaning that the whole-molecule and shape-based chirality components are

deeply intertwined and, therefore, the shape similarity components of regression equations will still figure prominently in QShAR work.

V.3.3 The Difference Density Shape Group Method and Methyl-Substituted bicyclo[2,2,1]heptan-2-one Derivatives

The application of the difference density Shape Group method to norcamphor and its derivatives required that alignment of the molecules be performed twice for each of the stereocentres, to allow for difference densities to be created for each of the molecular stereocentres relative to those from the reference molecule.

For the first stereocentre, the stereocentre was chosen to occupy the origin of the space. The positive x-axis was defined from the origin along the line given by stereocentre-carbonyl carbon bond. The positive y-axis was chosen by the definition of the plane created by the intersection of the x-axis and the line between the stereocentre and the bridging carbon. This resulted in the bridging carbon having a positive y coordinate. Finally the positive z-axis was defined by the position of the third ring-based carbon attached to the first stereocentre.

Alignment for the second stereocentre was accomplished in much the same way. The stereocentre was chosen as the origin, while the spatial axes were defined in order by the positions of the carbon atom next to the carbonyl carbon, the bridging carbon atom, and the remaining ring-based carbon atom.

Table V.3.3, much like Table V.3.1, presents the shape similarity and chirality based (difference density shape) data for the norcamphor and its related molecules. As in the array point handedness studies, four different regression equations were attempted based on the various data values.

Molecule	Expt. |R| (10^{-40} cgs)	Stereocentre 1 Shape	Stereocentre 1 DD Shape	Stereocentre 2 Shape	Stereocentre 2 DD Shape	3-Term Calc. |R| (10^{-40} cgs)	4-Term (S) Calc. |R| (10^{-40} cgs)	4-Term (D) Calc. |R| (10^{-40} cgs)	5-Term Calc. |R| (10^{-40} cgs)
A	6.26	1.0000	1.0000	1.0000	1.0000	6.25	6.54	6.33	6.69
B	4.60	0.6290	0.6311	0.8738	0.9846	3.04	3.25	2.76	2.96
C	4.47	0.8725	0.9907	0.8027	0.6453	2.27	2.28	3.47	3.73
D	2.17	0.5797	0.5972	0.7989	0.8241	1.27	1.27	1.43	1.47
E	1.43	0.7617	0.8432	0.8314	0.7940	1.63	2.77	0.67	1.88
F	1.25	0.7872	0.8503	0.7941	0.8096	2.62	1.83	2.24	1.21
G	0.83	0.7701	0.8258	0.7943	0.8226	2.63	1.79	2.23	1.13
H	0.45	0.5863	0.5981	0.8161	0.8255	1.77	1.73	2.31	2.38

Regression Equations		Regression Coefficients			
Basic eqn	BASE = A + B/(S1) + C/(DD1)	A:	11.4825	25.0527	4.3168	19.3959
3-Term	Calc. |R| = BASE	B:	-28.7861	-0.2592	46.2394	-22.0066
4-Term (S)	Calc. |R| = BASE + D/(S2)	C:	23.5546	-1.1357	-52.0854	20.7011
4-Term (D)	Calc. |R| = BASE + D/(DD2)	D:		-17.1206	7.8608	-20.8023
5-Term	Calc. |R| = BASE + D/(S2) + E/(DD2)	E:				9.4068
		R^2:	0.5317	0.6183	0.6277	0.7519

Legend	| R | is the absolute rotatory strength	DD Shape is the difference density shape
	S1 is the stereocentre 1 shape similarity	DD1 is the stereocentre 1 difference density shape
	S2 is the stereocentre 2 shape similarity	DD2 is the stereocentre 2 difference density shape

Table V.3.3 The shape similarity and difference density shape similarity values for the two stereocentres in the methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives. This data is used to create several inverse property regression equations used in the calculation of absolute rotatory strengths.

In the Table, it is seen that the trend in the difference density shape data for the electron densities of the first stereocentre mirrors the trend seen in the direct shape similarity comparisons. In those molecules where a methyl group is bonded to the stereocentre (B, D, and H) at the position where the reference molecule has a hydrogen atom, the difference density values can reasonably be expected to be smaller than in those molecules where the common hydrogen atom is present because the bonding region electron density between two carbon atoms is different from that between a carbon atom and a hydrogen atom. The secondary trend discussed in the previous section, based on whether the bridging carbon has two bonded methyl groups like the reference, can also be seen in the difference density numbers. Therefore molecules B and C have higher difference density shape

similarity numbers compared to the reference A than do the other molecules with the common hydrogen atom or methyl group bonded to the first stereocentre.

The same reasoning can be used in explaining the second stereocentre difference density shape similarity values. In molecule C, which is the only molecule where a methyl group is attached to the second stereocentre, the difference density shape similarity is the lowest. The difference density shape similarity is highest for molecule B, as would be expected based upon the commonality of the bridging carbon environment in molecule B compared to the reference molecule.

However, the mirroring of trends in the shape similarity data and the difference density shape similarity data foreshadow a problem. Essentially, a lack of commonality in the groups attached to the stereocentre must show up in both the difference density and shape similarity values. For instance, in the amino acid study, since there were three common groups on the stereocentre, the difference density shape was able to show the chirality differences because they mostly arose regionally from the one non-common group and its influence on the stereocentre electron density. Therefore, the two different shape measures should be able to extract the information that comes from this electron density influence. This is seen in Table V.1.5, where distinct mirroring of trends in the shape similarity and difference density shape similarity values of any of the amino acid studies is not pronounced. When there is little mirroring of information trends between the measures, there must be a separation of the chirality information from the shape information to accomplish this.

In the norcamphor-based study, however, the molecules have differing numbers of common groups on the stereocentres. This lack of common groups on the stereocentres means that the molecules with the lesser numbers of common groups will generally have lesser shape similarity values compared to the reference. Since each of these non-common groups exerts an influence on the electron density of the stereocentre in the bonding regions, each non-common group leads to large differences in electron density spatially in each of the bonding regions of the stereocentre electron densities. Since there are two or more regions of these differences the difference density method tends to be overwhelmed with information. Therefore, unlike the array point handedness method, which has the whole-molecule component of the measure to help make further differentiation of information, the difference density method cannot overcome the information glut, resulting in shape trends that are mirrored in both shape measures. Since the method cannot cope with all the information, it can be reasonably expected that the correlation with the difference density shape values will be lower than seen for the array point handedness case.

This is seen in the 5-Term Equation in the Table. With five separate terms to describe the information of eight data point comparisons, the R^2 value has to be considered to be very poor (0.75). As this is the best correlation that comes from the attempts at correlating the experimental and calculated data, it must be said that the difference density shape group method has little success in this particular study.

An examination of the regression coefficient contributions to the rotatory strengths of the 5-Term Equation confirms the idea that large amounts of information

are common between the shape and difference density shape information. For instance, the coefficients B and C, which apply to the shape and difference density shapes for the electron density fragments of the first stereocentre are of almost equal magnitude, but of opposite sign. This can be interpreted as a cancellation of much of the common information as seen in the mirrored trends. However, since the overall coefficient magnitudes are much larger than the absolute difference between the two values (the difference is about ten percent of either coefficients magnitude) it is possible that much useful information about relative chirality is being smothered while this cancellation is taking place. For coefficients D and E (second stereocentre) there is less net cancellation of information, probably because the general trend in the shape numbers of the stereocentre is less defined.

Generally, the difference density shape group method should have performed better for this set of molecules, based on the notion that the conformational inflexibility should benefit the method. This benefit would arise from the minimization of differences in density that arise from conformational variation. The fact that the method performed poorly points to the importance the lack of firm commonalities amongst the groups bonded to the stereocentres plays in the extraction of relative chirality information from the shape of the electron densities. Without these commonalities, the shape similarity and difference density shape similarity values are too similar in their presentation of information ("mirroring trends") for a good correlation to be established between shape features and experimental data.

V.3.4 Comparison of the Chirality Measures

Comparing the chirality measures, based on the poor performance of the difference density shape group method relative to the good performance of the array point handedness method, is not justified for the norcamphor-based molecules. However, for completeness, Figure V.3.3 presents the calculated data versus the experimental data for the array point handedness method (Table V.1.2 chirality-measure-only correlation) and the best regression (5-Term) for the difference density method (Table V.3.3).

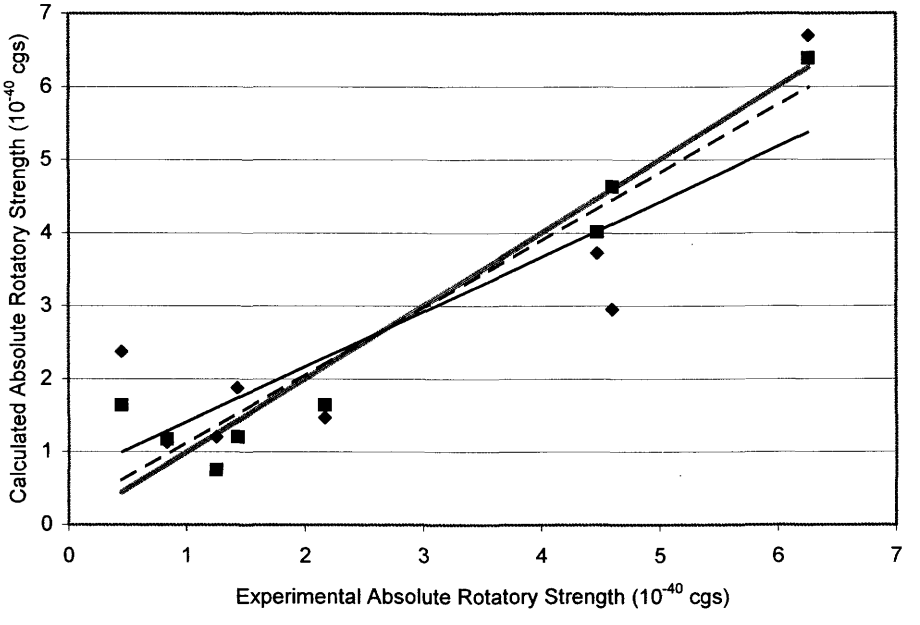


Figure V.3.3 Comparison of the calculated correlation data versus the experimental data for the eight methyl-substituted bicyclo[2,2,1]heptan-2-one derivatives. Array point handedness data are represented by squares (Table V.3.2) and the dashed best fit line, while the difference density data are presented as diamonds (5-Term Equation – Table V.3.3) and the black solid line. The grey line denotes a perfect correlation where the slope equals one.

VI CONCLUSION

This thesis has presented two separate methods for assessing relative chirality of molecules in a manner that can complement the Shape Group method. More specifically, both methods help overcome the problem of mirror image molecules that are considered to be "exactly" similar in shape due to how the Shape Group method mathematically describes shapes in terms of (a,b)-maps.

The first of these methods, the difference density Shape Group method, resolves "exactly" similar and "very" similar shapes by analyzing secondary shapes that are based on difference densities created by subtracting the electron density of a reference from the electron density of the molecule of interest. These secondary shapes are often quite different from each other, which is reflected in the difference density shape similarity numbers. This allows for differentiation of "very" similar shapes.

However, the difference density Shape Group method does not assess chirality *per se*, but rather it notes the differences in electron density that arise spatially when a chiral electron density file is subtracted from other electron densities. If the reference is not chiral, then no distinction between mirror image forms will be made. If the reference is chiral, then part of the difference density shape information mirrors the original information seen in the direct similarity numbers, and part of the information assesses the relative chirality differences of the molecules. However, as is seen in some of the studies performed for this thesis, accessing this relative

chirality information is not possible unless there are commonalities within all molecules of the study.

The second method of assessing chirality is the array point handedness assignment method. Array points within an electron density file are assigned a handedness value based on a rule set that uses the electron density values of neighbouring array points. In an achiral molecule, each point should have a mirror image partner with an opposing handedness. Therefore, the contribution of each point to the chirality measure is cancelled out by its partner. This leads to a consistent measure for all achiral molecules. Chiral molecules then show deviation from this achiral molecule measure based on two contributions: the "whole-molecule chirality" component that arises from the orientation of the molecule in the array space, and the "shape-based chirality" component, that assesses the distortion of the electron density shape from that of an achiral-source-equivalent electron density.

Both methods were used in attempts to correlate shape and relative chirality information of stereocentre electron densities to optical properties of chiral molecule classes. Overall, the array point handedness was more successful in these studies, as correlations were established between the data calculated for the molecules and experimental data. With the exception of a cysteine-referenced amino acid study, every study involving the array point handedness measures had R^2 values over 0.8. Special mention must be made of the spirobiindane study, though, for while the R^2 value of the study was over 0.9, the actual contribution to the correlation by the array point handedness measures was low. However, the method still provided a greater contribution than the difference density shape study for the same molecules.

The difference density Shape Group method appears to gain an advantage over the array point handedness method only in cases where a large number of commonalities exist between molecules, but conformational differences result in different subsets of "whole-molecule" chirality components in the array point handedness measure. The difference density method, which uses enforced orientations, does not look at whole-molecule chirality, and therefore, cannot be subjected to the problems inherent in different amounts of whole-molecule chirality contributions.

Therefore, future work involving the Shape Group method and complementary chirality assessment will probably be mainly carried out with the array point handedness method. In addition to the advantages seen in the studies of this thesis, the method is preferable because it can be performed on existing electron density files without fixed orientations. Also, the method can be easily changed to suit different data types, such as nuclear or electrostatic potentials. Most importantly, though not used in the studies of this thesis because of the lack of correlation between sign of optical activity and absolute configurations of molecules, the array point handedness measure truly differentiates between mirror image forms of molecules by assigning different signs to each member of the mirror pair. Such a distinction might turn out to be quite useful in Quantitative Shape-Activity Relationship studies where the sign difference might be associated with differing levels of activity.

REFERENCES

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

2. P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **14**, 127 (1987).

3. G.A. Arteca and P.G. Mezey, *Int. J. Quant. Chem., Quant. Chem. Symp.*, **23**, 305 (1989).

4. G.A. Arteca and P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **15**, 33 (1988).

5. G.A. Arteca and P.G. Mezey, *Int. J. Quant. Chem.*, **34**, 517 (1988).

6. G.A. Arteca and P.G. Mezey, *J. Comp. Chem.*, **9**, 554 (1988).

7. G.A. Arteca and P.G. Mezey, *J. Mol. Struct. (Theochem)*, **43**, 11 (1988).

8. G.A. Arteca and P.G. Mezey, *J. Math. Chem.*, **3**, 43 (1989).

9. G.A. Arteca and P.G. Mezey, *J. Phys. Chem.*, **93**, 4746 (1989).

10. X. Luo, G.A. Arteca, and P.G. Mezey, *Int. J. Quant. Chem., Quant. Chem. Symp.*, **25**, 335 (1991).

11. X. Luo, G.A. Arteca, and P.G. Mezey, *Int. J. Quant. Chem.*, **42**, 459 (1992).

12. P.G. Mezey, *J. Chem. Inf. Comput. Sci.*, **32**, 650 (1992).

13. G.A. Arteca, V.B. Jammal, P.G. Mezey, J.S. Yadav, M.A. Hermsmeier, and T.M. Gund, *J. Mol. Graphics*, **6**, 45 (1988).

14. P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **12**, 113 (1986).

15. I. Rozas, G.A. Arteca, and P.G. Mezey, *Int. J. Quant. Chem., Quant. Biol. Symp.*, **18**, 269 (1991).

16. G.A. Arteca, V.B. Jammal, and P.G. Mezey, *J. Comp. Chem.*, **9**, 608 (1988).

17. P.G. Mezey, *J. Comp. Chem.*, **8**, 462 (1987).

18. P.D. Walker, G.A. Arteca, and P.G. Mezey, *J. Comp. Chem.*, **12**, 220 (1991).

19. P.G. Mezey, *J. Chem. Inf. Comput. Sci.*, **34**, 244 (1994).

20. P.D. Walker, "Shape characterization of the electronic isodensity surfaces for small molecules": A Thesis, University of Saskatchewan (1992).

21. G.A. Arteca and P.G. Mezey, *J. Math. Chem.*, **10**, 329 (1992).

22. P.D. Walker, P.G. Mezey, G.M. Maggiora, M.A. Johnson, and J.D. Petke, *J. Comp. Chem.*, **16**, 1474 (1995).

23. G.A. Heal, P.D. Walker, M. Ramek, and P.G. Mezey, *Canad. J. Chem.*, **74**, 1660 (1996).

24. P.G. Mezey, Z. Zimpel, P.L. Warburton, P.D. Walker, D.G. Irvine, D.G. Dixon, and B.M. Greenberg, *J. Chem. Inf. Comput. Sci.*, **36**, 602 (1996).

25. P.G. Mezey, *Advances in Molecular Similarity*, **2**, 79 (1998).

26. P.D. Walker, G.M. Maggiora, M.A. Johnson, J.D. Petke, and P.G. Mezey, *J. Chem. Inf. Comput. Sci.*, **35**, 568 (1995).

27. P.G. Mezey, Z. Zimpel, P.L. Warburton, P.D. Walker, D.G. Irvine, X.D. Huang, D.G. Dixon, and B.M. Greenberg, *Environ. Toxicol. Chem.*, **17**, 1207 (1998).

28. P.G. Mezey, *J. Math. Chem.*, **27**, 61 (2000).

29. L.J. Wang, P.L. Warburton, and P.G. Mezey, *J. Phys. Chem. A,* **106**, 2748 (2002).

30. P.G. Mezey, *Math. Chem.,* **4**, 163 (1995).

31. P.G. Mezey, *J. Math. Chem.,* **12**, 365 (1993).

32. P.G. Mezey, *J. Math. Chem.,* **13**, 59 (1993).

33. P.D. Walker, G.A. Arteca, and P.G. Mezey, *J. Comp. Chem.,* **14**, 1172 (1993).

34. P.G. Mezey, *J. Math. Chem.,* **2**, 325 (1988).

35. P.G. Mezey, "Discrete Representations of Three-Dimensional Molecular Bodies and Their Shape Changes in Chemical Reactions", in Understanding Chemical Reactivity, D. Bonchev and O. Mekenyan, Eds., Kluwer Academic Publishers, Dordrecht, The Netherlands (1994).

36. P.G. Mezey, *Topics in Current Chemistry,* **173**, 63 (1995).

37. P.G. Mezey, *Canad. J. Chem.,* **72**, 928 (1994).

38. P.G. Mezey, *Structural Chemistry,* **6**, 261 (1995).

39. P.G. Mezey, *Computational Chemistry (Singapore),* **1**, 109 (1996).

40. P.G. Mezey, *Peptides 1998, Proceedings of the European Peptide Symposium, 25th, Budapest, Hungary,* 42 (1999).

41. P.G. Mezey, "Methods of Molecular Shape Similarity and Topological Shape Design", in Molecular Similarity in Drug Design, P.M. Dean, Ed., Chapman & Hall - Blackie Publishers, Glasgow, U.K. (1995).

42. Z. Zimpel and P.G. Mezey, *Int. J. Quant. Chem.,* **59**, 379 (1996).

43. P.G. Mezey, "Shape Analysis", in The Encyclopedia of Computational Chemistry, P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman,

H.J. Schaefer III, and P.R. Schreiner, Eds., John Wiley & Sons, Chichester (1998).

44. P.G. Mezey, *Math. Chem.,* **5**, 129 (1999).

45. P.G. Mezey, *J. Math. Chem.,* **7**, 39 (1991).

46. P.G. Mezey, *J. Math. Chem.,* **2**, 299 (1988).

47. P.G. Mezey, *Adv. Quant. Chem.,* **27**, 165 (1996).

48. S.F. Mason, Molecular Optical Activity and the Chiral Discriminations, University Press, Cambridge, (1982).

49. R. Carbó, L. Leyda, and M. Arnau, *Int. J. Quant. Chem.,* **17**, 1185 (1980).

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

51. R. Carbó and A.J. Farre, *Afinidad,* **39**, 205 (1982).

52. E.E. Hodgkin and W.G. Richards, *J. Chem. Soc. Chem. Commun.,* 1342 (1986).

53. E.E. Hodgkin and W.G. Richards, *Int. J. Quant. Chem., Quant. Biol. Symp.,* **14**, 105 (1987).

54. P.E. Bowen Jenkins and W.G. Richards, *Int. J. Quant. Chem.,* **30**, 763 (1986).

55. R. Carbó and L. Domingo, *Int. J. Quant. Chem.,* **32**, 517 (1987).

56. C. Burt, P. Huxley, and W.G. Richards, *J. Comp. Chem.,* **11**, 1139 (1990).

57. J. Cioslowski and E.D. Fleischmann, *J. Am. Chem. Soc.,* **113**, 64 (1991).

58. M. Manaut, F. Sanz, J. Jose, and M. Milesi, *J. Comput.-Aided Mol. Des.*, **5**, 371 (1991).

59. A.Y. Meyer and W.G. Richards, *J. Comput.-Aided Mol. Des.*, **5**, 426 (1991).

60. A.M. Richard, *J. Comp. Chem.*, **12**, 959 (1991).

61. P.L. Warburton, Z. Zimpel, and P.G. Mezey, unpublished work.

62. S. Golomb, Polyominoes, Springer, New York, (1965).

63. F. Harary, "The Cell Growth Problem and Its Attempted Solutions", in Beitrage Zur Graphentheorie, H. Sachs, H.J. Voss, and H. Walther, Eds., Teubner, Leipzig (1968).

64. G. Exoo and F. Harary, *Indian Nat. Acad. Sci. Lett.*, **10**, 67 (1987).

65. F. Harary and M. Lewinter, *Internat. J. Computer Math.*, **25**, 1 (1988).

66. N. Madras, C.E. Soteros, and S.G. Wittington, *J. Phys. A.: Math. Gen.*, **21**, 4617 (1988).

67. C.E. Soteros and S.G. Wittington, *J. Phys. A.: Math. Gen.*, **21**, 2187 (1988).

68. F. Harary and P.G. Mezey, *Theor. Chim. Acta*, **79**, 379 (1991).

69. F. Harary and P.G. Mezey, *Molecular Engineering*, **6**, 415 (1996).

70. F. Harary and P.G. Mezey, "Chiral and Achiral Square-Cell Configurations; the Degree of Chirality", in New Developments in Molecular Chirality, P.G. Mezey, Ed., Kluwer Academic Publishers, Dordrecht (1991).

71. CRC Handbook of Chemistry and Physics, R.C. Weast, D.R. Lide, M.J. Astle, and W.H. Beyer, Eds., CRC Press, Inc., Boca Raton, Florida, (1990).

72. Dictionary of Organic Compounds, J. Buckingham and F. Macdonald, Eds., Chapman & Hall, London, (1996).

73. W. Kauzmann, F.B. Clough, and I. Tobias, *Tetrahedron,* **13**, 57 (1961).

74. H. Neudek and K. Schlögl, *Chem. Ber.,* **110**, 2624 (1977).

75. H. Neudek, B. Richter, and K. Schlögl, *Mh. Chem.,* **110**, 931 (1979).

76. G. Derflinger, "Chirality and Group Theory", in Chirality - From Weak Bosons to the α-Helix, R. Janoschek, Ed., Springer-Verlag, Berlin (1991).

77. H. Tokiwa and M. Kamiya, *J. Mol. Struct. (Theochem),* **181**, 25 (1988).

78. S. Grimme, *Chem. Phys. Lett.,* **297**, 15 (1998).

79. M. Gardner, *Scientific American,* **240** (1979).

80. A.C. Good, E.E. Hodgkin, and W.G. Richards, *J. Chem. Inf. Comput. Sci.,* **32**, 188 (1992).

81. H. Zabrodsky, S. Peleg, and D. Avnir, *J. Am. Chem. Soc.,* **114**, 7843 (1992).

82. D. Avnir, H.Z. Hel-Or, and P.G. Mezey, "Symmetry and Chirality: Continuous Measures", in The Encyclopedia of Computational Chemistry, P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.J. Schaefer III, and P.R. Schreiner, Eds., John Wiley & Sons, Chichester (1998).

83. S. Keinan and D. Avnir, *J. Am. Chem. Soc.,* **120**, 6152 (1998).

84. Y. Salomon and D. Avnir, *J. Comp. Chem.,* **20**, 772 (1999).

85. Y. Pinto and D. Avnir, *Enantiomer,* **6**, 211 (2001).

86. A. Seri-Levy and W.G. Richards, *Tetrahedron: Asymmetry,* **4**, 1917 (1993).

87. A. Seri-Levy, S. West, and W.G. Richards, *J. Med. Chem.,* **37**, 1727 (1994).

88. M.S. Mestres, R. Carbó, F.J. Luque, and M. Orozco, *J. Phys. Chem.,* **100**, 606 (1996).

89. P.G. Mezey, R. Ponec, L. Amat, and R. Carbó-Dorca, *Enantiomer,* **4**, 371 (1999).

90. L. Amat, R. Carbó-Dorca, D.L. Cooper, and N.L. Allan, *Chem. Phys. Lett.,* **367**, 207 (2003).

91. G. Gilat, *J. Phys. A.: Math. Gen.,* **22**, L545-L550 (1989).

92. R.S. Cahn, C. Ingold, and V. Prelog, *Angew. Chem. Intern. Ed. Engl.,* **5**, 385 (1966).

93. D. Avnir and A.Y. Meyer, *J. Mol. Struct. (Theochem),* **72**, 211 (1991).

94. A.Y. Meyer and D. Avnir, *Structural Chemistry,* **2**, 475 (1991).

95. A.B. Buda and K. Mislow, *J. Am. Chem. Soc.,* **114**, 6006 (1992).

96. A. Golbraikh, D. Bonchev, and A. Tropsha, *J. Chem. Inf. Comput. Sci.,* **41**, 147 (2001).

97. M. Randic and M. Razinger, *J. Chem. Inf. Comput. Sci.,* **36**, 429 (1996).

98. R. Benigni, M. Cotta-Ramusino, G. Gallo, F. Giorgi, A. Giuliani, and M.R. Varî, *J. Med. Chem.,* **43**, 3699 (2000).

99. H. Zabrodsky and D. Avnir, *J. Am. Chem. Soc.,* **117**, 462 (1995).

100. K.B. Lipkowitz, D. Gao, and O. Katzenelson, *J. Am. Chem. Soc.,* **121**, 5559 (1999).

101. P.D. Walker and P.G. Mezey, *J. Am. Chem. Soc.,* **115**, 12423 (1993).

102. P.D. Walker and P.G. Mezey, *J. Am. Chem. Soc.,* **116**, 12022 (1994).

103. P.D. Walker and P.G. Mezey, *Canad. J. Chem.,* **72**, 2531 (1994).

104. P.D. Walker and P.G. Mezey, *J. Math. Chem.,* **17**, 203 (1995).

105. P.D. Walker and P.G. Mezey, *J. Comp. Chem.*, **16**, 1238 (1995).

106. R.W.F. Bader and T.T. Nguyen-Dang, *Adv. Quant. Chem.*, **14**, 63 (1981).

107. R.W.F. Bader, *Acc. Chem. Res.*, **9**, 18 (1985).

108. C. Chang and R.W.F. Bader, *J. Phys. Chem.*, **96**, 1654 (1992).

109. C.F. Matta and R.F.W. Bader, *Proteins: Structure, Function, and Genetics,* **40**, 310 (2000).

110. C.F. Matta and R.F.W. Bader, *Proteins: Structure, Function, and Genetics,* **48**, 519 (2002).

111. V. Pichon-Pesme, C. Lecompte, R. Wiest, and M. Bènard, *J. Am. Chem. Soc.*, **114**, 2713 (1992).

112. R. Wiest, V. Pichon-Pesme, M. Bènard, and C. Lecompte, *J. Chem. Phys.*, **98**, 1351 (1994).

113. J.H. Brewster, *Tetrahedron Letters,* 23 (1959).

114. J.H. Brewster, *J. Am. Chem. Soc.*, **81**, 5475 (1959).

115. J.H. Brewster, *J. Am. Chem. Soc.*, **81**, 5483 (1959).

116. J.H. Brewster, *J. Am. Chem. Soc.*, **81**, 5493 (1959).

117. D.D. Davis and F.R. Jensen, *Journal of Organic Chemistry,* **35**, 3410 (1970).

118. J.H. Brewster and R.T. Prudence, *J. Am. Chem. Soc.*, **95**, 1217 (1973).

119. R.D. Amos, *Chem. Phys. Lett.*, **87**, 23 (1982).

120. P.L. Polavarapu, *Tetrahedron: Asymmetry,* **8**, 3397 (1997).

121. P.L. Polavarapu, *Molecular Physics,* **91**, 551 (1997).

122. P.L. Polavarapu and D.K. Chakraborty, *J. Am. Chem. Soc.,* **120**, 6160 (1998).

123. P.L. Polavarapu and D.K. Chakraborty, *Chemical Physics,* **240**, 1 (1999).

124. S. Ribe, R.K. Kondru, D.N. Beratan, and P. Wipf, *J. Am. Chem. Soc.,* **122**, 4608 (2000).

125. R.K. Kondru, P. Wipf, and D.N. Beratan, *Science,* **282**, 2247 (1998).

126. P.L. Polavarapu and C. Zhao, *J. Am. Chem. Soc.,* **121**, 246 (1999).

127. J.R. Cheeseman, M.J. Frisch, F.J. Devlin, and P.J. Stephens, *J. Phys. Chem. A,* **104**, 1039 (2000).

128. P.L. Polavarapu, D.K. Chakraborty, and K. Ruud, *Chem. Phys. Lett.,* **319**, 595 (2000).

129. P.L. Polavarapu, *Angew. Chem. Intern. Ed. Engl.,* **41**, 4544 (2002).

130. R.K. Kondru, C.H.-T. Chen, D.P. Curran, D.N. Beratan, and P. Wipf, *Tetrahedron: Asymmetry,* **10**, 4143 (1999).

131. P.L. Polavarapu and C. Zhao, *Chem. Phys. Lett.,* **296**, 105 (1998).

132. R.K. Kondru, P. Wipf, and D.N. Beratan, *J. Phys. Chem. A,* **103**, 6603 (1999).

133. J. Olsen and P. Jørgensen, *J. Chem. Phys.,* **82**, 3235 (1985).

134. P. Jørgensen, H.J.A. Jensen, and J. Olsen, *J. Chem. Phys.,* **89**, 3654 (1988).

135. T. Helgaker, K. Ruud, K.L. Bak, P. Jørgensen, and J. Olsen, *Farad. Disc.,* **99**, 165 (1994).

136. K.L. Bak, A.E. Hansen, K. Ruud, T. Helgaker, J. Olsen, and P. Jørgensen, *Theor. Chim. Acta,* **90**, 441 (1995).

137. R.K. Kondru, P. Wipf, and D.N. Beratan, *J. Am. Chem. Soc.,* **120**, 2204 (1998).

138. R.K. Kondru, D.N. Beratan, G.K. Friestad, A.B. Smith, III, and P. Wipf, *Organic Letters,* **2**, 1509 (2000).

139. P.L. Polavarapu, *Chirality,* **14**, 768 (2002).

140. K. Ruud, P.R. Taylor, and P.-O. Åstrand, *Chem. Phys. Lett.,* **337**, 217 (2001).

141. R.K. Kondru, S. Lim, P. Wipf, and D.N. Beratan, *Chirality,* **9**, 469 (1997).

142. J.G. Kirkwood, *J. Chem. Phys.,* **5**, 479 (1937).

143. E.U. Condon, W. Altar, and H. Eyring, *J. Chem. Phys.,* **5**, 753 (1937).

144. S. Lang, Introduction to Linear Algebra, Springer-Verlag, New York, (1986).

145. D.F. Shriver, P.W. Atkins, and C.H. Langford, Inorganic Chemistry, W.H. Freeman and Company, New York, (1990).

146. R.K. Hill and D.A. Cullison, *J. Am. Chem. Soc.,* **95**, 1229 (1973).

147. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennuchi, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, and J.A. Pople, *Gaussian 98 (Revision A.9)*, Gaussian, Inc., Pittsburgh, PA (1998).

148. P.D. Walker, P.L. Warburton, and P.G. Mezey, *Rhocalc04*, University of Saskatchewan: Saskatoon, SK (2002).

149. P.L. Warburton, *Chiralizer*, University of Saskatchewan: Saskatoon, SK (2002).

150. P.L. Warburton, *Chirality*, University of Saskatchewan: Saskatoon, SK (2002).

151. M. Waldherr-Teshner, T. Goetze, W. Heiden, M. Knoblauch, H. Volhardt, and J. Brinkmann, "MOLCAD – Computer Aided Visualization and Manipulation of Models in Molecular Science", in Second Eurographics Workshop on Visualization in Scientific Computing, Delft, Netherlands (1991).

152. J. Brinkmann, M. Keil, T.E. Exner, R. Marhöfer, and G. Moeckel, "Molecular Models: Visualization", in The Encyclopedia of Computational Chemistry, P.v.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.J. Schaefer III, and P.R. Schreiner, Eds., John Wiley & Sons, Chichester (1998).

153. J. Brinkmann, M. Keil, T.E. Exner, and R. Marhöfer, *J. Mol. Model.*, **6**, 328 (2000).

154. *SYBYL 6.7*, Tripos, Inc., 1699 South Hamley Road, St. Louis, MO, 63144 (2000).

155. P.C. Hariharan and J.A. Pople, *Theor. Chim. Acta,* **28**, 213 (1973).

156. M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees, and J.A. Pople, *J. Chem. Phys.*, **77**, 3654 (1982).

157. V. Rassolov, J.A. Pople, M. Ratner, and T.L. Windus, *J. Chem. Phys.*, **109**, 1223 (1998).

158. P.D. Godfrey, S. Firth, L.D. Hatherley, R.D. Brown, and A.P. Pierlot, *J. Am. Chem. Soc.*, **115**, 9687 (1993).

159. S. Gronert and R.A.J. O'Hair, *J. Am. Chem. Soc.*, **117**, 2071 (1995).

160. A.G. Császár, *J. Phys. Chem.*, **100**, 3541 (1996).

161. S.G. Stepanian, I.D. Reva, E.D. Radchenko, and L. Adamowicz, *J. Phys. Chem. A,* **103**, 4404 (1999).

162. J. Rak, P. Skurski, J. Simons, and M. Gutowski, *J. Am. Chem. Soc.,* **123**, 11695 (2001).

163. P.G. Mezey, *J. Math. Chem.,* **23**, 65 (1998).

164. P.G. Mezey, *J. Chem. Inf. Comput. Sci.,* **39**, 224 (1999).

165. P.G. Mezey, *Advances in Biochirality, [Symposium on Biological Homochirality], Serramazzoni, Italy, 1998,* 35 (1999).

166. J.S. Binkley, J.A. Pople, and W.J. Hehre, *J. Am. Chem. Soc.,* **102**, 939 (1980).

167. M.S. Gordon, J.S. Binkley, J.A. Pople, W.J. Pietro, and W.J. Hehre, *J. Am. Chem. Soc.,* **104**, 2797 (1983).

APPENDIX

Program Chiralizer

```
      implicit double precision(a-h,q-z)
      implicit integer(i-p)
      real rho,chiral
      integer dirx,diry,dirz,sumdir
      dimension rho(300,300,300),chiral(300,300,300)
      character*40 fieldfile,datfile,sumfile,tempname,cfldfile
      character*55 junk
      character*80 junk2,garb

***   Zeroing of counters

 1    occhk=0
 2    pointtot=0
 3    rhotot=0.0
 4    rhoplus=0.0
 5    rhoneg=0.0
 6    rhomir=0.0
 7    pluspoint=0
 8    negpoint=0
 9    mirpoint=0
 10   bndlo=0.
 11   bndhi=10.
 12   igx=0
      igy=0
      igz=0

***   End of zeroing

***   Input of field file

 23   write(*,*) 'What is the field file name?'

      read(5,216) fieldfile
      if(fieldfile.eq.'STOP'.or.fieldfile.eq.'stop') goto 5300

 216  format(a40)
 222  format(a5,i5)
 41   format(a40)
 221  format(a11)
 999  format(a1)
 979  format(a6,f5.3)
 980  format(f5.3)
 969  format(a6,3(f8.5,2x))
 970  format(3(f8.5,2x))
```

```
989   format(a16,a35)

      open(13,file=fieldfile,status='old',form='formatted')
      read(13,221) junk

      if(junk.ne.'# AVS field')then
        write(*,*) ' File is not a field file,  try again'
        goto 23
      endif

      read(13,41) junk
      read(13,222) junk,igx
      read(13,222) junk,igy
      read(13,222)junk,igz
      write(*,*) igx,igy,igz

      do kkk=1,100
        read(13,999)junk
        if(junk.eq.'#')goto 121
      enddo

121   read(13,979) junk,res
      write(*,980) res
      read(13,969) junk,xst,yst,zst
      write(*,970) xst,yst,zst

      do kkk=1,100
        read(13,989) junk,datfile
          if(junk.eq.'variable 1 file=')then
            write(*,41) datfile
            goto 24
          endif
      enddo

24    close (13)

***   end of field file input

***   output file naming

207   write(*,*) 'Please input additional filename info (x for none)'
      read(*,*) tempname
      tempname='x'
      namelen=len(fieldfile)
      namelen2=len(tempname)
      j=-1

      do i=1,namelen
        j=j+1
        if (fieldfile(i:i).eq.'.') goto 208
      enddo

208   k=0

      do i=1,namelen2
        k=k+1
```

```
        if(tempname(1:1).eq.' '.or.tempname(1:1).eq.'x'.or.
     +    tempname(1:1).eq.'X')goto 209
      enddo

209   k=k-1
      if(k.eq.0) then
        garb=fieldfile(1:j)
        else
        garb=fieldfile(1:j)//tempname
      endif

***   Summary file name

      namelen=len(garb)
      j=0

      do i=1,namelen
        j=j+1
        if (garb(i:i).eq.' ') goto 1005
      enddo

1005  sumfile=garb(1:j-1)//'.chi'
      cfldfile=garb(1:j-1)//'chi.fld'
1010  write(*,*)'The output chirality file name is: ', sumfile
      open(17,file=cfldfile)
      junk='variable 1 file='//datfile
      junk2=junk//' filetype=binary skip=0'
      write(*,297) junk2
      write(17,296) igx,igy,igz

296   format('# AVS field',/, 'ndim=3',/,'dim1=',i3,/, 'dim2=',i3
     +   ,/,'dim3=',i3,/,'nspace=3',/,
     +   'veclen=2',/,
     +   'data=real',/,
     +   'field=uniform',/,
     +   '#')

      write(17,299) res,xst,yst,zst

299   format('# res ',f5.3,/,'# min ' ,3(f8.4,2x))
      write(17,297) junk2
      junk='variable 2 file='//sumfile
      junk2=junk//' filetype=binary skip=0'
      write(17,297) junk2

297   format(a80)

      close(17)

***   End of summary file name input

***   Reading of data file*

      do i=1,300
        do j=1,300
          do k=1,300
```

```
          rho(i,j,k)=0.
          chiral(i,j,k)=-99.
        enddo
      enddo
    enddo

1020 open(22,FILE=datfile,access='direct',form='unformatted',
  +   recl=igx*4)

    do k=1,igz
      do j=1,igy
        Read(22,rec=(k-1)*igy+j)  (rho(i,j,k),i=1,igx)
      enddo
    enddo

1021 close(22)

***   End of reading of data file

***   Analysis of chirality

***   This consists of several subsections

***   Loop reading of point, error check, orientation assignment

1600 mirpoint=0
    pluspoint=0
    negpoint=0
    rhoplus=0.
    rhoneg=0.
    rhomir=0.
    pointtot=0
    rhotot=0.
    pointthresh=0
    rhomax=0.

***   Start of loops

2000 do 4020,i=1,igx
2010   do 4010,j=1,igy
2020     do 4000,k=1,igz
          if(i.eq.1.or.i.eq.igx.or.j.eq.1.or.j.eq.igy.or.k.eq.1.or
  +         .k.eq.igz) then
            plusneg=99
            rhomax=max(rho(i,j,k),rhomax)
            goto 3000
          endif

***   Universal counter addition

2056 plusneg = 0

***   Determination of point properties


2060 pointthresh=pointthresh+1
```

```
2061 diffx = (rho((i-1),j,k)-rho((i+1),j,k))
2062 diffy = (rho(i,(j-1),k)-rho(i,(j+1),k))
2063 diffz = (rho(i,j,(k-1))-rho(i,j,(k+1)))

2066 diffxy = diffx-diffy
2067 diffxz = diffx-diffz
2068 diffyz = diffy-diffz

2070 slx = abs(diffx)
2071 sly = abs(diffy)
2072 slz = abs(diffz)

2073 slxy = abs(diffxy)
2074 slxz = abs(diffxz)
2075 slyz = abs(diffyz)

***   Finding largest and smallest face label differences

      ambmax=max(slx,sly)
      ambmin=min(slx,sly)
      ambmax=max(ambmax,slz)
      ambmin=min(ambmin,slz)

***   Assigning dir* values

2080  if(diffx.gt.0.) then
        bigx = rho((i-1),j,k)
        smallx = rho((i+1),j,k)
        dirx = -1
        else
        smallx = rho((i-1),j,k)
        bigx = rho((i+1),j,k)
        dirx = 1
      endif

2081  if(diffy.gt.0.) then
        bigy = rho(i,(j-1),k)
        smally = rho(i,(j+1),k)
        diry = -1
        else
        smally = rho(i,(j-1),k)
        bigy = rho(i,(j+1),k)
        diry = 1
      endif

2082  if(diffz.gt.0.) then
        bigz = rho(i,j,(k-1))
        smallz = rho(i,j,(k+1))
        dirz = -1
        else
        smallz = rho(i,j,(k-1))
        bigz = rho(i,j,(k+1))
        dirz = 1
      endif
```

```
2085 sumdir=dirx+diry+dirz

***     Point achirality within error check (mircode=1 achiral)

2090 mircode = 0
2091 ocode = 0

***     Diagonal "achirality" check

2100 temp1 = abs(slx-sly)
2101 temp2 = abs(slx-slz)
2102 temp3 = abs(sly-slz)

2104 if(temp1.le.0.) mircode=1
2105 if(temp2.le.0.) mircode=1
2106 if(temp3.le.0.) mircode=1

***     Normal "achirality" check

2107 if(slx.le.0.) then
       mircode=1
     endif

2108 if(sly.le.0.) then
       mircode=1
     endif

2109 if(slz.le.0.) then
       mircode=1
     endif

2115 if(mircode.eq.1) then
       plusneg=0
       ocode=0
       goto 2199
       else mircode=0
     endif

***     End of achirality check

***     Size sorting of face label differences and ocode assignment

2120 if(slx.gt.sly.and.sly.gt.slz) ocode=1
2121 if(slx.gt.slz.and.slz.gt.sly) ocode=2
2122 if(sly.gt.slx.and.slx.gt.slz) ocode=3
2123 if(sly.gt.slz.and.slz.gt.slx) ocode=4
2124 if(slz.gt.slx.and.slx.gt.sly) ocode=5
2125 if(slz.gt.sly.and.sly.gt.slx) ocode=6

***     End of size sorting

***     ocode=1 checks with plusnegcode assignment

2199 if(ocode.eq.0) occhk=occhk+1

2200 if(ocode.eq.1.and.mircode.eq.0) then
```

```
        if(sumdir.eq.-3.or.sumdir.eq.1) then
          plusneg=1
          else
          plusneg=2
        endif
      endif

***   ocode=2 checks with plusnegcode assignment

 2300 if(ocode.eq.2.and.mircode.eq.0) then
        if(sumdir.eq.-3.or.sumdir.eq.1) then
          plusneg=2
          else
          plusneg=1
        endif
      endif

***   ocode=3 checks with plusnegcode assignment

 2400 if(ocode.eq.3.and.mircode.eq.0) then
        if(sumdir.eq.-3.or.sumdir.eq.1) then
          plusneg = 2
          else
          plusneg=1
        endif
      endif

***   ocode=4 checks with plusnegcode assignment

 2500 if(ocode.eq.4.and.mircode.eq.0) then
        if(sumdir.eq.-3.or.sumdir.eq.1) then
          plusneg = 1
          else
          plusneg=2
        endif
      endif

***   ocode=5 checks with plusnegcode assignment

 2600 if(ocode.eq.5.and.mircode.eq.0) then
        if(sumdir.eq.-3.or.sumdir.eq.1) then
          plusneg = 1
          else
          plusneg=2
        endif
      endif

***   ocode=6 checks with plusnegcode assignment

 2700 if(ocode.eq.6.and.mircode.eq.0) then
        if(sumdir.eq.-3.or.sumdir.eq.1)then
          plusneg = 2
          else
          plusneg=1
        endif
```

```
      endif

***   plusneg=1 is "front"; plusneg=2 is "back"
***   "front" is positively handed
***   "back" is negatively handed
***   plusnegcode applied to points for output file

 3000 if(plusneg.eq.0) then
        chiral(i,j,k)=0.
      endif

      if(plusneg.eq.1) then
        chiral(i,j,k)=-1.
      endif

      if(plusneg.eq.2) then
        chiral(i,j,k)=1.
      endif

      if(plusneg.eq.99) then
        chiral(i,j,k)=99.
      endif

      if(chiral(i,j,k).eq.-99) then
        write(*,*) 'Error at ',i,j,k
        STOP
      endif

***   End of plusnegcode application

 4000    continue
 4010   continue
 4020 continue

***   End of loops

***   Output files write

      write(*,*) 'rhomax on edges is: ',rhomax

      open(22,file=
     +   sumfile ,recl=igx*4,
     +   form='unformatted')

      do k=1,igz
        do  j=1,igy
          write(22,rec=(k-1)*igy+j) (chiral(i,j,k),i=1,igx)
        enddo
      enddo

      close(22)

***   Repeat for multiple batch file assignment
 5250 goto 1

 5300 end
```