RELATIVE CHIRALITY INDEX: A NOVEL APPROACH TO THE CHARACTERIZATION OF MOLECULAR CHIRALITY

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Terrence Sherman Neumann

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Subhash Basak

Paul Siders

January 2011

Acknowledgements

I would like to thank Dr. Subhash Basak and Dr. Ramanathan Natarajan for their integral help in the development of this thesis. The support given by Mr. Brian Gute and Ms. Denise Mills also proved to be help in the advancement of my formal education.

A word of thanks to my family for their understanding and constant support in all my endeavors. Finally, I must thank my wife, Andrea, without whose support; this document would likely remain unfinished.

Abstract

Chiral compounds are being increasingly used as pharmaceuticals and agrochemicals to limit potential side effects of a drug, such as the thalidomide tragedy, and to limit the chemical load on the environment. Therefore it is important to be able to model properties of chiral compounds so such side effects can be avoided. A new method to encode information about the chirality of molecules is described herein. The Relative Chirality Index (RCI) was devised to encode chirality information. This index was shown to be useful to encode information about different molecules. The index is able to differentiate one stereoisomer from another in a logical manner. An application of this index to experimental data is also described.

Table of Contents

List of Tables

List of Figures

List of Examples

Relative Chirality Index: A Novel Approach to the Characterization of Molecular Chirality

Section 1. Introduction

The increased knowledge of biological systems has led to the realization that chirality is often an important factor in the receptor/substrate interaction.[1] Moreover, regulating authorities are encouraging the use of chiral compounds to increase the effect while reducing the dose or quantity of chemical needed to produce the effect. This has made necessary the advancement of synthetic methods to produce enantiomerically pure compounds as well as the technology to separate these compounds from their racemates. The term chiral is used to describe a molecule that is not super-imposable on its mirror image. These two mirror images are called enantiomers. Molecules containing two stereocenters have four stereoisomers, which can be paired into two sets of two enantiomers. Each stereoisomer that is not a mirror image of another stereoisomer is called a diastereomer. An equal mixture of enantiomers is said to be racemic.

Molecular chirality has become an important topic recently as the Food and Drug Administration (FDA) has begun regulating the production of racemic drugs due to tragedies such the Thalidomide case of the 1950s and 1960s.[2,3] Since each stereoisomer can have its own biological activity, it is important to be able to separate the stereoisomers from each other. The Thalidomide case is an example of this. One stereoisomer provided the antiemetic therapeutic effect and the other stereoisomer caused deformities in the fetus.[1,4]

Chemicals can be represented by molecular graphs from which various invariants can be calculated or the matrices associated with the molecular graph.[5,6] These descriptors are based on molecular structure and use theories from organic chemistry, quantum chemistry, information theory, graph theory, or other fields. Using these descriptors, one can predict various molecular properties or biological activities. These models have been used in a wide range of applications in fields like toxicology, medicinal chemistry, environmental chemistry, and physical chemistry, to mention just a few. One famous example of this was the use of the Wiener index to model the boiling point of hydrocarbons. By using the Wiener index to turn the molecular structure of the various hydrocarbons into numbers, Harry Wiener was able to develop a model for the boiling point that greatly reduced the variance of previous models.[5]

The use of chiral compounds as pharmaceuticals and agrochemicals continues to increase, warranting numerical characterization of chirality in order to develop structure-activity relationship models involving these compounds.[7] Conventional topological indices that describe the molecular topology, 3-D descriptors, and quantum chemical descriptors of energetics cannot differentiate enantiomers or diastereomers. There have been some attempts to develop topological indices to differentiate stereoisomers and enantiomers. The Schultz indices use a chiral factor (CF) of +1 for an R configuration and a CF of -1 for an S configuration following the Cahn-Ingold-Prelog (CIP) convention for designating absolute configuration.[7] The CIP convention assigns priority to groups attached to a stereocenter based on molecular weight of the substituent closest to the stereocenter. The lowest priority group is rotated to the position away from the viewer.

The absolute configuration for the stereocenter is assigned by connecting the groups by increasing priority in a clockwise manner for "R" or anti-clockwise manner for "S". Golbaikh, et al. have developed chirality descriptors from molecular topology. Using a chirality correction (+ or – system) that is added to the vertex degrees of the asymmetric atoms, these indices have adjusted existing indices to correct for asymmetric atoms.[8] Randic has developed a chirality index involving atomic asymmetry contributions. This approach assigns symmetric values with opposite signs for enantiomers and assigns a value of zero for all achiral systems.[9] Treating chirality as a discontinuous measure, however, has limitations in the application to quantitative structure-activity relationships (QSAR). These indices have a bias because the values for the enantiomers may not correctly follow the activity that is modeled. Also, the ordering of diasteromeric compounds has been shown to be biased when using discontinuous measures.[10-11] Other methods have been developed based on algebraic structure, but are not suitable for use in quantitative structure activity relationships.[12-14] Another approach using so called "chirality factors" modified traditional molecular descriptors to predict properties.[15]

A new method for calculating chirality indices has been proposed. These indices are called the Relative Chirality Indices (RCI). Using the Cahn-Ingold-Prelog rules as a basis for the calculations of the new chirality indices afforded the use of a three-point interaction model, which is often used in biodiscrimination models of enantiomers. A single molecular representation results from the priority ranking of the substituents attached to the chiral atom. In order to develop this index, the RCI was limited to the

representation of chirality around a carbon atom. Extension to 2D chirality and other atomic chiral centers would require modification of the RCI algorithm.[16]

This thesis will describe the calculation of the RCI, discuss the application of the RCI to a group of amino acids, and implement the RCI in a structure-activity relationship (SAR) for a set of mosquito repellants.[17,18]

Section 2. Methods

Chirality Indices. Calculating the Relative Chirality Index (RCI) for a given molecule includes multiple steps. First, the Cahn-Ingold-Prelog rules were applied to the molecule to determine priority of substituents on the chiral carbon atom.[16] Second, each substituent was assigned its corresponding group descriptor (δ_i) value if a single atom or calculated as a sum of descriptor values accounting for proximity to the chiral center via shortest topological (integer, through bond) distance.

$$\delta_i = \delta_{n1} + \left(\frac{\delta_{n2}}{2}\right) + \left(\frac{\delta_{n3}}{4}\right) + \left(\frac{\delta_{n4}}{8}\right) + \cdots$$

where *n1* is the adjacent atom (nearest neighbor), *n2* is separated by one atom, *n3* by two atoms, etc. This diminishing factor was included as atoms closer to the chiral center will affect the chiral carbon atom more than atoms further away from the chiral center. The scaling by 2^{ni-1} was an attempt to simulate this. Finally, relative chirality indices can be calculated by applying the appropriate equation for the R-isomer or S-isomer, respectively.

$$RCI_R = \delta_a + (\delta_a + \delta_a\delta_b) + (\delta_a + \delta_a\delta_b + \delta_a\delta_b\delta_c) + \delta_a\delta_b\delta_c\delta_d$$

$$RCI_S = \delta_a + (\delta_a + \delta_a\delta_c) + (\delta_a + \delta_a\delta_c + \delta_a\delta_b\delta_c) + \delta_a\delta_b\delta_c\delta_d$$

When a hydrogen atom is present as the lowest priority, "*d*" group, the final term in the above equations was omitted. This arises from the fact that topological descriptors neglect hydrogen atoms (thus canceling out the final term). These equations were chosen to simulate the method for which the Cahn-Ingold-Prelog rules determine absolute

configuration. The groupings for these equations represent the clockwise or anti-clockwise circle drawn to determine R- or S- configuration based on the Cahn-Ingold-Prelog rules. Figures 1 and 2 depict this method.

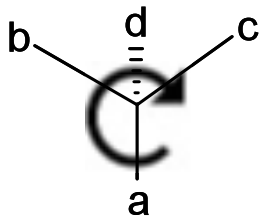


Figure 1. Depiction of simulation of the Cahn-Ingold-Prelog rules in calculating the RCI_R.

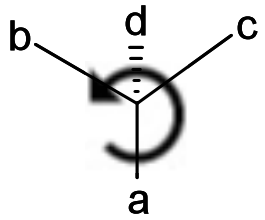


Figure 2. Depiction of simulation of the Cahn-Ingold-Prelog rules in calculating the RCI_S.

Calculating Simple Connectivity RCI (^{S}RCI). Using the hydrogen repressed skeleton, each atom was assigned a value according to the number of sigma or valence electrons present.[19,20]

$$\delta = \sigma - h$$

where δ is the value for the atom descriptor, σ is the number of valence electrons, and h is the number of hydrogen atoms. Therefore, δ is the count of the number of adjacent atoms in a hydrogen repressed graph. This descriptor encodes information about the connectivity of the atoms around the chiral center. Applying this to the RCI algorithm yields the simple connectivity RCI (^{S}RCI). Figure 3 shows the hydrogen repressed graph

of alanine and its chiral carbon. Example 1 depicts an example of the calculation involving the amino acid alanine.

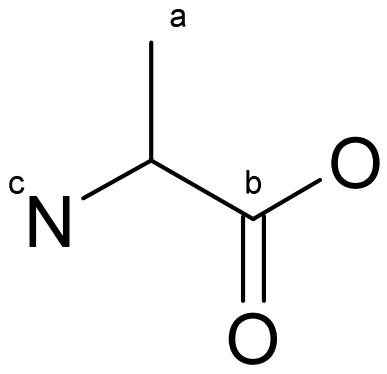


Figure 3. Hydrogen repressed graph of the amino acid alanine with RCI labeling convention.

Example 1. Calculation of ^{S}RCI for the amino acid alanine.

$$\delta_a = 1$$

$$\delta_b = 3 + \frac{1}{2} + \frac{1}{2} = 4$$

$$\delta_c = 1$$

$$^{S}RCI_R = 1 + \left(1 + 1(4)\right) + \left(1 + 1(4) + 1(4)1\right) = 15$$

$$^{S}RCI_S = 1 + \left(1 + 1(1)\right) + \left(1 + 1(1) + 1(1)4\right) = 9$$

Calculating Bond Connectivity RCI (^{B}RCI). Calculation of the bond connectivity for an atom is the sum of all its σ and π bonds in a hydrogen repressed graph. [19,20] This descriptor encodes information for the connectivity around the chiral center similar to that of the simple connectivity descriptor, but also includes information about double and triple bonds. Example 2 shows an example of the calculation for ^{B}RCI.

Example 2. Calculation of ^{B}RCI for the amino acid alanine.

$$\delta_a^B = 1$$

$$\delta_b^B = 4 + \frac{2}{2} + \frac{1}{2} = 5.5$$

$$\delta_c^B = 1$$

$$^{B}RCI_R = 1 + \left(1 + 1(5.5)\right) + \left(1 + 1(5.5) + 1(5.5)1\right) = 19.5$$

$$^{B}RCI_S = 1 + \left(1 + 1(1)\right) + \left(1 + 1(1) + 1(1)5.5\right) = 10.5$$

Calculating Valence Connectivity RCI (^{V}RCI). The valence connectivity descriptor takes into account the valences of all hetero atoms as well as the bond connectivity as shown by the following equation. [19,20]

$$\delta^V = Z^V - h$$

where δ^V is the valence connectivity atom descriptor and Z^V is the number of valence electrons. This descriptor has been shown to closely correlate to the Mulliken-Jaffe electronegativity.

Example **3** shows the calculation of ^{V}RCI for the amino acid alanine.

Example 3. Calculation of ^{V}RCI for the amino acid alanine.

$$\delta_a^V = 3$$

$$\delta_b^V = 4 + \frac{6}{2} + \frac{5}{2} = 9.5$$

$$\delta_c^V = 1$$

$$^{V}RCI_R = 3 + \left(3 + 3(9.5)\right) + \left(3 + 3(9.5) + 3(9.5)1\right) = 94.5$$

$$^{V}RCI_{S} = 3 + \left(3 + 3(1)\right) + \left(3 + 3(1) + 3(1)9.5\right) = 43.5$$

Calculating Intrinsic State RCI (^{I}RCI). The intrinsic state (*I*) of an atom was defined by Kier and Hall [19,20] to encapsulate two attributes:

1. A metric for describing the potential for intermolecular interaction by an atom or group.

2. The state that influences and is influenced by the topology of the group or atom. This descriptor uses both the δ and the δ^{V} to encode this information as show by the following equation.

$$I = \frac{\left(\delta^{V} + 1\right)}{\delta}$$

To simplify calculations, the intrinsic state for each atom was considered. Example 4 shows the calculation of ^{I}RCI for the amino acid alanine.

Example 4. Calculation of ^{I}RCI for the amino acid alanine.

$$\delta_{a}^{I} = 4$$

$$\delta_{b}^{I} = 1.667 + \frac{7}{2} + \frac{6}{2} = 8.167$$

$$\delta_{c}^{I} = 2$$

$$^{I}RCI_{R} = 4 + \left(4 + 4(8.167)\right) + \left(4 + 4(8.167) + 4(8.167)2\right) = 142.672$$

$$^{I}RCI_{S} = 4 + \left(4 + 4(2)\right) + \left(4 + 4(2) + 4(2)8.167\right) = 93.34$$

Calculating Formula Weight RCI (^{W}RCI). Among the most widely known molecular descriptor is formula weight. In an effort to encode information of the formula weight distribution around a chiral center, the formula weight of each group was calculated. The

formula weight of each group was then applied to the RCI algorithm without accounting for the proximity of the atom to the chiral center because this was not chemically intuitive. Example 5 shows the calculation of ^{W}RCI for alanine.

Example 5. Calculation of ^{W}RCI for the amino acid alanine.

$$\delta_a^W = 14.0 + 2(1.0) = 16.0$$

$$\delta_b^W = 12.0 + 2(16.0) + 1.0 = 45.0$$

$$\delta_c^W = 12.0 + 3(1.0) = 15.0$$

$$^{W}RCI_R = 16.0 + \left(16.0 + 16.0(45.0)\right) + \left(16.0 + 16.0(45.0) + 16.0(45.0)15.0\right) = 12{,}288$$

$$^{W}RCI_S = 16.0 + \left(16.0 + 16.0(15.0)\right) + \left(16.0 + 16.0(15.0) + 16.0(15.0)45.0\right) = 11{,}328$$

Calculating RCI for molecules containing more than one chiral center. Each chiral center was considered separately first. Then, a root-mean-square approach, as shown in this equation, was employed to calculate a RCI for a molecule containing more than one chiral center.

$$RCI = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(RCI_i\right)^2}$$

A root-mean-square average was used as an effort to encode information for each chiral center equally. Often in a biological setting for a molecule containing many chiral centers, one chiral center will have the dominating effect on activity while the other chiral centers will have a smaller effect. Since this chiral center is often not identified, it is not possible to account for this phenomenon.

Figure 4 shows the hydrogen repressed graph of the amino acid threonine and example 6 calculates its ^{V}RCI for one stereoisomer, D-allo-threonine.

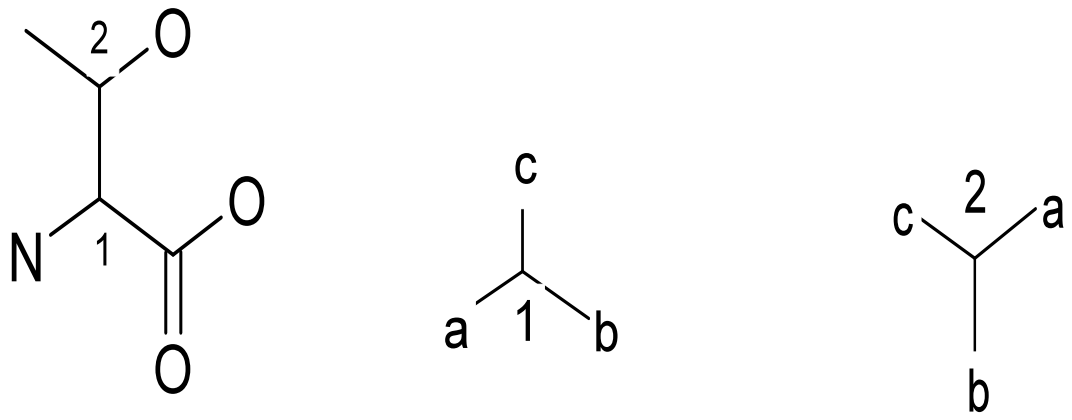


Figure 4. Hydrogen repressed graph of the amino acid threonine neglecting stereochemistry. The numbered carbons represent the chiral centers in the molecule. The Cahn-Ingold-Prelog priority is also shown above for each chiral center.

Example 6. Calculation of ^{V}RCI for the molecule of D-allo-threonine.

For chiral center 1:

$$\delta_a^V = 3$$

$$\delta_b^V = 4 + \frac{6}{2} + \frac{5}{2} = 9.5$$

$$\delta_c^V = 3 + \frac{1}{2} + \frac{5}{2} = 6$$

$$^{V}RCI_R = 3 + \left(3 + 3(9.5)\right) + \left(3 + 3(9.5) + 3(9.5)6\right) = 237$$

$$^{V}RCI_S = 3 + \left(3 + 3(6)\right) + \left(3 + 3(6) + 3(6)9.5\right) = 216$$

For chiral center 2:

$$\delta_a^V = 5$$

$$\delta_b^V = 3 + \frac{3}{2} + \frac{4}{2} + \frac{6}{4} + \frac{5}{4} = 9.25$$

$$\delta_c^V = 1$$

$${}^V RCI_R = 5 + \left(5 + 5(9.25)\right) + \left(5 + 5(9.25) + 5(9.25)1\right) = 153.75$$

$${}^V RCI_S = 5 + \left(5 + 5(1)\right) + \left(5 + 5(1) + 5(1)9.25\right) = 71.25$$

Combining both chiral centers, where the first 'R' in 'RR' represents chiral center 1 in Figure 4 and the second 'R' the second chiral center leads to the following:

$${}^V RCI_{RR} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(RCI_i\right)^2} = \sqrt{\frac{\left(237^2 + 153.75^2\right)}{2}} = 199.76$$

Section 3. Results

Understanding RCI. While developing a novel molecular descriptor, important factors must be considered. The descriptor should be able to discriminate molecules based on the factor it is designed to discriminate, for this study, chirality. It should also encode information about the molecule, in this case distribution of atoms around a chiral center. Therefore, it was necessary to apply the RCI algorithm to a well-known set of chiral molecules. The data set chosen was the 20 standard α-amino acids. Glycine was not included as it does not contain a chiral center. The 19 remaining standard α-amino acids represent an interesting data set for this particular study. Figure 5 shows the structure of 18 of 19 standard α-amino acids used for this study. Figure 6 shows the structure of proline. Note the residue backbonding onto the main amino acid framework. Therefore, the RCI algorithm was tested rigorously as many molecules were structurally similar. Also, by only varying the residue portion of the molecule, a qualitative understanding of the encoded information was gathered.

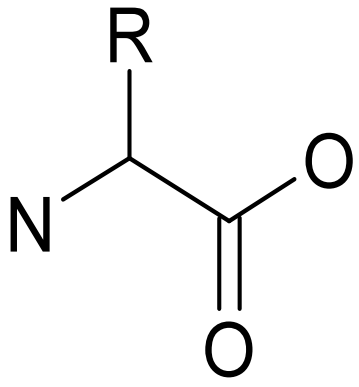


Figure 5. Hydrogen-repressed standard α-amino acid structure representative of 18 of 19 molecules for the data set.

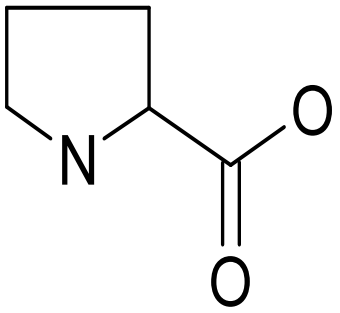


Figure 6. Hydrogen-repressed structure of proline.

Table 1 shows the RCI for various atomic descriptors for all standard α-amino acids containing one chiral center.

Table 1. RCI values for amino acids containing a single chiral center.

	^{S}RCI		^{B}RCI		^{V}RCI		^{I}RCI		^{W}RCI	
amino acid	R	S	R	S	R	S	R	S	R	S
alanine	15.00	9.00	19.50	10.50	94.50	43.50	142.67	93.34	23088	22128
serine	21.00	18.00	27.75	21.75	194.25	164.25	224.34	195.01	46128	45680
valine	27.00	27.00	36.00	33.00	151.50	112.50	186.22	147.55	61968	61872
cysteine	21.00	18.00	27.75	21.75	194.25	164.25	178.97	138.52	69312	69379
proline	69.00	63.00	91.50	76.50	252.50	187.50	122.56	84.10	75777	74791
leucine	27.00	27.00	36.00	33.00	180.00	147.00	180.78	140.78	83568	83952
asparagine	27.00	27.00	40.13	38.63	244.13	224.63	333.24	330.57	85008	85424
aspartic acid	27.00	27.00	40.13	38.63	258.38	241.88	259.74	239.07	86448	86896
gutamine	27.00	27.00	37.03	34.41	212.06	185.81	209.37	176.37	99408	100144
lysine	26.25	25.88	34.97	31.59	178.22	144.84	177.38	136.55	105168	106032
glutamic acid	27.00	27.00	37.03	34.41	219.19	194.44	217.54	186.54	106608	107504
methionine	25.50	24.75	33.94	30.19	197.81	168.56	173.98	132.31	109632	110595
histidine	35.00	39.00	47.00	48.00	258.38	241.88	216.18	184.84	118128	119280
phenylalanine	29.50	30.75	49.41	51.28	235.22	213.84	197.81	161.98	132528	134000
arginine	27.00	27.00	31.02	26.20	193.36	163.17	185.89	147.14	145488	147248
tyrosine	29.63	30.94	49.92	51.98	255.70	238.64	203.94	169.60	155568	157552
tryptophan	34.75	38.63	106.03	112.41	282.42	270.98	215.85	184.43	188688	191408

It should be noted that the conventional method for assigning configuration for the 19 chiral, standard α-amino acids is by comparison of the amino acid structure to that of D/L-glyceraldehyde. The L-amino acids are represented far more frequently than their D-amino acid counterparts in nature. Generally, the L-amino acids are characterized as having S- absolute configuration and D-amino acids are characterized as having R-absolute configuration.

The 19 standard α-amino acids, also, contained two molecules that have two chiral centers. Table 2 shows the RCI values for each of the stereoisomers of threonine and isoleucine.

Table 2. RCI values for amino acids containing two chiral centers.

amino acid	absolute configuration	^{S}RCI	^{B}RCI	^{V}RCI	^{I}RCI	^{W}RCI
D-allo-threonine	RR	23.55	29.74	199.76	225.49	54857.8
D-threonine	RS	20.48	26.67	174.99	200.55	54134.4
L-threonine	SR	23.55	27.95	187.48	213.07	54857.8
L-allo-threonine	SS	20.48	24.65	160.83	186.48	54134.4
D-isoleucine	RR	46.02	54.39	158.09	153.43	76583.3
D-allo-isoleucine	RS	36.07	43.63	150.61	151.17	75659.7
L-allo-isoleucine	SR	46.52	53.84	140.21	130.71	76793.0
L-isoleucine	SS	36.70	42.95	131.72	128.05	75872.0

Interpreting the values listed in the first two tables, it is reasonable to conclude that the RCI algorithm is able to distinguish enantiomers and diastereomers. However, meso compounds are neither enantiomers or diastereomers. Meso compounds are not optically active due to an internal plane of symmetry. Since meso compounds are not optically active, it is important that the RCI have degenerate values for each configuration that is the same. Three RCIs for tartaric acid were calculated to show that this was in fact true. This data is depicted in Table 3.

Table 3. RCI values for tartaric acid.

absolute configuration	^{V}RCI	^{I}RCI	^{W}RCI
RR	596.88	528.43	117861.0
RS	**600.64**	**529.93**	**118389.2**
SR	**600.64**	**529.93**	**118389.2**
SS	604.38	531.42	118915.0

Section 4. Discussion

A successful molecular descriptor should be able to discriminate molecules based on some aspects of molecular structure. To ensure that the various weights applied to atoms/vertices in the the RCI calculation algorithm are in fact encoding information while retaining their ability to discriminate, the single chiral center amino acids were ordered by their RCI values for the R- enantiomer. Information regarding discrimination, degeneracy, and inconsistencies were then more easily found.

By ordering the amino acids by their ^{S}RCI values, it is apparent that multiple degeneracies are present. Table 4 depicts the ^{S}RCI for the 17 amino acids.

Table 4. Ordered ^{S}RCI values for amino acids 17 amino acids.

	^{S}RCI	
amino acid	R	S
alanine	15.00	9.00
cysteine	**21.00**	**18.00**
serine	**21.00**	**18.00**
methionine	25.50	24.75
lysine	26.25	25.88
arginine	**27.00**	**27.00**
asparagine	**27.00**	**27.00**
aspartic acid	**27.00**	**27.00**
glutamic acid	**27.00**	**27.00**
gutamine	**27.00**	**27.00**
leucine	**27.00**	**27.00**
valine	**27.00**	**27.00**
phenylalanine	29.50	30.75
tyrosine	29.63	30.94
tryptophan	34.75	38.63
histidine	35.00	39.00
proline	69.00	63.00

Noticeably, there are two groups that stand out. By neglecting atom type, cysteine and serine both contain the same branching pattern. Similarly, by neglecting atom type, arginine, asparagine, aspartic acid, glutamic acid, glutamine, leucine, and valine all contain the same simple connectivity. Moreover when neglecting atom type, each of

these amino acids contains a pseudo plane of internal symmetry and are basically seen as not being chiral. Therefore, the ^{S}RCI cannot reliably discriminate between amino acids or enantiomers.

Incorporating bond type as well as simple connectivity, it is reasonable that some of the degeneracies present in the ^{S}RCI would be negated. Table 5 shows the ^{B}RCI for each of the 17 single-chiral-center amino acids. Cysteine and serine continue to exhibit degenerate values based on bond connectivity. The second group of amino acids that displayed degeneracies in the ^{S}RCI is reduced to leucine, valine, glutamic acid, gutamine, asparagine, and aspartic acid. However, the ^{B}RCI now successfully discriminates all pairs of enantiomers.

Table 5. Ordered ^{B}RCI values for amino acids 17 amino acids.

	^{B}RCI	
amino acid	R	S
alanine	19.50	10.50
cysteine	**27.75**	**21.75**
serine	**27.75**	**21.75**
arginine	31.02	26.20
methionine	33.94	30.19
lysine	34.97	31.59
leucine	**36.00**	**33.00**
valine	**36.00**	**33.00**
glutamic acid	**37.03**	**34.41**
gutamine	**37.03**	**34.41**
asparagine	**40.13**	**38.63**
aspartic acid	**40.13**	**38.63**
histidine	47.00	48.00
phenylalanine	49.41	51.28
tyrosine	49.92	51.98
proline	91.50	76.50
tryptophan	106.03	112.41

Introducing the atom type into the ^{B}RCI, the ^{V}RCI can be calculated. Table 6 shows the ordering of the 17 amino acids for the ^{V}RCI. There are two remaining degeneracies present. Cysteine and serine remain degenerate. That is because the only

difference between the cysteine and serine are the thiol and hydroxyl groups, respectively, and the δ^V formulation used in this study only encodes information regarding valence shell electrons and disregards core electrons. The second degeneracy is aspartic acid and histidine. This is a chance degeneracy since both aspartic acid and histidine have drastically different structures and have not exhibited the same degeneracy in either the ^{S}RCI or ^{B}RCI. Another interesting piece of data relates to proline. In Tables 4 and 5, both the R and S values for their respective RCI are listed in order. In Table 6, the S value for proline does not fall in order with those around it. It should be pointed out that since proline has a different structure than the other amino acids, namely the R group binding back onto the amino acid group, that this could explain the amino acids' order.

Table 6. Ordered ^{V}RCI values for amino acids 17 amino acids.

	^{V}RCI	
amino acid	R	S
alanine	94.50	43.50
valine	151.50	112.50
lysine	178.22	144.84
leucine	180.00	147.00
arginine	193.36	163.17
cysteine	**194.25**	**164.25**
serine	**194.25**	**164.25**
methionine	197.81	168.56
gutamine	212.06	185.81
glutamic acid	219.19	194.44
phenylalanine	235.22	213.84
asparagine	244.13	224.63
proline	252.50	187.50
tyrosine	255.70	238.64
aspartic acid	**258.38**	**241.88**
histidine	**258.38**	**241.88**
tryptophan	282.42	270.98

Using the intrinsic state the ^{I}RCI was calculated and the results are displayed in Table 7. The problem encountered with proline in the ^{V}RCI no longer is present and all

amino acids are ordered. Each amino acid has a unique value for the ^{I}RCI. Therefore, the ^{I}RCI can discriminate between enantiomers and other amino acids.

Table 7. Ordered ^{I}RCI values for amino acids 17 amino acids.

	^{I}RCI	
amino acid	R	S
proline	122.56	84.10
alanine	142.67	93.34
methionine	173.98	132.31
lysine	177.38	136.55
cysteine	178.97	138.52
leucine	180.78	140.78
arginine	185.89	147.14
valine	186.22	147.55
phenylalanine	197.81	161.98
tyrosine	203.94	169.60
gutamine	209.37	176.37
tryptophan	215.85	184.43
histidine	216.18	184.84
glutamic acid	217.54	186.54
serine	224.34	195.01
aspartic acid	259.74	239.07
asparagine	333.24	330.57

Switching the atomic descriptor used to formula weight, the distribution of formula weight around the chiral center was established. Table 8 depicts the ordering for the ^{W}RCI. This index also does not display any degeneracies and no amino acids have values that are out of order.

Table 8. Ordered ^{W}RCI values for amino acids 17 amino acids.

	^{W}RCI	
amino acid	R	S
alanine	23088	22128
serine	46128	45680
valine	61968	61872
cysteine	69312	69379
proline	75777	74791
leucine	83568	83952
asparagine	85008	85424
aspartic acid	86448	86896
gutamine	99408	100144
lysine	105168	106032

glutamic acid	106608	107504
methionine	109632	110595
histidine	118128	119280
phenylalanine	132528	134000
arginine	145488	147248
tyrosine	155568	157552
tryptophan	188688	191408

Combining the ordering of Tables 4, 5, 6, 7, and 8, some general trends can be established and shown in Table 9. Alanine, due to the residue being a methyl group, has the lowest values in four of the five indicies. Tryptophan with ten non-hydrogen atoms in its residue has the largest values in three of the five indices shown.

Table 9. Amino acids ordered by each RCI.

	^{S}RCI	^{B}RCI	^{V}RCI	^{I}RCI	^{W}RCI
1	alanine	alanine	alanine	proline	alanine
2	cysteine	cysteine	valine	alanine	serine
3	serine	serine	lysine	methionine	valine
4	methionine	arginine	leucine	lysine	cysteine
5	lysine	methionine	arginine	cysteine	proline
6	arginine	lysine	cysteine	leucine	leucine
7	asparagine	leucine	serine	arginine	asparagine
8	aspartic acid	valine	methionine	valine	aspartic acid
9	glutamic acid	glutamic acid	gutamine	phenylalanine	gutamine
10	gutamine	gutamine	glutamic acid	tyrosine	lysine
11	leucine	asparagine	phenylalanine	gutamine	glutamic acid
12	valine	aspartic acid	asparagine	tryptophan	methionine
13	phenylalanine	histidine	proline	histidine	histidine
14	tyrosine	phenylalanine	tyrosine	glutamic acid	phenylalanine
15	tryptophan	tyrosine	aspartic acid	serine	arginine
16	histidine	proline	histidine	aspartic acid	tyrosine
17	proline	tryptophan	tryptophan	asparagine	tryptophan

Overall, the RCI seems to discriminate enantiomers. The ^{S}RCI and ^{B}RCI have deficiencies in discriminating which are taken care of in the ^{V}RCI by adding atom type into the descriptor calculation.

RCI in SAR. To ensure that the RCI is actually encoding useful information, the RCI was applied to a series of mosquito repellants. Picaridin® and AI3-37220 (220) , shown in Figure 7, are mosquito repellants that have two chiral centers.[17,18]

AI3-37220-(220) Picaridin®

Figure 7. Mosquito repellants that have two chiral centers.

By calculating the single stereocenter RCI and including the adaptation for multiple chiral centers, a single chirality measure was constructed. Table 10 shows the ^{V}RCI, ^{I}RCI,^{W}RCI, and the biting proportion for the four stereoisomers for each compound. Biting proportion was obtained from the application of the compound to the skin of human volunteers and quantified by the Klun and Debboun modules.[27]

Table 10. RCI values for mosquito repellants containing two chiral centers. Biting proportions designated with a different letter signify experimental values that are significantly different from one another.

chemical	biting proportion	^{V}RCI	^{I}RCI	^{W}RCI
Picardin				
RR	0.22 a	242.38	121.92	407664.01
RS	0.18 a	230.49	112.82	407081.22
SR	0.40 b	234.29	119.08	406731.30
SS	0.44 b	221.96	109.74	406147.17
control	0.72 c			
220				
RR	0.56 c	249.32	100.19	332358.42
RS	0.32 b	233.86	96.59	329904.32
SR	0.51 c	242.07	100.89	332519.07
SS	0.18 a	302.44	119.41	330066.20
control	0.83 d			

The order of biting proportion for the diastereomers with the order of the RCI is reversed in the case of 220 and Picaridin. 220 has an increasing order of RCI that matches the

decreasing order of biting proportion for the diastereomers; for Picaridin diastereomers the lower the RCI the lower the biting proportion.

Section 5. Conclusion

The RCI algorithm was developed to discriminate chiral carbon compounds based on the Cahn-Ingold-Prelog rules for defining absolute configuration for these compounds. To this end, this study proposed an attempt to encode information about the distribution of the atoms surrounding a chiral center. The set of α-amino acids containing at least one chiral center was used to test the various features of the RCI models. Using these example calculations, the RCI established that it differentiates molecules from each other. The RCI, also, seems to encode important information about molecules as seen by the analysis of the order of the various RCI indices. By applying the RCI method to the mosquito repellant experimental data, it was shown that this index could be useful in the modeling of experimental data.

Section 6. References

1. Ariens, E. J.; Racemic therapeutics: problems all along the line. In *Chirality in Drug Design and Synthesis*, Brown, C., Ed.; Academic Press: New York, 1990; pp 29–43.

2. Ariëns, E. J.; Stereoselectivity in bioactive agents: general aspects. In *Stereochemistry and Biological Activity of Drugs*. Ariëns, E. J.; Soudijn, W.; Timmermans, P.B.M.W.M., Eds.; Balckwell Scientific Publications: Oxford, U.K., 1983; pp 11-32.

3. Kurihara, N.; Miyamoto, J., Eds. *Chirality in Agrochemicals,* Wiley Series in Agrochemicals and Plant Protection, John Wiley & Sons: New York, 1998; p 269.

4. Klun, J.A.; Schmidt, W.F.; Debboun, M. Stereochemical Effects in an Insect Repellent, *J. Med. Entomol*. **2001,** *38*, 809-812.

5. Trinajstic, N. *Chemical Graph Theory*, 2nd revised ed.; CRC Press: Boca Raton, FL 1992. ISBN: 0849342562.

6. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoninformatics*. Vol. II; Wiley-VCH verlag GmbH & Co. KGaA: Weinheim, 2009; ISBN: 9783527318520.

7. Schultz, H.P.; Schultz, E.B.; Schultz, T.P. Topological Organic Chemistry. 9. Graph Theory and Molecular Topological Indices of Stereoisomeric Organic Compounds. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 864-870.

8. Golbraikh, A.; Bonchev, D.; Tropsha A. Novel Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 147-158.

9. Randic, M. Graph Theoretical Descriptors of Two-Dimensional Chirality with Possible Extension to Three-Dimensional Chirality. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 639-649.

10. Aires-de-Sousa, J.; Gasteiger, J. New Description of Molecular Chirality and Its Applications to the Prediction of the Preferred Enantiomer in Stereoselective Reactions. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 369-375.

11. Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidovi_, D. Chirality Code and Molecular Structure. *J. Chem. Inf. Comput. Sci*. **2004**, *44*, 831-836.

12. Capozziello, S.; Lattanzi, A. Algebraic Structure of Central Molecular Chirality Starting From Fischer Projections. *Chirality*. **2003**, *15*, 466-471.

13. Capozziello, S.; Lattanzi, A. Molecular Approach to Central Molecular Chirality: A Chirality Selection Rule. *Chirality*. **2003**, *15*, 227-230.

14. Capozziello, S.; Lattanzi, A. Description of chiral tetrahedral molecules via an Aufbau approach. *Theochem*. **2003**, *671*, 205-209.

15. Yang, C.; Zhong, C. Chirality Factors and Their Application to QSAR Studies of

Chiral Molecules, *QSAR Comb. Sci.* **2005**, *24*, 1047 – 1055.

16. Harris, A.B.; Kamien, R.D.; Lubensky, T.C. Molecular chirality and chiral parameters. *Rev. Mod. Phys.* **1999**. *71*, 1745-1757.

17. Basak, S.C., Natarajan, R., Nowak, W., Miszta, P, Klun, J.A, Three Dimensional Structure-activity Relationships (3D-QSAR) for Insect Repellency of Diastereoisomeric Compounds: A Hierarchical Molecular Overlay Approach. *SAR QSAR Environ. Res.* **2007** in press

18. Natarajan, R.; Basak, S.C; Balaban, A.T.; Klun, J.A.; Schmidt, W.F. Chirality Index, Molecular Overlay and Biological Activity of Diastereoisomeric Mosquito Repellents. *Pest Manag. Sci.* **2005**, *61*, 1193-1201.

19. Kier, L.B.; Hall, L. *Molecular Connectivity Analysis*, Research Studies Press: Letchworth, UK, 1986; pp 1-26.

20. Kier, L.B.; Hall, L.H *Molecular Structure Description*, Academic Press: New York, 1999; pp 13-36.

Appendix

Novel Approach for the Numerical Characterization of Molecular Chirality†

Ramanathan Natarajan,‡ Subhash C. Basak,*,‡ and Terrence S. Neumann§

Natural Resources Research Institute, University of Minnesota Duluth, 5013 Miller Trunk Highway, Duluth, Minnesota 55811, and Department of Chemistry and Biochemistry, University of Minnesota Duluth, 1039 University Drive, Duluth, Minnesota 55812

Received December 1, 2006

The use of chiral compounds as pharmaceuticals and agrochemicals continues to increase, warranting numerical characterization of chirality in order to develop structure−activity relationship models involving these compounds. Enantiomers are identical in all scalar properties and, hence, are not differentiated by topological indices and 3-D descriptors. Three distinct measures of chirality were developed to discriminate diastereomers and enantiomers. The novel topological indices treat chirality as a continuous measure, and hence we prefer to call it the Relative Chirality Index (RCI). Application of RCI in developing SAR is illustrated with the repellency data for the diastereomers of picaridin and AI3-37220.

INTRODUCTION

Biological discrimination of stereoisomers is well-known and has been illustrated with examples such as (*R*)-(+)-limonene which has orange-like odor, while the (*S*)-(−)-limonene has the smell of lemon. Several reviews and monographs highlight the pharmacological implications of chiral discrimination.[1,2] However, one of the incidents that had a very high impact and changed the line of thinking of the pharmaceutical industry and the policy makers was the thalidomide episode. Racemic thalidomide was introduced as a sedative and an antinausea agent. The teratogenicity of (*S*)-(−)thalidomide led to the high incidence of fetal abnormalities and children born to women who used thalidomide had deformities of limbs. The less potent or the inactive enantiomer present in a racemate affects nontarget cells and to avoid this risk enantiopure or single isomer drugs should be used. The Food and Drug Administration (FDA) requires that both enantiomers of a racemate be studied in detail before marketing a racemate. Several pharmaceutical companies are going for the redevelopment in single-isomer forms of chiral drugs that were originally approved for marketing as racemates. This practice is called a racemic switch. Other important factors that gave impetus in using enantiopure drugs are the new analytical tools to separate enantiomers and improved methodologies to synthesize enantiopure compounds. In addition to this, there are methods available to monitor the therapeutic action of the chiral drugs and their metabolites. In the case of agrochemicals also there is an increasing trend to use single isomer or enantiopure compounds,[3] and this reduces loading the environment with isomers that have low or no efficacy. Recently Klun et al.[4] demonstrated the stereospecificity of insect repellants in the case of AI3-37220 and picaridin.

Computational chemists are now facing the challenge of developing models to predict the beneficial as well as adverse effects of diastereoisomers (diastereomers). In such situations quantitative stereochemical structure−activity relationship modeling (QSSAR) approach is necessary rather than simple quantitative structure−activity relationship (QSAR). Conventional QSAR modeling using simple computed molecular descriptors or physicochemical properties fail in handling compounds that exhibit polychiral diastereomerism. The reason for the limitation of the molecular descriptors derived from adjacency and distance matrices of molecular graphs in differentiating enantiomers is their identical scalar properties. In other words, enantiomers are isometric, and for each distance between two given atoms in one isomer there is a corresponding identical distance between a pair of atoms in the other. Thus, distance matrices for the enantiomers have identical entries, and consequently the various topological indices derived from the distance matrices could not differentiate enantiomers. The same is true for three-dimensional (3-D) distance matrices of enantiomers.

In order to develop QSSAR, descriptors that encode the stereochemical features of a molecule are the primary requisite. Some attempts have been made by Schultz and co-workers[5] to develop topological indices to differentiate *cis*(*Z*) and *trans*(*E*) isomers as well as diastereomers arising out of chirality of molecules. They used vertex weighted distance matrices to compute descriptors and then added to the descriptors correction terms for different diastereomers. Golbraikh, Bonchev, and Tropsha[6] used a similar approach, applying chirality corrections to Zagreb indices, molecular connectivity indices, extended connectivity indices, overall connectivity indices, and topological charge indices. They implemented these chirality indices in QSAR modeling of ecdysteroids and compared the results with that of Comparative Molecular Field Analysis (CoMFA). Randić's[7] approach to the problem of chirality was confined to two-dimensional chirality and is applicable to benzenoid hydrocarbons. Recently, Fujita[8−10] suggested sphericity indices for the enumeration of stereoisomers. Aires-de-Sousa et al.[11−12]

† Dedicated to Professor Nenad Trinajstić on the occasion of his 70th birthday.

* Corresponding author phone: (218)720-4230; fax: (218)720-4328; e-mail: sbasak@nrri.umn.edu.

‡ NRRI-UMD.

§ UMD.

10.1021/ci600542b CCC: $37.00
Published on Web 04/05/2007

Figure 1.

attempted to describe chirality using a "chirality code". A chirality-sensitive flexibility (CSF) descriptor based on the distance between a pharmacophore point and a plane defined by three pharmacophore points was introduced by Dervarics et al.[13] Capozziello and Lattanzi[14] derived algebraic structures of molecular chirality from simple Fischer projections. They used the approach to define a chirality index[15,16] to characterize the global chirality of a structure and suggested an aufbau (built up) method to predict the same property for a new compound. Yang and Zhong[17] suggested that a chirality factor be applied to several commonly used topological indices to calculate molecular descriptors for chiral molecules and tested their applicability in QSAR studies. However, these indices are calculated by applying the chiral correction to the commonly used topological indices. Hence, they depend on the computation of other topological indices.

In several of these studies, chirality is treated as an either—or property, i.e., on a black or white scale. Activities of many enzymes and biological activities of chiral molecules are not either—or properties, and so, chirality must be treated as a continuous measure in developing structure—activity relationships. This could be substantiated with the insect repellency of picaridin and AI3-37220[5,18,19] and the HIV protease inhibition activity of chiral sulfonamide-substituted cyclooctylpyranones.[20] Hence, there is a need to develop molecular descriptors that could not only differentiate enantiomers and diastereomers but also treat chirality as a continuous measure. Natarajan and Basak[21] attempted to develop such indices using ^{1}H NMR chemical shift values by converting a NMR spectrum into a linear graph. However, this method requires NMR data for a compound under investigation. The present study describes the calculation of new chirality descriptors to discriminate enantiomers and diastereoisomers. Application of the new chirality descriptors in SAR modeling is illustrated with mosquito repellency of the diastereomers of picaridin and AI3-37220.

METHODS

Calculation of Chirality Indices. Three-point interaction models are always proposed for biodiscrimination of enantiomers, and this formed the basis for the calculation of the new indices. Instead of considering all possible spatial orientations for a given configuration we preferred to use the single representation of spatial arrangement of groups around a given chiral center (Figure 1) in calculating the chirality measure. The reasons for our preference to use a single representation are as follows: Although there are 24 (4! permutations) possible planar projections (Fischer projections) for the four chemical groups (12 for each of the enantiomers, *R* or *S*), three-point correlation is important. The importance of three point correlation for tetrahedral chiral centers was discussed in detail by Harris et al.[22−24] According to the Cahn-Ingold-Prelog rule, different degrees of priorities are assigned to the four chemical groups attached to the chiral carbon, "*a*" being given highest priority, then "*b*", etc. Difference in the disposition of the groups *a*, *b*, *c*, and *d* around the asymmetric carbon is given in Figure 1. The least important chemical group (*d*) is placed at the rear, and the clockwise or the anticlockwise arrangement of the other three groups (*a*, *b*, *c*) is considered to assign the configuration as *R* or *S*. So, all the 12 projections of an enantiomer are reduced to only one representation to assign the absolute configuration.

In the approach developed in this paper, the three groups of highest priority attached to a chiral center were viewed from a reference point to calculate the new chirality metric. The groups/atoms *a*, *b*, *c*, and *d* are then assigned valence delta values of atoms (δ^v) according to the method of Hall and Kier.[25] When the group has more than one atom, δ^v for the group *a*, *b*, or *c* is calculated considering the relative proximities of the atoms to the chiral center, and decreasing importance with increasing topological distance (through bond) was assigned while calculating the contribution of atoms other than hydrogen in a group. The group delta value for any group (δ^v_i) attached to a chiral carbon is calculated as

$$\delta^v_i = \delta^v_{n1} + (\delta^v_{n2}/2) + (\delta^v_{n3}/4) + (\delta^v_{n4}/8) + ...$$

where n_1 is the atom attached directly to the chiral center (nearest neighbor), n_2 is separated by one atom, n_3 by two atoms, etc. Relative chirality indices (^{p}RCI) for a pair of enantiomers are calculated as

$$^{p}\mathrm{RCI}_R = \delta^v_a + (\delta^v_a + \delta^v_a\delta^v_b) + (\delta^v_a + \delta^v_a\delta^v_b + \delta^v_a\delta^v_b\delta^v_c) + \delta^v_a\delta^v_b\delta^v_c\delta^v_d$$

$$^{p}\mathrm{RCI}_S = \delta^v_a + (\delta^v_a + \delta^v_a\delta^v_c) + (\delta^v_a + \delta^v_a\delta^v_c + \delta^v_a\delta^v_c\delta^v_b) + \delta^v_a\delta^v_c\delta^v_b\delta^v_d$$

For example,

$$^{p}\mathrm{RCI}_R = 5 + (5 + 5(3)) + (5 + 5(3) + 5(3)(1)) + 0 = 60$$

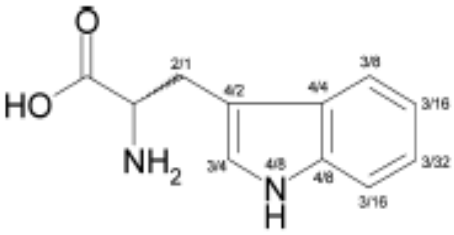
$$^{p}\mathrm{RCI}_S = 5 + (5 + 5(1)) + (5 + 5(1) + 5(1)(3)) + 0 = 40$$

To obtain ^{p}RCI for molecules containing more than one chiral center, root-mean-square of ^{p}RCI for all the chiral centers is taken.

$$^{p}\mathrm{RCI} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(^{p}\mathrm{RCI}_i)^2}$$

Calculation of group δ^v values for cyclic structures is illustrated with tryptophan (Figure 2). Figure 2 examples the calculation of ^{p}RCIs for *R* and *S* isomers of tryptophan. Information regarding the fourth group is encoded in the new index by the fourth term $\delta_a\delta_b\delta_c\delta_d$. When "*d*" is hydrogen, $\delta_a\delta_b\delta_c\delta_d$ becomes zero; otherwise it contributes to the RCI for the chiral center.

In addition to the valence connectivity, we also used the formula weights of the groups and the electrotopological state of the various atoms[26] and groups in *a*, *b*, *c*, and *d* to calculate ^{w}RCI and ^{h}RCI, respectively. In order to calculate ^{h}RCI, one



$$\delta_a = 3$$

$$\delta_b = \frac{4}{1} + \frac{6}{2} + \frac{5}{2} = 9.5$$

$$\delta_c = \frac{2}{1} + \frac{4}{2} + \frac{3}{4} + \frac{4}{4} + \frac{3}{8} + \frac{4}{8} + \frac{4}{8} + \frac{3}{16} + \frac{3}{16} + \frac{3}{32} = 7.59375$$

$${}^{V}RCI_R = 288.42$$

$${}^{V}RCI_S = 270.98$$

Figure 2. Calculating of group δ and RCI for tryptophan. $\delta_a = 3$, $\delta_b = 4/1 + 6/2 + 5/2 = 9.5$, $\delta_c = 2/1 + 4/2 + 3/4 + 4/4 + 3/8 + 4/8 + 4/8 + 3/16 + 3/16 + 3/32 = 7.59375$, ${}^{V}RCI_R = 288.42$, ${}^{V}RCI_S = 270.98$.

has to replace δ^v of the groups *a*, *b*, *c*, and *d* by their δ^1 (electrotopological state) in the formula to calculate ^{V}RCIs. In the same way, ^{W}RCI for enantiomers and diastereomers can be calculated using the formula weights of *a*, *b*, *c*, and *d*. However, while calculating formula weights for groups we considered the whole substituent instead of using diminishing importance of contribution of atoms based on their neighborhood. It is to be noted that simple connectivity and bond order connectivity of vertices (atoms) were not useful in generating RCI due to their inability to differentiate atom types and this consequently gave degenerate values for several structures.

RESULTS AND DISCUSSION

Calculation of RCI for α-Amino Acids. The RCI calculated from all three methods for the 17 biologically important chiral α-amino acids are given in Table 1. In the case of α-amino acids having 2 chiral centers the values for the two pairs of enantiomers are given in Table 2. RCI can be calculated considering chirality at a given center in the case of a polychiral molecule, and this will be very useful in finding out the relative importance of chirality of the various centers with respect to one another. Hence, it is possible to calculate a set of RCI for a polychiral system taking into consideration one chiral center only at a time. This is illustrated in the case of isoleucine and threonine, and the values are given in Table 2. Calculation of the new chiral metric considering specific chiral centers is expected to bring out the importance of chirality at a particular chiral atom. Distribution of atoms according to their mass around a chiral center is encoded in ^{W}RCI, while ^{P}RCI and ^{S}RCI quantify the distribution of groups based on branching and electrotopological states, respectively. Thus the "handedness" based on three different measures are encoded by the chirality indices. For example, bioisosteric amino acids, serine and cysteine have the same set of ^{P}RCI (${}^{P}RCI_R = 194.250$; ${}^{P}RCI_S = 164.250$). However, the other two RCI values are different for these two amino acids indicating their difference with respect to chemical nature and other aspects of chirality. ^{V}-RCI is degenerate for aspartic acid and histidine, and this type of degeneracy can be handled by using correction factors for cyclic structures. RCI values for the diastereomers of tartaric acid (Table 2) showed that the two *meso* forms have the same RCI values indicating that they are nondistinguishable. The primary objective of a molecular descriptor is to map the structural information into a set of real numbers, and the descriptors thus generated must be able to discriminate different structures. The above discussion shows that the new chirality indices satisfy this requirement. However, their applicability in SAR modeling is the next most important test of their usefulness.

Table 1. RCI Values for Amino Acids Containing a Single Chiral Center

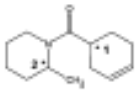
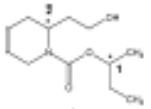
		^{W}RCI		^{P}RCI		^{S}RCI	
no.	amino acid	*R*	*S*	*R*	*S*	*R*	*S*
1	alanine	23088	22128	94.50	43.50	142.67	93.34
2	arginine	145488	147248	193.36	163.17	185.89	147.14
3	asparagine	85008	85424	244.13	224.63	333.24	330.57
4	aspartic acid	85448	86896	258.38	241.88	259.74	239.07
5	cysteine	69312	69879	194.25	164.25	178.97	138.52
6	glutamic acid	106608	107504	219.19	194.44	217.54	186.54
7	glutamine	99408	100144	212.06	185.81	209.37	176.37
8	histidine	118128	119280	258.38	241.88	216.18	184.84
9	leucine	83568	83952	180.00	147.00	180.78	140.78
10	lysine	105168	106032	178.22	144.84	177.38	136.55
11	methionine	109632	110595	197.81	168.56	173.98	132.31
12	phenylalanine	132528	134000	235.22	213.84	197.81	161.98
13	proline	75777	74791	252.50	187.50	122.56	84.10
14	serine	46128	45680	194.25	164.25	224.34	195.01
15	tryptophan	188688	191408	282.42	270.98	215.85	184.43
16	tyrosine	155568	157552	255.70	238.64	208.94	169.60
17	valine	61968	61872	186.22	147.55	151.50	112.50

Table 2. RCI Values for Amino Acids Containing Two Chiral Centers and Tartaric Acid

amino acid	absolute configuration	^{W}RCI	^{P}RCI	^{S}RCI
D-allo-threonine	*RR*	54857.8	199.76	225.49
D-threonine	*RS*	54134.4	174.99	200.55
L-threonine	*SR*	54857.8	187.48	213.07
L-allo-threonine	*SS*	54134.4	160.83	186.48
D-isoleucine	*RR*	76883.3	158.09	153.43
D-allo-isoleucine	*RS*	75659.7	150.61	151.17
L-allo-isoleucine	*SR*	76793.0	140.21	130.71
L-isoleucine	*SS*	75872.0	131.72	128.05
tartaric acid	*RR*	117861.0	596.88	528.43
	RS	115539.2	600.64	529.93
	SR	115539.2	600.64	529.93
	SS	118915.0	604.38	531.42

Application of RCI in SAR. In order to test the application of RCI in structure−activity relationship modeling we considered the insect repellency data for diastereomers of AI3-37220 and picaridin. Both picaridin and AI3-37220 have two asymmetric centers (structures are given in Table 3), and the four diastereoisomers (*1R,2′R*, *1R,2′S*, *1S,2′R*, and *1S,2′S*) of each compound were shown to have differing degrees of mosquito-repellent activity.[5] The order of repellency (proportion of biting) of the diastereomers of AI3-37220 and picaridin are given in Table 3. In situations such as this where activities of diastereomers are to be modeled, conventional QSAR using calculated molecular descriptors failed to discriminate chemicals. In our earlier attempt[19] to

Table 3. RCI and Repellency Data for AI3-37220 and Picaridin Diastereomers

chemical	biting[a] proportion	[a]RCI	[b]RCI	[c]RCI
AI3-37220				
				
1*R*,2′*R*	0.56c	332358.4	249.32	100.19
1*R*,2′*S*	0.32b	329904.3	233.86	96.59
1*S*,2′*R*	0.51c	332519.1	242.07	100.89
1*S*,2′*S*	0.18a	330066.2	226.11	97.31
control	0.83d			
picaridin				
				
1*R*,2′*R*	0.22a	407664.0	242.38	121.92
1*R*,2′*S*	0.18a	407081.2	230.49	112.82
1*S*,2′*R*	0.40b	406731.3	234.29	119.08
1*S*,2′*S*	0.44b	406147.2	221.96	109.74
control	0.72c			

[a] Proportions followed by the different letters are significantly different from one another at $P = 0.05$. Repellency data are from refs 4 and 15. Lower the biting proportion higher the repellency.

use the Schultz method of discriminating diastereomers, we found that the *RR* isomer always had the largest numerical value of the index and the *SS* the smallest irrespective of the substituents attached to the asymmetric carbon. This limitation in using the Schultz approach is eliminated in the new indices described in this paper. However, the order of repellency of diastereomers with the order of RCI is reversed in the case of AI3-37220 and picaridin (see Table 3). In the case of AI3-37220, the increasing order of RCI matches with the increasing order of repellency of the diastereomers; for picaridin diastereomers, the lower the RCI the higher the repellency. It is to be noted that AI3-37220 is an amide and picaridin is a carbamate. The difference in their chemical nature might be one of the possible reasons for the inverse relation observed between their repellencies. In SARs we try to find the mathematical association of structure/properties of molecules and chemical/ biological activities. The key to identifying the most relevant molecular descriptors depends on the understanding of the mechanism of action. Correlation of a descriptor to a biological activity does not guarantee causality (mechanism of action). Moreover, there is no reason to assume a priori that steric effects will totally determine bioactivity. The CIP rule uses atomic number (≈ size) to distinguish different substituents if one follows the approach to develop quantitative indices that will lead to numerical discrimination of diastereomers and enantiomers. What a particular biological receptor senses in a chiral molecule may vary from one receptor to another. If the ordering of a set of chiral molecules by calculated indices parallels the ordering of them by the receptor, then and only then can we expect a correlation between the calculated RCI and the bioactivity. Otherwise, the indices will only discriminate the structures without any necessary correlation with biological function.

CONCLUSION

The three distinct classes of indices of chirality take into account the distribution of groups around a chiral center based on different measures of "handedness". They are able to discriminate diastereomers and enantiomers very well. Singularity of *meso* compounds is also addressed because the *meso* compounds have the same values for all three RCIs. As they are computed based on a three-point fitting in receptors, they are expected to perform well in SAR modeling of biological activities of diastereomers. The new indices might be used along with the other commonly used topological indices to handle enantiomeric and diastereomeric compounds. The new chiral indices are conceptually very simple and can be computed easily for simple molecules. However, for large molecules like peptides and proteins we need to develop a computer program.

ACKNOWLEDGMENT

Research was supported by Grant F49620-02-1-0138 from the United States Air Force. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official views, policies, or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government. We acknowledge Natural Resources Research Institute, University of Minnesota Duluth for the financial support for one of the authors (R.N.). This paper is contribution number 444 from the Center for Water and the Environment, Natural Resources Research Institute, University of Minnesota Duluth, Duluth, U.S.A.

REFERENCES AND NOTES

(1) Ariens, E. J. Racemic therapeutics: problems all along the line. In *Chirality in Drug Design and Synthesis*; Brown, C., Ed.; Academic Press: New York, 1990; pp 29−43.
(2) Ariëns, E. J. Stereoselectivity in bioactive agents: general aspects. In *Stereochemistry and Biological Activity of Drugs*; Ariëns, E. J., Soudijn, W., Timmermans, P. B. M. W. M., Eds.; Blackwell Scientific Publications: Oxford, U.K., 1983; pp 11−32.
(3) *Chirality in Agrochemicals*; Kurihara, N., Miyamoto, J., Eds.; *Wiley* Series in Agrochemicals and Plant Protection, John Wiley & Sons: New York, 1998; p 269.
(4) Khan, J. A.; Schmidt, W. F.; Debboun, M. Stereochemical Effects in an Insect Repellent. *J. Med. Entomol.* **2001**, *38*, 809−812.
(5) Schultz, H. P.; Schultz, E. B.; Schultz, T. P. Topological Organic Chemistry. 9. Graph Theory and Molecular Topological Indices of Stereoisomeric Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 864−870.
(6) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147−158.
(7) Randić, M. Graph Theoretical Descriptors of Two-Dimensional Chirality with Possible Extension to Three-Dimensional Chirality. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 639−649.
(8) Fujita, S. Graphs to Chemical Structures 1. Sphericity Indices of Cycles for Stereochemical Extension of Polya's Theorem. *Theor. Chem. Acc.* **2005**, *113*, 73−79.
(9) Fujita, S. Graphs to Chemical Structures 2. Extended Sphericity Indices of Cycles for Stereochemical Extension of Polaya's Coronas. *Theor. Chem. Acc.* **2005**, *113*, 80−86.
(10) Fujita, S. Graphs to Chemical Structures 3. General Theorems with the Use of Different Sets of Sphericity Indices for Combinatorial Enumeration of Nonrigid Stereoisomers. *Theor. Chem. Acc.* **2006**, *115*, 37−53.
(11) Aires-de-Sousa, J.; Gasteiger, J. New Description of Molecular Chirality and Its Application to the Prediction of the Preferred

Enantiomer in Stereoselective Reactions. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 369−375.
(12) Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidović, D. Chirality Code and Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 831−836.
(13) Dervarics, M.; Otyes, F.; Martinek, T. A. Development of a Chirality-Sensitive Flexibility Descriptor for 3+3D-QSAR. *J. Chem. Inf. Model.* **2006**, *46*, 1431−1438.
(14) Caporziello, S.; Lattanzi, A. Algebraic Structure of Central Molecular Chirality Starting From Fischer Projections. *Chirality* **2003**, *15*, 466−471.
(15) Caporziello, S.; Lattanzi, A. Molecular Approach to Central Molecular Chirality: A Chirality Selection Rule. *Chirality* **2003**, *15*, 227−230.
(16) Caporziello, S.; Lattanzi, A. Description of chiral tetrahedral molecules via an Aufbau approach. *Theochem* **2003**, *671*, 205−209.
(17) Yang, C.; Zhong, C. Chirality Factors and Their Application to QSAR Studies of Chiral Molecules. *QSAR Comb. Sci.* **2005**, *24*, 1047−1055.
(18) Basak, S. C.; Natarajan, R.; Nowak, W.; Mazza, P.; Klun, J. A. Three Dimensional Structure-activity Relationships (3D-QSAR) for Insect Repellency of Diastereoisomeric Compounds: A Hierarchical Molecular Overlay Approach. *SAR QSAR Environ. Res.* **2007**, in press.
(19) Natarajan, R.; Basak, S. C.; Balaban, A. T.; Klun, J. A.; Schmidt, W. F. Chirality Index Molecular Overlay and Biological Activity of Diastereoisomeric Mosquito Repellents. *Pest Manage. Sci.* **2005**, *61*, 1193−1201.
(20) Skulnick, H. I.; Johnson, P. D.; Aristoff, P. A.; Morris, J. K.; Lovasz, K. D.; Howe, W. J.; Watenpaugh, K. D.; Janakiraman, M. N.; Anderson, D. J.; Reischer, R. J.; Schwartz, T. M.; Banitt, L. S.; Tomich, P. K.; Lynn, J. C.; Horng, M.-M.; Chong, K.-T.; Hinshaw, R. R.; Dolak, L. A.; Seest, E. P.; Schwende, F. J.; Rush, B. D.; Howard, G. M.; Toth, L. M.; Wilkinson, K. R.; Kaltuk, T. J.; Johnson, C. W.; Cole, S. L.; Zaya, R. M.; Zipp, G. L.; Possert, P. L.; Dalga, R. J.; Zhong, W. Z.; Williams, M. G.; Romines, K. R. Sulfonamide-Substituted Cyclooctylpyranones. *J. Med. Chem.* **1997**, *40*, 1149−1164.
(21) Natarajan, R.; Basak, S. C. NMR Spectral Invariants − A New Class of Descriptors for Diastereomers and Enantiomers. *Croat. Chem. Acta* **2007** (submitted).
(22) Harris, A. B.; Kamien, R. D.; Lubensky, T. C. Molecular chirality and chiral parameters. *Rev. Mod. Phys.* **1999**, *71*, 1745−1757.
(23) Harris, A. B.; Kamien, R. D.; Lubensky, T. C. Microscopic Origin of Cholesteric Pitch. *Phys. Rev. Lett.* **1997**, *78*, 1476−1479.
(24) Harris, A. B.; Kamien, R. D.; Lubensky, T. C. Microscopic Origin of Cholesteric Pitch. *Phys. Rev. Lett.* **1997**, *78*, 2867(erratum).
(25) Kier, L. B.; Hall, L. *Molecular Connectivity Analysis*; Research Studies Press: Letchworth, U.K., 1986; pp 1−26.
(26) Kier, L. B.; Hall, L. H *Molecular Structure Description*; Academic Press: New York, 1999; pp 13−36.

CI600542B