Hypercube Stereoisograms as Efficient Graphs to Represent the Stereoisomeric Relationships of Organic Molecules Containing 1 to 4 Stereogenic Centers

Rosa Elena Arroyo-Carmona^{a,*}, Sylvain Bernès^b, Jenaro L. Varela Caselis^c, Artemio Martínez-Alvarado^d, G. F. Pérez-Sánchez^e, Aarón Pérez-Benítez^{a,*}

^a Facultad de Ciencias Químicas. Benemérita Universidad Autónoma de Puebla. 18 sur y av. San Claudio, col. San Manuel. C. P. 72570. Puebla, Pue. MEXICO

^bInstituto de Física Luis Rivera Terrazas-BUAP. C. P. 72570. Puebla, Pue. MEXICO

^cCUVyT-BUAP. 24 Sur y Av. San Claudio, Col. San Manuel C.P. 72570. Puebla, Pue. MEXICO

^dPreparatoria Tlatlauquitepec, Complejo Regional Nororiental. Av. Revolución No. 35, centro. Tlatlauquitepec. C. P. 73900. Puebla. MEXICO

^eCentro de Investigación en Fisicoquímica de Materiales-ICUAP. C. P. 72000. Puebla. MEXICO

rosa.arroyo@correo.buap.mx, sylvain.bernes@correo.buap.mx, genaro.varela@correo.buap.mx, artemio.martinez@correo.buap.mx, francisco.perezsanchez@correo.buap.mx, aaron.perez@correo.buap.mx

(Received October 23, 2023)

 $^{^{*}}$ Corresponding author.

Abstract

The *n*-dimensional unit hypercubes, Q_n (n = 1 to 4), are used as *n*-dimensional graphs (Stereoisograms) to describe the stereoisomeric (enantiomeric and/or diastereomeric) relationships of molecules containing one to four RS-stereogenic centers. To achieve this goal, the strings of 0's and 1's, used for labelling the vertices of unit hypercubes, are substituted by strings of R's and S's-descriptors coming from the CIP priority rules. The development of those ndimensional unit hypercubes along ordered axes and the application of combinatorial techniques allow to locate the enantiomeric pairs at antipodal vertices (points related by a chirality center, which satisfies the bitwise operation known as "NOT"); whereas diastereomers are related by mirror planes, whose reflection results in a partial permutation of the RS-descriptors. These facts are in good agreement with the literature and the principles of the combinatorial theory applied to such hypercubes. On the other hand, because the maximum number of stereoisomers 2^n is not always held, "pseudo hypercubes" containing one or more "ghost vertices" are proposed to include the meso (achiral) stereoisomers.

1 Introduction

Although the foundations of stereochemistry date back to 1874, with the works of Jacobus H. van 't Hoff Jr. [1] and Joseph A. Le Bel [2], the formal representation of the stereoisomeric relationships that exist among a set of stereoisomers is currently under examination [3, 4]. They established independently that the four substituents connected to a carbon atom can be arranged in space in two different ways, labeled later by R. S. Cahn, C. Ingold and V. Prelog with the chiral descriptors R (R from Rectus = right) and S (S from Sinister = left) (Figure 1) [5].



Figure 1. The chiral descriptors R (rectus = right) and S (sinister = left), coming from the CIP-system, are used to describe the two possible arrangements in space (clockwise and counterclockwise, respectively), of the four different substituents around a tetrahedral stereogenic center.

If one tetrahedral carbon atom carrying four different substituents generates two possible orientations of its substituents, then a molecular skeleton composed of n chiral carbons is possible to exist in 2^n combinatorial arrangements of R's and S's called by van 't Hoff and Le Bel as stereoisomers (Le Bel-van 't Hoff rule) [6].

About this, a breakthrough was achieved in the field of mathematical stereochemistry, thanks to the contributions by Shinsaku Fujita, who a few years ago proposed the use of one or more types of five possible stereoisograms based on a new arrow language [4,7,8]. For example, the set of stereoisomers (2R, 3R)-, (2R, 3S)-, (2S, 3R)- and (2S, 3S)-2,3dihydroxybutanoic acids can be drawn as in Figure 2a, where *C*- and *S*-axis describe the chirality and sclerality relationships, respectively [4].

In this stereoisogram, the red double-headed arrows at the vertical axis describe a mathematical operation that drives one steroisomer to convert in the corresponding enantiomer (meaning a "Full R/S permutations"), whereas the blue double-headed arrows describe the conversion of one steroisomer to one or more diastereomers (meaning "Partial R/S permutations") [4,7,8].

To follow mathematizing the stereochemistry of these types of compounds, in a previous paper we carried out a change of the binary descriptors R and S by 0 and 1, respectively, and with the aid of the language coined by Fujita, both a set of matrices and polygonal stereoisograms were developed; for example, the full set of stereoisomers of 2,3dihydroxybutanoic acid could be rewritten and redrawn as in Figure 2b [9].

Note that the enantiomeric pairs are located in opposed corners of that square stereoisogram and can be recognized because the string of R's/S's of one enantiomer is fully permuted in its pair: (RR) vs (SS) and (RS) vs (SR). On the other hand, in diastereomers, one or more, but not all of R's/S's are permuted; so, in the current example, blue arrows indicate that (SR) and (RS) are diastereomers of (RR).

Figure 3 illustrates, as a second example, the octagonal stereoisogram for a compound possessing three stereogenic centers, for which $2^3 = 8$ stereoisomers are expected; in particular, the eight possible stereoisomers of 2,3,4-trihydroxypentanoic acid are presented. Note again, in Figure



Figure 2. (a) Fujita's stereoisogram showing the stereoisomeric relationships among the four possible stereoisomers of 2,3dihydroxybutanoic acid [4]; (b) Binary matrices and polygonal stereoisogram approach indicating the full (red arrows) and partial (blue arrows) permutations of the binary strings based on R's/S's and 0's/1's: unlike diastereomers, enantiomers are located at opposite corners [9].

3a, that the enantiomeric pairs are located at antipodal vertices; however, this polygonal stereoisogram turns more complicated if we try to point the diasteromeric relationships among the eight stereoisomers; thus, in Figure 3b is presented only the six diastereomers of (2R, 3R, 4R)-2,3,4trihydroxypentanoic acid.

This drawback led us to look for *n*-dimensional diagrams to illustrate the complete series of stereoisomers and their corresponding stereoisomeric relationship, and we found a good correlation between our matrix approach (composed of 0's and 1's) with those schemes developed in the graph theory; specifically, we refer to the *n*-dimensional unit hypercubes (also known as Q_n or *n*-cubes), because they are composed precisely by 0's and 1's.



Figure 3. (a) Octagonal stereoisogram showing the four enantiomeric pairs of 2,3,4-trihydroxypentanoic acid. In (b), the polygonal stereoisogram was modified from reference [10], to represent only the six diastereomers of (2R, 3R, 4R)-2,3,4trihydroxypentanoic acid. For simplicity, the locators at Fischer's projections are omitted in the labels that indicate the stereochemistry of chiral carbon atoms C₂, C₃ and C₄.

2 Graph theory and its application in chemistry

Also in 1874, the famous British mathematician Arthur Cayley used the graph theory for trying to determine a formula for counting the number of isomers of acyclic alkanes [10]. In his approach, Cayley built acyclic connected graphs (molecular graphs) and named them "plerogram" and "kenogram", being the last one a hydrogen-suppressed molecular graph

of the corresponding alkane (Figure 4). The success of these mathematical trees in several fields of science derives from the fact that the user can assign to them the meaning he needs. Thus, in the case of Cayley's kenograms (Figure 4b), the classification of carbon atoms as primary (C_1 , C_6 and C_{a-d}), secondary (C_4), tertiary (C_3 and C_5) and quaternary (C_2) depends on the number of their connections with other carbons.



Figure 4. Plerogram (a) and kenogram (b) for 2,2,3,5tetramethylhexane. Open and filled circles represent hydrogen and carbon atoms, respectively

In fact, in the words of A. T. Balaban: "All structural formulas of covalently bonded compounds are graphs: they are therefore called molecular graphs, or better, constitutional graphs" [11]. Going further, these formulas evolved favorably to also describe the position of the substituents in space, giving raise to "Configurational graphs" or molecular projections (i.e. Fischer's projections in Figures 2 and 3). Because of the plethora of symbols that are currently being used by chemists to draw molecular structures, we suggested naming them as "*Pure and Hybrid Geometric Molecular Formulas*" [12], containing the last ones both different kinds of lines and alphanumeric elements such as points, asterisks, letters, numbers, etc.; in any case, they are "molecular trees or molecular graphs".

3 The *n*-dimensional unit hypercube graphs

The *n*-dimensional unit hypercubes $(Q_n \text{ or } n\text{-cubes})$ are Cayley's graphs whose vertex sets are restricted to the binary alphabet $\{0, 1\}$. This means they contain 2^n vertices (labeled with *n*-bit binary strings) and $n(2^{n-1})$ edges; moreover, every edge links two vertices whenever their labels differ in a single bit. Hypercubes Q_n for n = 1 to 4 are represented respectively as a line, a square, a cube and a hypercube (better known as tesseract) (Figure 5a-d).



Figure 5. The first four *n*-dimensional unit hypercubes Q_n or *n*-cubes (for n = 1 to 4) (a-d) and the obtaining of Q_3 (cube) from the displacement of Q_2 (square) in one unit (e). The strings of 0's and 1's correspond to the coordinate values of (x), (x, y), (x, y, z) or (x, y, z, w).

One of the most popular ways for constructing a hypercube Q_n is to use the previous hypercube Q_{n-1} as a generator: two copies of the geometric entity labeled Q_{n-1} are displayed in parallel, at a distance equal to a vector unit [13, 14]. Then, the corresponding vertices are joined using lines, in such a way that each vertex satisfies the condition mentioned above. For example, the building of Q_3 (cube) starting from Q_2 (square) is illustrated in Figure 5e, and an animated construction of Q_1 to Q_4 can be watched on open websites [15]. Note that the growing process implies the use of a new coordinate and the insertion of a new bit in the corresponding string (e.g. $(00) \rightarrow (000)$ in the building of Q_3 starting from Q_2).

Whereas the first three hypercubes Q_1 , Q_2 and Q_3 can be easily conceptualized in our three-dimensional world (a line, a square and a cube, respectively), the building and bidimensional representation of Q_4 , which

requires a 4-coordinate system (x, y, z, w), is a little tricky; it is schematized, for example, as two concentric cubes joined by edges, which form the lateral faces of six new skeletal cubes, to give a total amount of eight cubes (See 6th column for n = 4 in Table 1). Although the eight cubes of the hypercube Q_4 have the same size, the inner cube seems to be smaller than the external one, and the six middle cubes appear as truncated square pyramids (Figure 5d).

Table 1. Summary of the geometric features of the first four hyper-
cubes Q_n (n = 1 to 4).Dimension |Hypercube| Corners | Edges | Squares | Cubes | Hypercubes|

Dimension	Hypercube	Corners	Edges	Squares	Cubes	Hypercubes
(n)	Graph					
	(Q_n)					
1	Line	2	1	0	0	0
	(Q_1)					
2	Square	4	4	1	0	0
	(Q_2)					
3	Cube	8	12	6	1	0
	(Q_3)					
4	Hypercube	16	32	24	8	1
	(Q_4)					
Formula		$\binom{n}{0}2^n$	$\binom{n}{1}2^{n-1}$	$\binom{n}{2}2^{n-2}$	$\binom{n}{3}2^{n-3}$	$\binom{n}{4}2^{n-4}$

Where $\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}$. For example, the number of cubes in a hypercube Q_4 is: $\binom{n}{3}2^{n-3} = \binom{4}{3}2^{4-3} = 8$. Moreover, the column "corners" is written in bold letters because they represent the number of stereoisomers mentioned in the current approach.

The hypercube Q_4 is the most famous of all of hypercubes, because it has driven the imagination of many scientists, technicians and artists from around the world. The word "Hypercube" has been taken as brand name for the scientific software company Hypercube Inc. [16]; the polyhedron net of Q_4 was used by Salvador Dalí in his 1954 oil-on-canvas painting known as "Crucifixión" of Jesus or "Corpus Hypercubus" [17] and among many scientific papers devoted to the usage or application of Q_4 , its "concentric cubes" version (or tesseract) has been used by Erik A. Schultes *et al.*, for describing sequenomics as a combinatorial approach in biological building blocks (e.g. set string length-4 $\{A, G\}^4$) [18].

3.1 The *n*-dimensional unit hypercube stereoisograms, Q_n or *n*-cube stereoisograms

In our previous work, after the replacement of R's/S's by 0's/1's, a set of stereoisomers was represented as a matrix, whose rows are bit strings length-n of 0's and 1's (Figure 2b). Currently, after the replacement of 0's/1's by R's/S's, a set of stereoisomers is represented as a unit ndimensional hypercube; this means that the number of stereogenic centers is also the number of coordinates (x, y, z, w, ...) necessary to describe the corresponding unit n-hypercube. Thus, Q_1 (a line), Q_2 (a square), Q_3 (a cube) and Q_4 (a hypercube) are used to host 2^n stereoisomers of compounds possessing n = 1 to 4 stereogenic centers, respectively (See Table 1). Besides, if we keep the arrows notation coined by S. Fujita, it is possible to observe, at glance, the stereoisomeric relationships among them (Figure 6):

1. Q_1 has two vertices to host enantiomers R and S (Figure 6a). In the language of bits, the red arrow indicates the permutation of 0 and 1, denoted as $\begin{pmatrix} 0\\1 \end{pmatrix}$. Additionally, there are not diastereomers, because for n = 1 we have $2^n - 2 = 2^1 - 2 = 0$.

2. Q_2 has four vertices to host the two pairs of enantiomers RR/SS and RS/SR (Figure 6b). They are located at opposed square's corners and are connected by red arrows that indicate, for each pair of enantiomers, a full permutation of R's/S's, that in permutation notation of bits are $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Those arrows intersect at the square's center, meaning that there is a chirality center, as it is the line's midpoint in Q_1 . With blue arrows that indicate a partial permutation of R's/S's, each stereoisomer is connected along edges with its two diastereomers $(2^n - 2 = 2^2 - 2 = 2)$. In matrix notation of bits, the partial permutation of (RR), that generates its diastereomers (RS) and (SR), can be written as $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, respectively.

3. Q_3 has eight vertices to host the four pairs of enantiomers RRR/SSS, RRS/SSR, RSR/SRS and SRR/RSS. These enantiomeric pairs are located at antipodal vertices and are connected by colored arrows that intersect at the cube's center, meaning again, that the hypercube's center is a chirality center (Figure 6c). On the other hand, two kinds of diastere-omers can be identified in Q_3 . Firstly, type-I diastereomers are the three ones connected along the edges by blue arrows, that produce the permutation of only one bit. Secondly, type-II diastereomers are the three ones connected along the face's diagonals by green arrows, that produce the permutation in type-I diastereomers of (2S, 3S, 4S) = (111) are $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$; whereas the two-bit permutations in type-II diastereomers are $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$.



Figure 6. The substitution of R's/S's by 0's/1's allows to use the *n*-dimensional unit hypercubes as hypercube stereoisograms. Enantiomeric pairs are located at antipodal vertices and connected by red arrows in (a-b) and by colored arrows in (c). Diastereomers are connected by blue arrows in (b) and by blue and green arrows in (d) and (e). For the sake of clarity, enantiomers and diastereomers of Q_3 are presented separately in (c-d) and even more, only diastereomers of (2S, 3S, 4S) are illustrated in (e) as follow: type-I diastereomers (one-bit permutation) are connected by blue arrows along the edges, whereas type-II diastereomers (two-bit permutations) are connected by green arrows along face's diagonals.

3.2 Mathematical properties of *n*-dimensional unit hypercube stereoisograms

1. The *n*-dimensional unit hypercubes possess 2^n vertices, the same as the maximum number of stereoisomers of compounds possessing *n* stereogenic centers. For taking advantage of this approach, the binary alphabet $\{0,1\}$ is changed by the binary alphabet of chiral descriptors $\{R,S\}$, or even better, the labels of both alphabets are kept. This change converts *n*dimensional hypercubes into *n*-dimensional hypercube stereoisograms that host a set of 2^n stereoisomers.

2. The enantiomeric pairs $(2^n/2 = 2^{(n-1)})$ are located at antipodal vertices. They are related by a chirality center that produces the full permutation of 0's/1's or R's/S's (Figure 6c). The bitwise operation that affords this result is known as "NOT". For example, to obtain the enantiomeric pair of (001):

NOT $(001) = (110) \equiv \text{NOT} (RRS) = (SSR).$

And the resulting bit string is the complement of the initial, meaning that their sum is equal to the unit vector length-n [9]:

(001) + (110) = ((0+1) (0+1) (1+0)) = (111)

or: (RRS) + (SSR) = ((R+S) (R+S) (S+R)) = (SSS).

Observe that R (or 0) is neutral on addition operation.

3. In a hypercube stereoisogram, except for the arrow connecting an enantiomer to its couple, the number of diastereomers of a given stereoisomer is determined by the number of arrows that connect it with the remaining vertices $(2^n - 2)$. For example, there are 0, 2 and 6 connections (diastereomers), for stereoisomers (S) in Q_1 , (SS) in Q_2 and (SSS) in Q_3 (Figure 6).

Thus, these arrows indicate the permutation of one or more bits, but not all, and are vectors normal to mirror planes that produce such permutations. In turn, some of these planes are Cartesian (Figure 7a-b for Q_2 and Q_3), while other ones are dihedral planes (Figure 7c, for Q_3).



Figure 7. Diastereomers are related by Cartesian planes that permute one bit in Q_2 and Q_3 (a-b) and dihedral planes that permute two bits in Q_3 . For the sake of clarity, Cartesian (b) and dihedral planes (c) are presented separately and, in each case, only one reflection operation is schematized at right.

4 Pseudohypercube stereoisograms for a set of stereoisomers containing meso-type compounds

It is very well known that a compound containing n stereogenic centers do not always accomplishes the rule of the maximum number of stereoisomers, 2^n . If an improper symmetry element (a symmetry plane, σ , an inversion center, i, or an improper rotation axis, $S_m : m \ge 3$) is present in one of the compounds belonging to a series of stereoisomers, then it is not chiral and therefore does not has an enantiomeric couple; however, it must be included in the hypercube stereoisogram as a "ghost node", that connects only with its antipodal vertex through Fujita type-arrow corresponding to homostereoisomers (Figure 8).



Figure 8. Although n = 2, the maximum number of stereoisomers, 2^n , of 2,3-dihydroxybutanodioic acid is not 4 but 3. This is due to stereoisomer (2R, 3S), which possesses a symmetry plane and, after a rotation of 180° around z axis, its mirror image (2S, 3R) is congruent with it (a-b). This fact is registered in a pseudohypercube of Q_2 as a "ghost node" and with Fujitatype arrow for homostereomers between (2R, 3S) and (2S, 3R) (c).

5 Conclusions

The substitution of the binary language $\{0, 1\}$ by $\{R, S\}$ in the unit *n*dimensional hypercubes resulted in a successful representation for the 2^n stereoisomers of compounds possessing *n* stereogenic centers. The enantiomers are hosted in antipodal vertices, whereas, with exception of its antipodal vertex, each stereoisomer connects with the remaining vertices that are their $2^n - 2$ diastereomers. Thus, in the hypercube stereoisograms, enantiomers are related by the hypercube's center which acts as a chirality center, whereas diastereomers are related by mirror planes (diastereomeric planes). Paying attention to the permuting elements into the binary strings, the diastereomeric planes result in different types or classes, one of them is for the Cartesian planes that permute only the bit vector orthogonal to each one of those planes. On the other hand, for a set of stereoisomers containing an achiral (meso) compound, it is proposed the building of a pseudo hypercube that includes a "ghost node", which must be equalized with its antipodal vertex.

Acknowledgment: We thank VIEP for financial support to Proyecto 00030 de grupos de investigación interdisciplinaria 2023.

References

- J. H. Van't Hoff, Sur les formules de structure dans l'espace, Arch. Néerland. Sci. Exact. Natur. 9 (1874) 445–454.
- [2] J. A. Le Bel, Sur les relations qui existent entre les formules atomiques des corps organiques et le pouvoir rotatoire de leurs dissolutions, *Bull.* Soc. Chim. Paris 22 (1874) 337–347.
- [3] S. Fujita, Stereogenicity revisited. Proposal of Holantimers for Comprehending the relationship between stereogenicity and chirality, J. Org. Chem. 69 (2004) 3158–3165.
- [4] S. Fujita, Stereoisograms for specifying chirality and RSstereogenicity. A versatile tool for avoiding the apparent inconsistency between geometrical features and *RS*-nomenclature in stereochemistry, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 11–38.
- [5] R. S Cahn, C. K. Ingold, V. Prelog, The specification of asymmetric configuration in organic chemistry, *Experientia* 12 (1956) 81–94.
- [6] C. Goedecke, 90th anniversary: Death of Joseph Achille Le Bel, *Chemistry Views*, August 6, 2020.
- [7] S. Fujita, Type-itemized enumeration of quadruplets of RSstereoisomers. I. Cycle indices with chirality fittingness modulated by type-IV quadruplets, J. Math. Chem. 54 (2016) 286–309.
- [8] S. Fujita, Asymmetry revisited by Fujita's stereoisogram approach. Part 1: Asymmetry under point-group symmetry and under RSpermutation-group symmetry, *Tetrahedron Asymm.* 27 (2016) 675– 686.

- [9] R. E. Arroyo-Carmona, S. Hernandez-Anzaldo, A. Perez-Benitez, S. E. Pineda-Sanabria, M. A. Velazquez-Carmona, Y. Reyes-Ortega, J. J. Vazquez Bravo, Matrix representation of stereoisomers and new polygonal stereoisograms, *MATCH Commun. Math. Comput. Chem.* 81 (2019) 603–620.
- [10] A. Cayley, On the mathematical theory of isomers, *Phil. Magazine* 67 (1874) 444–447.
- [11] T. Balaban, Application of graph theory in chemistry, J. Chem. Inf. Comput. Sci. 25 (1985) 334–343.
- [12] R. M. Aguilar-Garduño, R. E. Arroyo-Carmona, J. J. Cerón Luna, E. González-Vergara, A. Pérez-Benítez, Precisions and comments on the article: "Assessment of learning in molecular representations bond-line formulas of organic compounds. A case study", *Educ. Quím.* 24 (2013) 270–277.
- [13] M. Gardner, *Mathematical Carnival*, Am. Math. Soc., 2020, pp. 21.
- [14] M. Al Achhab, O. El Hichami, B. E. El Mohajir, M. Naoum, Hypercube graph decomposition for boolean simplification: An optimization of business process verification, *Int. J. Adv. Comput. Sci. Appl.* 10 (2019) 467–473.
- [15] V. Ostrosablin, An animation showing how to create a tesseract from a point. In hypercube, Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Hypercube#/media/File: From_Point_to_Tesseract_(Looped_Version).gif
- [16] Hypercube, Custom Software Development, https://hypercubesoft.com/
- [17] S. Dali, Crucifixion (Corpus Hypercubus). Oil on canvas painting at "The Metropolitan Museum of Art, NY", https://www.metmuseum. org/art/collection/search/488880
- [18] P. T. Hraber, T. H. LaBean, E. A. Schultes, No molecule is an island: Molecular evolution and the study of sequence space, in: A. Condon, D. Harel, J. N. Kok, A. Salomaa, E. Winfree (Eds.), *Algorithmic Bioprocesses*, Springer-Verlag, Berlin, 2009, pp. 675–704.
- [19] D. B. Priest, Hypercubes, hyperwindows and hyperstars, Math. Teacher 70 (1977) 606–609.