On a Graph Theoretical Characterization of Cis/Trans Isomers

Lionello Pogliani[†]

Centro de Quimica-Fisica Molecular, Instituto Superior Técnico, 1096 Lisboa Codex, Portugal

A novel graph theoretical approach to characterization of cis/trans isomerism is outlined. The given approach allows the definition of a new topological cis connectivity ^nX_c index by the aid of which it is possible to construct a cis/trans connectivity X_{ct} descriptor. This last descriptor performs very well in describing the first moments of hexatriene molecules and in distinguishing cis/trans isomers in olefins. The basis of the approach is the definition of virtual ring fragments which can be formed by cis structures when the delta values of the two cis points of these fragments are increased by 1. A descriptor based on the previous index that encodes also the different positions of the double bond in an olefin is suggested and tested.

INTRODUCTION

Some years ago Randić[1-3] proposed, by using topographic (metric) rather than topological (graph theoretical) matrices, a new kind of indices that could differentiate between different conformers such as cis and trans isomers. These indices, which describe fixed metric properties such as lengths and angles and can characterize three-dimensional structures, are qualitatively and quantitatively different from the normal topological indices. These last indices, which contain information concerning only the components and the connectivity relations of the chemical graph to be studied, are independent of the metric aspects of a chemical structure, which viewed as a connected graph has no dimension.

Up to date the most widely used topological indices are the indices obtained by the extension approach of the Randić topological branching index given by Kier and Hall.[4,5] These indices and the recently introduced D index[6,7] rooted on the cardinal branching δ number cannot distinguish between different cis/trans isomers. In fact, only the points and the connections viewed in terms of atoms and bonds, respectively, are topologically significant in connected graphs used to represent molecular structures; the neighborhood relations in such graphs are given in terms of edges connecting the points. Nevertheless, it is rather unsatisfying that there is no way to encode a cis/trans isomerism by the aid of topological indices, as one of the most important properties of a topological space is the neighborhood relation between points of the space.[8] Here we will present an approach to the characterization of cis/trans isomers based on an extension of the concept of connectivity which will lead to the definition of a new molecular topological connectivity index to be added to the set of molecular connectivity indices (MCI) developed by Kier and Hall.[4,5] We will test its validity against the three-dimensional XX connectivity descriptor developed by Randić[3] for the six conformers of the hexatriene and against the newly proposed GAI index,[9] based on quantum chemistry considerations and the validity of which has been tested on a series of different physicochemical properties of cis and trans olefins.

METHOD

We will start just as Randić[2,3] did by thinking of the different hexatriene conformational isomers of Figure 1 as embedded on an idealized graphite grid (see Figure 1 in ref 2) and noticing
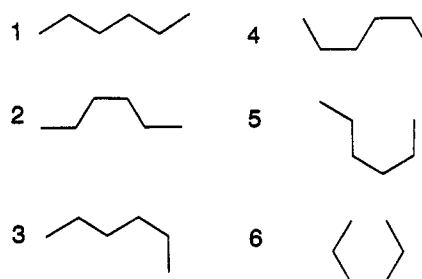
[†] On sabbatical leave from the Dipartimento di Chimica, Università della Calabria, 87030 Rende (CS), Italy.

Figure 1. The six conformers of hexatriene.

that if we increase, in the chemical graphs 2–6, the δ valence of two cis points by 1 and connect them by an edge, they can form four-membered rings and, precisely, one four-membered ring for graphs 2 and 3, two for graphs 4 and 5, and three for graph 6. The six-membered embedding rings can be considered a stable embedding form. The newly formed rings, a sort of virtual rings, with the raised δ^r valence of the newly connected cis points, can be seen as virtual ring fragments for which we can define the following cis ^nX_c connectivity index (c stands for cis and here $n = 4$; something like the $^nX_{ch}$ index of Kier and Hall[4,5]):

$$^nX_c = \sum (\delta^r_1 \delta_2 \dots \delta^r_n)^{-6/n} \tag{1}$$

where 6 and n in the exponent are the number of edges of the embedding ($n = 4$) rings, respectively, while the numbering at the different delta values denotes the different vertices of the ring fragment and the limits of the summation are the number of n-membered-ring fragments that can be formed inside the embedding rings. If $\delta = \delta^r = 2$, eq 1 has a constant solution for every n ($^nX_c = 0.01563$). The proposed index (i) has no meaning for trans structures, where no cis points are present, and furthermore (ii) cannot encode branching. These two reasons bring us to define a more global molecular X_{ct} connectivity index for olefins and more generally for chemical graphs. This index should be defined in a way that its maximum value should correspond to the all-trans isomer, while it decreases the more entangled ("cis-rich") and the more branched the chain gets; that is, our new index should be a kind of corrected ^1X index (the value of this index decreases, in fact, with increasing branching[4]). Its mathematical expression should then be

$$X_{ct} = {}^1X - {}^nX_c \tag{2}$$

(here $n = 4$) where ^1X is the well-known first-order MCI.[4,5]

Table 1. ^{1}X, X_{ct}, and Physical Constants of Cis/Trans Olefins[a]

olefin	^{1}X	X_{ct}	bp	n_D^{20}	d_4^{20}	MR_D
trans-2-butene	1.914 21	1.914 21	0.88			
cis-2-butene	1.914 21	1.898 59	3.7			
trans-2-pentene	2.414 21	2.414 21		1.3792	0.6475	
cis-2-pentene	2.414 21	2.405 71		1.3828	0.6554	
trans-3-hexene	2.914 21	2.914 21	67.5	1.3938	0.6779	29.72
cis-3-hexene	2.914 21	2.909 58	66.85	1.3934	0.6796	29.61
trans-2-octene	3.914 21	3.914 21	123.4	1.4128		
cis-2-octene	3.914 21	3.905 71	124.6	1.4139		
trans-3-octene	3.914 21	3.914 21	122.4	1.4124	0.7163	39.04
cis-3-octene	3.914 21	3.909 58	122.3	1.4125	0.7189	38.91
trans-4-octene	3.914 21	3.914 21	121.4	1.4116	0.7147	39.05
cis-4-octene	3.914 21	3.909 58	121.7	1.4136	0.7205	38.92
trans-5-decene	4.914 21	4.914 21	170.2	1.4235	0.7401	48.34
cis-5-decene	4.914 21	4.909 58	169.5	1.4252	0.7445	48.22

[a] The physical constants are taken from ref 8.

In all-trans graphs we have $^nX_c = 0$ and

$$X_{ct} = {}^1X = \sum (\delta_i \delta_j)^{-1/2} \qquad (3)$$

where i and j denote adjacent atoms and the limits of the summation are all the adjacent atom pairs. The given definition of X_{ct} should allow us to encode not only different cis and trans olefins but, in the future, different types of chemical graphs such as alkanes and alkenes. The ^nX_c index, instead, could only be used inside a specific set of olefins such as the given trienes, but it would be useless, for example, to differentiate between cis-2-butene and cis-2-octene (see Table 1 and the following discussion).

DISCUSSION

While the δ row vector of the hexatriene **1–6** graphs used to derive the ^{1}X values is

$$\delta = (1, 2, 2, 2, 2, 1)$$

the fragment δ^r row vectors of cis **2–6** conformers (first and last δ values are the raised δ values) are the following:

$$\delta^r(2) = (3, 2, 2, 3)$$

$$\delta^r(3) = (3, 2, 2, 2)$$

$$\delta^r(4) = (2, 2, 3, 3)$$

In this last case due to the simultaneous formation of two four-membered rings also the third δ value is a δ^r value. Compound **5** can form two four-membered rings, the δ^r matrices of which are

$$\delta^r(5) = (3, 2, 2, 3) \text{ and } (3, 2, 2, 2)$$

and compound **6**, which can form three four-membered rings, has the following δ^r matrices (the first vector should be counted twice):

$$\delta^r(6) = (2, 2, 2, 3) \text{ and } (3, 2, 2, 3)$$

By the aid of eqs 2 and 3, with the given δ^r and δ matrices, we obtain the following X_{ct} values for compounds **1–6**:

	1	**2**	**3**	**4**	**5**	**6**
X_{ct}	2.914 21	2.909 58	2.905 71	2.904 95	2.901 08	2.892 57

This index and the three-dimensional topographic XX connectivity index derived by Randić[3] are related by the
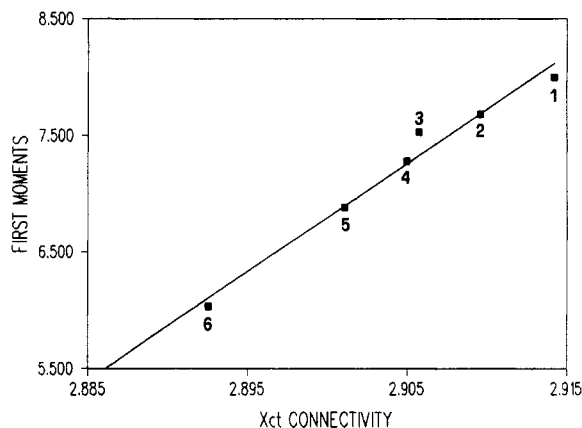
Figure 2. A plot of the X_{ct} connectivity index and the first molecular moments of hexatrienes of Figure 1.

following relationship:

$$XX = -1.7097 X_{ct} + 7.932 \qquad (4)$$

$$R = 0.9787, \quad S = 0.0030, \quad F = 91, \quad n = 6$$

The values of the statistical parameters (especially S) show that the two indices have similar descriptor capabilities. Randić[3] illustrated the validity of the XX indices showing in a figure (Figure 3 of ref 3) the very good linear correlation between the first molecular moments and the XX connectivity values. In Figure 2 the obtained X_{ct} values are plotted against the same first molecular moments; the resulting linearity between X_{ct} and the first moments can hardly be distinguished from the one obtained by Randić except that they have, as expected, slopes of differing signs.

We will test, now, the derived connectivity index with some physicochemical properties (P) of 14 olefin isomers: boiling points (bp) ($n = 12$), refractive index (n_D^{20}) ($n = 12$), density (d_4^{20}) ($n = 10$), and molar refractivity (MR_D) ($n = 8$). These properties were used by Xu, Wang, and Su[9] to test their GAI index and have been collected in Table 1 together with the corresponding ^{1}X and X_{ct} values. The correlations between the GAI index and given properties show the following statistical parameters:

	R	S	F	n
bp	0.9960	5.2	1244	12
n_D^{20}	0.9798	0.0032	240	12
d_4^{20}	0.9805	0.0071	199	10
MR_D	0.9941	0.83	500	8

while the statistical parameters of the correlations between the given properties and the ^{1}X values are

	R	S	F	n
bp	0.9969	4.6	1593	12
n_D^{20}	0.9817	0.0030	266	12
d_4^{20}	0.9831	0.0066	231	10
MR_D	0.999 96	0.071	684 71	8

Comparing the statistical parameters of the GAI and ^{1}X indices, we notice that the ^{1}X index, which does not differentiate between cis and trans isomers, scores better than the GAI index. This fact is due to the very small and sometimes compensating d_{ct} differences between the P values of the cis and the corresponding trans isomer. For example, d_{ct} ranges from –0.7 to 1.2 for the boiling points (mean d_{ct} = –0.05, excluding the consistent difference between trans- and cis-butene) and from –0.014 to 0.0017 with a mean of –0.0016 for n_D^{20}, while d_{ct} has a mean of 0.0045 for d_4^{20} and a mean

of –0.1225 for MR_D. A better description of these differences than the one given by ^1X (which acts here practically as a kind of high-performing mean cis/trans index) should then be a rather formidable task.

The statistical parameters of the correlations between the same properties and the X_{ct} values are

	R	S	F	n
bp	0.9969	4.6	1600	12
n_D^{20}	0.9815	0.0030	263	12
d_4^{20}	0.9830	0.0067	229	10
MR_D	0.999 98	0.046	160 426	8

Comparing the statistics for ^1X and X_{ct}, we notice that while there is no significant difference between results for bp, n_D^{20}, and d_4^{20}, the X_{ct} index is somewhat better for MR_D. The connectivity X_{ct} descriptor, derived by the aid of considerations on hexatriene isomers only, then (i) scores better than the GAI index, (ii) has the same encoding capability of the ^1X index relative to three P properties (all of them showing vanishingly small mean d_{ct} values: –0.05, –0.0016, and 0.0048, respectively; see above), and (iii) describes better than the ^1X index the rather small but noticeable (mean d_{ct} = –0.12) differences in MR_D of a set of cis/trans olefins.

In Table 1 we notice that *trans*-2-octene, *trans*-3-octene, and *trans*-4-octene as well as *cis*-3-octene and *cis*-4-octene isomers show the same X_{ct} values. This fact means that the X_{ct} connectivity index is not able to distinguish between the different positional isomers of the given octenes. We will test now the following positional cis/trans connectivity index $X_{ct}(p)$, defined in eq 5, which takes into account the different positions of the double bond in compounds of Table 1; such a test has only an orientating value as the properties of three different octene isomers only are known:

$$X_{ct}(p) = X_{ct} - (1/p)^n \qquad (5)$$

where p stands for position and n for the length of the chain (i.e., number of carbon atoms). Compounds in Table 1 show the following $(1/p)^n$ values:

	2-butene	2-pentene	3-hexene	2-octene
$(1/p)^n$	0.0625	0.031 25	0.001 37	0.003 906

	3-octene	4-octene	5-decene
$(1/p)^n$	0.000 152	0.000 015	10^{-7}

From eq 5 and from the found $(1/p)^n$ values we notice that the more the double bond shifts toward the molecule midpoint, the more $(1/p)^n$ decreases and the more $X_{ct}(p)$ corresponds to X_{ct}. The statistical parameters of the correlations between the $X_{ct}(p)$ index and the P values are

	R	S	F	n
bp	0.9977	3.9	2184	12
n_D^{20}	0.9827	0.0029	281	12
d_4^{20}	0.9842	0.0064	248	10
MR_D	0.999 981	0.046	161 922	8

Corresponding $^1X(p)$ values show similar scores with F(bp) = 2189, $F(n_D^{20})$ = 284, $F(d_4^{20})$ = 250, and $F(MR_D)$ = 68769. Introduction of the positional factor $(1/p)^n$ improves nicely both ^1X and X_{ct} indices. Index $^1X_{ct}$ maintains, also in this case, its better score over the ^1X index in the description of the only property, MR_D, which shows small but noticeable d_{ct} differences in the P values of the cis and trans isomers.

A deeper examination of Figures 1 and 2 allows us to introduce two other kinds of X_{ct} indices, which do not affect obtained results with compounds of Table 1. An X_{ct}' index
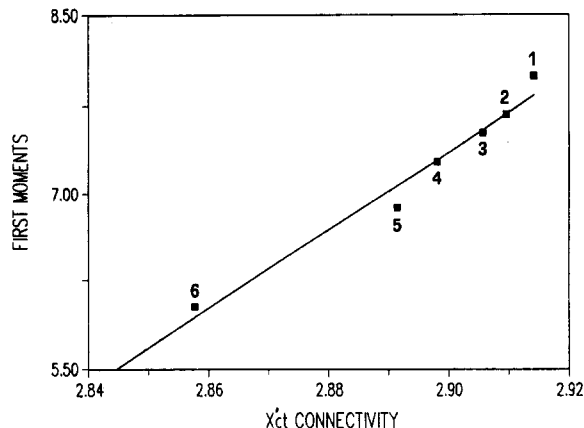

Figure 3. A plot of the X_{ct}' connectivity index and the first molecular moments of hexatrienes of Figure 1.

could be defined just by taking into consideration also the n = 5 (one for compound **5** and two for compound **6**) and n = 6 ring fragments (one for each compound **4** and **6**), and in this case ^nX_c of eq 2 will be

$$^nX_c = \sum{}^iX_c = {}^4X_c + {}^5X_c + {}^6X_c \qquad (6)$$

each ^iX_c (i = 4, 5, and 6) being defined by eq 1. Clearly, the obtained results for olefins in Table 1 are unchanged as for these olefins: $^nX_c = {}^4X_c$.

The fragment δ^r row matrices of our hexatrienes for n = 5 are

$$\delta^r(5) = (3, 2, 2, 2, 2)$$

$$\delta^r(6) = (2, 2, 2, 2, 3)$$

while for n = 6 are

$$\delta^r(4) = (3, 2, 3, 3, 2, 3)$$

$$\delta^r(6) = (2, 2, 2, 2, 2, 2)$$

The corresponding X_{ct}' values for compounds **1–6** will be

	1	2	3	4	5	6
X_{ct}'	2.914 21	2.909 58	2.905 71	2.898 01	2.891 47	2.8577

Correlation between these indices and the aforementioned three-dimensional XX Randić index shows the following statistical parameters: R = 0.9902, S = 0.0020, F = 201, n = 6. These statistical parameters are somewhat better than the corresponding R, S, and F values relating XX and X_{ct}, while the plot (see Figure 3) of the first (FM) moments vs X_{ct}' is nearly as good as the plot of FM vs X_{ct}. This last index shows, nevertheless, redundancies in its definition, as, to define the type of cis structure in graphs **2–6** is sufficient, the number of four-membered ring fragments with raised valence.

Another type of X_{ct}'' index could be defined taking into consideration only the ring fragment with the maximum number of bonds (excluding bicyclo rings), i.e., n = 4 for compounds **1–4**, n = 5 for compound **5**, and n = 6 for compound **6**. This new index is a poorer descriptor of XX and FM than the two previous indices even if the last word about this, and the aforementioned indices should be left to their power to fit different physical constants of a wide set of compounds.

Before concluding a word should be said about cis fragments with n > 6. These rings with seven or more edges cannot be embedded in a graphite-like grid, as the seventh edge will be out of plane of the graphite grid. The growing of a third

dimension in the molecular graph should be considered a stabilizing condition, and in this case the corresponding ^nX_c value for $n > 6$ should, then, be added to X_{ct}.

The attempt to define a valence ^nX_c based on δ^v values and consequently a valence $^vX_{ct}$ index failed, as such an index was a poorer descriptor of the physical (FM) and mathematical (XX) properties of the hexatriene molecules and of the given properties of olefins of Table 1. Instead, the following index,

$$X_{ct}(p^v) = X_{ct} - (1/\delta^v)^p \qquad (7)$$

where exponent p stands for the position of the double bond and $\delta^v = 3$ for olefins of Table 1, could satisfactorily accommodate the variability of the given database (but poorer than $X_{ct}(p)$); in fact we have $F(\mathrm{bp}) = 1251$, $F(n_D^{20}) = 241$, $F(d_4^{20}) = 265$, and $F(\mathrm{MR_D}) = 32616$, while the corresponding $^1X(p^v)$ shows the following F values: $F(\mathrm{bp}) = 1220$, $F(n_D^{20}) = 244$, $F(d_4^{20}) = 268$, and $F(\mathrm{MR_D}) = 25\,690$.

CONCLUSION

Proposed ^nX_c and X_{ct} graph theoretical descriptors, which could be considered virtual molecular connectivity indices based on δ and δ^r values of virtual ring fragments formed by embedded cis structures, are not based on empirical data and are derived solely from the information on the molecular structures like the other MCI proposed by Kier and Hall.[4,5] Furthermore, they obey the list of properties that are desirable for topological (graph theoretical) indices proposed by Randić.[10] The central finding, that the X_{ct} index provides an accurate description (i) of the three-dimensional topographic XX Randić index, (ii) of the first molecular (FM) moments of the hexatriene molecules, and (iii) of four physicochemical properties of olefins for one of which, $\mathrm{MR_D}$, provides the best overall description, appears to be impressive, providing the limited number of observations of the five physical properties and a topographic property do not play a role. Noteworthy is the fact that the $^1X_{ct}$ index describes better than the ^1X index the only property, $\mathrm{MR_D}$, which does not show vanishingly small d_{ct} differences in the P property of the cis and the corresponding trans olefin. Certainly, there is a need to work with larger data sets to confirm, refine, or even reject the outlined descriptor, as, quoting Randić,[11] each time a new descriptor is proposed there is always the possibility that yet unexplored descriptors may surpass in quality those considered hitherto. Index $X_{ct}(p)$ should be considered as just a suggestion, as the three different positional octene isomers, the properties of which are known, are too small of a data set to allow for a test. We should also not forget an important side result of this study concerning the good performance of the ^1X index, which performs better than the GAI index in describing the given properties of olefins. The proposed approach to design a topological connectivity index, which encodes information on cis/trans isomerism, should also help us to design new topological descriptors for the rotational isomers around a single σ bond in saturated hydrocarbon fragments (the well-known Newman projections) provided a good P database is given to test it.

REFERENCES AND NOTES

(1) Randić, M.; Jerman-Blazić, B.; Trinajstić, N. Computation of 3-dimensional Molecular Descriptors. *Comput. Chem.* **1990**, *14*, 237–246.
(2) Randić, M. Molecular Topographic Descriptors. *Stud. Phys. Theor. Chem.* **1988**, *54*, 101–108.
(3) Randić, M. On Characterization of Three-Dimensional Structures. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1988**, *15*, 201–208.
(4) Kier, L. B.; Hall, L. B. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, 1986, and references therein cited.
(5) Hall, L. H. Computational Aspects of Molecular Connectivity and its role in Structure-Property Modeling. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Press: New York, 1990; Chapter 8, pp 201–233.
(6) Pogliani, L. Molecular Connectivity Model for Determination of Isoelectric Point of Amino Acids. *J. Pharm. Sci.* **1992**, *81*, 334–336.
(7) Pogliani, L. Molecular Connectivity Model for Determination of Physicochemical Properties of α-Amino Acids. *J. Phys. Chem.* **1993**, *97*, 6731–6736.
(8) Turro, N. J. Geometric and Topological Thinking in Organic Chemistry. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 882–901.
(9) Xu, L.; Wang, H. Y.; Su, Q. A Newly Proposed Molecular Topological Index for the Discrimination of CIS/TRANS Isomers and for the Studies of QSAR/QSPR. *Comput. Chem.* **1992**, *16*, 187–194, and references therein cited.
(10) Randić, M. Generalized Molecular Descriptors. *J. Math. Chem.* **1991**, *7*, 155–168.
(11) Randić, M. Novel graph theoretical approach to heteroatoms in quantitative structure-activity relationships. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213–227.