

ENUMERATION AND SYMMETRIES OF THE STEREO-ISOMERS OF PARAFFINIC MOLECULES

Denis Guillaume[†], Nik Lygeros^{‡‡}, Paul-Valère Marchand^{††},

Marc Massot^{††}, Daniel Schweich^{‡‡}

[†]IFP Énergies nouvelles, Rond-point de l'échangeur de Solaize, BP 3, 69360 Solaize, France

^{††}MAPLY – UMR 5585, Laboratoire de Mathématiques Appliquées de Lyon,
Université Claude Bernard, Lyon I
69622 Villeurbanne Cedex, France

^{‡‡}EM2C - UPR 288 – CNRS, Laboratoire Énergétique Moléculaire et Macroscopique, Combustion
École Centrale Paris,
Grande Voie des Vignes
92 295 CHATENAY-MALABRY Cedex

^{‡‡}LGPC – UMR 2214, Laboratoire de Génie des Procédés catalytiques,
École Supérieure de Chimie Physique Électronique de Lyon
BP 2077, 69616 Villeurbanne Cedex, France

Abstract: Microkinetic modelling of large reaction networks usually necessitates the generation of molecules and the computation of some of their physical properties and especially their symmetries. The graph theory is an interesting tool for achieving these goals and yields new insights to these problems providing both certificates, via generating functions, and ways to easily generate molecules and properties. Results presented here are based on alkanes (paraffins) but can be extended to other types of molecules.

Keywords: microkinetics, molecule enumeration, graph theory

1 Introduction

Kinetic modeling of reactions involving complex mixtures of long-chain hydrocarbons can now be based on computer calculations using concepts of molecular chemistry. One of these approaches is the “single event theory”^{1,2,3}. The method assumes that the kinetic constants of the elementary steps can be estimated on the basis of intrinsic properties (nature of the active center, contribution of the atoms or groups that compose the molecule, etc.) and of characteristics of the symmetry of the reactant, product or activated complex. Alternative methods exist that takes into account Statistical Factors instead of Symmetries⁴. Statistical Factors counts the number of different possible products, by labeling all the identical atoms. Bishop and Laidler established the equivalence between Symmetries and Statistical Factors⁵, as the ratio between forward and backward reactions are the same with both descriptions.

Chemical acid catalysed reactions usually involves paraffins, olefins ions and activated complex³¹. For a sake of simplicity, we will limit the present study to paraffins, but the approach hereafter can be extended to other families of molecules.

Since we intend to bring a new approach in the evaluation of these characteristics for a class of paraffin molecules, an interesting tool to this purpose is graph theory. A correspondence between graphs and chemical categories has found numerous applications in chemistry: a graph corresponds to a molecule, i.e. points symbolize atoms and lines symbolize chemical bonds. Graphs gave Cayley the incentive to develop a procedure for counting the structural isomers of alkanes as early as in 1874⁶. In the framework of acyclic alkane molecules, hereafter denoted paraffins, there is a one-to-one correspondence between the various isomers of the molecule where only carbon atoms and C-C bonds are represented (the lacking atoms being hydrogen) and mathematical graphs. Each bound between two carbon atoms, the vertices, is represented as a line relating two points; thus every structural formula of paraffin is a graph whose points have at most degree 4 and every structural formula of paraffin is a tree with the same property (see Figure 2 and reference 7).

Figure 1 illustrates a typical example, although simplified, of the problematic that motivates this article. Excluding the S_3 symmetry axes of the three rotating methyl groups, reactants A and C have an “internal” symmetry of order two. Conversely, product B has an S_2 axis of “external” symmetry and product D contains a chiral carbon (star). According to rules of the single event theory, the rate constants are those of the Figure 1 where k is proportional to the single event rate constant.

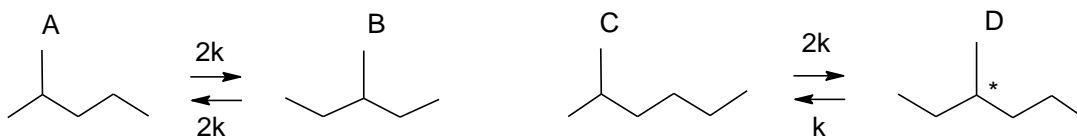


FIG. 1 – Elementary step rate constants according to the symmetry of the reactant and product

For both reactions, the forward rate constants are the same because the formation of the activated complexes is assumed to be independent of the length of the linear chain (in A or C). Conversely, the backward rate constants depend on the symmetry of B or D. At equilibrium, the composition of the mixture of C and D (both stereo-isomers) can be obtained from the group contribution method of Benson⁸. The enthalpies of formation of C and D are identical whereas the entropies are related by $S^\circ(C) = S^\circ(D) - R \ln(2)$ because of the chiral carbon atom and of the “internal” symmetry. As a result the ratio of D to C at equilibrium is 2, in agreement with the equality of the forward and backward step rates. The same reasoning can be applied to the second reaction with $S^\circ(A) = S^\circ(B)$ because of the “external” symmetry of B.

For molecules more complex than those of figure 1, one must calculate a symmetry number σ . Because a given paraffin molecule can have identical lateral branches and symmetries, its graph is left invariant by a set of permutations of vertices of the graph which leave the relation between these vertices unchanged. In the mathematical playground, this set of transformations is called the automorphism group of the graph. It purely represents the degree of symmetry of a structural isomer of the molecule and does not bring any information about its spatial configuration. One has then to be careful when using the representation of Figure 1 that there is no 2D structure left in the underlying use of 120 degree angles.

It is justified to claim that the enumeration of structural isomers and their symmetries has received a satisfactory answer⁹. Usually, the number of symmetries is decomposed into the internal and external symmetries, a decomposition which is only intrinsic through the product which corresponds to the order of the associated automorphism group.

It is not the case as far as stereo-isomers are concerned. Actually, when optical activity is taken into account, the enumeration and the evaluation of the symmetry numbers are only usually performed approximately. For instance, when steric hindrance is excluded, Benson proposed:

$$\sigma = \frac{\sigma_{\text{graph}} 3^M}{2^N} \quad (1)$$

where σ_{graph} is the number of indistinguishable configurations due to internal and external symmetries when chirality is ignored, N is the number of chiral carbon atoms and M is the number of terminal methyl groups. Benson’s formula implicitly assumes that the number of stereo-isomers is 2^N . Figure 2 illustrates 3,4 dimethyl-hexane which is a simple counter-example.

When 3D representation of the molecule is required, such as in this case, we embed the paraffin molecules in a centered network of non-regular octahedral paving the three-dimensional Euclidean space⁷. Subsequently, we consider the equivalence class up to discrete carbon-carbon rotations which is supported by the network and consider a natural representative of this class obtained by placing the longest chain in a plan by locating the position of the possible branches up and down with respect to this plan (see Figure 2).

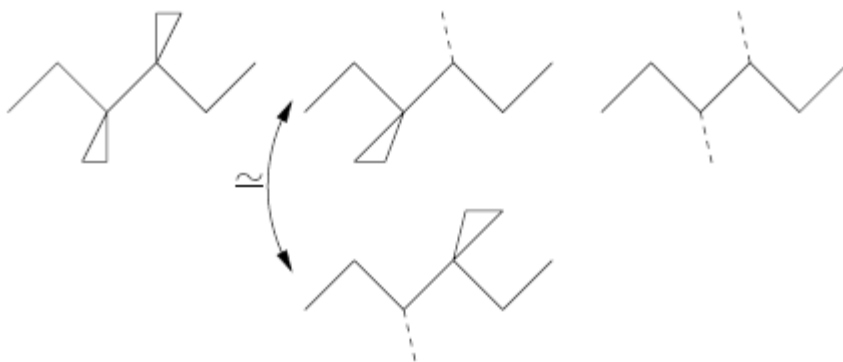


FIG. 2 – Stereo-isomers of 3,4 dimethyl-hexane

This alkane molecule contains two chiral carbon atoms so that the “usual” number of stereo-isomers is given to be $2^N = 4$. However, as represented in Figure 2, two of the four three-dimensional configurations are isomorphic (meso form) since the chirality interferes with the global symmetry of the molecule. Consequently only three stereo-isomers are to be enumerated! Besides, the symmetry properties of the three-isomers considered as objects in the three-dimensional Euclidean oriented space, will not be identical and will then differ from the original symmetry number associated to the graph of the molecule. It becomes then clear that, first the number of stereo-isomers as well as their symmetry properties are not simply provided by the number of chiral carbon atoms and the order of the automorphism group of the graph associated to the molecule when chirality and symmetry interferes; and second, the symmetry properties in such a case have to be evaluated in the 3D oriented Euclidean space.

The systematic enumeration of stereo-isomers and their degree of symmetry appears as a necessary investigation in order to exactly evaluate thermodynamical and kinetic properties. Since it is not possible to tackle the problem in its full generality, we will concentrate on a generic class of paraffins. In many practical applications ranging from chemical engineering, thermodynamics, up to combustion problems such as Massot and co-workers^{10,11}, the branching complexity of the paraffins remains limited.

As a consequence, we restrict ourselves, in this paper, to paraffins being constituted of a main linear chain, on which only methyl ($-\text{CH}_3$) and ethyl ($-\text{C}_2\text{H}_5$) radicals can be connected; in the following this condition will be denoted *Froment's condition*¹, and the corresponding molecules will be denoted paraffins under Froment's condition (paraffins UFC). Let us mention that estimating the symmetry number of a paraffin is sufficient for calculating its thermodynamical properties using Benson's method. Conversely, it is not sufficient for calculating the kinetic rate constant of an elementary step because the latter involves the symmetry number of the intermediate activated complex that does not satisfy Froment's condition. However, the tools developed in this study can be extended in order to tackle the case of the activated complexes even if this is out of the scope of the present paper.

The purpose of the paper is then to present a way to obtain exactly the number and symmetry properties of stereo-isomers of paraffins UFC and to show how the global symmetry of molecules can interfere with chirality. We use the denomination “global symmetry” since for the paraffins UFC with a longest chain of at least 6 carbon atoms, if we exclude the internal symmetries associated to the free CH_3 radicals, the other internal symmetries can only take place at the extremities of the

molecule as presented in Figure 3-(a). In such a case, there exists a unique decomposition of the automorphism group into the internal symmetries S_2 and/or S_3 , corresponding to the four cases presented in Figure 3-(a), and the external symmetry which can only be S_2 so that we have:

$$\sigma_{graph} = \sigma_{glob} \sigma_{int} \quad (2)$$

It is worth noticing the degenerated situations presented in Figure 3-(b), that are generic of what can be encountered for general paraffins not UFC. In such a case, it is for example impossible to decide, except by a conventional choice if the S_3 symmetry is internal or external since considered as an external symmetry, it leave the other branches invariant. Consequently, except for some degenerated situations, the “global symmetry” will precisely denote the S_2 symmetry of the automorphism group associated to the longest chain of the paraffin UFC (for the graph representation of the molecules).

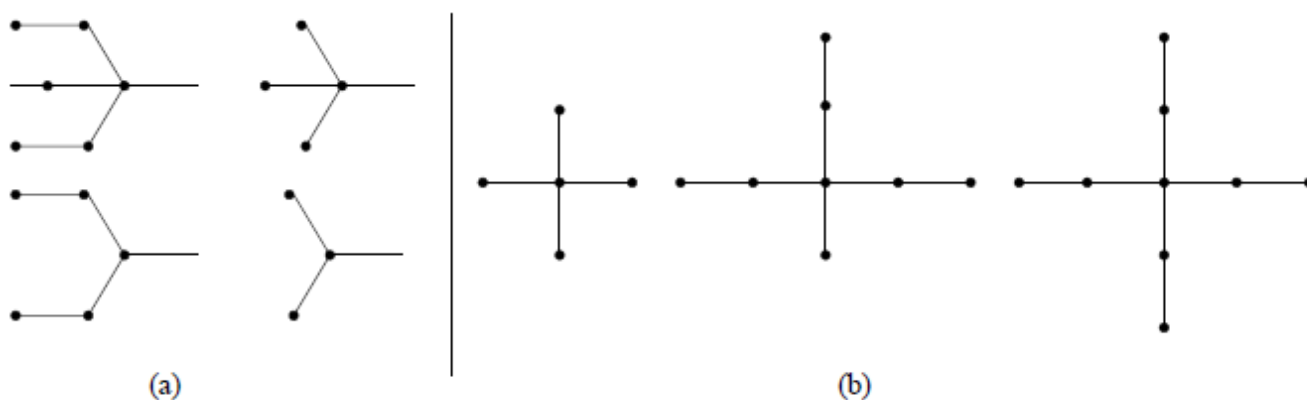


FIG 3 – Possible internal symmetry for paraffins UFC (a), and degenerated situations (b)

The natural mathematical starting point for the enumeration process in the context of graphs, where we want to take into account their possible symmetry, is the theory of generating functions introduced by Pòlya. It was later extended by Read in a very nice way in order to enumerate the isomers and stereo-isomers of various molecules^{12,13,14}. The first part of the paper is then devoted to the application of Read’s theory in the framework of paraffins UFC.

However, this mathematical tool provides the number of structural and stereo-isomers in an implicit manner but does not explicitly yield their symmetry properties. The only way to obtain the numbers we need for the applications is to go through an exhaustive generation of the various topological structures and to calculate from there the associated stereo-isomer numbers with their group of symmetry. To this purpose, we present the coding of the paraffins UFC in a computer algebra code as well as the associated generation algorithms. It is checked that the resulting global numbers of isomers match the results obtained by the generating function theory which then offer a nice certificate. From the exhaustive generation, we can get a number of recurrence properties from which the information we need can be extracted up to an arbitrary number of carbon atoms

since the exhaustive generation is only easily tractable up to 23 carbon atoms as opposed to the generating function approach presented in the first part ⁷.

In the third part, we show the ratio of the number of symmetries of the molecule to the number of stereo-isomers (Benson's formula 1) is valid only when chirality and the group of symmetry of the molecule do not interfere. We describe, using arguments from statistical physics related to the mixing entropy, how the entropy of formation given by Benson ⁸ has to be modified in some situations in order to take into account this interference.

As a conclusion, we finally explain how the result obtained in the present paper can be generalized and can bring exact evaluation of chirality properties in the framework of results obtained by Muller, Scacchi and Come on structural isomers ⁹.

2 Generating functions: the certificate

In this section we enumerate the structural isomers of paraffins UFC as well as their degree of symmetry (in the sense of graphs which means without any spatial representation) using the generating function concept. We then deduce the enumeration and symmetries of stereo-isomers defined by the distinguishable three-dimensional structures owing to the tetrahedral organization around carbon atoms ⁷.

The automorphism group is at the heart of the strategy introduced by Pòlya and extended by Read ^{7,12-16}. This strategy makes the link between the enumeration problem and the automorphism group of the graph associated to the molecule. Making use of Read's study ¹⁴, we state the following theorem and provide only the sketch of the proof, the comprehensive proof being given in the Appendix.

Theorem. – *The number of structural isomers of paraffin molecules with n carbon atoms under Froment's condition is given the coefficient of x^n in*

$$\frac{S_i^{neq}(x)b(x)}{1-b(x)} + \frac{S_i^{eq}(x)}{2} \left[\frac{1}{1-b(x)} + \frac{1+b(x)}{1-b(x^2)} \right] + P_i(x);$$

Where

$$b(x) = x + x^2 + 2x^3 + x^4 + x^5$$

$$S_i^{neq} = x^5(1+x+x^2)$$

$$S_i^{eq} = x^4(1+x^2+x^4)$$

$$\text{and } P_i(x) = x + x^2 + x^3 + x^4 + 2x^5 + x^6 + x^7$$

-The number of stereo-isomers of paraffin molecules with n carbon atoms under Froment's condition is given by the coefficient of x^n in

$$\frac{S_s^{neq}(x)}{1-xc^2(x)} + \frac{S_s^{eq}(x)}{2} \left[\frac{1}{1-xc^2(x)} + \frac{1}{1-xc(x^2)} \right] + P_s(x)$$

where $c(x) = 1 + x + x^2$

$$S_s^{neq} = x^7(3 + 8x + 21x^2 + 34x^3 + 50x^4 + 52x^5 + 49x^6 + 32x^7 + 18x^8 + 7x^9 + 2x^{10})$$

$$S_s^{eq} = x^6(1 + 3x^2 + 5x^4 + 6x^6 + 6x^8 + 2x^{10} + x^{12})$$

and $P_s(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 4x^6 + 7x^7 + 8x^8 + 9x^9 + 7x^{10} + 6x^{11}$

Sketch of the Proof:

We start by considering the problem of counting structural isomers. In the following, the *skeleton* of a paraffin UFC will denote two radicals X and Y joined by a chain of carbon atoms. Thus, these paraffin molecules will be composed by a skeleton on which radicals are connected. For the purpose of counting structural isomers, X and Y will take their value in the set \mathcal{F} , whereas they will take their value in an other set, denoted \mathcal{F}_s , for the purpose of counting stereo-isomers in the second part of the proof. In Figure 4-(a), X, Y are in \mathcal{F} , Q_i and R_i are radicals such that:

$$\mathcal{F} = \{-C_2H_5, -CH(-CH_3)_2, -C(-CH_3)_3\}, \quad Q_i, R_i \in \{-H, -CH_3, -C_2H_5\}.$$

This extension of the concept of main chain will then be coherent with Read's method and allows us to control isomorphism phenomena at extremities. Furthermore, the similarity between the structure of paraffins UFC and the structure of di-substituted hydrocarbons ($C_nH_{2n}XY$), enumerated by Read, is obvious.

Let us first suppose that X and Y are different. Consider any pair of 'boxes' Q_i and R_i . Since we are counting structural isomers as presented in Figure 4-(a), we can permute these radicals independently from the others, and it follows that the group of permutation is $S_2 \times S_2 \times \dots \times S_2$ with p factors. We then make use of Pólya's Theorem, generate the counting series and take into

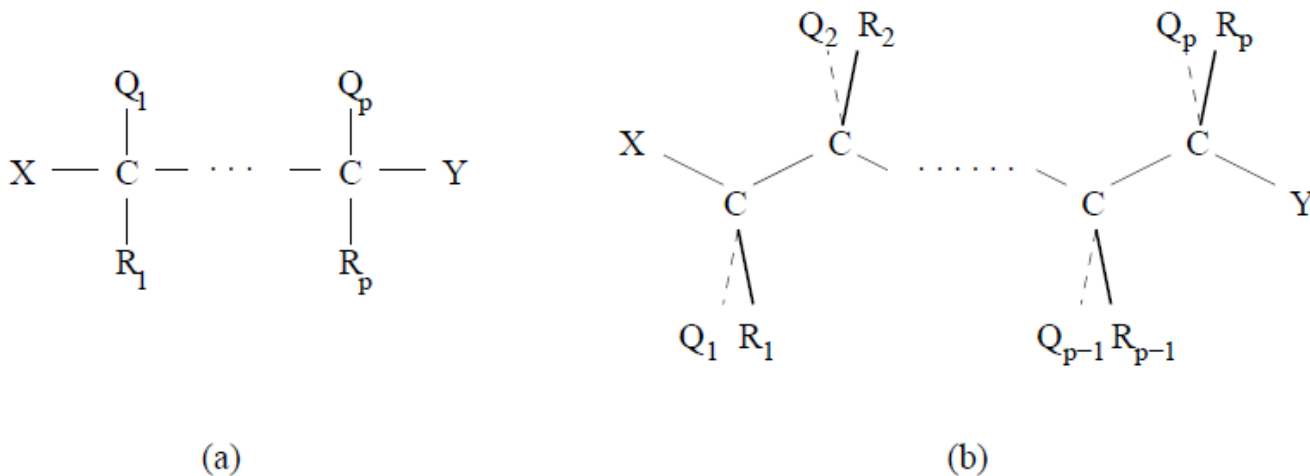


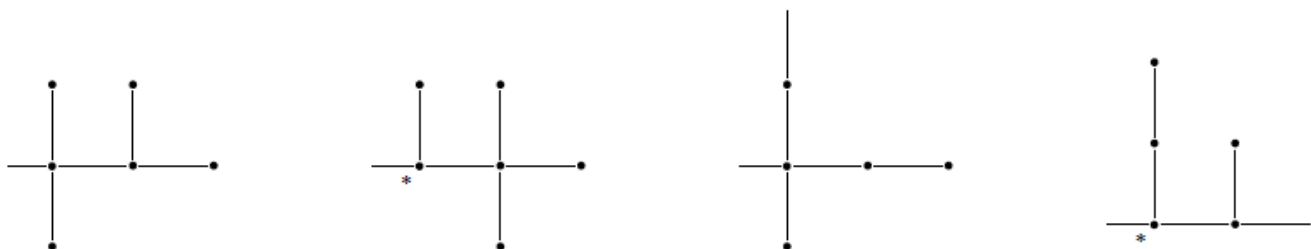
FIG. 4 – Representation of paraffins UFC

account the extra carbon atoms added to the skeleton through X and Y by multiplying by the ‘shift’ polynomial, S_i^{neq} , related to the number of added carbon atoms in the extremities. All the possible configurations of the two extremities have to be determined and are given in details in the Appendix.

When X and Y are the same, the skeleton admits a supplementary global symmetry. If $p = 2k$ is even then, the group of permutations is now $S_2 \times S_2 [S_2 \times \dots \times S_2]$ (k factors). We then use the same kind of argument. The details of the obtained counting series are given in the Appendix.

The last step to explicitly express the counting polynomial $P_i(x)$ which enumerates paraffins for small values of n , n up to 7, which cannot be obtained by the previous method because of the structures of \mathcal{F} and consequently of the skeleton. Thus, the counting series of structural isomers of paraffins UFC is obtained by summing the counting series and $P_i(x)$.

To enumerate the paraffins, having regard to stereo-isomers, we use the same method as before. Taking into account the chirality, we have to consider a 3D representation as illustrated in Figure 4-(b). However a new difficulty arises. Actually, the set of extremities cannot be taken to be \mathcal{F} any more because of the possible chirality of the third carbon starting from one end of the main chain, which will depend on the branches R_1 and Q_1 . In order to avoid these interferences, the length of the main chain involved in the extremities has to be 3 carbon atoms. Here is, for example, the possibilities when X or Y have 6 carbon atoms:



Alkyl radicals with 6 carbon atoms. (star denote the chiral character of a carbon atom).

If X and Y are different, the group is E_{2p} (identity group). Therefore, the counting series are given by Pölya’s theorem and we now have to take into account the extra carbon atoms in X and Y . There are many possible combinations; they are given in Table V of the Appendix which allow to determine the “shift” polynomial, S_8^{neq} . If X and Y are the same, the group is $S_2[E_p]$. As in the case $X \neq Y$, we provide the numbers of possibilities for the extremity with table VI in the Appendix. We deduce the “shift” polynomial, S_8^{eq} . Finally, the counting polynomial $P_8(x)$ is the analogous of $P_1(x)$ as far as stereo-isomers are concerned.

The configuration counting series which gives us the number of stereo-isomers of paraffins UFC is finally obtained on summing the counting series and $P_8(x)$.

The computer algebra of Maple allows finding very quickly the coefficient of x^n for large values of n . As an example, the coefficient of x^{50} in the counting series of stereo-isomers is given

by 598 411 199 178 385 732. The numbers of structural isomers and stereo-isomers of paraffins UFC and paraffins after Read ¹⁴ are tabulated in Table I.

n	isomers			stereo-isomers		
	paraffins UFC	paraffins	%	paraffins UFC	paraffins	%
1	1	1	100	1	1	100
2	1	1	100	1	1	100
3	1	1	100	1	1	100
4	2	2	100	2	2	100
5	3	5	100	3	3	100
6	5	5	100	5	5	100
7	9	9	100	11	11	100
8	18	18	100	24	24	100
9	35	35	100	55	55	100
10	71	75	94.7	132	136	97.1
11	143	159	89.9	321	345	93.0
12	295	355	83.1	780	900	86.7
13	606	802	75.6	1 915	2412	79.4
14	1261	1858	67.9	4 705	6563	71.7
15	2616	4347	60.2	11 567	18127	63.8
16	5461	10359	52.7	28 468	50 699	56.2
17	11 385	24 894	45.7	70 108	143 255	48.9
18	23 806	60 523	39.3	172 721	408 429	42.3
19	49 748	148 284	33.6	425 631	1 173 770	36.3
20	104 109	366 319	28.4	1 049 038	3 396 844	30.9
21	217 808	910 726	23.9	2 585 793	9 892 302	26.1
22	455 993	2 278 658	20.0	6 374 186	28 972 302	22.0
23	954 512	5 731 580	16.6	15 713 531	85 289 390	18.4
24	1 998 699	14 490 245	13.8	38 737 748	252 260 276	15.4
25	4 184 892	36 797 588	11.4	95 499 763	749 329 719	12.7
30	168 559 829	4 111 846 763	4.1	8 697 332 533	182 896 187 256	4.8
40	273 611 306 905	62 481 801 147 341	0.4	72 142 591 078 702	13 180 446 189 326 135	0.6

Table I: Numbers of structural isomers and stereo-isomers of paraffins.

Figure 5 illustrates the percentage of paraffins UFC among the set of any paraffins with n carbon atoms obtained from Read ¹⁴. It is shown to decrease exponentially at infinity and it quickly reaches low values even for moderate n . $n = 30$ can be considered to be relevant of industrial hydrocracking operations. Figure 5 shows that kinetic modeling of the hydrocracking process under Froment's condition deal with about 4% of the whole set of paraffins that can be involved. The trees satisfying Froment's condition then asymptotically represent only a family of zero density as the whole of the paraffin molecules which decrease exponentially at infinity. Valery³ studied experimentally the hydrocracking of n-hexadecane, mixtures of long-chain normal paraffins ($n = 20$ to 30) and a squalane (2,6,10,15,19,23-hexamethyltetracosane). He found some discrepancies

between experimental and simulation results for the higher molecular weights and especially with squalane that is a branched reactant. This could be attributed to the restricted framework of Froment's condition, besides the challenge of analyzing a mixture of numerous isomers. Finally, Figure 5 illustrates the percentage in number; using Benson's method, it would be nice to obtain the percentage in weight at chemical equilibrium even if this is out of the scope of this paper.

We next focus on the explicit and exhaustive generation of the structures enumerated in this section, and further evaluate their degree of symmetry and geometrical properties.

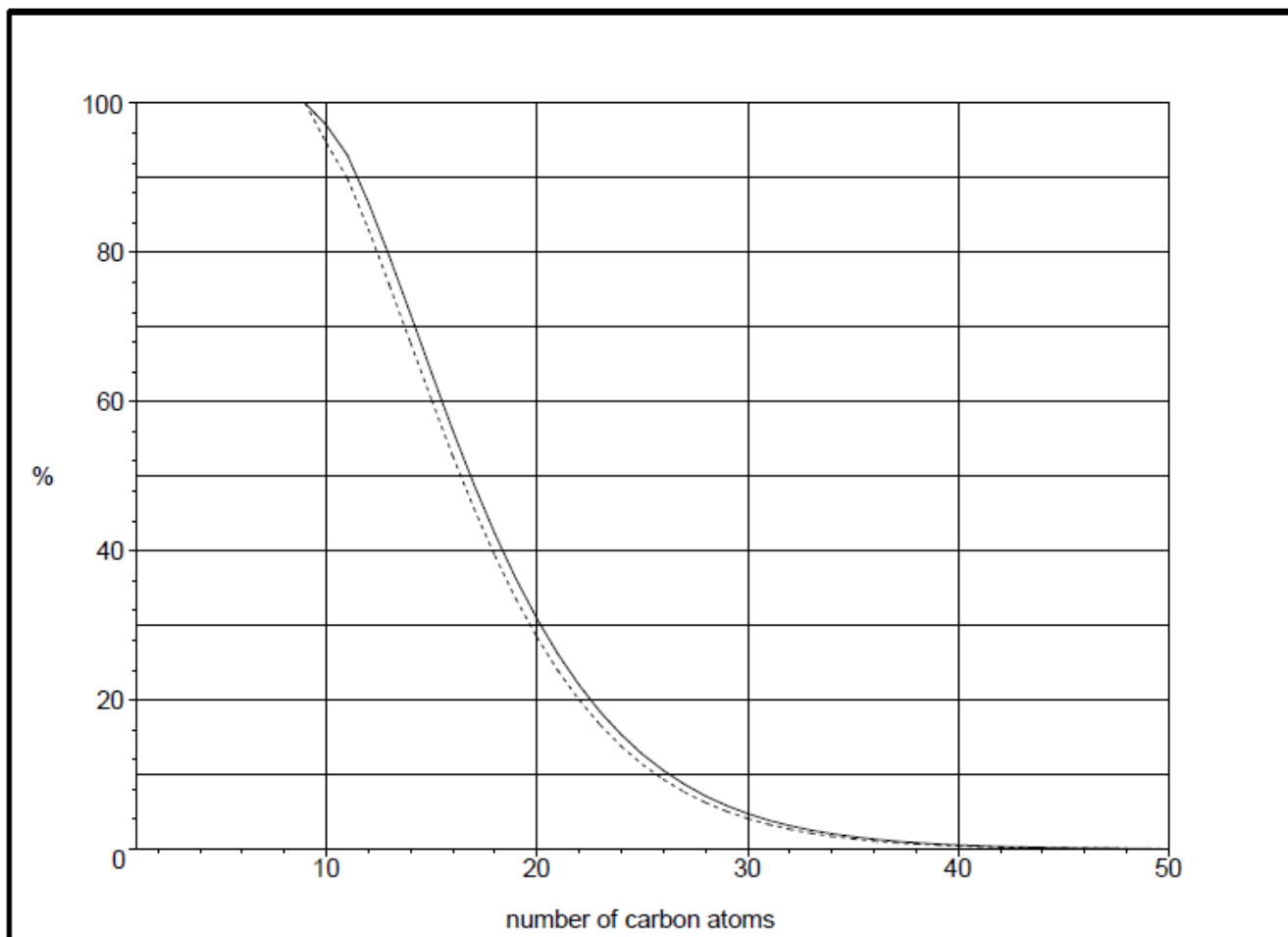


FIG. 5 – Proportion of paraffins UFC; continuous: stereo-isomers; dotted: structural isomers.

3 Explicit generation

The approach used by Read revealed itself as a breakthrough in the field; however, it is implicit in the sense that it only yields one global numerical information. Some more information is needed for evaluating the thermodynamical and kinetic properties of paraffins UFC. We propose a new approach which is explicit and granular, and thus fundamentally different from the preceding one; it is rendered possible by the developments of the power of computer algebra^{7,17-19}. The granular character of the approach means that we may generate sub-classes of paraffins UFC thus avoiding partially the explosive character of the generation.

The base of our strategy is to embed the paraffin molecules in a centered network of non-regular octahedra paving the three-dimensional Euclidean space. This network allows an easy coding of the molecule as well as an effective isomorphism test through discrete rotations associated to the network. The generation consists of three successive steps. We first of all generate structural isomers and classify them in families according to their topological properties (number and type of branches on the main linear chain); each family is then partitioned into subfamilies based on the automorphism group associated to the molecule (symmetry property of the graph – no spatial information). Finally we rigidify the obtained structures, place them in the centered network of non-regular octahedra and generate the whole set of stereo-isomers thanks to the exhaustive study of chirality⁷.

3.1 Fundamentals of the coding-structural isomers

The selected coding consists of encased lists. This type of object allows a fast access to the data with Maple. The coding only has three operands on the first level as far as the first two steps of the enumeration are concerned:

- t , size of the principal chain (the number of vertices);
- $[m_1, m_2, \dots, m_p]$, the list of the ranks of the connected methyls;
- $[e_1, e_2, \dots, e_q]$, the list of the ranks of the connected ethyls.

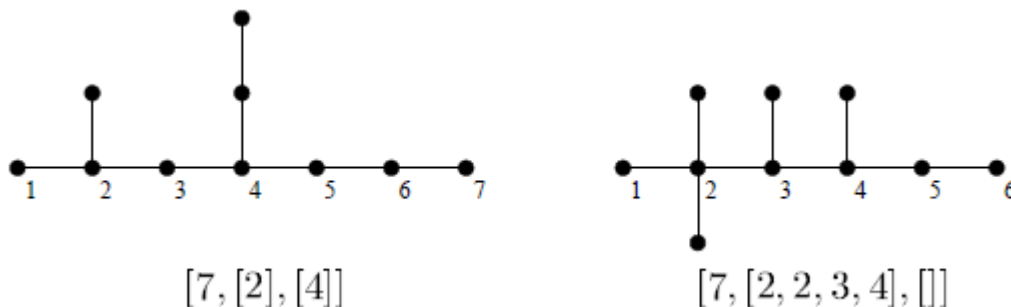


FIG. 6 – Two non isomorphic configuration of $C_{10}H_{22}$

3.2 The Weight of a coding

A significant question, leading to an orientation in the programming, was to know how to avoid generating a too great number of molecules. Indeed, as we want to enumerate non isomorphic configurations, the two graphs of Figure 7, for example, will have to be regarded as two representations of the same molecule. A way of generating only one of these two configurations is to introduce a function *weight*. The weight of a coding will be defined by the function

$$f : [t, [m_1, m_2, \dots, m_p], [e_1, e_2, \dots, e_q]] \rightarrow \sum_{i=1}^p m_i + 2 \sum_{j=1}^q e_j$$

This function calculates a certain barycenter on the connections (a weight of 1 is allocated to methyls and a weight of 2 with ethyls). The program begins and ‘sets’ the connections on the left of the chain. In this way, it avoids counting twice the same molecule, i.e. the identical molecules according to whether one reads them starting from the right-hand side or from the left-hand side.

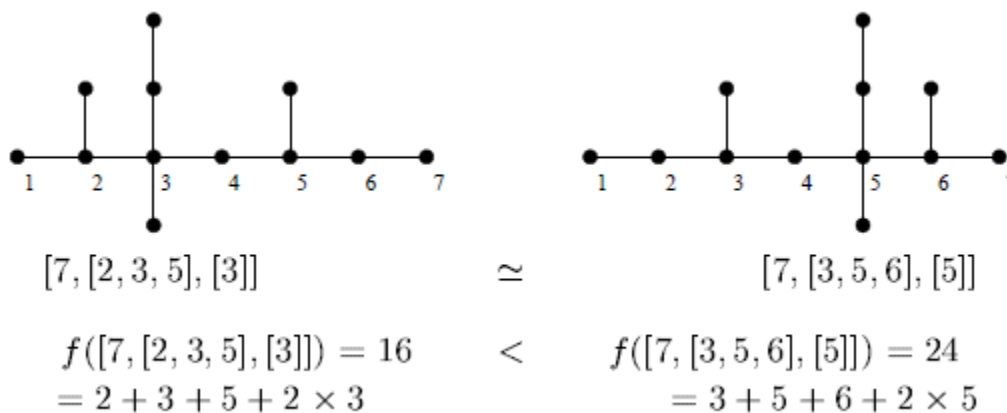


FIG. 7 – Two isomorphic configurations with different weights

This weight function coupled to the coding allows us to generate very efficiently the structural isomers. Once this generation is performed, we are ready to deal with the three-dimensional nature of the molecules and the chirality effects.

3.3 The 3D-coding – Stereo-isomers

The molecules are embedded in the 3D network of non-regular centered octahedra and we consider the equivalence class up to discrete carbon-carbon rotations; it is supported by the network. A natural representative of this class, because of the Froment’s condition, is obtained by choosing a molecule the longest chain of which lies in a plane and by locating the position of the possible branches by “U” for up and “D” for down with respect to this plan. It is valid since, if one takes into account thermal agitation, the molecules undergo rotations around any carbon-carbon axis. Among all space conformations of a given molecule, one will retain that whose main chain is planar (see Figure 8). The geometrical representations of the molecules will be thus in three dimensions in order to account for chirality.

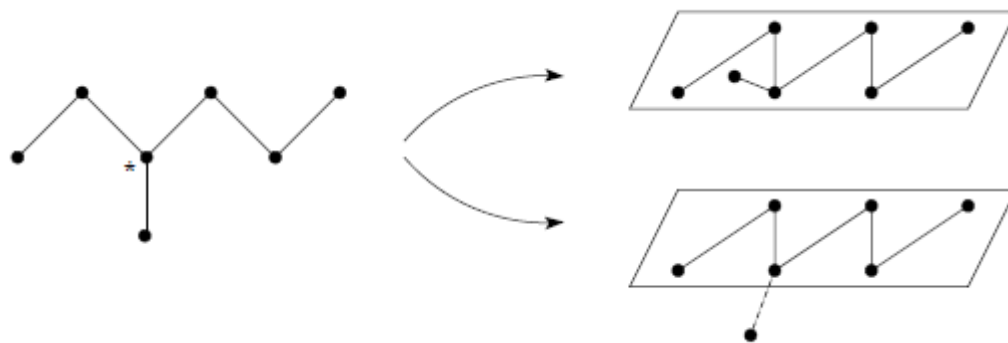


FIG. 8 – One graph, two stereo-isomers

Thus the spatial configuration of the molecule is coded in a canonical way.

t , size of the principal chain (number of vertices);

$[m_1, m_2, \dots, m_p]$, list of ranks of the connected methyl radicals;

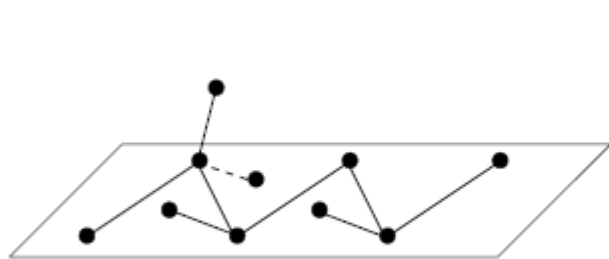
$[\varepsilon_1^m, \varepsilon_2^m, \dots, \varepsilon_p^m]$, list of position of methyl radicals, where ε_i^m belongs to the set $\{U, D\}$. It is chosen according to whether the i -th methyl branch is above the plan of the principal chain or below

$[e_1, e_2, \dots, e_q]$, list of ranks of the connected ethyl radicals.

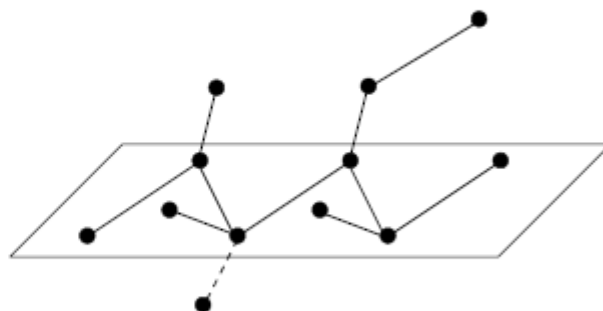
$[\varepsilon_1^e, \varepsilon_2^e, \dots, \varepsilon_q^e]$ list of position of ethyl radicals, where $\varepsilon_i^e \in \{U, D\}$

The Figure 9 shows two examples of the coding.

•



$[6, [2, 2, 3, 5], [U, D, U, U], [], []]$



$[6, [2, 3, 3, 5], [U, U, D, U], [4], [U]]$

FIG. 9 – Example of 3D codings

The detection of chiral atoms and isomorphy are done through simple tests on the introduced 3D coding, thus rendering the approach very efficient.

3.4 Exhaustive generation

The enumeration procedure encounters a combinatorics explosion as predicted by the theory of generating functions. Consequently, the global generation and storage of *all* configurations rapidly becomes computationally expensive and requires a lot of memory storage.

To avoid this problem, we take profit of the granular character of the approach. It consists in solving sub-problems defined by supplementary constraints. For instance, we may study the set, $G_{p,q}(n)$, of paraffins UFC with n carbon atoms, p methyl and q ethyl side chains, n , p and q being fixed and setting the topology of the molecule. Let us notice that this type of set is used in the application of the single event theory because any tractable chemical analysis of a paraffin mixture allows determining isomers up to $p + q = 3$ at most. The parallelization of the program over various p and q is then straightforward since $G(n) = \bigcup_{p,q} G_{p,q}(n)$ and $\bigcap_{p,q} G_{p,q} = \emptyset$ where $G(n)$ is

the full set of paraffins UFC with n carbon atoms. The granularity can even be finer since the sets $G_{p,q}(n)$ themselves can be partitioned with respect to the automorphism group of the graph of the molecule. Let us mention that S_1 is the identity which means that the molecule has no symmetry. The order of the automorphism group is the cardinal of this group and is representative of the degree of indistinguishability of the atoms constituting the molecule. As explained in the introduction, it can be viewed as the product of external and internal symmetries⁸ provided that the free rotation of terminal $-\text{CH}_3$ is excluded. This is because our approach focuses on carbon atoms and ignores hydrogen atoms. If there are M terminal methyl groups, the “chemical global symmetry” is $\sigma_{\text{global}} = 3^M \sigma$ when chirality is ignored (structural isomers generation). Table III provides characteristics of $G_{2,2}(n)$ up to $n = 25$ carbon atoms when chirality is not considered. Inspecting the $(S_2)^3$, $S_2 \times S_3$, $(S_2)^4$ and $(S_3)^2$ columns immediately suggests some recurrence formulae. These formulae depend on p , q and the automorphism group, and they must be derived case by case; this can be done systematically for all the columns in Table III. Eventually, they yield the number of isomers for arbitrary n without requiring the exhaustive generation. Some recurrences appear to be so “obvious” that rigorous demonstrations are useless from a pragmatic point of view. Such recurrences are explicitly described for charged molecules (ions) by Valéry et al²⁰. and by Guillaume et al²¹. From the numbers obtained, we can easily compute the associated numbers for stereo-isomers presented in Table III-bis by using simple isomorphism tests.

p=2 q=2	automorphism group								sum
	S_1	S_2	$(S_2)^2$	S_3	$(S_2)^3$	$S_2 \times S_3$	$(S_2)^4$	$(S_3)^2$	
n									
11	0	0	0	0	0	0	1	1	2
12	0	0	3	0	3	4	0	1	11
13	0	6	15	2	3	8	2	1	37

14	2	30	32	6	7	12	0	1	90
15	14	74	62	12	7	16	2	1	188
16	42	158	93	20	11	20	0	1	345
17	102	276	141	30	11	24	2	1	587
18	204	456	186	42	15	28	0	1	932
19	372	684	252	56	15	32	2	1	1414
20	620	996	311	72	19	36	0	1	2055
21	980	1370	395	90	19	40	2	1	2897
22	1470	1850	468	110	23	44	0	1	3966
23	2130	2406	570	132	23	48	2	1	5312
24	2982	3090	657	156	27	52	0	1	6965
25	4074	3864	77	182	27	56	2	1	8983

Table III: Distribution of isomers according to the automorphism group. The lateral branches are composed by 2 methyls and 2 ethyls

p=2 q=2	automorphism group								sum
	S_1	S_2	$(S_2)^2$	S_3	$(S_2)^3$	$S_2 \times S_3$	$(S_2)^4$	$(S_3)^2$	
11	0	0	0	0	0	0	1	1	2
12	0	0	6	0	5	6	0	1	18
13	0	26	36	8	5	12	2	1	90
14	16	153	77	24	13	18	0	1	302
15	140	403	155	48	13	24	2	1	786
16	456	902	232	80	21	30	0	1	1722
17	1184	1616	358	120	21	36	2	1	3338
18	2472	2727	471	168	29	42	0	1	5910
19	4668	4145	645	224	29	48	2	1	9762
20	7984	6108	794	288	37	54	0	1	15266
21	12 896	8470	1016	360	37	60	2	1	22842
22	19 680	11 525	1201	440	45	66	0	1	32958
23	28 940	15 071	1471	528	45	72	2	1	46130
24	41 016	19 458	1692	624	53	78	0	1	62922
25	56 640	24 428	2010	782	53	84	2	1	83946

Table III-bis: *Distribution of stereo-isomers according to the automorphism group. The lateral branches are composed by 2 methyls and 2 ethyls.*

In Figure 10, we give a few examples of molecules with $n = 11, 12$ and 13 having as automorphism groups $S_2 \times S_3$, $(S_2)^4$ and $(S_3)^2$. The calculation of the number of stereo-isomers performed here can be considered as an extension of the method by Muller et al.⁹ since the chirality properties are properly taken into account. We will come back to this point in the conclusion and see how the methods presented in this paper can be extended to a more general class of molecules than the paraffins UFC.

However, the diversity of such families $G_{p,q}$ increases with n as noticed by Lygeros et al.⁷. A second method of generation called “mixed” consists in using the already known topological structure and the previously developed recurrence formulae and in generating explicitly the new families created as n increases⁷.

It is worth insisting on the fact that the knowledge of the detailed structure of all the isomers and stereo-isomers in a very precious information in order to conduct accurate thermochemical and kinetic calculations. This will be illustrated in the next section

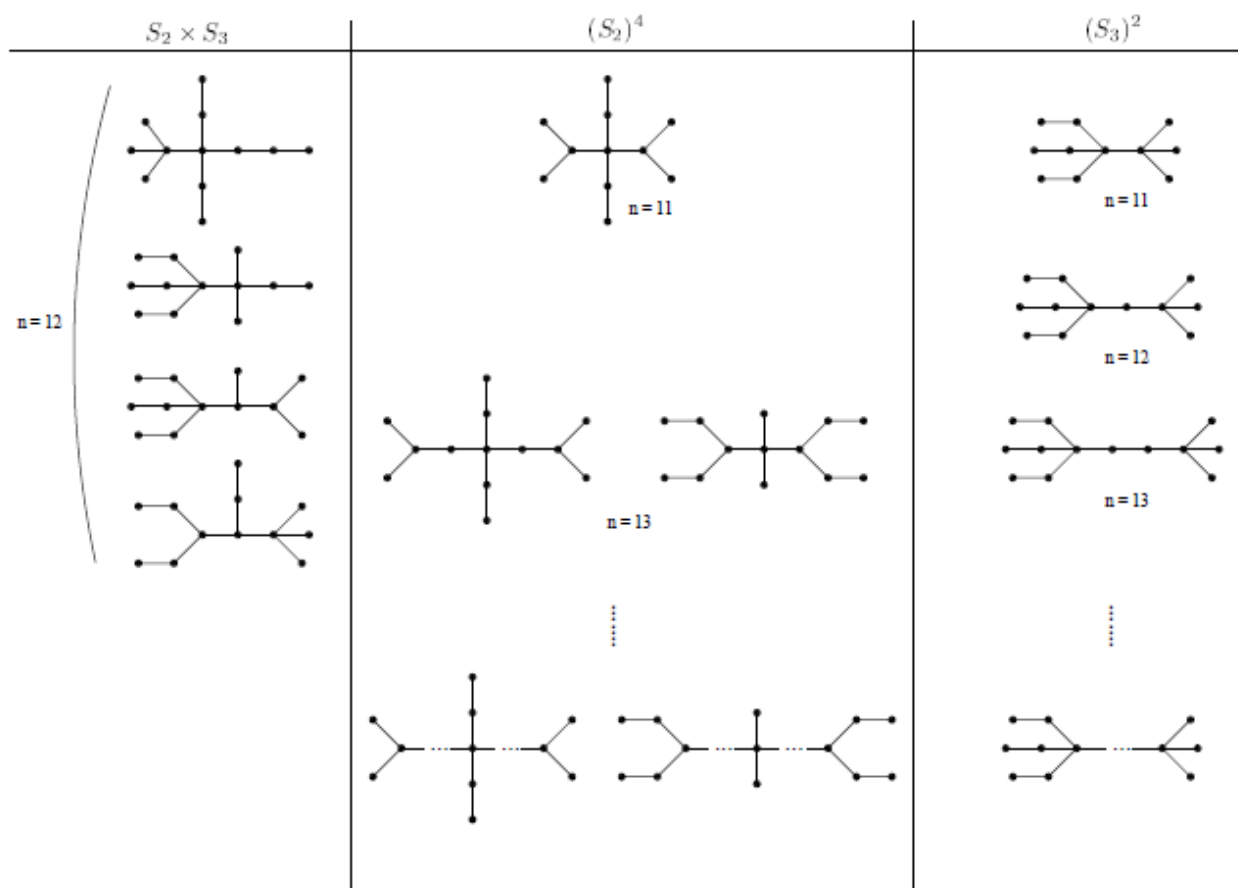


FIG 10 – Examples of molecules having $S_2 \times S_3$, $(S_2)^4$ and $(S_3)^2$ as automorphism group.

4 Applications

We will here measure the impact of the enumeration and generation on applications. The first subsection will be devoted to the enumeration of stereo-isomers when the molecule admits at least one chiral carbon atom, a situation where the number of stereo-isomers is usually approximated by 2^N , N being the number of chiral carbon atoms. The second subsection will deal with Benson evaluation of entropy when chirality and global symmetry interfere. Using statistical physics and the results presented above, a general symmetry correction for entropy will be demonstrated for paraffins UFC.

In this section, since we consider paraffins UFC, the only possibility of getting a factor of three in the number of indistinguishable configurations due to the symmetry of the molecule is to be found at the extremities and can not interfere with chirality as already mentioned (See Figure 14 at the end of this section for a richer situation). Consequently, according to Benson’s method ⁸, the entropy of formation of a given structural isomer is given by:

$$S^0 = S_{contrib}^0 + S_{corr} \quad (3)$$

where $S_{contrib}^0$ is an “intrinsic” term due to the groups that compose the molecule, and S_{corr} is a term accounting for symmetries. In the framework of paraffins UFC, if we exclude some particular degenerate situations that can be treated separately, we have

$$S_{corr} = S_{corr}^{glob} + S_{corr}^{int} + S_{corr}^M$$

where the second term is purely due to S_2 or S_3 symmetries at the extremities due to C-C rotations (internal symmetries), where the last one is due to terminal methyl groups and where the first term accounts for the global symmetry of order two eventually coupled to chirality.

In this section, we then focus on the *exact* evaluation of the number of stereo-isomers and on the interference of chirality and global symmetries of the molecules.

4.1 Symmetry, Chirality and the number of stereo-isomers

The explicit enumeration of stereo-isomers is conducted starting from the generated set of structural isomers with the automorphism group associated to the graph of each molecule. At this level global symmetry and chirality can interfere as in the example of Figure 2 provided in the introduction.

In fact, each graph, symbolizing a structural isomer, yields one more stereo-isomer configurations according to the number of chiral vertices it contains. In the case of Figure 8, the presence of a chiral atom generates two stereo-isomers. It is related to the number N_{oi} introduced in Benson for the calculation of the formation entropy ⁸. In the case of N chiral atoms, if we use the notation introduced in the 3D-coding of the molecules, the graph, symbolizing a structural isomer generated 2^N AB-configurations (depending on the position of the two branches: Above or Below) such as the 16 AB configurations represented in Figure 11.

The number of stereo-isomers is clearly 2^N if the graph does not have a global symmetry of order two in his automorphism group. Indeed, let us consider the symmetrical graph which contains 4 chiral carbon atoms represented in Figure 11. Out of the $2^4 = 16$ AB-configurations, several are one-to-one isomorphic and joined by arrows. Thus there are 16 AB-configurations to be enumerated but only 10 distinguishable stereo-isomers as represented in Figure 11. Besides the various stereo-isomers can have various group of 3D symmetries preserving the orientation (rotations but not planar symmetries); in Figure 11, the stereo-isomers corresponding to the AB-configurations at the corners admit a symmetry of order 2, whereas the other ones do not admit any symmetry. At a general level, let us remark that this situation can be encountered only when N is even (when N is odd, the molecule cannot be symmetric) and let us conduct the enumeration of the stereo-isomers and degree of symmetry for an arbitrary even N in the following lemma.

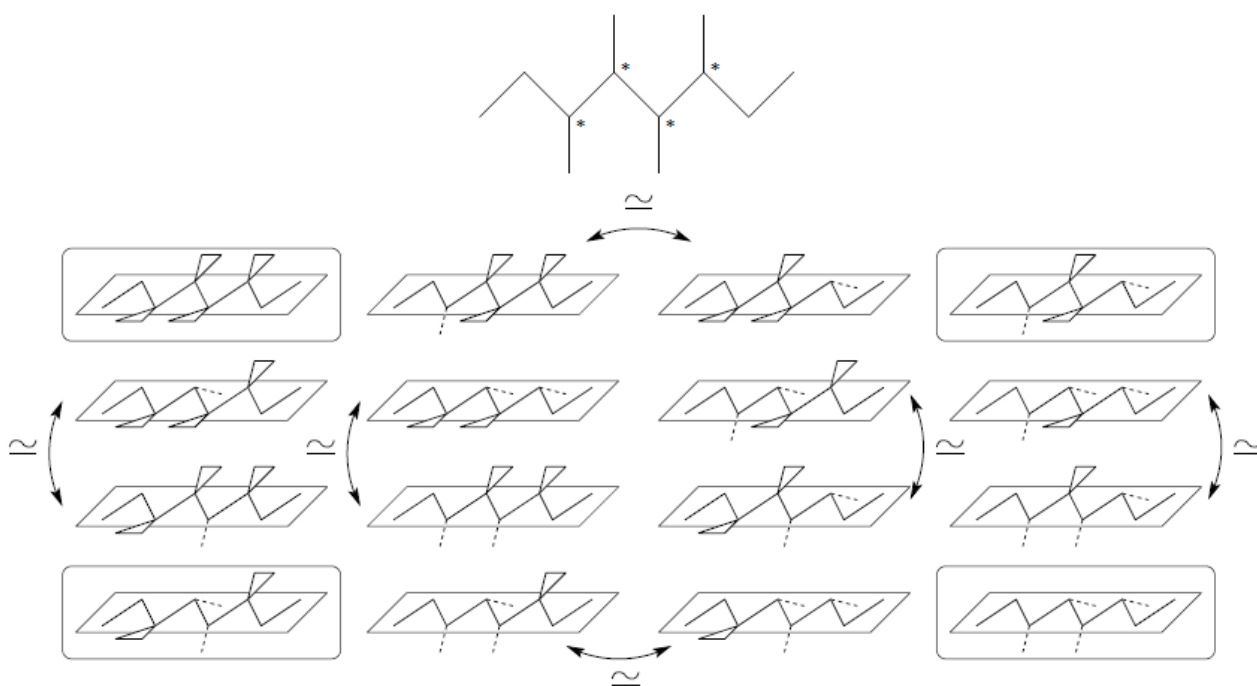


FIG. 11 – A paraffin molecule UFC and its various AB-configurations, with a global symmetry and four chiral carbon atoms.

Lemma. – Let us consider a paraffin UFC which admits a global symmetry and a number of N chiral atoms, N even.

Case I: assume that it satisfies one of the two conditions:

- (i) the main chain admits an even number of carbon atoms,
- (ii) the main chain admits an odd number of carbon atoms and the remaining two branches of the central carbon atom are the same.

Then there are $2^{N-1} + 2^{N/2-1}$ stereo-isomers; $2^{N/2}$ have a symmetry of order two and the other $(2^N - 2^{N/2})/2$ do not have any symmetry.

Case II: *the main chain admits an odd number of carbon atoms and the remaining two branches of the central carbon atom are different. Then the number of stereo-isomers is 2^N without any symmetry.*

Proof: Indeed, each side branch is either ‘Above’ or ‘Below’ the plan of the main chain. From a given graph, let us consider the set E_N formed by the 2^N configurations obtained by combinations of the positions of the side branches (the unit E_4 is represented figure 11). The main chain is the same for any element of E_N and admits symmetry by rotation of axis (orthogonal to the plan if the chain has an even number of carbons, in the plan if not). Let us consider the case (i), the other case (ii) being essentially the same. The image of an element E_N from this rotation belongs to E_N (this operation amounts to the reversing certain connections). The image and the antecedent are confused in E_N if the configuration is invariant by rotation. These configurations are $2^{N/2}$. The total number of elements of E_N non-isomorphic one-to-one is then given by:

$$\begin{aligned} 2^{N/2} + \frac{2^N - 2^{N/2}}{2} &= 2^{N/2} + 2^{N-1} - 2^{N/2-1} \\ &= 2^{N/2-1} + 2^{N-1} \end{aligned}$$

In the situation of Case II, for example if the central carbon atom, which belongs to the axis of symmetry has H and CH_3 as the two other branches, we can consider the same reasoning as before forgetting the branches on the central carbon atom and then see the influence of these branches.

It is easy to see that each of the $2^{N/2}$ symmetrical configurations generates 2 configurations which are isomorphic to each other, thus generating the same number of stereo-isomers. As far as the other $\frac{2^N - 2^{N/2}}{2}$ configurations are concerned, each of them generates two stereo-isomers so that finally they sum of the two reaches 2^N .

Some examples of the two cases investigated in the Lemma are presented in Figure 12 and Figure 13. In case II, remark that a substituted central carbon atom is never chiral although the two halves of the main chain can be in different optical configurations.

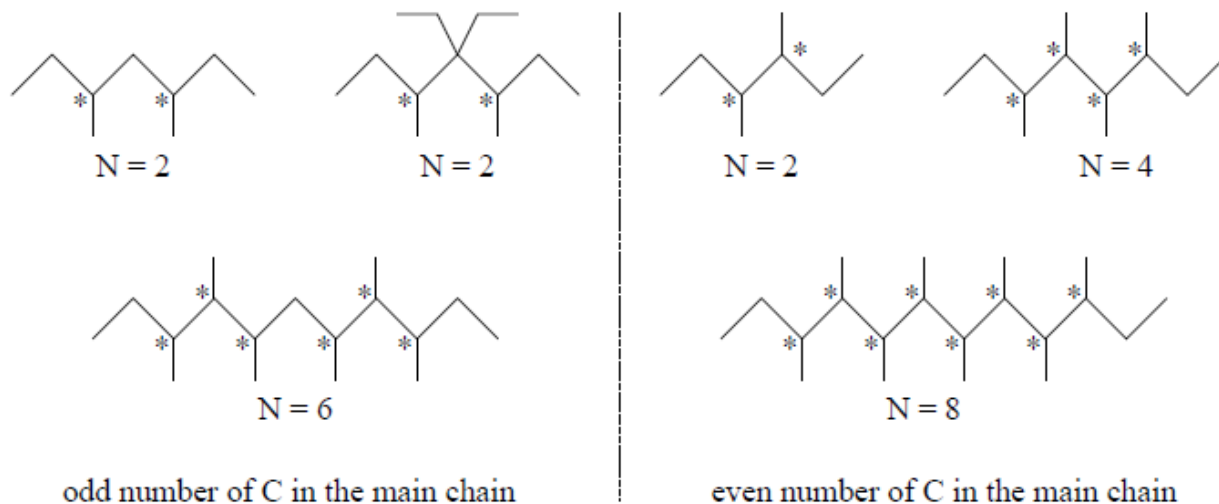


FIG 12 – Examples of configurations where symmetry and chirality interfere – Case I

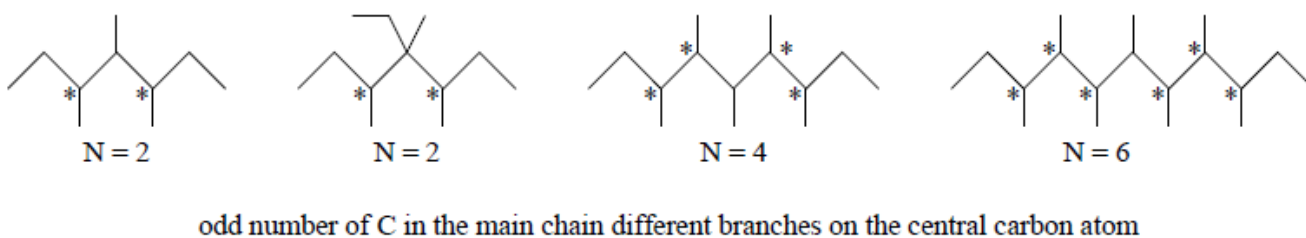


FIG 13 – Examples of configurations where symmetry and chirality interfere – Case II

Table IV gives the number of paraffin stereo-isomers according to N in Case I. From the lemma, we see that using 2^N asymptotically overestimate the number of stereo-isomers by a factor 2 for large N .

N	Number of stereo-isomers	2^N	Percentage of errors
2	3	4	33.3%
4	10	16	60.0%
6	36	64	77.8%
8	136	256	88.2%
10	528	1024	93.9%

Table IV: Number of paraffin stereo-isomers according to N in Case I.

Now that we have the exact number of stereo-isomers, it is tempting to use Benton's formula with this number instead of 2^N and σ_{graph} as defined in the introduction. However, things are not that simple as we will see in the next sub-section.

4.2 Entropy of formation – back to statistical physics

The purpose of the present sub-section is to demonstrate that the usual formula used by Benson for the evaluation of the entropy correction can fail in the sense of statistical physics in the case when chirality does interfere with the automorphism group of the graph associated to the structural isomer. If no interference is to be found, the number of stereo-isomers is 2^N , N being the number of chiral carbon atoms and the entropy correction due to chirality and global symmetry of order 2 is ⁸

$$S_{\text{corr}}^{\text{glob}} = \log\left(\frac{2^N}{2}\right) = \log(2^{N-1})$$

In the situation where the paraffin molecule UFC has a global symmetry in Case I or II, we define the racemic mixture by assuming that all the AB-configurations have equal chance to be represented thus defining various mole fractions for the various stereo-isomers depending on how many AB-configuration they represent. The previous formula is still valid in Case I, but as we will see in the following, a compensation is taking place, but ceases to be valid in Case II and the correct value is given by the following Theorem.

Theorem. – *Let us consider a paraffin UFC with a global symmetry and N chiral atoms, N even. In Case I, there are $2^{N-1} + 2^{N/2-1}$ stereo-isomers and the entropy correction in the sense of Benson due to chirality and global symmetry of order 2 is given by*

$$S_{\text{corr}}^{\text{glob}*} = \log(2^{N-1})$$

In case II, there are 2^N stereo-isomers and this entropy correction is given by:

$$S_{\text{corr}}^{\text{glob}*} = \log(2^N)$$

Proof: In the present situation, we consider the racemic mixture of the stereo-isomers. As explained in the previous Lemma, if conditions (i) or (ii) are satisfied, we have two types of stereo-isomers, the ones which have a global 3D symmetry of order $\sigma_0 = 2$ (they are denoted by type a , their set is noted A and there are $2^{N/2}$ molecules) and the ones for which the global symmetry is broken by the chiral effects, but which are generated under two isomorphic forms (they are denoted by type b , their set is noted B and there are $2^{N-1} - 2^{N/2-1}$ molecules)

We know from statistical physics ²² that the entropy correction is due to both the mixing entropy of the various stereo-isomers and the symmetric character of stereo-isomers of type a :

$$S_{\text{corr}}^{\text{glob}*} = \sum_{k \in A} X_k \log\left(\frac{1}{\sigma_0 X_k}\right) + \sum_{k \in B} X_k \log\left(\frac{1}{X_k}\right)$$

Where X_k denotes the molar fraction of the k^{th} stereo-isomer. However, $X_k = 1/2^N$, for $k \in A$ and $X_k = 1/2^{N-1}$, for $k \in B$. Thus

$$S_{corr}^{glob*} = \frac{2^{N/2}}{2^N} \log(2^{N-1}) + \frac{2^N - 2^{N/2}}{2^N} \log(2^{N-1}) = \log(2^{N-1})$$

We then recover the formula we were looking for in Case I. In Case II, following the same sort of arguments, we obtain:

$$S_{corr}^{glob*} = \frac{2^{N/2}}{2^N} \log(2^N) + \frac{2^N - 2^{N/2}}{2^N} \log(2^N) = \log(2^N)$$

Consequently, we have provided a general justification of Benson's formula in the situation when chirality and global symmetry interfere for paraffins UFC in case I and identified a situation where it fails since one should find $\log(2^{N-1})$ too in case II. One has also to be careful not to extend Benson's formula by taking $S_{corr}^{glob**} = \log\left(\frac{\eta}{\sigma}\right)$ where η denotes the number of stereo-isomers as calculated in the Lemma.

Let us insist on the fact that the chirality only interferes with global or "external" symmetries is purely due to the constraint associated with the paraffins UFC. In the following we provide a complementary example presented in Figure 14 which does not fulfill Froment's condition; it admits an internal symmetry of order three which interferes with the chirality present in the R

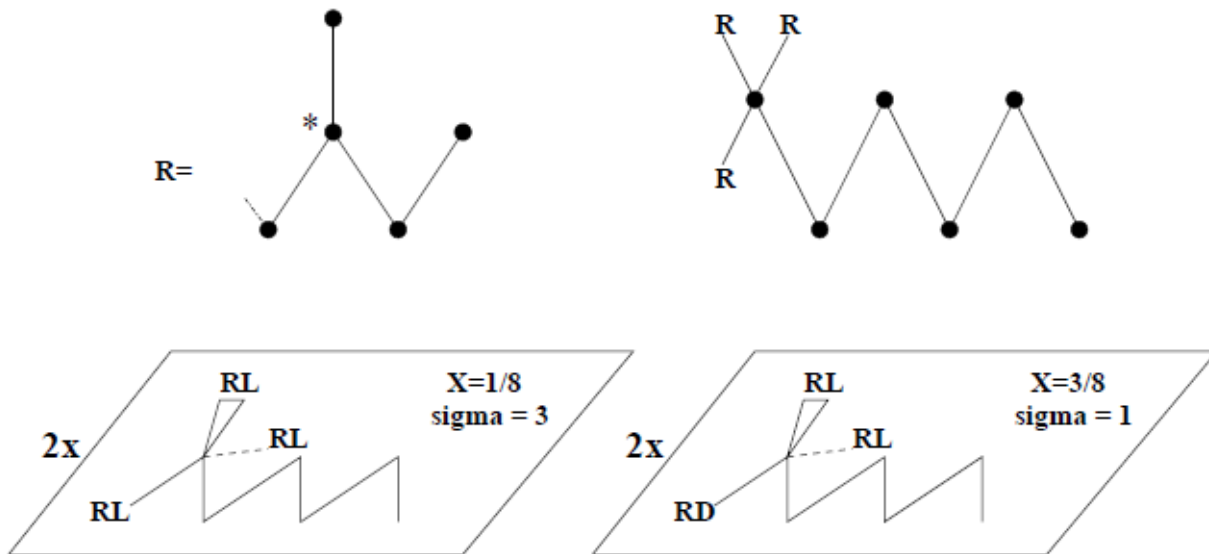


FIG 14 – A configuration for which chirality interferes with internal symmetries (RD denotes the radical R in its dextrogyre configuration and RL, the radical R in its levogyre configuration).

radical. The purpose of this example is two-fold: first it proves that chirality can also interfere with internal symmetries and second it provides a case where Benson's formula still holds.

The entropy correction then reads:

$$S_{corr}^{glob*} = 2 \frac{3}{8} \log\left(\frac{8}{3}\right) + 2 \frac{1}{8} \log\left(\frac{8}{3}\right)$$

and is coherent with Benson's formula since the structure of the internal symmetry does not fulfill Froment's condition. It is represented in Figures 15 and 16.

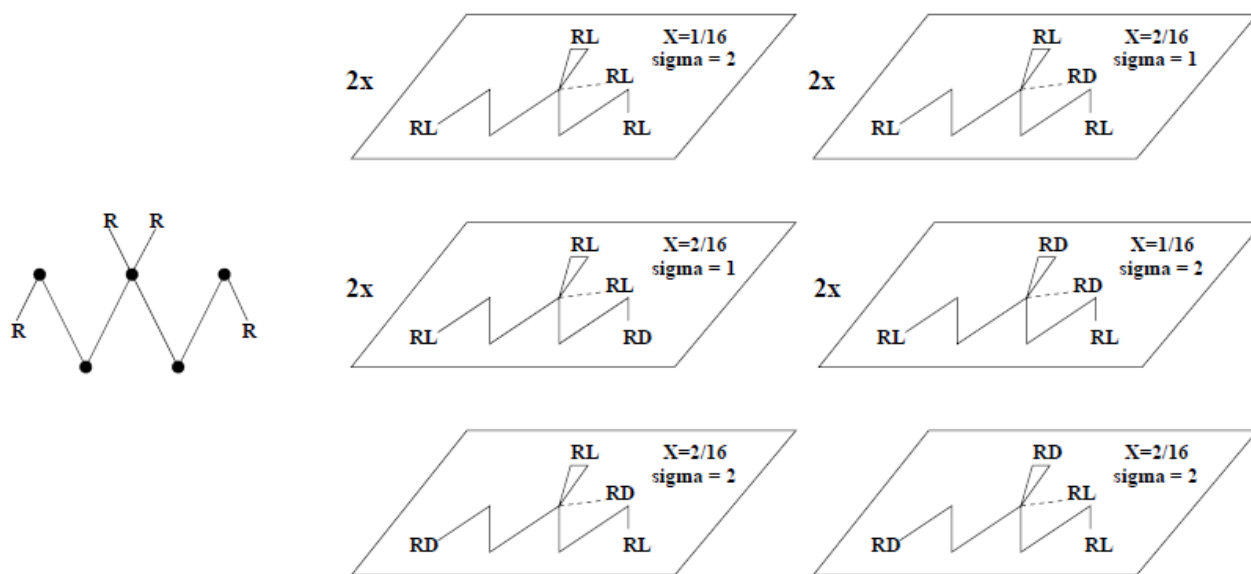


FIG 15 – A configuration for which chirality interferes with symmetries and creates a possible new chiral carbon atom (the 2x suggests the presence of two enantiomers).

The entropy correction for this first example reads:

$$S_{corr}^{glob*} = 2 \frac{1}{16} \log\left(\frac{16}{1*2}\right) + 2 \frac{2}{16} \log\left(\frac{16}{2}\right) + 2 \frac{2}{16} \log\left(\frac{16}{2}\right) + 2 \frac{1}{16} \log\left(\frac{16}{1*2}\right) + 2 \frac{2}{16} \log\left(\frac{16}{2*2}\right)$$

$$S_{corr}^{glob*} = \frac{3}{4} \log(8) + \frac{1}{2} \log(2) = 2 \log(2) + \frac{3}{4} \log(2)$$

whereas Benson's formula predicts $3 \log(2)$. It is interesting to note that in Figure 15, in the last row, the central carbon atom does not get any chiral character (the "2x" sign mentioning the presence of two enantiomers is not to be found in such a case). If the planar symmetry which results in $\sigma_{glob} = 2$ were to be eliminated, we would recover Benson's formula in the case where the symmetry number is exactly the one associated to the order of the automorphism group of the corresponding structural isomer.

In the theorem of section 4.2, Froment's condition imposed a single possibility for Benson's formula not to be valid and this scenario was the appearance of a new chiral carbon atom due to the presence of other optically active centers. This is a more peculiar case as compared to the previous one since the central carbon atom can be chiral or not depending on the preserved

symmetry (see Figure 15). This situation can be relatively rich as observed on the next example of Figure 16.

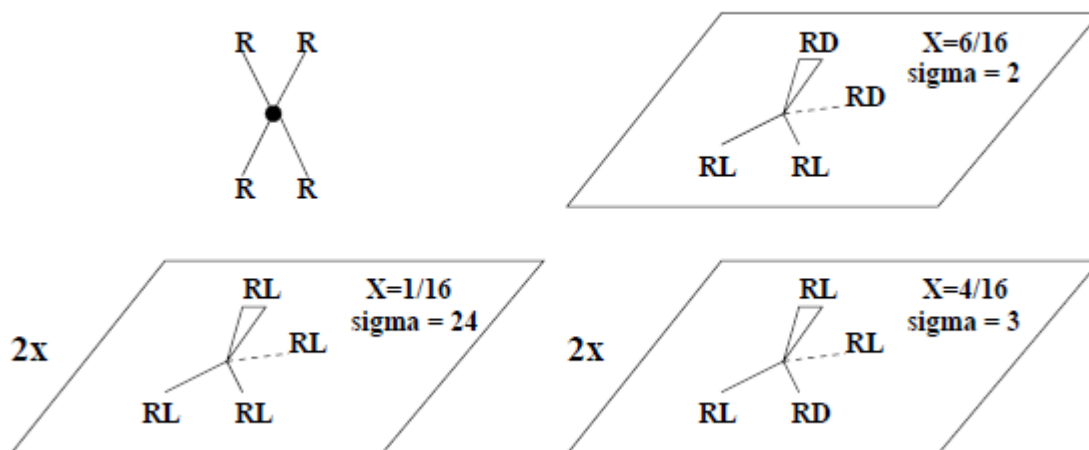


FIG. 16 – A configuration for which chirality interferes with symmetries and creates a possible new chiral carbon atom.

The entropy correction for this first example reads:

$$S_{corr}^{glob*} = 2 \frac{1}{16} \log\left(\frac{16}{12}\right) + 2 \frac{4}{16} \log\left(\frac{16}{4*3}\right) + \frac{6}{16} \log\left(\frac{16}{6*4}\right)$$

$$S_{corr}^{glob*} = \frac{1}{2} \left(\log\left(\frac{2}{3}\right) + \log\left(\frac{4}{3}\right) \right) = \log\left(\frac{2}{3}\right) + \frac{1}{2} \log(2)$$

whereas Benson's formula predicts $\log\left(\frac{2}{3}\right)$.

These two examples illustrate in a very convincing fashion the possible interferences between chirality and symmetries and suggest a possible general formula in order to correct Benson's formula.

It is also then interesting to measure the impact of the exhaustive generation conducted in this paper and to realize that the knowledge of the various configurations of the stereo-isomers with their associated symmetries yields a better understanding of the interaction of global symmetry and chirality.

5 Conclusion

Our effective approach of the problem of the isomorphic enumeration of stereo-isomers is basically different from Pòlya's method and Read's method because it is based on an exhaustive generation of the geometric structure as well as optical activity and degree of symmetry of the alkanes and not simply the global number of stereo-isomers.

In our approach, the geometrical structure, which can be interpreted for an alkane molecule as an embedding into a centered network of non-regular octahedral paving the 3D space, does not only result in the definition of formal series and finally in a way of calculating the global number of stereo-isomers as in Read, but all the non-isomorphic stereo-isomers are generated with their degree of symmetry.

Besides, this exhaustive approach can be adapted to specific graph subfamilies of alkane molecules which are mostly encountered in practical applications. It is worth noting that the enumeration of stereo-isomers of such subfamilies with the approach of generating functions provides a certificate.

Moreover, this new approach provides an important new feature: it becomes possible to use a combinatorial approach to obtain physical or chemical properties of mixtures of stereo-isomers as the emergence of geometrical properties of the various configurations.

As a perspective of the present work, it is worth making the link with the paper by Muller et al.⁹, where classes of much more complex molecules are considered as compared with the class of alkanes UFC. However, chirality is not taken into account in details so that the evaluation of the number of stereo-isomers, the associated degree of symmetry of the molecules and the corresponding entropy correction can only be done approximately, even if it is done with a good approximation. It is clear that the approach proposed in this paper can be generalized to such classes of molecules following the techniques developed by Muller et al.⁹ and relating it to the tools we have developed. In the present work, the determination of the focus and of the resulting decomposition⁹ is very simple because of the structure of the alkane molecules UFC and it is equivalent to what is done in section 3 during the tests for isomorphy in order to determinate the automorphism group of the structural isomers.

The mathematical frame presented here enlightens the results of the recurrence formulae that were established for the modeling of hydrocracking of long paraffins by the single event kinetic methodology^{20,21}.

Thus, this approach opens new perspectives in the field of organic and statistical chemistry since it is intrinsically adequate for problems in which the spatial configurations of the various molecules have a key role in the global properties of the mixture of stereo-isomers.

6 Acknowledgments

We wish to thank Michel Forestier and Isabelle Pitault, LGPC – UMR 2214, Laboratoire de Génie des Procédés Catalytiques, École Supérieure de Chimie Physique Électronique de Lyon, and Jan Verstraete from IFP Énergies nouvelles for some insightful comments and fruitful discussions.

Références

[1] M. A. Baltanas and G. F. Froment. Computer generation of reaction networks and calculation of product distributions in the hydroisomerisation and hydrocracking of paraffins on pt-containing bi-functional catalysis. *Comp. Chem. Eng.*, **1985**, 9(1): 71-81.

- [2] E. Vynckier and G. F. Froment. Modeling of the kinetics of complex processes based upon elementary steps. In G. Astarita and S. I. Sandler, editors, *Kinetic and thermodynamic lumping of multicomponent mixtures*, pages 131-161. Elsevier, 1991.
- [3] E. Valéry. *Applications de la théorie des événements constitutifs à l'hydrocraquage de paraffines lourdes*. PhD thesis, Université Claude Bernard, Lyon 1, Mai 2002.
- [4] D. M. Bishop and K. J. Laidler, Statistical Factors for Chemical Reactions, *Trans. Faraday Soc.*, **1969**, 66,1685-1687.
- [5] D. M. Bishop and K. J. Laidler, Symmetry Numbers and Statistical Factors in Rate Theory, *J. Chem. Phys.*, **1965**, 42(5), 1688-1691.
- [6] N. L. Biggs, E. K. Lloyd, and R. J. Wilson. *Graph theory. 1736-1936*. The Clarendon Press Oxford University Press, New York, second edition, 1986.
- [7] N. Lygeros, P.-V. Marchand, and M. Massot. Enumeration and 3d-representation of the stereoisomers of alkane molecules. Technical Report, Prepublication of the Laboratory MAPLY - UMR 5585, 2003.
- [8] S. W. Benson. *Thermochemical kinetics*. John Wiley & Sons, New York, 1976.
- [9] C. Muller, G. Scacchi, and G. M. Côme. A topological method for determining the external symmetry number of molecules. *Comp. Chem.*, **1991**, 15(1): 17-27.
- [10] M. Massot, M. Kumar, A. Gomez, and M. D. Smooke. Counterflow spray diffusion flames of heptanes: Computations and experiments. In *27th Symposium (International) on Combustion*, pages 1975-1983. The Combustion Institute, 1998.
- [11] F. Laurent and M. Massot. Multi-fluid modeling of laminar poly-dispersed spray flames: Origin, assumptions and comparison of sectional and sampling methods. *Combust. Theor. Model.*, **2001**, 5(4) :537-572.
- [12] G. Pölya and R. C. Read. *Combinatorial enumeration of groups, graphs, and chemical compounds*. Springer-Verlag, New-York, 1987. Pölya's contribution translated from German by Dorothee Aeppli.
- [13] R. C. Read. Some recent results in chemical enumeration. In *Graph theory and applications (Proc. Conf., Western Michigan Univ., Kalamazoo, Mich. 1972; dedicated to the memory of J. W. T. Youngs)*, pages 243-259. Lecture Notes in Math., Vol. 303. Springer, Berlin, 1972.
- [14] R. C. Read. *The enumeration of Acyclic Chemical Compounds*, chapter 4, pages 25-62. Academic Press [Harcourt Brace Jovanovich Publishers], London, 1976. A. T. Balaban Ed.
- [15] G Pölya. Kombinatorische anzahlbestimmung für gruppen, graphen und chemische verbindungen. *Acta. Math.*, **1937**, 68: 145-254.
- [16] A. T. Balaban, editor. *Chemical applications of graph theory*. Academic Press [Harcourt Brace Jovanovich Publishers], London, 1976.

[17] C. Chaunier and N. Lygeros. The number of orders with thirteen elements. *Order*, **1992**, 9(3) :203-204.

[18] C. Chaunier and N. Lygeros. Le nombre de posets à isomorphie près ayant 12 éléments. *Theoret. Comput. Sci.*, **123 (1)** :89-94, 199. Number theory, combinatorics and applications to computer science (Marseille, 1991).

[19] N. Lygeros and M. Mizony. Construction de posets dont le groupe d'automorphismes est isomorphe à un groupe donné. *C. R. Acad. Sci. Paris Sér. I Math.*, **322(3)** :203-206, 1996.

[20] E. Valéry, Kinetic modeling of acid catalyzed hydrocracking of heavy molecules: Application to squalane, *Ind. Eng. Chem. Res.*, **46(14)**, 4755-4763, 2007.

[21] D. Guillaume, E. Valéry, J. Verstraete, K. Surla, P. Galtier, D. Schweich, Single Event Kinetic Modelling without Explicit Generation of Large Networks: Application to Hydrocracking of Long Paraffins, *Oil Gas Sci. Technol.*, **66(3)**, 399-422, 2011.

[22] R. Bamian. *Physique Statistique*, volume II. Cours de l'École Polytechnique, 1986.

7 Appendix: Details of the proof for generating functions

Let us recall the details of the Theorem:

Theorem. – The number of structural isomers of paraffin molecules with n carbon atoms under Froment's condition is given by the coefficient of x^n in

$$\frac{S_i^{neq} b(x)}{1-b(x)} + \frac{S_i^{eq}(x)}{2} \left[\frac{1}{1-b(x)} + \frac{1+b(x)}{1-b(x^2)} \right] + P_i(x);$$

$$b(x) = x + x^2 + 2x^3 + x^4 + x^5$$

$$S_i^{neq} = x^5(1+x+x^2)$$

$$S_i^{eq} = x^4(1+x+x^4)$$

And
$$P_i(x) = x + x^2 + x^3 + x^4 + 2x^5 + x^6 + x^7$$

- The number of stereo-isomers of paraffin molecules with n carbon atoms under Froment's condition is given by the coefficient of x^n in

$$\frac{S_s^{neq}(x)}{1-xc^2(x)} + \frac{S_s^{eq}(x)}{2} \left[\frac{1}{1-xc^2(x)} + \frac{1}{1-xc(x^2)} \right] + P_s(x);$$

$$c(x) = 1 + x + x^2$$

$$S_s^{neq} = x^7(3 + 8x + 21x^2 + 34x^3 + 50x^4 + 52x^5 + 49x^6 + 32x^7 + 18x^8 + 7x^9 + 2x^{10})$$

$$S_s^{eq} = x^6(1 + 3x^2 + 5x^4 + 6x^6 + 6x^8 + 2x^{10} + x^{12})$$

And
$$P_s(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 4x^6 + 7x^7 + 8x^8 + 9x^9 + 7x^{10} + 6x^{11}$$

Proof:

We start by considering the problem of counting structural isomers. In the following, the *skeleton* of a paraffin UFC will denote two radicals X and Y joined by a chain of carbon atoms. Thus, these paraffin molecules will be composed by a skeleton on which radicals are connected. For the purpose of counting structural isomers, X and Y will take their value in the set F , whereas they will take their value in an other set, denoted F_s , for the purpose of counting stereo-isomers in the second part of the proof. In Figure 4-left, X, Y are in F, Q_i and R_i radicals such that:

$$F = \{-C_2H_5, -CH(-CH_3)_2, -C(-CH_3)_3\}, \quad Q_i, R_i \in \{-H, -CH_3, -C_2H_5\}$$

This extension of the concept of main chain will then be coherent with Read's method and allows us to control isomorphism phenomena at extremities. Furthermore, the similarity between the structure of paraffins UFC and the structure of di-substituted hydrocarbons ($C_nH_{2n}XY$), enumerated by Read, is now obvious.

Let us first suppose that X and Y are different. Consider any pair of 'boxes' Q_i and R_i . Since we are counting structural isomers as presented in Figure 4-(a), we can permute these radicals independently from the others, and it follows that the group of permutation is $S_2 \times S_2 \times \dots \times S_2$ with p factors. Thus, the cycle index is the polynomial in the variables s_1, s_2 given by the following expression

$$\{Z(S_2)\}^p = \frac{1}{2^p} (s_1^2 + s_2)^p$$

Let $c(x)$ be the figure counting series; we have $c(x) = 1 + x + x^2$ where the 1 indicates that the position (Q_i or R_i) is taken by hydrogen, while x by a methyl radical, and x^2 by an ethyl radical.

Thus, with Pólya's Theorem the counting series is $\frac{1}{2^p} (c^2(x) + c(x^2))^p$. Taking into account the

extra carbon atoms in the skeleton, we obtain the 'shift' polynomial $x^p S_i^{neq} = x^5(1 + x + x^2)x^p$, and we see that the number of ways of allocating alkyl radicals to $2p$ boxes is therefore given by the coefficient of x^n in

$$S_i^{neq} \frac{x^p [c^2(x) + c(x^2)]^p}{2^p} = S_i^{neq} b^p(x),$$

$$b(x) = \frac{x}{2} [c^2(x) + c(x^2)] = x + x^2 + 2x^3 + x^4 + x^5$$

To find the total number of compounds when X and Y are different we must sum $p \in \{1, 2, \dots\}$ and we obtain

$$S_i^{neq} \sum_{p=1}^{\infty} b^p(x) = \frac{S_i^{neq} b(x)}{1-b(x)}$$

When X and Y are the same, the skeleton admits a supplementary global symmetry. If $p=2k$ is even then, the group of permutations is now $S_2[S_2 \times \dots \times S_2]$ (k factors), whose cycle index is

$$\frac{1}{2} \left\{ \left[\frac{1}{2} (s_1^2 + s_2) \right]^{2k} + \left[\frac{1}{2} (s_2^2 + s_4) \right]^k \right\}$$

We obtain the series

$$\frac{x^4(1+x^2+x^4)}{2} \{b^{2k}(x) + b^k(x^2)\} = \frac{S_i^{eq}}{2} \{b^{2k}(x) + b^k(x^2)\};$$

where $b(x)$ is given by (8) and $S_i^{eq} = x^4(1+x^2+x^4)$.

If $p = 2k + 1$ is odd then, the group of permutations is $S_2 \times S_2 [S_2 \times \dots \times S_2]$ (k factors) which eventually gives us the counting series

$$\frac{S_i^{eq}}{2} b(x) \{b^{2k}(x) + b^k(x^2)\}$$

We now sum (10) and (11) over all values of k :

$$\frac{S_i^{eq}}{2} \left\{ \sum_{k=0}^{\infty} [b^{2k}(x) + b^k(x^2)] + \sum_{k=0}^{\infty} b(x) [b^{2k}(x) + b^k(x^2)] \right\} = \frac{S_i^{eq}}{2} \left\{ \frac{1}{1-b(x)} + \frac{1+b(x)}{1-b(x^2)} \right\}$$

The last step is to explicitly express the counting polynomial $P_i(x)$ which enumerates paraffins for small values on n which can not be obtained by the previous method because of the structure of F and consequently of the skeleton. For n up to 7, we obtain

$$P_i(x) = x + x^2 + x^3 + x^4 + 2x^5 + x^6 + x^7$$

Thus, this counting series of structural isomers of Paraffins UFC is obtained by summing (9), (12) and (13).

To enumerate the paraffins, having regard to stereo-isomers, we use the same method as before. Taking into account the chirality, we have to consider a 3D representation as illustrated in Figure 4-(b). However a new difficulty arises. Actually, the set of extremities can not be taken to the \mathcal{F} any

more because of the possible chirality of the third carbon starting from one end of the main chain, which will depend on the branches R_1 and Q_1 . In order to avoid these interferences, the length of the main chain involved in the extremities has to be 3 carbon atoms. Here is, for example, the possibilities when X or Y have 6 carbon atoms: The figure counting series is the same as in the first case and is still given by $c(x) = 1 + x + x^2$. If X and Y are different, the group is E_{2p} (identity group) and the associated cycle index is z_p . Therefore, the counting series given by Pòlya's theorem is $c^{2p}(x)$. Now we have to take into account the extra carbon atoms in X and Y . The possible combinations are given in Table I.

Number of carbon atoms	7	8	9	10	11	12	13	14	15	16	17
Number of configurations	3	8	21	34	51	55	54	38	24	9	3

Table V. Number of configurations for different radicals X and Y with n carbon atoms as a whole.

These numbers allow to determinate the "shift" polynomial S_s^{neq} :

$$S_s^{neq}(x) = x^7(3 + 8x + 21x^2 + 34x^3 + 51x^4 + 55x^5 + 54x^6 + 38x^7 + 24x^8 + 9x^9 + 3x^{10})$$

Thus we obtain

$$\begin{aligned} \sum_{p=1}^{\infty} S_s^{neq}(x) x^p c^{2p}(x) &= S_s^{neq}(x) \sum_{p=0}^{\infty} x^p c^{2p} \\ &= \frac{S_s^{neq}(x)}{1 - xc^2(x)} \end{aligned}$$

If X and Y are the same, the group is $S_2[E_p]$ whose cycle index is $\frac{1}{2}(s_1^{2p} + s_2^p)$. The counting series is then

$$\frac{1}{2} [c^{2p}(x) + c^p(x^2)]$$

As in the case $X \neq Y$, we provide the table which contains the numbers of possibilities for the extremity:

Number of atoms carbon	6	8	10	12	14	16	18
Number of configurations	1	3	5	6	6	3	1

Table VI. Number of configurations for radical $X = Y$ with , carbon atoms as a whole.

We deduce the “shift” polynomial, S_s^{eq}

$$S_s^{eq}(x) = x^6(1 + 3x^2 + 5x^4 + 6x^6 + 6x^8 + 3x^{10} + x^{12})$$

We now sum for $p \geq 0$

$$\begin{aligned} & \sum_{p=0}^{\infty} \frac{x^p}{2} [c^{2p}(x) + c^p(x^2)] S_x^{eq}(x) \\ &= S_s^{eq}(x) \sum_{p=0}^{\infty} \frac{x^p}{2} [c^{2p}(x) + c^p(x^2)] \\ &= \frac{S_s^{eq}(x)}{2} \left[\frac{1}{1 - xc^2(x)} + \frac{1}{1 - xc(x^2)} \right] \end{aligned}$$

Finally, the counting polynomial $P_s(x)$, analogous of $P_i(x)$ (13) is given by

$$P_s(x) = x + x^2 + x^3 + 2x^4 + 3x^5 + 4x^6 + 7x^7 + 8x^8 + 9x^9 + 7x^{10} + 6x^{11}.$$

The configuration counting series which gives us the number of stereo-isomers of paraffins UFC is finally obtained on summing (14), (16) and (17).