ESSAY

Small Molecules as Exemplars of Emergent Properties and Diversification into the ‘Adjacent Possible’

by Bernard Testa*, Giulio Vistolib) and Alessandro Pedrettib)

*) Department of Pharmacy, Lausanne University Hospital (CHUV), Rue du Bugnon, CH-1011 Lausanne (e-mail: Bernard.Testa@chuv.ch)
b) Dipartimento di Scienze Farmaceutiche ‘Pietro Pratesi’, Facoltà di Farmacia, Università degli Studi di Milano, Via Mangiagalli, 25, I-20133 Milano

‘A property of something made of parts is emergent if it would not make sense when attributed to any of the parts.’

Lee Smolin [1]

‘Emergence means complex organizational structure growing out of simple rules […]
Emergence means unpredictability,
in the sense of small events causing great and qualitative changes in larger ones […]
Emergence is a law of nature […].’

Robert B. Laughlin [2]

Contents

1. Introduction. Objectives and Context
   1.1. Objectives
   1.2. Molecular Properties and Property Spaces
   1.3. Symmetry Breaks
2. Emerging Structural and Stereochemical Properties in Alkanes
   2.1. Introduction
   2.2. Acyclic Alkanes
   2.3. Acyclic Monochloroalkanes
   2.4. Discussion
3. Lipophilicity as a Property Space
   3.1. Introduction
   3.2. The Compared Lipophilicity Spaces of C1-to-C5 Alkanes and Alkanols
   3.3. The Compared Lipophilicity Spaces of Heptanes and Mono-Substituted Hexanes
   3.4. Discussion
4. Conclusions
   4.1. An Overview
   4.2. Property Spaces and/or the ‘Adjacent Possible’?
   4.3. The Emergence of Biodiversity

© 2014 Verlag Helvetica Chimica Acta AG, Zürich
5. Computational Methods
   5.1. Conformational Analyses
   5.2. Molecular Lipophilicity Potential (MLP)

1. Introduction: Objectives and Context. – 1.1. Objectives. This essay looks at molecular properties from an unusual angle, and it does so with two complementary objectives in mind. First, it aims to show that molecules and their representations can offer a convenient means to observe and investigate the appearance of emergent properties [3] in chemical systems (molecules) of increasing complexity. And second, this essay will offer a broader view and propose an interpretation based on Kauffman’s concept of the ‘adjacent possible’ [4][5]. Stated differently, molecules are taken as exemplars illustrating how increased molecular complexity implies properties non-existent in simpler chemical systems, a phenomenon opening the door to diversification into new territories of propensities [6][7] and especially to biodiversity [8].

Since a few decades, the phenomenon of emergence enjoys great currency among complexity scientists and other researchers [2][5][9–27]. A number of definitions have been proposed, particularly meaningful ones in our view being those given by the theoretical physicists Lee Smolin [1] and Robert B. Laughlin [2] and quoted above.

Investigating the emergence of properties necessitates typical examples (exemplars), yet their complex nature, for example in biology, renders them difficult to describe, disembroil, solve, and interpret. Here, the emergent molecular properties to be considered are structural-stereochemical and physicochemical ones, namely constitutional, configurational, and conformational isomerism, prostereoisomerism (Sect. 2), and lipophilicity spaces (Sect. 3). A number of distinguished authors have illustrated emergent properties in the chemical domain [28–32]. However, their arguments and examples bear little resemblance with the approach followed here.

1.2. Molecular Properties and Property Spaces. As part of this Introduction, we begin by reflecting on properties and property spaces. Their meaning may seem clear-cut and different, but this is not necessarily the case. First, there exist yes/no cases; for example whether a given three-dimensional (3D) molecular structure is chiral or not, in which case one speaks of a 1D property space. Furthermore, one finds a wealth of properties compiled in huge databases, e.g., melting points, boiling points, ionization constants, and partition coefficients. The values of these properties are usually reported as single numbers, but no (bio)chemist needs to be reminded that their value often depends on experimental conditions such as pressure or solvent. Given such values, it becomes possible to plot a given property \( Y \) (taken as the dependent variable) against an independent variable \( X \) such as pressure or temperature, all other influencing factors being kept constant (usually under standard conditions).

This is exemplified here with the well-known density of \( \text{H}_2\text{O} \) at 100 kPa (1 bar = standard pressure), whose variation as a function of temperature is shown in Fig. 1 [33]. The point to be emphasized is that such a plot represents a finite two-dimensional (2D) property space. Plots analogous to this one, be they 2D or 3D (one dependent and two independent variables), allow a clear visualization and better understanding of the property spaces of a given compound. This will be illustrated with molecular dynamics (MD) simulations giving access to the conformational space of a given molecule by systematically varying the torsion angles of rotatable bonds (rotors), followed by the
computation of the resulting Gibbs energies [34–36]. Another example would be the lipophilicity space of the compounds, obtained as illustrated below by using a quenched MonteCarlo procedure to perform exhaustive conformational simulations, thus generating for each molecule 1,000 conformers whose Molecular Lipophilicity Potential (MLP; also known as ‘virtual log $P$’) was then calculated [35][37–41].

The property spaces of a given compound should not be confused with what is usually called a ‘chemical universe’ or a ‘chemical space’, namely a multi-dimensional descriptor space including the set of all possible compounds belonging to a given chemical class [42–46]. The examples below are based on small sets of this nature.

1.3. Symmetry Breaks. A further graphical tool to be used here is that of symmetry breaking as investigated by a number of authors, Prigogine and co-workers in particular [22][47–51], and as exemplified in Fig. 2. At a given value of the control (i.e., independent) variable $X$, a qualitative change such as a phase transition occurs which renders unstable the unique solution of the dependent variable $Y$. At this point, a phenomenon known as bifurcation occurs with two or more new solutions emerging, each of which may be stable or not. Symmetry breaking is indeed considered by most if not all physicists and biochemists as one of the most fundamental phenomena in the complexification of matter and life.

2. Emerging Structural and Stereochemical Properties in Alkanes. – 2.1. Introduction. Each compound in this Sect. was examined for the possession, or non-possession, of the following structural and stereochemical properties:

i) Existence of a vibrational space (vibrations in bond lengths and valency angles);
ii) Existence of a conformational space (presence of rotors defined as a chain of four atoms joined by single bonds, X-C-Y-Z, in this writing with X = C or H; Y = C or O; and Z = C, Cl, O, or H) [52];

iii) Existence of a constitutional space, namely the possibility of two or more constitutional isomer(s) [53];

iv) Existence of prostereoisomerism, e.g., the presence of one or more center(s) of prochirality, namely a position featuring either two stereoheterotopic H-atoms A–CH₂–B (A ≠ B ≠ H) or two stereoheterotopic groups A–CR₂–B (A ≠ B ≠ R) [54]. When several centers carrying stereoheterotopic H-atoms of groups were found in a given molecule, a single one was arbitrarily chosen to be marked with a blue star;

v) Existence of a configurational space, namely of two or more stereoisomers which may be enantiomers and/or diastereoisomers [55][56]. In such molecules, all stereogenic centers are marked with a red star.

As this list shows, the terms 'structural and stereochemical properties' are meant to cover those molecular features connected to the molecule’s three-dimensional (spatial) geometry. The latter is defined as the position of each atom relative to that of all other atoms in the molecule, and relative to an external observer when defining chirality.

2.2. Acyclic Alkanes. Our exploration of molecular structures begins with the series of all homologs and constitutional isomers in acyclic alkanes from C₁ to C₈ [57][58]. Fig. 3 shows the 2D structure of all acyclic alkanes in the C₁ to C₈ domain. The red arrows indicate filliations (i.e., the addition of a CH₂ unit), but this does not imply that all possible paths are shown.
Fig. 3. Emergent structural properties in the C$_1$–C$_8$ series of acyclic alkanes. The uppercase letters refer to the following structural properties possessed by a one or more isomer(s) in a given series. Properties already present in the lower homolog(s) are in parentheses. The red arrows indicate ‘filiation’, but no attempt was made to show all possibilities of generating a given structure by the addition of a C$_1$ unit. A) Existence of a vibrational space; B) existence of a conformational space; C) existence of a constitutional space (constitutional isomerism); D) existence of prosteroisomerism (when several centers carrying stereoheterotropic groups exist in a given molecule, a single one was arbitrarily chosen to be marked with a blue star*); E) existence of a configurational space (stereoisomerism), all stereogenic centers being marked with a red star.*
Clearly, a vibrational space (indicated as ‘A’) emerges with CH₄, the first member of the series, unless one takes the simplest of all molecules, C₂H₂ (i.e., H₂), as the lowest member in the series CₙH₂ₙ₊₂. The possession of a conformational space, B emerges with ethane [52]. Given the existence of a single rotor (H–C–C–H) in this molecule, its conformational space is a 2D one (Fig. 4, upper part). A more complex situation is encountered with propane and its two rotors, whose full conformational profile needs a 3D potential-energy surface to be represented (Fig. 5) [59].

The possibility of constitutional isomerism, C makes its appearance with C₄, namely with the two butane isomers. The total numbers of constitutional isomers in the series, i.e., the constitutional spaces from C₁ to C₈, are 1-1-2-3-5-9-18. These numbers seem initially to resemble the Fibonacci sequence (0-1-1-2-3-5-8-13), before rapidly overtaking it. The calculation of these numbers using topological arguments and algorithms has fascinated chemists and mathematicians, beginning with the pioneering (and not completely correct) work of Cayley of 1875 [60][61], even though the recently revised approaches are anything but simple [62].

Fig. 4. Upper part: The 2D conformational space of ethane, showing how the relative Gibbs energy of the molecule varies with the torsion angle of the single rotor. The lowest-energy conformers (i.e., the staggered, synclinal (sc) ones) and the highest-energy conformations (i.e., the eclipsed, syneriplanar (sp) ones) have achiral geometries. In contrast, the transitional conformations between the energy minima and maxima have a chiral structure. The lower part exemplifies two chiral conformations among the innumerable existing ones. The (P)- and (M)- descriptors of helical chirality allow their unambiguous designation [56], with the ranges of (P)- and (M)- conformations indicated by black and red dotted lines, respectively, in the upper part of the Figure.
Another structural property emerges with butane, namely prochirality. \textbf{D}\cite{54}. Indeed, butane features two identical centers of prochirality, as defined above. But the series has to grow to C\textsubscript{7} for the occurrence of stereoisomerism, \textbf{E} to emerge \cite{55}\cite{56}, since two of the nine constitutional isomers of heptane are chiral and hence exist as two enantiomers.

When reaching the 18 constitutional isomers of octane, we see that 17 among them feature one or more centers of prochirality. Noteworthy is the fact that four constitutional isomers have a stereogenic center and thus display enantiomerism, while a fifth (3,4-dimethylhexane) has two such centers. This allows for enantiomerism plus diastereoisomerism, with the restriction in this particular case that the two asymmetric centers are constitutionally symmetrical and allow for only three rather than the four expected stereoisomers. The stereoisomers of 3,4-dimethylhexane are

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{propane_conformations.png}
\caption{The conformational space of propane, whose two dihedral angles \(\varphi_1\) and \(\varphi_2\) serve as independent variables, while the dependent variable is the relative Gibbs energy of the computed conformations. These conformations were obtained by systematically rotating the two monitored torsions by 10\degree increments, thus generating 36 \times 36 = 1296 conformational forms whose potential energy was computed at the PM7 semi-empirical level of theory \cite{59}.}
\end{figure}
thus the two enantiomers (3R,4R) and (3S,4S), both of which are diastereoisomers of the meso-(R,S)-form [55].

The stereoisomers identified among the C7 and C8 isomers are stable, meaning that their racemization is a high-energy process necessitating bond cleavage. However, and as cogently discussed by Heilbronner and Dunitz [63], the duration of an observation (i.e., the ‘exposure time’) must also be taken into account when examining an object for its symmetry properties. Specifically, flexible molecules may also display transient stereoisomeric features, but these features will be conformational rather than configurational ones [53]. The discrimination between these two types of stereoisomers is generally straightforward, conformational isomers being separated by a low transition barrier of only a few or some kJ/mol.

Returning to ethane, it can be seen that its 2D-conformational space contains hidden information worth considering. Indeed, the lower part of Fig. 4 shows two chiral and enantiomeric conformations taken as examples among innumerable pairs of this type. The (P)- and (M)-descriptors of helical chirality allow their unambiguous designation [56], with the approximate range of (P)- and (M)-conformations indicated by the black and red dotted lines, respectively, in the upper part of the Fig. 4. These ranges all correspond to transitional forms between the staggered (synclinal, sc) preferred conformations and the eclipsed (synperiplanar, sp) higher-energy conformations, the isomerization time between two staggered forms being in the order of a few picoseconds (ps) [64].

When applying the concept pioneered by Prigogine [22][50] to ethane conformations, it proves possible to represent their symmetry break (Fig. 6). Under experimental (i.e., instrumental) conditions, a detection of this symmetry break is impossible due to chiral conformations in each pair having identical probability. At the level of in silico conformational simulations of a single molecule, however, one can witness how intramolecular rotation causes ethane to occupy a succession of fleeting chiral conformations captured in a series of frames. The same of course applies to higher homologs and to innumerable flexible compounds, but this aspect will not be considered further here.

2.3. Acyclic Monochloroalkanes. The discussion now turns to a series of somewhat more complex molecules, namely acyclic monochloroalkanes. Given the number of possible isomers, the argument is restricted to the series CH3Cl to C6H13Cl.

Fig. 7 (with X = Cl) depicts the emergence of structural properties (3rd row) in monochloroalkanes (4th to 8th rows) as compared to their parent acyclic alkane (2nd row). Remarkably, prostereoisomerism, D, which appeared with butane, emerges earlier in this series, namely with monochloroethane (three heavy atoms). Constitutional isomerism, C, which, in the alkanes, appeared at C4 (two isomers), is seen to emerge with monochloropropane (also four heavy atoms and two isomers). Furthermore, the number of constitutional isomers in the CnH2n+1Cl series increases faster (1-1-2-4-8-17) compared to the CnH2n+2 series; for example, there are nine constitutional isomers for C7H16 vs. 17 for C6H13Cl (also seven heavy atoms).

As for stereoisomerism, E, it was first seen in the C10 series with 2,3-dimethylpentane and 3-methylhexane (Fig. 3), whereas, in the monochloroalkanes, it also appears earlier, namely with 2-chlorobutane (five heavy atoms; Fig. 7). This is seen even better in Fig. 8 which contrasts the many occurrences of symmetry breaks
Fig. 6. A representation of the unstable (and unobservable) symmetry breaks in ethane, with the fleeting emergence of transient chiral conformations which average out, as soon as the exposure time is longer than their evanescent detection in MD simulations.

Fig. 7. Emergent structural properties in acyclic monosubstituted alkanes. See Fig. 3 for further details.
in the complete series of the 17 monochlorinated hexanes with the five constitutional isomers of $\text{C}_6\text{H}_{14}$. Indeed, the $\text{C}_6\text{H}_{13}\text{Cl}$ series has nine of its constitutional isomers endowed with stereoisomerism, compared to zero and two in the $\text{C}_6\text{H}_{14}$ and $\text{C}_7\text{H}_{16}$ series, respectively. Just like the number of atoms in a molecule, the nature of its atoms is seen to play a role in the emergence of structural and stereochemical properties, but no trend is apparent.

2.4. Discussion. A number of structural properties have been considered in this Sect. (Table):

i) Vibrational spaces have been mentioned but were not monitored here. This is not denying the significance of vibrations in bond lengths and valency angles, but this feature emerges with the simplest of molecules ($\text{H}_2$) and not with a member of our series of alkanes.

ii) That a conformational space (i.e., conformational isomerism) emerges when adding one $\text{CH}_2$ moiety to $\text{CH}_4$ is a trivial statement. Its observation here serves to contrast the comparably simple conformational behavior of ethane (a 2D-plot) to that of propane (a 3D-plot). This focus is meant to offer a glimpse on the intractable

Fig. 8. Symmetry breaks seen when going from hexane isomers to their monochlorinated derivatives, including emerging stereoisomers

[49] in the complete series of the 17 monochlorinated hexanes with the five constitutional isomers of $\text{C}_6\text{H}_{14}$. Indeed, the $\text{C}_6\text{H}_{13}\text{Cl}$ series has nine of its constitutional isomers endowed with stereoisomerism, compared to zero and two in the $\text{C}_6\text{H}_{14}$ and $\text{C}_7\text{H}_{16}$ series, respectively. Just like the number of atoms in a molecule, the nature of its atoms is seen to play a role in the emergence of structural and stereochemical properties, but no trend is apparent.

2.4. Discussion. A number of structural properties have been considered in this Sect. (Table):

i) Vibrational spaces have been mentioned but were not monitored here. This is not denying the significance of vibrations in bond lengths and valency angles, but this feature emerges with the simplest of molecules ($\text{H}_2$) and not with a member of our series of alkanes.

ii) That a conformational space (i.e., conformational isomerism) emerges when adding one $\text{CH}_2$ moiety to $\text{CH}_4$ is a trivial statement. Its observation here serves to contrast the comparably simple conformational behavior of ethane (a 2D-plot) to that of propane (a 3D-plot). This focus is meant to offer a glimpse on the intractable
multidimensionality of the complete conformational space of medium- and large-size compounds.

**iii** In the $\text{C}_n\text{H}_{2n+2}$ series, a constitutional space (i.e., constitutional isomerism) appeared with butane and isobutane. As noted above, the number of constitutional isomers was seen to increase faster than a Fibonacci sequence in alkanes, and even faster in mono-substituted alkanes.

**iv** Prostereoisomerism is a genuine structural feature in molecules, but one whose role and significance become meaningful only if and when the compound interacts with a chiral probe, e.g., when it undergoes product enantioselective biotransformation, a concept pioneered by Prelog [54].

**v** A configurational space (i.e., stereoisomerism) exists only in molecules having one or more stereogenic elements, namely stereogenic centers, axes and/or planes, or helicity. A significant problem, here, is the configurational stability of the relevant stereogenic element(s). Our discussion of the conformational behavior of ethane, for example, has revealed the highly labile helicity of many of its conformations. But only stable stereogenic elements are relevant when searching for configurational enantiomerism or diastereoisomerism.

The observations made with acyclic and mono-substituted alkanes are compiled in the *Table*. Clearly, the structural property spaces considered up to this point fall into two categories, depending whether their representation is continuous (e.g., a multidimensional surface) or limited to a discrete series of points. But before returning to this issue, we need to move beyond the spaces of structural properties to examine physicochemical spaces, as exemplified here with lipophilicity spaces.
3. Lipophilicity as a Property Space. – 3.1. Introduction. The various properties and property spaces considered up to this point and summarized in the Table are structural and stereochemical ones whose detection can occur at the graphical and conceptual level. The present Sect. turns to physicochemical spaces whose relevance can be microscopic (in *in silico* simulations) or macroscopic, and probabilistic when physical samples are investigated *in vitro*. Thus, when an ionizable molecule is considered, it is seen to possess a discrete ionization space formed from the individual electrical forms the molecule can exist in (neutral, mono-anionic, mono-cationic, zwitterionic, etc.). A continuous, conformation-dependent space appears more appropriate in our context, hence a focus on lipophilicity in what follows.

Lipophilicity, as expressed quantitatively by the log of a given partition coefficient \( P \), is a solvent-dependent physicochemical property of major significance in processes of intermolecular recognition, be it in biochemistry, medicinal chemistry, molecular pharmacology, and the supramolecular chemistry of self-assembling systems. Two popular solvent systems used to determine \( \log P \) are octan-1-ol/H\(_2\)O and heptane/H\(_2\)O. Innumerable studies have confirmed that lipophilicity expresses the contributions of two types of recognition forces. On the one hand, it contains a cavity term quantifying the hydrophobicity of the molecule. On the other hand, it encodes polar (electrostatic) forces, *i.e.*, a capacity for ionic interactions (for ionized compounds), H-bond donation and acceptance, and the various *Van der Waals* forces (dipole–dipole, dipole–induced dipole, dipole–instantaneous dipole) [34].

The dependence of lipophilicity on conformation has been documented in kinetic NMR studies, which established that conformers differ in their octan-1-ol/H\(_2\)O partition coefficient [65–68]. Well-validated pieces of software exist to compute the conformation-dependent molecular lipophilicity potential (MLP) parameterized from octan-1-ol/H\(_2\)O values. The procedure followed here is explained under ‘Computational Methods’. It involved the use of a quenched Monte Carlo procedure to generate 1,000 3D-molecular geometries for each compound. The MLP of each 3D geometry was then calculated, yielding 1,000 \( \log P_{\text{MLP}} \) values which formed the lipophilicity space of each compound [34–36]. While non-weighted means were calculated, the outcome of interest was the range of \( \log P_{\text{MLP}} \) values spanned by each compound. Such a range indeed reflects the relative capacity of each compound to adapt partly to the physicochemical features of its molecular environment, be it a solvent, a membrane, or partner molecules.

3.2. The Compared Lipophilicity Spaces of C\(_1\)-to-C\(_5\) Alkanes and Alkanols. Beginning with the series of eight ‘low’ alkanes running from \( \text{CH}_4 \) to the pentanes (Fig. 3), their mean \( \log P \) values are given in Fig. 9, according to the number of heavy-atom rotors they possess. As expected, the mean \( \log P \) values are seen to increase as a function of molecular weight, with steric factors (*e.g.*, proximity effects) playing a modest role.

The possession of a range of lipophilicity emerges together with a conformational space, namely with ethane, the first compound in the series to possess a rotor (see Sect. 2.2). In the series of alkanes, the individual ranges (in \( \log P \) units) vary from 0.23 (isobutane) to 0.41 (pentane). In other words, the series covers a total range of 0.18 \( \log P \) units.
Fig. 9. Lipophilicity (mean log $P_{MLP}$ values) and lipophilicity range, as calculated by the Molecular Lipophilicity Potential (MLP), for all C$_2$-to-C$_5$ alkanes (Fig. 9,a; $n=8$) and all C$_2$-to-C$_5$ alkanols (Fig. 9,b; $n=16$). The compounds are classified according to the number of heavy-atom rotors (C-C-Z; Z=C or O).
In the series of the 16 ‘low’ alkanols running from MeOH to the pentanols (Fig. 7; \(X = \text{OH}\)), the \(\log P\) ranges (Fig. 9.b) again vary as a function of molecular weight and steric factors. Here, MeOH differs from CH\(_3\)Cl in that the latter does not possess a conformational space while the former does thanks to the H-C–O-H rotor. In this series, the individual ranges (in \(\log P\) units) vary from 0.11 (EtOH) to 0.29 (pentane-1-ol). This again represents a total range of 0.18 \(\log P\) units.

Inspecting Fig. 9.a and b, would suggest that the individual \(\log P\) ranges increase with the number of heavy-atom rotors (C-C–C-C or C-C–C-O). However, this trend fails to meet statistical significance probably due to the diversity of molecular weights.

3.3. **The Compared Lipophilicity Spaces of Heptanes and Mono-Substituted Hexanes.** Having looked at the lipophilicity spaces in two series of homologous and isomeric alkanes and alkanols, respectively, we now turn to three series of constitutional isomers, namely heptanes (nine isomers), hexanols (17 isomers), and mono-chlorohexanes (17 isomers). Their mean \(\log P_{\text{MLP}}\) values are shown in Fig. 10.a, the compounds being categorized according to the numbers of heavy-atom rotors they feature. Clearly, the lipophilicity spaces in each series (Fig. 10.a) vary in a modest and seemingly unexplainable way. The last column presents the means of the means, which again do not inspire specific comments.

The same does not appear to be true for the lipophilicity ranges (Fig. 10.b). First, when the two extreme values over all constitutional isomers in a series are considered, a very large range is found to be covered (see the three columns on the far right), namely (in \(\log P\) units) 1.31 for the heptanes, 1.22 for the chlorohexanes, and 0.94 for the hexanols. Furthermore, inspection of Fig. 10.b, suggests that the ranges (dependent variable) increase with increasing number of heavy-atom rotors (independent variable). This trend is verified by univariate statistical analyses of the heptanes (linear correlation \(R^2 = 0.59; n = 9\)), and the chlorohexanes and the hexanols (quadratic correlations \(R^2 = 0.79\) and 0.54, resp.; \(n = 17\)). In other words, the number of heavy-atom rotors in these series accounts for ca. 50 and 75% of the variance in lipophilicity ranges. Additional factors such as proximity effects are assumed to play a secondary role.

3.4. **Discussion.** To make sense, the above correlations between number of rotors and lipophilicity space must be based on a causal chain-linking flexibility, conformational space, and lipophilicity space. This begs the question of the description of property spaces [69]. As presented in the Table, some property spaces are clearly continuous or discrete, but many other experimental or conceptual spaces are often undefined in terms of their nature and extent. An interesting and relevant example is that of ‘morphospace’, a term coined by Stephen Jay Gould to describe the space of possible morphologies for a given biological species [48][70]. Expressing the full information content of property spaces often proves impossible, hence the use of convenient parameters implying an acceptable loss of information. This may explain why property spaces are sometimes called ‘parameter spaces’.

As far as flexibility and conformational space are concerned, four convenient and informative parameters have been established as [36][71][72]:

i) the number of rotors, a useful index of flexibility;

ii) the radius of gyration, which encodes molecular shape, and whose range offers a good estimate of molecular flexibility;
Fig. 10. Lipophilicity as calculated by the Molecular Lipophilicity Potential (MLP) for all isomers of heptane ($n = 9$), chlorohexane ($n = 17$) and hexanols ($n = 17$). The mean log $P_{\text{MLP}}$ values of the 1,000 conformers obtained by a Quenched Monte Carlo procedure. The ranges covered by the individual log $P_{\text{MLP}}$ values of each conformer. The compounds are classified according to the number of heavy-atom rotors (C=C–C=Z; Z = C, Cl, or O). The compounds are (from left to right): heptanes: 1,2,3-triMe-Bu//2,2-diMe-Pen/3,3-diMe-Pen//2,3-diMe-Pen//2,4-diMe-Pen//2-Me-Hex/3-Me-Hex/3-Et-Pen//Hep; chlorohexanes (X = Cl) and hexanols (X = OH): 2-X-2,3-diMe-Bu//2-X-3,3-diMe-Bu//1-X-2,2-diMe-Bu//1-X-3,3-diMe-Bu//1-X-2,3-diMe-Bu//2-X-2,2-diMe-Bu//2-X-3,3-diMe-Bu//1-X-2,2-diMe-Bu//1-X-3,3-diMe-Bu//2-X-2,3-diMe-Bu//2-X-4-Me-Pen//2-X-2-Me-Pen//3-X-3-Me-Pen//2-X-Hex//3-X-Hex//1-X-2-Me-Pen//1-X-3-Me-Pen//1-X-4-Me-Pen//1-X-2-Et-Bu//1-X-Hex.
iii) the root-mean-square deviation (RMSD [Å]), whose mean value is an index of molecular flexibility and a descriptor of conformational space, highly correlated with the number of rotors.

iv) A related parameter is sensitivity, namely the ratio between the range of a given physicochemical property and an index of flexibility, for example, between the range in log \( P_{MLP} \) and the number of rotors. Thus, molecules are said to be sensitive when their log \( P_{MLP} \) is markedly influenced by even small conformational changes, whereas the log \( P_{MLP} \) of insensitive molecules changes only modestly even during major geometric fluctuations. In other words, molecular sensitivity expressed how much variations in the 3D geometry of a given compound affect variations in its physicochemical property. Sensitivity is also described as the coefficient in a function relating for example conformational space and lipophilicity space [36]. In the isomeric series of heptanes, chlorohexanes and hexanols, the average sensitivities (log \( P_{MLP} \) range divided by the total number of rotors) were found to be 0.073 (± 0.016; range 0.057–0.114), 0.079 (± 0.015; range 0.060–0.121), and 0.049 (± 0.008; range 0.039–0.068), respectively.

Our analysis of the heptanes, chlorohexanes, and hexanols (Fig. 10,b) thus confirms the relation between flexibility and lipophilicity ranges, with up to twofold variations within series which are assumed to reflect compound-specific intramolecular effects. This relation in turn satisfies common sense and gives some tangibility to the concepts of conformational spaces, lipophilicity spaces, and more generally property spaces, as they emerge when chemical complexity crosses a threshold.

4. Conclusions. – 4.1. An Overview. In the above presentation, simple molecules have illustrated how emergent properties depend on the number and nature of their components (atoms as nodes), on their connectivity (nature and location of bonds), and ultimately also on the stereochemical features of the molecular system [73]. Translated into chemical language, the above molecules illustrate the contribution of composition, constitution, configuration, and conformation (the four ‘C-words’ [74]) to the emergence of new structural and physicochemical properties. In a conceptual transition from one set of constitutional isomers to the higher (homologous) set of isomers, each of these four structural features contributes to the emergence of new properties and hence to increased complexity [34].

One may wonder whether there is a way to describe and interpret the emergence of new molecular properties along, for example, growing homologous series? In particular, it seems legitimate to question the ‘predictability’ of emergent properties. To this end, it may help to discriminate between two types of algorithmic prediction, namely i) simulations, which in our context consists in the \textit{in silico} (or ‘in cerebro’) construction of all realistic constitutional isomers or stereoisomers derivable from a given molecular or structural formula, and ii) computations based on mathematical models and first principles. In neither approach, however, are the so-called emergent properties ‘predicted’ in the narrow sense of the word, but they are discovered in the computed outcomes. This is the case, for example, in the computation of the number of constitutional isomers in successive series of alkanes isomers [60][62], as alluded to in Sect. 2.2. Similarly, the well-known \textit{Corina}™ algorithm [75] is a powerful generator and displayer of 3D molecular structures, being endowed with the capacity to decode
encoded structural information and interpret the resulting 3D outcomes in terms of flexibility and stereoisomerism.

4.2. Property Spaces and/or the ‘Adjacent Possible’? One aspect of emergence to repeat here is the fact that each molecular property exemplified in this essay, structural or physicochemical, can be conceived as a property space. Some of such spaces (Table) are restricted to a single bit of information, e.g., chiral/achiral? Others are bidimensional and can be depicted as a line (continuous or discrete) or a set of points. Other still can be envisaged as discrete or continuous hypersurfaces in a multidimensional space, for example, a conformational space or a lipophilicity space. These two spaces were chosen purposefully given their one-to-one correspondence (each conformation corresponding to a given property value) and statistical relationship, as expressed by the parameter of molecular sensitivity. But does there exist in the literature an alternative conceptual approach with which the concept of property spaces could be compared or even identified?

In the context of complex systems, we consider as particularly pregnant and inspiring Stuart Kauffman’s conceptualization of the ‘adjacent possible’ [5]. As far as chemistry is concerned, Kauffman defined the adjacent possible as consisting in ‘all molecular species that do not exist in our biosphere’ but await their creation in the chemically adjacent possible by being ‘one reaction step away from the current actual’. More generally, the adjacent possible can be viewed as a molecular, morphological, organizational, and even behavioral attractor. In other words, adjacent possibles are spaces of possibilities, or better propensities to use Karl Popper’s term [6], while also setting limits to the directly accessible space.

The chemically adjacent possible as defined by Kauffman exactly reflects our reasoning in this essay. But there is more, since the concept of the ‘adjacent possible’ appears identical to the concept of emerging properties developed here. Kauffman [5] writes of the adjacent possible as an ‘abstract shape space’ and a ‘phase space’; these are terms which evoke the same ‘spaces’ described in these pages. To repeat, an emerging property is part of the package acquired when entering an adjacent possible, and as such the two can be described by 1D-, 2D-, or multidimensional property spaces. Remarkably, Kauffman went as far as formalizing the adjacent possible using the concept of a 6N-dimensional phase space (three positional variables and three velocity variables, hence six numbers per particle).

Such descriptions bring us one step closer to their information content [24][76]. Indeed, discrete series of values as well as hypersurfaces can be described by strings of bits. This is equivalent to stating that what basically emerges during the conceptual transition from the current to the adjacent possible is information – property spaces are the realization of this information, the concrete form under which we perceive and investigate it.

4.3. The Emergence of Biodiversity. Emergence is a property of complex systems such as atoms, organisms, ecosystems, or galaxies. It arises from the collective, coordinated behavior of many entities, be they identical or (more often than not) different. But for a collective behavior to arise, the agents to be integrated must first recognize each other and interact. This process is facilitated by the fact that material systems fluctuate within their property spaces in a probabilistic manner, for example, within the confines of their conformational and electronic spaces. Such a drive to
fluctuate has been called ‘dispersal’ or ‘dissipation’ or ‘expansion’ [5][24][77], and is the expression of entropy. Yet, it is a creative process in that it multiplies the number of encounters among agents, and so increases the scope and variety of their interactions. This is assumed to facilitate the aggregation or merger of agents, which thereby can move into their adjacent possible while forming a functional unit at a higher-level of complexity.

Current science is engaged at three frontiers, namely the accessible confines of smallness (the quantum world and Planck’s limits), of immensity (cosmology and the multiverse), and of complexity (life, the biosphere, and the noosphere). In this view, the adjacent possible is necessarily expandable within these confines. Once a complex system have been realized in the current adjacent possible, a new adjacent possible becomes available. So has life emerged, not as a property of matter per se, but as the outcome of the organization of that matter [27]. And as an inevitable consequence of matter’s drive to experiment with new associations and combinations [78], biosystems of ever increasing complexity have kept evolving. To quote Kauffman’s [5] insightful sentence, ‘The past four billion years has seen a persistent flow of the biosphere into its adjacent possible’.

Such are the Laws of Nature whose outcome we know, study, and admire under the name of Biodiversity [8][79].

5. Computational Methods. – 5.1. Conformational Analyses. The conformational profile of ethane was investigated by systematically rotating the single rotor by 10° steps and generating 36 rotamers. A similar approach was used for propane by systematically rotating the two monitored torsions and generating 36 rotamers for each torsion so obtaining 1296 conformations. The potential energy of these conformations was computed at the PM7 semiempirical level of theory [59].

The conformational profiles of all constitutional isomers of C1-to-C5-alkanes, heptane, monochlorohexane, and C1-to-C6-alkanols were investigated by a quenched Monte Carlo procedure which generated 1,000 conformers by randomly rotating the rotors [80]. All generated conformers are minimized to avoid high-energy conformations. Their molecular lipophilicity potentials were then calculated as explained just below.

5.2. Molecular Lipophilicity Potential (MLP). The virtual log P (i.e., the virtual lipophilicity) of each of the 1,000 conformers obtained for all constitutional isomers of C1-to-C5-alkanes, heptane, monochlorohexane, and C1-to-C6-alkanols was computed by a Molecular Lipophilicity Potential (MLP) using the VEGA suite of programs [80]. For each investigated constitutional isomer, the results reported herein are the mean and range of its computed 1,000 log PMLP values.

REFERENCES


Received Mai 8, 2014