Structure, shape, topology: entangled concepts in molecular chemistry

(Article begins on next page)

Structure, shape, topology:

entangled concepts in molecular chemistry

Elena Ghibaudi[1], Luigi Cerruti[2], Giovanni Villani[3]*

1 Department of Chemistry, University of Torino, Via Giuria 7, 10125 Torino, Italy

2 Viale delle Acacie 12, 10148 Vinovo, Italy

3 Istituto di Chimica dei Composti OrganoMetallici, ICCOM-CNR (UOS Pisa), Via G. Moruzzi 1, 56124 Pisa, Italy, giovanni.villani@pi.iccom.cnr.it

Abstract

The concepts of *molecular structure* and *molecular shape* are ubiquitous in the chemical literature, where they are often taken as synonyms, with unavoidable drawbacks in chemistry teaching. A third concept, *molecular topology,* is less frequent but it is a reference term in molecular research domains such as Quantitative Structure-Activity Relationships. The present paper proposes an epistemological analysis of these three notions, aimed at clarifying the nature of their relationship, as well as the contiguities and differences between them. At first, we discuss the various acceptations of the terms *molecular structure* and *molecular shape*. Then, we examine some crucial milestones in the history of these concepts and we analyse the relationship between structure, shape and topology from an epistemological viewpoint. We point out the distinguishing features of each concept and we show that their semantic openness, that may be fruitful in a specialized context, turns into a source of incoherence and inaccuracy in the teaching context, fostered by the misleading use of these terms made by textbooks. Eventually, we propose a criterion fit to discriminate between the conceptual domains of *molecular shape, molecular structure* and *molecular topology.*

Keywords

Molecular structure, molecular shape, molecular topology, systemic approach, chemical education.

Introduction

The concepts of *shape* and *structure* are commonly used in molecular sciences and are ubiquitous in the chemical literature, where they are often taken as synonyms. A third concept, *molecular topology,* is less frequently encountered, albeit its being a reference term in both QSAR (Quantitative Structure-Activity Relationship) and graph theory (Albert 2000; Balaban 1985; Watts 2000). Topology is defined as the mathematical discipline investigating the properties of spaces that are preserved under continuous deformations. Its application to molecular issues involves graph theory and focuses on "the way the nodes (or atoms) are connected" (Teixeira, 2013). In such perspective, bond angles and bond lengths, i.e. geometrical or stereochemical features, are neglected. Only adjacency is taken into account.

The terms *molecular structure*, *molecular shape* and *molecular topology* bear each a peculiar philosophical burden; they are also characterised by distinct epistemic procedures. In the present work, we analyse the connection between *structure* and *shape*, taken as the expression of distinct molecular properties; the concept of *topology* will serve as correlation tool between *structure* and *shape*.

Molecular structure, molecular topology and *molecular shape* refer to distinct models of representation of the atomic-molecular realm. The epistemological implications of these terms are related with the nature and use of models in experimental sciences (Zeidler, 2000, Klein 2003). However, these are also specialized expressions used by experts that - based on their own approach to the physical realm - assign them specific meanings (e.g. an NMR structurist, an expert in heterogeneous catalysis and a QSAR theoretical expert use different acceptations of these

terms). Our analysis focuses on the models which these expressions refer to, and the linguistic expressions in themselves.

First, we argue about the systemic character of the concept of molecular structure, its dynamic quality and its relationship with Quantum Mechanics (QM). Then we examine the different acceptations of the term *molecular shape*. The close relationship between *molecular structure* and *molecular shape* is deeply rooted in history. Based on the analysis of a few, significant historical milestones, we argue that the first *structural formulas* are in fact molecular topologies that were progressively enriched with further information. An epistemic analysis of the concepts of *molecular structure*, *molecular shape* and *molecular topology,* through the discussion of a conceptual scheme aimed at clarifying the nature of their relationship, is presented in the final part of this work.

The systemic concept of molecular structure

The relationship between the concept of *molecular structure* and Quantum Chemistry (QC) is widely debated in the literature, and it is part of the more general issue concerning the meaning of the notion of *structure* in Quantum Mechanics (QM) (see for example Hendry 2011 and references therein). The common idea is that the application of quantum mechanics to the motion of electrons and their interactions in atoms and molecules accounts for the explanation of bonding and structure in a molecule. Gonzáles and coauthors (Gonzáles, 2018) have recently highlighted the problems related with talking of the structure of quantum 'composite' systems, whose components cannot be split out from the system ("*after the interaction, a quantum 'composite' system is not really composite since it does not have objective components*"). This is true for molecular structure as well as for atomic structure, even in the simplest case of hydrogen. In order to overcome this problem, some authors have tried general approaches (Fortin, 2017 and references therein), but the issue remains largely unsolved. For example, Fortin and coauthors (Fortin, 2017) suggest to rely on both Bohmian mechanics and standard QM. Hendry (Hendry, 2011) argues that the ontological emergence of molecular structure, with respect to quantum-mechanical systems of nuclei and electrons interacting via Coulomb forces, is at least as much supported by the available scientific evidence as its ontological reducibility is. This is a consequence of the fact that the solutions of the exact (Coulombic) Schrödinger equations should be spherically symmetrical. But polyatomic molecules cannot be spherically symmetrical, for their lower symmetries are crucial for explaining their behaviour. Hendry argues that the symmetry properties of the molecular wavefunction are removed in the process of applying the Born-Oppenheimer approximation. Hence, exact quantum mechanics cannot recover the structure of real molecules. For example, in the case of isomers, quantum mechanics cannot distinguish between two different molecules, for it assigns the same wavefunction to two distinct molecular structures. Practically, Hendry (Hendry 2010) takes a clear systemic stance: he suggests to treat molecular structure as an emergence and to move toward a systemic approach. The systemic approach has been extensively discussed and endorsed by one of us (Villani, 2001, 2008, 2017a): in the present paper, the concept of molecular structure is analysed in such a perspective.

Molecular structure is a systemic concept in that it conveys information on the internal organization of an entity (a molecule) whose constituents (atoms) coalesce to form a system[1]. Chemists are aware that a molecule is a new, distinct entity as compared to its constituent atoms, with peculiar properties and a specific name. Further, each possible organisation of the same constituents bears a distinct name. A molecular structure provides information on the distribution of inter-atomic interactions. Here, we point out the non-equivalence of the terms *interaction* and *chemical bond*.

These are a few implications of the systemic view of molecular structure:

(1) The systemic character of the molecule entails the overall interaction between all its constituent atoms. Nevertheless, some of the atoms interact strongly, others interact so weakly that may be approximately considered as non-interacting. The so-called *structural formula* codifies these 'preferential' interactions as direct links between adjacent atoms (e.g., the structural formula of a water molecule shows the oxygen atom linked to both hydrogen atoms, whereas these latter ones are not interlinked).

---

[1]   Here, we use this term in the acceptation of Systemic Science

(2) A molecule is not merely an ordered aggregate of atoms: atoms within the molecule are different entities as compared to their initial condition. They are distinct from isolated atoms and even from atoms of the same element within other molecules. QM highlights this condition of being part of a system through the assignment of fractional charges to each atom within the molecule: this aspect will be discussed further on. Here, we just wish to point out that atoms within the molecule must be considered *in situ* (i.e. in a peculiar relational context) as regards both their intrinsic and relational properties. This means that atoms of the same element may be different, even if they are designated by the same symbol: they display an individual character, related with their location within the molecule. The molecular environment concurs in defining their chemical identity by contributing to their distinct reactivity. Let's compare a set of H-containing molecules: H_2O, CH_4, C_6H_6, C_2H_5OH. Hydrogen atoms are more acidic in water than methane: they display peculiar features, despite their belonging to the same element, because they are part of distinct molecular environments. This may be true even for atoms of the same element within the same molecule: ethanol exhibits hydrogen atoms with distinct chemical behaviour, depending on their being linked to Carbon or Oxygen. In the systemic perspective, one has to take into account both the interactions amongst the constituents of the system and the micro-emergences that modify such constituents, and makes them different from their former condition of unbound (or differently bound) atoms (Villani 2017a).

The idea that an atom is modified by its becoming part of a molecule is an achievement of modern science, far from the classical concept of atom; in fact, it is a consequence of the impressive evolution undergone by the notion of atom. The atom conceived by Greek philosophers was constitutively unchangeable: material changes were interpreted in terms of aggregation and separation of atoms, thought as rigidly invariant parts of matter. The idea of atom's immutability was first challenged by Dalton's atomic theory and changed gradually over the 19th century, as witnessed by Rocke (Rocke 1993). Notwithstanding, this misconception is still somehow conveyed by symbolic chemical formulas[2] (including usual *structural formulas*) where the *same* symbol designates atoms of the same element in *distinct* relational contexts. Likewise, in organic synthesis schemes, functional groups (e.g. a methyl) may be transferred from one molecule to another, without apparently undergoing any change. In fact, organic chemists are well aware of the influence exerted by the molecular context on functional groups: this is the meaning of partial charges (δ^+ and δ^-) found in structural formulas.

All in all, atoms, as well as functional groups, have an intrinsic *identity* that is preserved upon combination or transfer from one molecular environment to another. Conversely, the *individual character* of an atom or a functional group may change, as it depends on both intrinsic and relational features. Such character is highlighted by its reactivity.

(3) Molecular structure endures small perturbations. This resilience is crucial, as it marks a boundary between the molecule and its environment, i.e. it allows to identify chemical individuals with specific properties instead of a continuous energy landscape where changes occur continuously. Hence, any representation of molecular structure catches a still instant of its endless internal periodic vibrations. This view corresponds to the so-called *skeleton model*, that sees molecular structure as a permanent scaffold undergoing translational, rotational, and vibrational motion, like a skeleton in motion[3]. QM associates such endurance with the fact that bonded electronic states are quantized, i.e. they are separated by a finite energy gap. Structural changes occur whenever the extent of the environmental perturbation allows the system to overcome the energy gap that separates one structure from the other.

---

[2] This is especially relevant at the teaching level, as the symbolic language may induce misconceptions in pupils. For a discussion on this educational topic see (Jensen 1998).

[3] In the following paragraph, we will discuss two cases ($H_5O_2^+$ and the base pairs of DNA) where this model does not hold, as the skeleton may change or break without any considerable energy expense, giving rise to different structures. The existence of such systems marks the limits of the skeleton model (Cerro 2009).

The concept of molecular structure is based on the assumption that this molecular property persists over a finite length of time. The extent of such interval depends on: i) the molecular features that are to be investigated and/or, ii) the experimental technique chosen for the investigation. Today, advanced experimental techniques allow the detection of chemical species that were traditionally considered unstable (Chamizo 2017).

The structure of a molecule is not static: it changes with time. If this modification occurs within a defined and limited energy range, it is possible to identify the *average* structure of such molecule. This is the case of molecular vibrations, that imply periodic oscillations of atoms around a space point. Whenever the molecule is investigated through a technique whose average detection time length is longer than the half-life of oscillations, the resulting (average) structure corresponds to the static representation of the molecular structure. Should one observe the molecule within the range of vibrational time (femtoseconds) then, structural changes occurring at that scale would be highlighted.

The non-static nature of molecular structure is well illustrated by the peculiar case of molecules in pre-dissociative states, i.e. metastable molecular states evolving towards final states that are structurally distinct from the starting ones. In this case, the molecular structure is ostensibly in evolution: statements of its existence or non-existence depends on the process lifetime, the technological apparatus available for its investigation, and even the final aim of the investigator (e.g. a chemical synthesis).

A final example of the relationship between molecular structure and time is provided by water, taken as a reference system. Even though the structure of a single water molecule is well known, the problem of defining the structure of 'water' as a substance - at the molecular level - remains open. The relevance is related with the explanation of peculiar properties of the substance 'water', such as its electric conducibility or its physical state at RT. Water molecules are not isolated entities, as they interact through hydrogen bonding. It is possible to compute the average persistence of hydrogen bonding between two water molecules, as well as to count the number of water molecules bound at a single molecule at the (still) time *t*. It remains that the structure of water is intrinsically dynamic and *it exists independently from the single molecular objects whereof it is formed* (Lami 1995a).
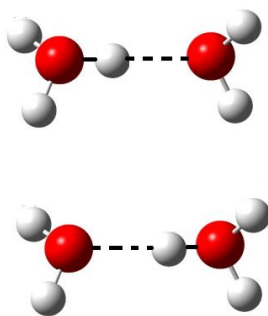

Figure 1 – Representation of two structures of system $H_5O_2^+$

Let us consider the $H_5O_2^+$ system, made of one H_3O^+ ion and one H_2O molecule bound through hydrogen bonding (Figure 1) (Lami 1995b). First of all, one could question whether the $H_5O_2^+$ system is comparable to a molecule with its own molecular structure. Again, the answer depends on the properties that are to be investigated. H-transfer from H_3O^+ to H_2O occurs in about 40 fs (Lami 1995b). Depending on the properties to be investigated, this is either an inner event of the system or an exchange process between two distinct molecular entities.

Issues related with H-bonding are not exclusive of water: they are crucial in biochemistry and molecular biology. The DNA base-pairs (Adenine-Timine and Guanine-Citosine) exists in two or three tautomeric forms (Figure 2) (Villani 2005, 2006, 2010a, 2010b). It is the job of theoretical chemists to take into account the fact that these are models of complex systems (Villani 2012, 2013a, 2013b, 2013c).[4]
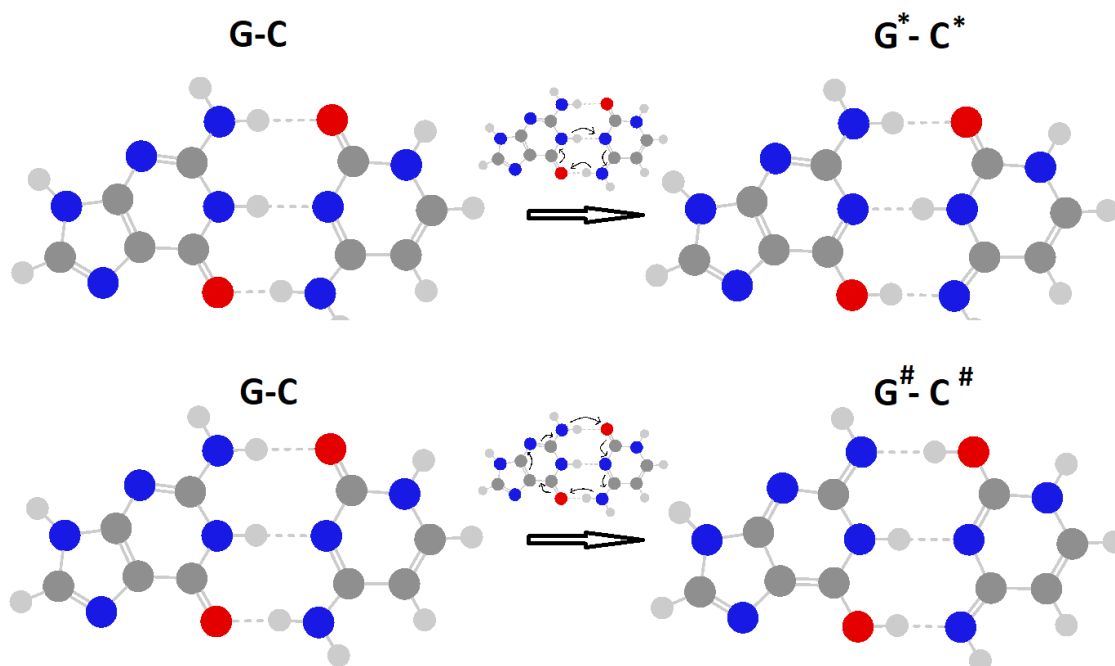

Figure 2 – Tautomers of the Guanine-Citosine base pair. Left panel: representations of the G-C (Watson-Crick pair). Right panel: G^*-C^* and $G^\#$-$C^\#$ imino-enols pairs. The mechanism of shifting from the Watson-Crick pair to the imino-enols G^*-C^* and $G^\#$-$C^\#$ pairs, respectively, is reported above the arrows.

    b.   Molecular structure and Quantum Mechanics

We have already mentioned the problem of justifying molecular structures based on QM. This is a consequence of Heisenberg's uncertainty principle as well as the breaking of the Hamiltonian spherical symmetry, in order to identify linkages between specific atoms in the molecule. It is worth pointing out that a systemic view allows such problem to be considered under a wider, and less problematic, perspective. The analysis of the issue of molecular structure in QC is often reduced to the analysis of the Born-Oppenheimer (BO) approximation. In fact, in a systemic context, the problem of considering atomic nuclei as fixed, so to focus solely on the electron motion, in a static field

---

[4]   Theoretical analysis always implies models. In molecular quantum calculations, the model focuses on the isolated molecule, with the assumption that the description of complex experimental situations may rely exclusively on the intrinsic properties of such molecule. The model may be further developed by adding other entities in interaction with the original single molecule. For example, the theoretical investigation of DNA base pairs coupling was carried out by adding solvation to the PCM (Polarizable Continuum Model) modelisation (Villani 2012, 2013a); the system was expanded to include dimers of base-pairs and the stacking interaction between them (Villani 2013b and c).

is not crucial[5,6]. The essential requirement for molecular structure to be defined in QC is the persistence of a net of interactions all over the molecule, over the time range of the experimental investigation.

Other relevant aspects concerning the notion of molecular structure in QC may be pointed out. The input ~~for~~ needed by *ab initio* quantum calculations of a molecule consists in the atomic coordinates and the specification of the kind of component atoms. There is no need for information on the presence of strong or weak interactions between atoms, because the potential energy term of the Hamiltonian takes into account all possible interactions between pairs of particles inside the system. The QM approach considers all particles within the system as interacting, albeit to different extents: there are no priviledged interactions between atoms or totally disconnected atoms. This is not devoid of consequences (Earley 2012). Let's examine a system made by 4 hydrogen atoms and 2 oxygen atoms. Quantum calculations predict two probable configurations, that are both stable, albeit quite different from the chemical viewpoint: i) the presence of 2 water molecules (2 H_2O) or ii) one hydrogen peroxide and one dihydrogen molecule (H_2O_2 e H_2). The quantum formalism needs an external input (i.e. atomic coordinates) in order to discriminate between these two configurations because quantum calculations do not identify chemical bonds; hence they cannot predict whether the two oxygen atoms belong to the same molecule or not. In order to get this piece of information, one must operate on the molecular orbital delocalized over the whole molecule, and localize it between specific pairs of atoms. At the beginning of QC, the Valence Bond approach was proposed instead: this approach requires info on directly bonded atoms as part of the initial input (i.e. topological information). Today, only a few research groups still apply this approach. Conversely, the Molecular Orbital approach takes the molecule as a whole: it calculates the overall properties of the system without connecting them to the former atomic components. Molecular properties can be projected back onto each atom of the system only *a posteriori*. A typical case is the assignment of atomic charges, that requires the partition of the (previously computed) overall electron density over each atom in the molecule. This implies a critical choice as regards the meaning of 'atomic volume': in fact, molecular volume can be splitted into atomic volumes in several ways. This results in several distinct definitions of atomic charges (Mulliken atomic charge, natural charge, Bader approach, etc.). A detailed discussion of these definitions is beyond the goal of the present work. Here, we just wish to point out that the lack of univocal definitions of atomic properties in a molecule stems from the general problem of spotting out the components of the system and defining the relationship between the whole and its parts.[7]

The concept of shape at the molecular level

The concept of *molecular shape* is widely used in molecular sciences; nevertheless, providing an unambiguous definition of it is difficult. The issue is complicated by the widespread use of *molecular shape* as a synonym of *molecular structure.* The below-reported excerpt is a clear example of such confusion. In spite of their mixing up, *shape* and *structure* remain key-concepts in chemistry as they allow interpreting and predicting the behaviour of molecules in a reactive or relational context. This is why chemists are interested in clarifying structure-function relationships in molecules:

"Those sciences that are concerned with the molecular aspects of the properties of matter, principally chemistry, but also molecular physics and biochemistry, are founded on the belief that all experiments

---

[5] The main reason for adopting the BO approximation is practical: the mass disparity between nuclei and electrons allows disentangling their respective motions and simplifies calculation tremendously. The BO approximation applies to most chemically relevant cases, although the calculation carried out without approximation would end up almost to the same result. The BO approximation is not applicable whenever two or more electronic states are quasi- or completely degenerate. In this case one must rely on conical intersections. This kind of systems has been investigated either through model systems (Ferretti 1996; Lami 2004) or biologically relevant system, like rhodopsin (Andruniów 2004).

[6] Besides, linking a foundational chemical concept, such as molecular structure, to an approximation has always raised philosophical problems.

[7] The widely used Gaussian software (Gaussian 2016) encompasses three possibilities: Mulliken's charges, Hirshfeld's charges and NBO (Natural Bond Orbital)

involving molecules can be understood in terms of the relative dispositions of the constituent atoms in the molecules. This idea of **molecular structure** (or "**molecular shape**") has been fundamental to the development of our understanding of the physicochemical properties of matter, and is now so familiar and deeply ingrained in our thinking that it is usually taken for granted - *it is the central dogma of molecular science.*" (Wolley 1978) (Italics in the text; bold character added)

We will now focus on the notion of *shape* and later discuss the relation between molecular shape and structure.

The term *shape* is part of natural language. According to the Oxford English Dictionary[8], shape is:

1. The external *form*, contours, or outline of someone or something.

2. A geometric *figure* such as a square, triangle, or rectangle.

3. The correct or original *form* or contours of something (italics added).

However, a lexical definition of shape is intrinsically ambiguous because it contains equivalent terms (e.g. *form, figure*) that need to be defined, in turn. Further, this concept is language-dependent: the semantic field of the Italian *forma* is different from that of the German *Form* or the English *shape*.

All three definitions proposed by the Oxford Dictionary make explicit or implicit reference to the contour (lines, surfaces or volumes). Whenever the shape shows regularities, it acquires a *geometrical* connotation and it possibly finds a collocation amongst the usual 2D and 3D geometrical figures. The geometrical approach recognizes entities with the same shape, based on peculiar relationships (coherence, similarity, etc.) between the objects whose shape is compared. Hence, the shape of an object is the result of spatial relations between its parts: these relations are preserved as soon as the object preserves its own shape. This sounds like an *operational* definition of shape; unfortunately, it has a circular character as it relies on the same concept that is to be defined.

Identifying the contour of a macroscopic object is (almost) a trivial problem[9], as the object (e.g. a chair) occupies a finite, perceptible space. The external (continuous) surface of the object may be represented by sampling a suitable number of (discrete) points on the contour of the object and by charting their spatial coordinates. A macroscopic object is representable; its shape may be recognized through a perceptive approach and represented in the 3D space.

Triviality vanishes upon shifting to the microscopic level, namely the atomic-molecular level. In fact, what is the shape of a molecule? Is it representable? How? According to Nicholls, it is still possible to do so by relying on the relationship between volume and surface: "*So what do we really mean by shape? There is a simple, universal meaning to the concept as the coincidence of volumes [...] that can also be extended to surfaces*" (Nicholls 2010).

However, how to define the volume or surface of a molecular object? The gap between macroscopic and microscopic objects is clear-cut. First of all, it has to be established whether treating a microscopic entity as a microscopic object is legitimate or not. There is a vast debate on this issue, whose boundaries extend from pure realism (a molecule is a real object, *tout court*) to pure idealism (a molecule is a pure concept). Our position lies somewhere in the middle and stems from these two premises:

1. Molecules are real, microscopic entities displaying measurable properties (e.g. spectroscopic properties)

2. Any representation of microscopic entities requires the mediation of models.

The choice of distinct models for representing a given molecule allows highlighting distinct aspects of the molecule. Any statement concerning the molecule is, in fact, referred to a specific model of such a molecule. The concept of *molecular shape* makes no exception: it is *model-dependent* and holds within definite theoretical frames (with their boundary conditions and approximations).

---

8   URL: https://en.oxforddictionaries.com/definition/shape.
9   At least, this is true with objects whose dimensional scale is directly accessible to humans. In all other cases, the problem is complex (e.g. how a coastal profile can be defined?).

This leads the discussion towards the model field, where it is possible to seek for an operational/computational definition of *shape.*

Papers dealing with molecular shape underline that the relation between shape and surface is recalled in all those (numerical) methods, aimed at determining and representing molecular shape, that have been *"developed with the goal of describing molecular shapes accurately as surfaces with properties such as polarity"* (Nicholls 2010). The contour surface of a molecule – meant as an entity made up of interrelated numerable entities (atomic nuclei, electrons) – can be defined only with reference to a chosen model. There is not a univocal algorithm for molecular shape: shape is always the outcome of conventional choices. One can say that *shape* results from the iconographic representation of sets of numbers related with the trend of one (or more) molecular properties. The choice of the model depends on the research goals of the operator and cannot be defined *a priori:* this highlights the heuristic value of shape. In line with this view, Ramsey states that *"Like all concepts, shape is an interpretive, approximate concept"* and, opposite to molecular structure, *"Shape is a response property rather than an intrinsic property"* (Ramsey 1997). By the way, IUPAC seems to take a different stance as it defines shape as a molecular attribute, i.e. an intrinsic property of molecules: "*The molecular shape is an attribute of a molecule dealing with spatial extension, form, framework, or geometry. It is often described by, e.g., principal axes, ovality, or connectivity indices.*" (IUPAC 1997)

The conjectural and interpretive aspects of shape have their logical-mathematical roots in the concept of *parametric shape.*[10] A rigorous definition of this latter notion may be found in the field of computer graphic (Schulz 2017), but here we just mention Terry Knight: "A parametric shape is an indefinite shape. It is a shape with some details or characteristics that are fixed and others that vary" (Knight 2003). The 'characteristics that are fixed' are constraints to be observed in the specific context of use, e.g. a definite area to be circumscribed by a polygonal line or a definite volume to be defined by plane surfaces.

Examples of the operational/computational definitions of *shape* are well illustrated by Figure 3, which is taken from a computer modelling work applied to pharmacological research. Figure 3 shows several *in silico* representations of a molecule's shape. Based on the properties they refer to, they are classified as:

(1) moment-based, (2) gnomic-based, (3) volume-based, and (4) surface-based (Putta 2007).

Each representation is based on different algorithms and choices. For example, Putta and Beroza explain that "*the moment-based methods represent a shape as a set of multipole moments of inertia [...] This method reduces a shape to a set of ordered numbers, which replaces any three-dimensional representation*" (Putta 2007).

Differently from the previous case, in gnomic-based representations "*A molecule's shape is mapped to a simple shape (e.g., a polyhedron) that encodes the complexities of the original shape. Points on the simple shape (e.g., the polyhedron's vertices) encode an approximation of the original shape with a set of values*" (Putta 2007). In this case, shape is the result of a mapping action that – according to the authors – reduces the original complexity and promotes the recognition of similarities between distinct molecules. The linguistic circularity of this definition is striking: the technical definition of shape cannot help but rely on a spontaneous, intuitive idea of shape. The outcome of the mapping action is abstract and decidedly not intuitive, as the operation consists in correlating two sets of data through a morphism.

The conventional nature of choices whereon the representation of molecular shape is based is especially evident in the definition of volume-based representations of molecular shape: *"each atom is treated as a hard sphere with a radius determined by its van der Waals radius".* (Putta 2007). Besides *"Shape volumes can also be represented by steric grids, in which each grid point has a value associated with its relationship to the three-dimensional shape"* (Putta 2007), according to a conceptual procedure similar to the previous case.

---

[10] For a formal use of the notion of parametric shape it is possible to refer to the literature of research fields related with space, as architecture and territorial planning, or to research field related with the graphical representation of shapes (Grasl 2013)

Finally, surface-based methods are shown. These are based on the relation between surface and shape: "*focus on the interface between the volume interior and exterior. In one method, this surface is represented by a shell of finite width [..]. The other two approaches reported in the literature treat a molecular shape as a set of patches on the surface of the shape [..]. The three-dimensional shape is replaced with a distribution of surface patches in space.*" (Putta 2007)

All in all, these examples show that a single molecule may exhibit distinct shapes: it depends on the model chosen by the investigator.

We have already stressed the heuristic character of molecular shape. Let's discuss two pertinent examples: a) the identification of similar molecules in the frame of QSAR and the search for pharmacologically active molecules; b) the iconographic representation of the shape of biological macromolecules.

    a.   <u>The quest for similarity</u>

Molecular similarity is a very active research line in the molecular computation field, with applicative purposes in the medical and pharmacological domain. Molecular similarity is strictly related with molecular shape. Similarity is, by definition, a binary relationship between two (real or abstract) objects. The terms of such relation are specified through 'similarity indexes' expressed by algorithms; the closer are the values of such indexes, the higher the similarity. The variety of available algorithms shares three main features (Todeschini 2009, pp. 693-694):

(a) the use of molecular descriptors, that may obey to different criteria, as witnessed by Figure 3;

(b) the weight assigned to each element in the description;

(c) similarity indexes

The final outcome of the whole search for similarity depends critically on the evaluation of the distance (d_{st}) between the target structure t and a reference structure s registered in a data bank. In 2012, Todeschini and coauthors reports the existence of 51 distinct similarity coefficients*:*

> "This paper reports an analysis and comparison of the use of 51 different similarity coefficients for computing the similarities between binary fingerprints for both simulated and real chemical data sets. Five pairs and a triplet of coefficients were found to yield identical similarity values, leading to the elimination of seven of the coefficients." (Todeschini 2012)

The above-mentioned reduction of their number stems from the statistical character of component (c), that Maggiora and coauthors define as:

> "a similarity function (also called a similarity coefficient) that combines the information contained in the representations to yield an appropriate similarity. This value usually lies between '0' and '1', where '1' results from the complete identity of the molecular representations (but not necessarily the compounds)." (Maggiora 2014)

Maggiora's remark clarifies why molecular similarity is tentative: indexing concerns the information found in molecular representations and not in molecular objects. Hence, the sameness of representations does not necessarily imply that the corresponding molecular objects are identical.

The abstract and conventional character of shape is evident in that the assessment of molecular similarity, aimed at predicting the quality of interactions between a molecule and a target, is based, in turn, on conventional descriptions of the investigated systems*: it operates on models, not on real objects*. The relevance of these procedures stems from the effectiveness of their predictions, that must always find validation in the real system (i.e. the material substance) and cannot be a mere result of calculations. In fact, according to Paul Mezey:

> "The main difficulty lies in the complexity of molecules and their behavior: similarity may refer to one or another particular type of molecular property or process. Whether two molecules are judged similar or dissimilar is dependent on the context: the molecules of water and methane are similar in size, yet their chemical properties are very different. Even if one is concerned with a limited aspect of similarity such as molecular shape similarity, still some ambiguities prevail, since, depending on the

context, different aspects of shape may be important. An assessment of similarity depends on the relative importance of these shape features." (Mezey 1994, p. 138)

Fig. (2). Examples of the 4 categories of *in silico* representation of molecular shape. (A) Moment-based: The positions of the atoms in the molecule's conformer are used to define principal axes that represent the molecule. (B) Gnomic: The shape is mapped onto a fixed set of points on a simpler shape, in this case a sphere. (C1) Volume-based (grid): A grid is superimposed on the molecule and grid points are assigned values based on their geometric relationship to the molecule's shape. (C2) Volume-based (Gaussian): A set of Gaussian functions, centered on the atomic nuclei, represent the molecule's shape. (D) Surface-based: Patches of the molecular surface (shown in red) are chosen to represent the molecule's shape.
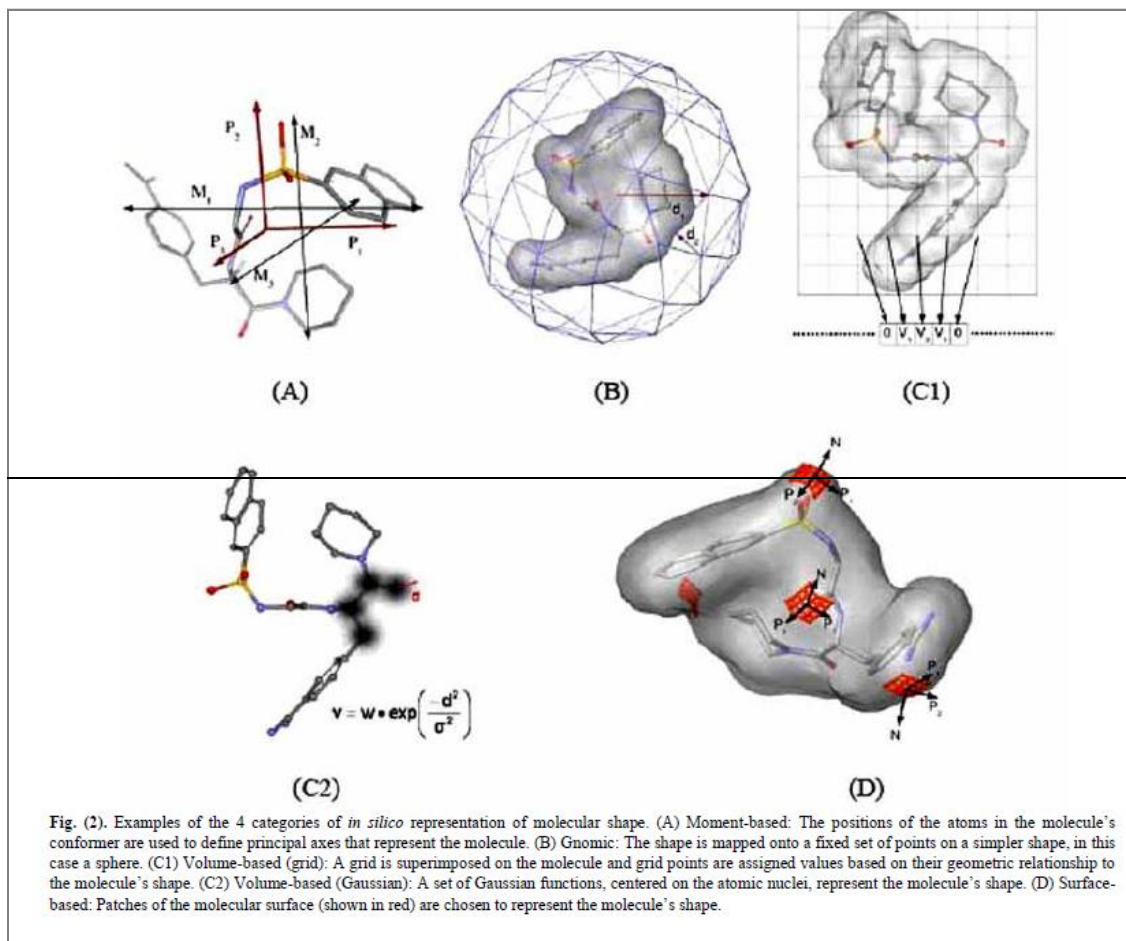
Figure 3 – *In silico* representation of the shape of a molecule (Putta 2007; Reprinted by permission from Bentham Science Publishers Ltd)

b. The iconographic representation of biological macromolecules' shape

A research domain where the concept of molecular shape plays a key role in the prediction and understanding of the system's behaviour is the domain of biological macromolecules. Figure 4 reports 4 distinct representations of the Hybrid 2 of G4 DNA found in human telomere. G4 is a Quadruplex DNA, a supramolecular architecture characterized by the coupling of 4 guanine residues (Villani 2017b). Figure 4a differs substantially from the others, in that it is just a scheme showing the G4 architecture. It is not a detailed structural representation of G4: it aims at highlighting the *topological relations* between guanine residues on different planes. The other three representations derive specifically from information available on the DNA sequence of hybrid 2 and its structure: in fact, they imply knowledge of atomic coordinates. All three figures show a molecular shape that is referred to the same molecule: shapes are different because they are generated by distinct algorithms that compute distinct parameters: Figure 4b is a stick-and-balls representation, Figure 4c displays Van der Waals radii and Figure 4d shows the solvent-accessible surface of G4. Three distinct theoretical framework, referring to distinct conventions, with different aims: the choice is up to the user and depends on the problem to be solved. The interactions found inside the G4 molecule are better investigated through the stick-and-balls representation. Conversely, Figure 4d is more helpful in the prediction of the interactions between G4 and other molecules in the cellular environment, as it shows grooves, protrusions, etc. Which picture of these best represents the molecule's shape? Undoubtedly each of them, depending on the distinct

(and equally legitimate) conventions adopted for defining the shape. Which of them is more useful? It depends on the aim of the user. Once again, their heuristic function stands out.
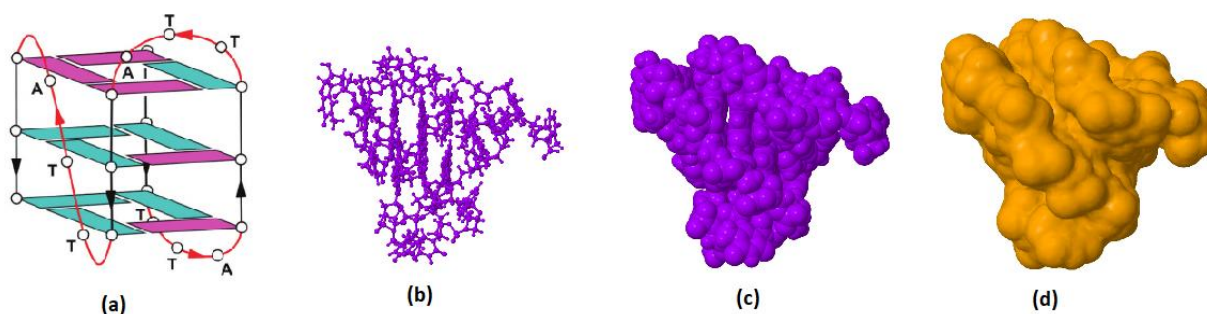

Figure 4 – Iconic representations of the shape of hybrid 2 of G4 DNA found in human telomere (PDB entry 2JPZ): (a) Schematic picture, (b) stick and ball, (c) space fill and (d) solvent accessible surface.

Ultimately, these figures are attempts to "make abstractions visible" (Cerruti forthcoming), an expression of our need for analogic reasoning that exploits the categories of the macroscopic realm and perception in order to interpret and predict the behaviour of the microscopic realm (Del Re 2000).

There is a further acceptation of *molecular shape* that is widely used in both chemistry teaching and practice: it is related with the geometrical distribution of bonds inside the molecule and deserves some discussion.

c.   A further acceptation of shape at the microscopic level

Chemistry textbooks often reports statements like "*the shape of a methane molecule is tetrahedrical*"; organic chemistry textbooks often refer to *tetrahedrical* C atoms for interpreting the properties of saturated organic compounds and to account for optical isomery. Now, a tetrahedron is a regular polyhedron and the adjective *tetrahedrical* recalls undoubtedly a *geometrical shape*. Notwithstanding, this acceptation is closer to the concept of (molecular and electronic) structures than to the concept of shape.

Speaking of which, it is worth recalling Gilbert N. Lewis and its *cubical atom*. In 1916, Lewis proposed two electronic models for atomic structure and chemical bond, that were based on the following assumptions, respectively: i) the stability of the electron octet within each atom; ii) the chemical bond as a pair of electrons shared by two atoms. Lewis built the *cubical atom* model, where each vertex is an electron. This spatial arrangement of eight electrons did not display the flexibility required by the chemical knowledge already available at that time:

> "When we consider only known chemical phenomena, and their best interpretation in terms of atomic structure, we are led to assume a somewhat different arrangement of the group of eight electrons." (Lewis 1916)
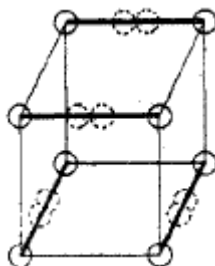
Figure 5 – Lewis' model of cubical atom (Lewis 1916)

In fact, Lewis refers to Figure 5 as follows:

> "The nature of this arrangement is shown in Fig.5. The cube representing the electron structure that we have hitherto assumed for the carbon atom is joined to four other atoms, which are not shown in the figure, but which are attached to the carbon atom each by a pair of electrons. These pairs are indicated by being joined by heavy lines. Assuming now, at least in such very small atoms as that of carbon, that each pair of electrons has a tendency to be drawn together, perhaps by magnetic force if the magneton theory is correct, or perhaps by other forces which become appreciable at small distances, to occupy positions indicated by the dotted circles, we then have a model which is admirably suited to portray all of the characteristics of the carbon atom. […]. On the other hand, the group of eight electrons in which the *pairs* are symmetrically placed about the center gives identically the model of the **tetrahedral carbon atom** which has been of such signal utility throughout the whole of organic chemistry". (Lewis 1916) (Italics in the text, bold character added)

Which relationship between the tetrahedron and molecular shape? A comparison between a water and a methane molecule may help finding an answer. No chemistry freshman would ever state that methane and water share the same structure or shape, as the atom distribution in the two molecules corresponds to distinct geometries. On the other hand, the application of the Lewis model (or, more precisely, the more recent VSEPR theory applied to molecular Lewis models) leads to the conclusion that both molecules have tetrahedral *electronic geometries*. This term indicates that bonding and non-bonding electron pairs point towards the vertexes of a tetrahedron. A deeper theoretical treatment than VSEPR theory relies on the concept of hybridation, thus providing the expression *tetrahedral carbon* with a meaning. It is clear that the focus is put on single atoms inside the molecule instead of the molecule as a whole or isolated atoms (which are not *tetrahedral*). The model description of their interaction with other atoms identifies the geometry of distribution of electron density around these atoms. Then, the 4 C-H bonding electron pairs of methane pinpoint a tetrahedron. In water, a (distorted) tetrahedron comes out from two O-H bonding and two non-bonding electron pairs. The context is relational (atoms are bound to each other), but the reference is to the internal structure of single atoms (C or O, respectively. On the other hand, the tetrahedron clearly does not represent a structure (in the sense that we discussed previously): in an analogical context, it identifies a *geometry*. Its purpose is to point out the anisotropic distribution of electron density around the bound C or O atom, in contrast to the isotropic distribution typical of an isolated atom. This *geometry* is distinct from *molecular shape:* e.g. the O atom in water is tetrahedrical, whereas the water molecule is not. Similarly, a propane molecule contains three tetrahedral C atoms: this is unrelated with the shape of the whole molecule.

Hence, whenever textbooks state that the shape of methane is tetrahedrical, we are faced to a misleading use of the term *shape*, that should be avoided in chemical education.

As a conclusion, we maintain that:

- The shape of a molecule is essentially a descriptive property. It cannot have a univocal identity because it is the outcome of *ad hoc* algorithms, that are tools at the service of the researcher's goals. Molecular shape is not an intrinsic molecular property: it has a pure conventional character.

- Such conventionality highlights the heuristic function of molecular shape. This aspect is especially evident in the use of molecular shape in biomedical and pharmacological research. Nicholls and coauthors underline that "*shape can play an essential role in guiding project goals*" and insist on the fact that "*shape as a volume or as a surface is a vibrant and useful concept when applied to drug discovery*". (Nicholls 2010) A heuristic procedure is part of a process of assignment of meaning, that is aimed at making conjectures on a specific research object. In fact, at the microscopic level, shape is nothing but a metaphor that allows assigning a meaning to microscopic objects, in the frame of a relational picture.

Topology to Structure: a Return Ticket

Molecular structure and *shape* seems conceptually far, albeit related with each other. A third concept, *molecular topology*, may act as a bridge between them as it is related to both structure and shape.

In order to analyse their relationship, it is useful to recall some crucial steps in the history of these three concepts. History shows that molecular topology, structure and shape came out gradually as a theoretical response to new interpretive needs sprang from experimental practices.

The first step of our historical analysis is the proposition of structural diagrams by the Scottish chemist Couper in 1858 (Couper 1858). Then, we discuss the pivotal contribution by van'Hoff (van'Hoff 1874a, 1874b, 1875, 1877). The final step is the interpretation of enzyme activity through the renown key/lock metaphor by the German chemist Emil Fischer (Fischer 1894). The dates *a quo* and *ad quem* cover a long period, almost four decades, during which the structural thought in organic chemistry was established. A selective choice was then compelling.[11]

In 1858, Couper proposed to the *Philosophical Magazine* the first ever published structural diagrams, as an expression of epistemological clarity. The first part of Couper's paper reports a cogent epistemological analysis of various chemical theories adopted at the time. He takes *affinity* – the property of atoms to bind to each other - as the foundation of his own theory:[12]

> "affinity [...] is an inherent property common to all elements, by the removal of which the chemical character of an element will be destroyed, and by virtue of which an element finds its place marked out in a complex body." (Couper 1858)

According to Couper, this is the base for explaining the peculiar ability of Carbon atoms to bind to each other. Couper's purely topological reasoning is visualized by diagrams like the one reported in Figure 6.
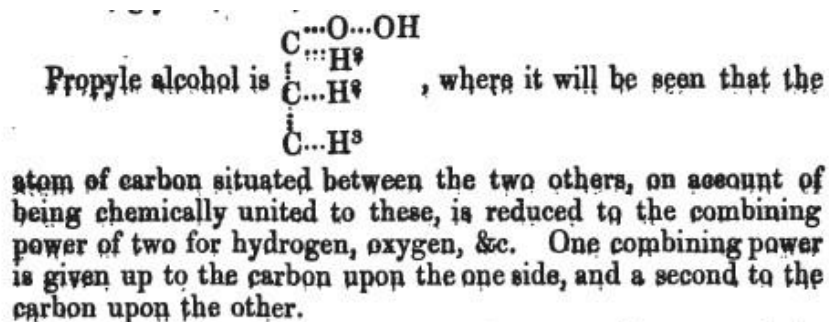

Figure 6 – A typical occurrence of a molecular diagram in (Couper 1858)

Couper merge his brand-new formulas with the text, in a colloquial mode, within a sentence expressed in the natural language. It must be recalled that Couper assigns Oxygen an atomic weight of 8, whereas Hydrogen's atomic weight is 1. Couper draws his attention to the carbon scaffold; he states that, while moving from the atomic to the molecular level, affinity finds its expression as 'combining power', whose value is 4 for Carbon. Couper comments the structural diagram as follows:

---

[11] The overall development of stereochemistry is described by important historiographic works (Ramsay 1981; Ramberg 2017).

[12] The term 'affinity' wold be subsequently replaced by 'valence'; here we report the original terms employed by the authors.

"It will be seen that the atom of carbon *situated between* the two others, on account of being chemically united to these, is reduced to the combining power of two for hydrogen, oxygen, &c. One combining power is given up to the carbon *upon the one side*, and a second to the carbon *upon the other*." (Couper 1858) (italics added)

It is doubtless that Couper's diagrams are the graphical transcript of purely topological arguments.

Couper's structural diagrams and topological reasoning found a good match in the theoretical work by the Russian organic chemist Alexander Mikhailovich Butlerov, whose concept of structure is expressed in the following terms:

"In truth, we do not know what kind of interdependence exists between the chemical action, reciprocally exerted by the atoms inside a compound molecule, and their mutual mechanical position (*mechanischen Lage*). We do not even know if two atoms of a compound molecule, one acting chemically and directly towards the other, are placed right next to each other. However, we can neither completely ignore the concept of *physical* atom, nor hide the fact that the chemical properties of a body are specifically conditioned by the chemical interrelation of its constituent elements. Starting from the assumption that each chemical atom employs only a given and limited quantity of chemical force to take part in the formation of a body, I could define the type and mode of reciprocal bonding (*gegenseitigen Bindung*) of atoms in a compound body with the name of chemical structure (*chemischen Structur*)." (Butlerov 1861)[13] (italics in the text).

The systemic aspect of the term *Structur* must not overshadow the fact that Butlerov refers to *gegenseitigen Bindung:* the adjective *gegenseitig* indicates that the chemical bond is the outcome of an action carried out by both atoms involved in it. Butlerov's systemic view does not conflict with the privileged relationship between contiguous atoms in a molecule; this aspect is further clarified by Butlerov in the conclusions of his article:

"I am very far from thinking that here I have proposed a new theory: I rather believe that I'm expressing ideas that really belong to many chemists. I must also note that Couper's conception and formulas, whose absolute and overly exclusive conclusions right now I contest, were based on a similar but not sufficiently clear and well understood thought." (Butlerov 1861)

If at the base (*zu Grunde*) of Couper's and Butlerov's explanations there is a 'similar thought' (*ein ähnlicher Gedanke*), then it is clear that Couper's topological perspective caught the attention and inspired the Russian chemist. In fact, Butlerov seems to describe more of a topology than a structure ("the type and mode of reciprocal bonding of atoms in a compound"), as such description does not (and could not) contain information on bond lengths and angles.

The new horizon of stereochemistry came out several years later. In 1874, a young Dutch chemist, Jacobus H. van't Hoff, published a breakthrough work. The editorial events of van't Hoff writings were complicated: in 1874 he first published a *pamphlet* in Dutch (van't Hoff 1874a), followed by its French translation that came out on a Dutch scientific journal (van't Hoff 1874b). As both writing did not raise the expected attention, in 1875 van't Hoff published a new work entitled *La Chemie dans l'Espace*. (van't Hoff 1875). This paper differs from the previous ones because van't Hoff realised that a purely graphical (bidimensional) communication was not effective. So he made up paperboard models (*modèles en carton*) of his structures, in double specular pairs, and sent them to eight renown chemists, namely August Kekulé, Alexander Butlerov, Adolf Baeyer and Johannes Wislicenus (van't Hoff 1875). Wislicenus was persuaded by van't Hoff theory and proposed him to translate the French monography into German. The translation was made by his student Felix Herrmann and he wrote the foreword (van't Hoff 1877, pp. VII-X). This assured van't Hoff's theory a great impact on the scientific community, to the point that his contributions are still exemplary, not least for the diversified use of theoretical, iconic and material models.

---

[13] The *Zeitschrift für Chemie* was published in German; the *Zeitschrift* was managed by German and Russian chemists.

van't Hoff 1874 writing shows already a theoretical model grounded on the conjecture that a carbon atom displays 4 affinities (*affiniteiten*) pointing at the vertexes of a tetrahedron, with the C atom in the middle. This hypothesis allows to make a correct prediction of the number of isomers of substituted methane; further, in case the four univalent substituents are different from each other (*univalente groepen*) something relevant comes out:

> "In case the four affinities of a carbon atom are saturated by four mutually different univalent groups, two and not more than two different tetrahedra can be obtained, which are each other's mirror images, but never can be thought that they cover each other, i.e. one has to do with two isomeric structural formulas in space (*twee isomere structurformules in de ruimte*)." (van't Hoff 1874, p.4)

A semantic innovation follows immediately: "a [carbon atom] bound to four mutually different univalent groups, in the following is referred to as asymmetrical (*asymmetrisch*)". This theoretical model is illustrated by pictures that would become renown, such as the iconic model of Fig. 7 (van't Hoff 1874).[14] At the beginning of his French paper, van't Hoff tackles a crucial issue. The research aim being the investigation of the mutual position of atoms in a molecule, one has to take into account that a molecule is a dynamic system; hence "*each movement changes the shape of a system that has been defined*". Nevertheless "*atoms must undergo a periodic movement in a molecule: hence it will be possible to represent the relative position of these atoms in a specific phase of their movement.*" (van't Hoff 1875, p.5) [15]
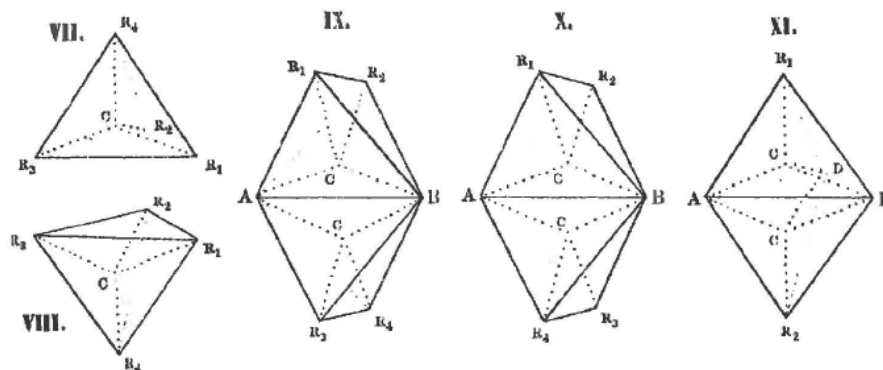

Figure 7 – Part of the figures reported in (van't Hoff 1874). The discussion on the asymmetric carbon is based on Images VII and VIII.

A thorough reading of van't Hoff works published between 1874 and 1877 brings to light some relevant epistemological aspects. First of all, the introduction of geometrical elements (valence angles) marks the evolution of these structural diagrams from purely topological representations to geometrical representations (Figure 7): this is an historical determining step. In the second place, the use of material models becomes increasingly important in van't Hoff communication strategy: these are absent from the Dutch pamphlet, they were sent to a selected number of chemists together with the French paper and, finally, they were offered to the whole readership in the shape of graphical models that could be traced on cardboard and properly folded. For these readers, material models display an argumentative – rather than heuristic – function; but their design had no doubt a relevant epistemic function in van't Hoff thought.[16] Finally, we remark that the mention of *shape* occurs within a dynamic description of
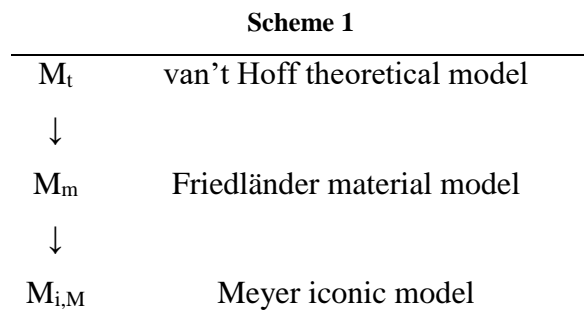
---

[14] The page of (van't Hoff 1874) where the icons are displayed is available at URL: http://sciencepenguin.com/jacobus-henricus-van-t-hoff/.

[15] In other two passages of the text, van't Hoff refers to his molecular models as *forme du tétrahèdre* (van't Hoff 1875, p. 15, 16)

[16] The cognitive pathway followed by van't Hoff in devising his material models is effectively retraced in (Friedman 2016)

molecules, whose atoms change their position instantaneously, in a periodic mode. This detail discloses van't Hoff's idea that his diagrams are representative of the 'average' position of atoms in the molecular space; consequently, shape appears as a theoretical construct.

Sixteen years after van'Hoff's first *pamphlet, Stereochemie* had become a recognized sub-discipline of chemistry. On January 28th 1890 Victor Meyer gave a talk in Berlin: he spoke about the results achieved by stereochemistry and the future perspectives. During that conference, the speaker showed material models of molecular structures, mentioned Paul Friedländer, professor in Karlsruhe, as the proposer of such models and an Austrian firm as the producer of commercial kits. The models of molecules bearing asymmetrical carbon atoms served as material supports, helping Meyer to explain the kind of projection he was using. This articulate use of models is schematized in Scheme 1.

Scheme 1

M_t	van't Hoff theoretical model
$\downarrow$	
M_m	Friedländer material model
$\downarrow$	
$M_{i,M}$	Meyer iconic model

It's worth underlining that the geometrical content (bond angles) of M_t is conserved in M_m, but gets lost in $M_{i,M}$.[17] The projection is a geometrical procedure whose starting point is the 3D material model; the resulting graphical representation information about angles is lost and becomes dependent on the procedure followed for getting the projection, a detail that Meyer does not specify in the text. As a consequence of this loss of information, Meyer structural diagrams are - once again - *topological*, albeit their being enriched with a relevant convention. Figure 8 reports Meyer's structural diagram of lactic acid (Meyer 1890). He states that the central C atom must be "thought of as tetrahedrical" (*tetraëdrisch gedachte*); hence, the symbol C – in this graphical representation – summarizes the usual properties of a C atom (e.g. atomic weight, tetravalence, a specific atomic volume, etc.) *plus* the property of being tetrahedral. The conventional character of such property is enhanced by the fact that – in principle – the other two C atoms should be tetrahedral as well, but this detail is clearly judged irrelevant by Meyer, in the frame of this discussion.
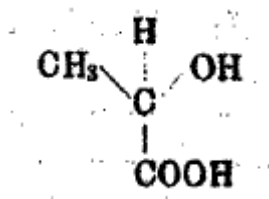

Figure 8 – Meyer's structural diagram of lactic acid (Meyer 1890)

In addition, Meyer employs the terms *Configuration* (configuration) and *Form* (shape) in the text, as if they were interchangeable.

---

[17] We use the double index ($M_{i,M}$) to distinguish this model from Fischer iconic model, designated as ($M_{i,F}$), that is mentioned further on

Let's consider the last relevant contributions in this historical *excursus*. In years 1891-1894 Emil Fischer wrote three foundational papers that stand out amongst his vast scientific production (Fischer 1891a, 1891b, 1894). In 1891 the German chemist accomplished his researches on the configuration of sugar molecules and proposed the projections that are named after him. In 1894 Fischer proposed the key/lock metaphor to describe the boundaries of the enzyme activity on sugar molecules. These two research lines are strictly connected.

In his first paper (Fischer 1891a) Fischer organizes the impressive amount of experimental data obtained in his work on natural and synthetic sugars. Through a patient and smart work, he retraced the 'genealogy' of monosaccharides and established the contribution of each asymmetric C atom to the optical activity of the investigated molecular structures. A pivotal argument is found in the excerpt reported in Figure 9(a): "*Since what is designated by + and - is equivalent, I have arbitrarily (willkürlich) given d-glucaric acid the formula ...*" (Fischer 1891a). The formula of Figure 9(a) is doubtless a topological one; the atoms indexed with + and – are *tetrahedral,* just like those of Meyer's diagrams; nevertheless, despite their conventional nature, these signs designate a specific – individual! – contribution to the overall molecular optical activity. The systemic character of the relationship between each atom and the molecular properties is here quite evident.

When Fischer assigned a conventional meaning to the + and – signs in his formulas, he had already decided to follow Meyer's suggestion as regards material models and projections. A second paper, published shortly after the first one, Fischer applies his arguments at the formula (*Formel*) of grape sugar' written according to the conventions of Figure 9(a). The author expresses his dissatisfaction for this kind of formulas, due to their potential ambiguity: *"So, since the above expression is ambiguous for the configuration of the grape sugar, it seems to me convenient to illustrate it by the following images (Bilder)*" (Fischer 1891b). In order to get unambiguous representations, Fischer suggests the reader to use Friedländer's models and build the 3D structures of dextro-, levo- and optically inactive tartaric acid. Then he invites the reader to: *"place them on the plane of the paper so that the four carbon atoms are in a straight line and that the hydrogens and hydroxyls of interest are above the plane of the paper. By projection you get the following drawings ..."* (Fischer 1891b), i.e the icons of Figure 9(b).
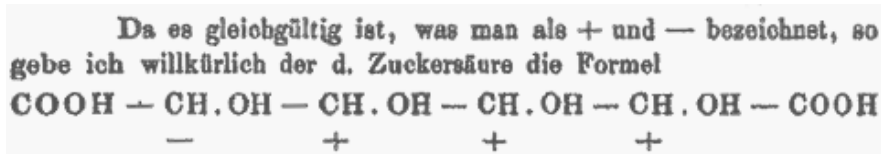
(a)
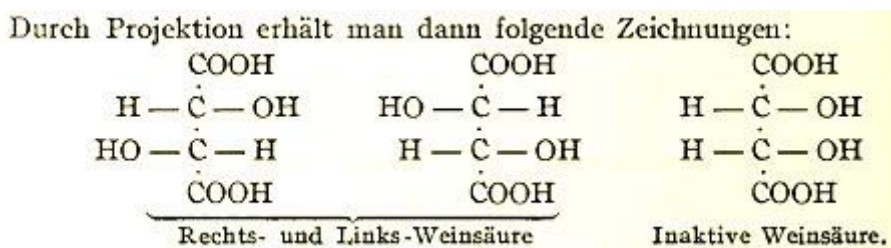
(b)

Figure 9 – Excerpts from (a) Fischer 1891a and (b) Fischer 1891b

The historical and epistemological meaning of these first Fischer projections may be better understood by using the symbols of Scheme 1. Scheme 2 compares the use of theoretical, material and iconic symbols made by Meyer and Fischer, respectively.

Scheme 2

| Victor Meyer: | $M_t \rightarrow M_m \rightarrow M_{i,M}$ |
| Emil Fischer: | $M_t \rightarrow M_m \rightarrow M_{i,F}$ |

Scheme 2 highlights that both scientists shared van't Hoff theoretical model and Friedländer material models, but they diverged in their use of projections. There are several reasons behind the scarce reception of Meyer's iconic model ($M_{i,M}$) and the success of Fischer's ($M_{i,F}$); here we will just remark that Meyer provided no instructions to his readers about how to get to a given icon starting from the same molecule. Conversely, Fischer describes a simple and univocal procedure that allows all readers to speculate over the same icon.

Fischer gets to his renown key/lock metaphor without breaks between his researches on sugars and those on enzymes. The metaphor is mentioned in several papers: we refer to the most renown one (Fischer 1894). The experimental part of this work illustrates the distinct activity of *Invertin* (a yeast extract) and *Emulsin* (a preparation from almonds) on alpha-methyl glucoside and beta-methyl glucoside. In fact, *Invertin* hydrolised only the alpha isomer, whereas *Emulsin* was active on the beta isomer. In the final part of the paper, Fischer addresses the enzyme activity as follows:

> "Their limited effect on the glucosides could therefore also be explained by the assumption that only with a similar geometric construction (*ähnlichem geometrischen Bau*) can the approach of the molecules take place, which is necessary for the triggering off of the chemical process. In order to use an image (*Bild*), I want to say that enzyme and glucoside must fit together like a lock and a key (*wie Schloss und Schlüssel*) in order to be able to exert a chemical effect on each other." (Fischer 1894)

The words *geometrischen Bau* are demanding, on the epistemological plane; notwithstanding, Fischer had no information on the molecular structure of these enzymes, even though he had a mental representation of how glucoside was made. So, his mental representation of the relationship between substrate (inside) and enzyme (outside) is undoubtedly a topological metaphor.

The possible epistemic connections between topology, structure and shape are discussed in the next section, but this historical survey allows us to forerun some considerations. Let's consider the timeline of publication of the texts analysed in this section:

Couper	Butlerov	van't Hoff	Fischer	Fischer
1858	1861	1874-1877	1891	1894
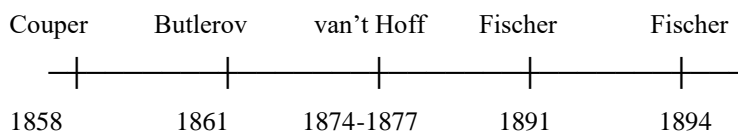
This temporal outline highlights the crucial function of van't Hoff's contribution, both as regards chronology and the epistemic content of the procedures employed by chemists. Couper's and Butlerov's proposal share the common base of a topological thought that gives a prominent position to the connections between neighbouring atoms: Butlerov's structure consists of the combination of these connnections. van't Hoff's drawing upon material models had a crucial epistemological outcome: molecular diagrams drawn on paper were perceived differently, in that they became *projections* of atoms' positions inside the molecular space. Despite the lack of experimental evidence concerning bond distances, van't Hoff conclusively argues that the four simple bonds of a carbon atom may not lie on the same plane; for symmetry reasons, he invokes a tetrahedral arrangement. In his works on the genealogy of hexoses, Fischer exploits the topological conception in all those synthetic processes aimed at extending the Carbon atom chains; in parallel, he exploits van't Hoff structural thought in order to assign to each Carbon atom its specific contribution to optical activity. Fischer's representations - through the choice of the mathematical term *projection* - inherently suggest a 3D geometrical view of molecular structure.

A distinct epistemological content connotes the key/lock metaphor. In fact, despite Fischer's mentioning a 'geometrical construction by analogy', the inside/outside concepts are essentially topological and may be rigorously defined as such (Egenhofer 1953). In a sense, Fischer gets to molecular shape by getting back to the topological thought: he comes full circle of chemists' theoretical thought, that started from topology and virtuously gets back to it after picking up geometrical concepts like symmetry and projection.[18]

Analysing the connections between molecular structure, shape and topology

The relation between the concepts of *molecular structure, shape* and *topology*, their contiguities and their peculiarities can now be discussed in the light of the content of the previous sections.

A recent work by Ochiai (Ochiai 2017) argues that "*Neither shape nor structure should be taken as attributes of real entities, but as theoretical constructs*". We agree with Ochiai as regards molecular shape (understood as a pure theoretical construct and a contribute of the subject), but have a divergent position as regards structure. We pointed out earlier that we think of molecules as real microscopic entities, with measurable properties that are accessible to experimental investigation. This position is close to Del Re (Del Re 1998), who conceives structure as an order principle belonging to the reality of a molecule:

> "Molecular structure is a *static topological and geometrical order principle* which belongs to the reality of a molecule, indeed is what distinguishes a particular molecule from all other clusters and molecules consisting of the same atoms. It is a 'principle' in the following sense: it is a unitary [..] entity to which *a variety of observable molecular properties belong.*" (Del Re 1998) (italics added)

Del Re was a theoretical chemist; thus, his urgency on the fact that structure implies both topological and geometrical information is intriguing. In fact, we have seen that the notion of structure, in the frame of the systemic view typical of QM, does not necessarily imply a topology. On the other hand, any chemist knows that an operational use of the outcome of a quantum-mechanical calculation in the domain of chemical reactivity does need to point out the way atoms are interconnected and which privileged relationships occur inside a molecule. In such a chemical perspective, it seems impossible to pass over molecular topology while dealing with structure.

Still, in the excerpt by Del Re we notice a striking coherence between the adjective 'static' referred to the structure and the previously reported excerpt by van't Hoff, where he pointed out that any representation of molecular structure catches a still instant of its endless internal periodic vibrations.

Molecular shape, on the other hand, is a pure conceptual product and a subjective representation (Ochiai 2017) – a derivative concept with respect to structure and topology, too. It is not by chance that, unlike structure, molecular shape is generally referred to in analogical terms (Del Re 2000), to the point that we recognize star-shaped, V-shaped, tree-shaped molecules, chair or boat conformations, or yet supramolecular systems characterised by channels, cavities, groves, etc.[19]. The heuristic aspect of the notion of molecular shape is here especially clear.

So, which connection between molecular structure and shape?

In a chemical perspective, aiming at pointing out structure-function relationships between microscopic entities and macroscopic behaviour of the corresponding substances in various environments, topology may be seen as a *pivot* between structure and shape. Molecular topology encompasses less information than structure, as its (algebraic or

---

[18] According to Leslie Glasser: "Being an essentially geometrical concept, symmetry is especially suitable for study by means of diagrams and models, the use of which can enhance that ability to visualize in three dimensions" (Glasser 1967)

[19] An official site of the US Government collects several papers on this issue, under the common title "Sample records for molecular shape amphiphiles". The abstract reports expressions like: "wedge-shaped amphiphilic molecules", "H-shaped supra-amphiphiles", "dumbbell-shaped giant hybrid molecule", "Y-shaped amphiphilic block polyurethane (PUG) copolymers", "Gear-shaped amphiphile molecules", "Amphiphilic crescent-moon-shaped microparticles", V-shaped polyaromatic amphiphiles.
URL: https://www.science.gov/topicpages/m/molecular+shape+ amphiphiles.html

graphical) expression does not convey any quantitative information concerning bond lengths, angles or the charge distribution within the molecule. But, unlike shape, it contains information on the connectivity of atoms, that is directly related with chemical reactivity. Molecular shape - unlike molecular structure and topology - is not univocal, due to its conventional character. This is supported by Ochiai, according to whom molecular structure "*shows not only the spatial arrangement of atoms, but also the linkage between atoms*" whereas "*Shape does not connote these details. It is concerned with the extension of (parts and the whole of) a molecule*" (Ochiai 217).

The relational value of these three terms is also different. We see structure as an intrinsic property that provides an inner insight into the molecule. Shape is rather projected towards the outside, as it is crucial for predicting interactions between the molecule and its chemical environment. Topology is half way, in that it bears information related with the inside (connectivity) and the outside (reactivity) of the molecule.

Our view of the relationship between molecular structure, shape and topology is schematized in Table 1.

Table 1 – Schematic representation of the relationship between molecular structure, shape and topology

STRUCTURE		SHAPE
↓		↓
CONSTITUTIONAL PROPERTY	↔ TOPOLOGY ↔	DESCRIPTIVE PROPERTY
↕		↕
ONTOLOGY		EPISTEMOLOGY
A systemic concept		A relational concept
Focused on the interior of the molecule		Focused on the exterior of a molecule

In this perspective, it would be mistaken to assign the same epistemological status to shape and structure and reduce both of them to pure theoretical constructs. If the molecule is a real microscopic entity, molecular structure is an inherent property, identified through observables related with such entity: in other words, structure is related with molecular ontology. Conversely, the extrinsic and conventional character designates molecular shape as an epistemological outcome: shape is a descriptive property. This difference is mirrored by the cognitive procedure leading to the identification of structure or shape. As a matter of fact, the identification of molecular structure or shape requires radically distinct cognitive procedures. Structure requires the experimental determination of bond lengths and angles as well as charge densities (through either X-rays diffraction, NMR, etc. or theoretical calculations based on experimental evidence). Shape is the outcome of calculations: it has conjectural character and is a pure interpretive elaboration. The procedures leading to the identification of molecular shape are purely epistemic, in that they have purely explicative and/or predictive goals; in fact, they are strongly dependent on the research aims of the subject.

Once again, topology lies in between: it relates with both molecular constitution and description, with both the intrinsic properties of the molecule (atomic connectivities) and the extrinsic ones (reactivity). Hence it has both ontological and epistemological character.

Another aspect that deserves attention is the distinction between the systemic character of structure and the relational character of shape.

In a previous section we have thoroughly discussed the meaning that we assign to the term *systemic,* referred to molecular structure, and its implications. However, the systemic character of molecular structure can be examined from other viewpoints. The most straightforward is the relationship between structure and activity. A good example is provided by mono-substituted benzene that may undergo electrophilic aromatic substitutions, whose outcome depends on the presence of ortho-, para- or meta directors on the molecule. The distinct reactivity observed in these cases is an 'internal' business of molecular structure: these directors are responsible for imparting the molecule a

specific reactivity, by affecting the electron density distribution *over the entire* benzene molecule; these changes occur in the molecule through internal processes.

Before discussing the relational character of shape, it is worth clarifying the interior/external duality in atomic-molecular systems, at the operational level. It is *interior* to the molecule any entity that belongs to that very same molecule on the basis of a conservation principle. For example, if the total charge of a molecular object is zero, the sum of the atomic charges must be zero, no matter which is the actual distribution of charges over all (ground and excited) energy states. Similarly, the mass conservation principle[20] and the conservation of the atomic number witness the fact that atomic nuclei and their specific elementary identity belong to the molecular system.

The identification of molecular boundaries (the *exterior* of the molecule*)* requires quite different treatments. At the quantum level, the functions that describe the spatial distribution of electron density are asymptotic: consequently, the circumscription of the internal molecular space implies the choice of limiting surfaces. A similar move is made in chemistry teaching, as regards *orbital shapes*. Another example is the identification of van der Waals radii: a choice is mandatory between the variety of experimental methods that provide numerical data used for representing an isolated atom or the set of atoms found inside a molecule. In addition, van der Waals radii are always determined with reference to non-bonded contact distances between *pairs* of identical atoms: here again, we find the relational character of shape (Scott Rowland 1996).

A further example of this character comes from the biological world and, more specifically, from the representations of molecular surfaces of biological macromolecules. These representations (see Figure 4) are often aimed at identifying points of interaction between the macromolecule and its chemical environment (e.g. water molecules, substrates, effectors, etc.). Again, this is clearly related with the *exterior* of the molecule.

Let's go back to topology as a mediator between structure and shape. In current use, topology is rarely used in its simplest form (a pure indication of inter-atomic connectivity). It often bears further information, such as parametrical data concerning atomic orbital hybridation or standard distances between atomic pairs. This is also the case with many QSAR descriptors. Consequently, the arrow between topology and structure in Table 1 gains a specific epistemic meaning: in a chemically meaningful, reactive domain, topology informs structure about interatomic connectivity. On the other side, structure provides bond lengths and angles (through hybridation) to molecular topology.

The relation between structure, shape and topology is especially intriguing in biological macromolecules, exhibiting primary sequences, secondary and tertiary structures, and shape. Primary sequence bears topological information and may be associated with various conformers (i.e. structures), characterised by distinct molecular shapes. Here we find a direct relation between molecular shape and structure: proteins' shape may change without affecting topology. On the other hand, structural predictions of biological macromolecules may be solely based on (topological) primary sequences. Here the transition from topology to shape is straight, without the mediation of an experimentally determined structure.

The ensemble of these aspects helps pointing out that the concepts of molecular structure, topology and shape serve different functions. Structure informs on the internal situation of the molecule. Topology allows predicting molecular reactivity (e.g., the presence of a carboxyl group opens up to specific reactions).[21] Shape is the fuzzier, but also the more open notion. The inspection of molecular shape suggests how the molecule may interact with its chemical surroundings: *the relational aspect of shape is the main reason for its use.*

---

[20] The principle is valid within the energetic scale of fundamental and excited states and in non-radioactive systems.

[21] As regards the topological character of the notion of 'functional group' we may cite the Nobel Laureate Corey: "Modern synthetic chemistry is a multifaceted discipline that greatly benefits from the development of unifying concepts. One of the most useful of these is the idea of the "functional group," generally considered to be a specific collection of connected atoms that occur frequently in organic structures and that exhibit well defined and characteristic chemical behavior." (Corey 2007, p. vii)

The last issue to be discussed as regards the three terms *molecular shape, structure* and *topology* is lexical. Most textbooks use them almost as synonyms, thus implying that the boundaries between them are not sharp. Wittgenstein's lesson *"The meaning of a word is its use in the language"* (Wittgenstein 1968, p.20) sounds here like a warning: the inaccurate use of a word may result in misconceptions, and this is especially relevant in chemical education. The teaching domain is marked by the asymmetry of interpretive tools available to teachers and students. Experts escape misunderstandings because they can contextualise technical terms, and they are able to properly decode them. Further, the flexibility of their semantic fields contributes to generate further semantic shades. The historical survey has shown that the meaning of s*tructure* has changed over time, enriching this term with new connotations.

Conversely, students (at any scholar level, from school to university) do not own the same cognitive baggage or linguistic expertise as their teachers do. As a consequence, a fruitful conceptual openness in a specialized context may turn into a source of incoherence and inaccuracy in the teaching context, fostered by the misleading use of these terms made by textbooks.

In this respect, it is the task of epistemology to propose a criterion fit to discriminate between the conceptual domains of *molecular shape, molecular structure* and *molecular topology*. A dimensional analysis points out the distinctive physical quantities for each notion. The suggestion comes from the conclusion of the remarkable paper by Nicholls and coauthors: "*shape as a volume or as a surface* is a vibrant and useful concept when applied to drug discovery" (italics added) (Nicholls 2010). Let's focus on the physical dimension length [L] and have a look to Table 2. Along this paper we have seen that topological representations do not include bond lengths; hence, from the dimensional viewpoint, they are characterised by L^0. Structural representations obtained, for example, by X-ray diffraction are characterised by atomic coordinates, that allow identifying bond lengths [L^1]. Eventually, the representation of shape is characterised by either surfaces or volumes, that is [L^2] or [L^3].

Table 2 – Dimensional analysis of physical representations of molecules

Molecular Representation		Physical Dimension [L] Length	Relevant Chemical Terms
Topology		L^0	Functional groups
Structure		L^1	Bond length
Shape	Surface	L^2	Accessible surface area
	Volume	L^3	van der Waals volume

The criterion of Table 2 finds immediate application in the measurement of the *accessible surface area*, i.e. the surface area of a biomolecule that is accessible to a solvent and is expressed in $Å^2$. According to this criterion, any reference to electron density relates to the notion of *shape* of the molecular object: this quantity is measured in $e/Bohr^3$. The criterion of Table 2 applies even to less obvious cases, such as the moment-based methods employed for *in silico* representations of molecular shape (Putta 2007). In fact, the inertial moments mentioned by Putta and co-authors have dimension ML^2.

In conclusion, this paper aimed at analysing the concepts of *molecular shape*, *molecular structure* and *molecular topology* and their mutual relations. After discussing the implications and distinct acceptations of these terms, as well as their historical roots, we suggest that the notion of topology acts as a mediator between those of structure and shape. These latter pertain to the ontological and epistemological levels, respectively, and they are accessible through radically distinct cognitive practices. Unfortunately, the three notions are often mixed up in textbooks: this can result in misconceptions and inaccuracies in the chemical education domain. A simple criterion based on dimensional analysis may help in discriminating between the conceptual domains of *molecular shape, molecular structure* and *molecular topology*.

References

Albert, R., Jeong, H., Barabasi, A.L.: Error and Attack Tolerance of Complex Networks. Nature 406, 378–382 (2000)

Andruniów, T., Ferré, N., Olivucci M.: Structure, initial excited-state relaxation, and energy storage of rhodopsin resolved at the multiconfigurational perturbation theory level. Proc. Natl. Acad. Sci. U.S.A 101 17908-17913 (2004)

Balaban A.T.: Graph theory and theoretical chemistry. J. Mol. Struct. 12O, 117-142 (l985)

Butlerov, A. M.: Einiges über die chemische Structur der Körper. Zeitschrift für Chemie, 4, 549–560 (1861)

Cerro, M., Merino, G.: ¿Moléculas sin esqueleto?: La oportunidad perfecta para revisar el concepto de estructura molecular (Molecules without framework? The perfect opportunity to review the concept of molecular structure). Education Quìmica 20, 187-191 (2009)

Cerruti, L., Ghibaudi, E.: Peirce's Semiosis and the Representation of Protein Molecules. ISPC 2017 Conference paper, forthcoming (forthcoming)

Chamizo, J.A.: The fifth chemical revolution: 1973–1999. Found. Chem. 19, 157–179 (2017)

Corey, E.J: Foreword. In: Li, J.J., Corey, E. J. (eds.) Name Reactions of Functional Group Transformations. Wiley, Hoboken (NJ) (2007)

Couper, A. S.: On a new chemical theory. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 16, 104-116 (1858)

Del Re, G.: Ontological Status of Molecular Structure. HYLE 4, 81-103 (1998)

Del Re, G.: Models and analogies in science. HYLE 6, 5-15 (2000)

Earley, J.: A neglected aspect of the puzzle of chemical structure: how history helps. Found. Chem. 14, 235–243 (2012)

Egenhofer, M.J.: A model for detailed binary topological relationships. Geomatica, 47, 261-273 (1993)

Ferretti, A., Granucci, G., Lami, A., Persico G., Villani, G.: Quantum mechanical and semiclassical dynamics at a conical intersection. J. Chem. Phys. 104, 5517-5527 (1996)

Fischer, E.: Über die Configuration des Traubenzuckers und seiner Isomeren. I. Berichte der Deutschen Chemischen Gesellschaft 24, 1836-1845 (1891a)

Fischer, E.: Über die Configuration des Traubenzuckers und seiner Isomeren. II. Berichte der Deutschen Chemischen Gesellschaft 24, 2683-2687 (1891b)

Fischer, E.: Einfluss der Configuration auf die Wirkung der Enzyme. Berichte der Deutschen Chemischen Gesellschaft 27, 2985–2993 (1894)

Fortin, S., Lombardi, O., González, J.C.M.: The relationship between chemistry and physics from the perspective of Bohmian mechanics. Found. Chem. 19. 43-59 (2017)

Friedman, M.: A Failed Encounter in Mathematics and Chemistry: The Folded Models of van 't Hoff and Sachse. Teorie vědy / Theory of Science 38, 359-386 (2016)

Gaussian 16, Revision B.01, Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M. A., Marenich, V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, K. Toyota, M., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K. A., Rendell, P. J., Burant,

C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., Fox, D. J. Gaussian, Inc., Wallingford CT (2016)

Glasser, L.: Teaching Symmetry: the use of decorations. J. Chem. Ed. 44, 502−511 (1967)

González, J.C.M., Fortin, S., Lombardi, O.: Why molecular structure cannot be strictly reduced to quantum mechanics. Found. Chem. (2018) doi.org/10.1007/s10698-018-9310-2

Grasl, T., Economou, A.: From topologies to shapes: parametric shape grammars implemented by graphs. Environment and Planning B: Planning and Design 40, 905 – 922 (2013)

Hendry, R.F.: Chemistry: emergence vs. reduction. In: Macdonald, C., Macdonald, G. (eds.) Emergence in Mind. Oxford University Press, Oxford (2010)

Hendry, R.F.: Reduction, Emergence and Physicalism. In Hendry, R.F., Needham, P & Woody, A (eds.) Philosophy of Chemistry. Elsevier, Amsterdam (2011)

IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by McNaught, A. D. and Wilkinson, A. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by Nic, M., Jirat, J., Kosata, B., updates compiled by Jenkins, A. doi.org/10.1351/goldbook. Last update: 2014-02-24; version: 2.3.3. Last access: May 2018.

Jensen, B.W.: Logic, History, and the Chemistry Textbook II. Can We Unmuddle the Chemistry Textbook? J. Chem. Educ. 75, 817-828 (1998)

Klein, U.: Experiments, models, paper tools: Cultures of Organic Chemistry in the Nineteenth Century. Stanford University Press (2003)

Knight, T.: Computing with emergence. Environment and Planning B: Planning and Design 30, 125-155 (2003)

Lami, A., Villani, G.: Quantum dynamics of proton transfer in H_3O^+-H_2O complex. Chem. Phys. Lett. 238 137-142 (1995a)

Lami, A., Villani, G.: A model study of proton transfer in $H_5O_2^+$ complex on a bidimensional potential energy surface. J. Mol. Struct. (Theochem) 330, 307-312 (1995b)

Lami, A., Villani, G.: Model studies of the dynamics at conical intersections. In: Domcke, W., Yarkony, D. R., Köppel, H. (eds.) Conical intersections: Electronic structure, Dynamics and Spectroscopy, Advanced series in Physical Chemistry, vol.15, ch. 8. World Scientific, Singapore (2004)

Lewis, G.N.: The atom and the molecule, J. Am. Chem. Soc. 38, 762–785 (1916)

Maggiora, G., Vogt, M., Stumpfe, D., Bajorath, J.: Molecular Similarity in Medicinal Chemistry. J. Med. Chem. 57, 3186-3204 (2014)

Meyer, V.: Ergebnisse und Ziele der stereochemischen Forschung. Berichte der Deutschen Chemischen Gesellschaft 23, 567-619 (1890)

Mezey, P.G.: Shape in Chemistry. An Introduction to Molecular Shape and Topology. VCH, New York (1993)

Nicholls, A., Good, A.C., Warren, G., Mathieu, M., Muchmore, S.W., Brown, S.P., Grant, J.A., Haigh, J.A., Nevins, N., Jain, A.N., Kelley, B.: Molecular Shape and Medicinal Chemistry: A Perspective. J. Med. Chem. 53, 3862–3886 (2010)

Ochiai, H.: Does a molecule have a structure? Found. Chem. 19, 197–207 (2017)

Putta, S., Beroza, P.: Shapes of Things: Computer Modeling of Molecular Shape in Drug Discovery. Current Topics in Medicinal Chemistry 7, 1514-1524 (2007)

Ramsay, O.B.: Stereochemistry. Heyden, London (1981)

Ramberg, J.L.: Chemical Structure, Spatial Arrangement: The Early History of Stereochemistry, 1874–1914. Routledge, London (2017)

Ramsey, J.L.: Molecular Shape, Reduction, Explanation and Approximate Concepts. Synthese 111, 233-251 (1997)

Rocke, A.: The quiet revolution of the 1850s: social and empirical sources of scientific theory. In: Mauskopf, S.H. (ed.) Chemical Sciences in the Modern World. University of Pennsylvania Press, Philadelphia (1993)

Scott Rowland, R., Taylor, R.: Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. J. Phys. Chem. 100, 7384–739 (1996)

Schulz, A., Shamir, A., Baran, I., Levin, D.I.W., Sitthiamorn, P., Matusik, W.: Retrieval on Parametric Shape Collections. ACM Transactions on Graphics 36, Article No.11 (2017)

Teixeira A.L., Leal, J.P., Falcao, A.O.: Automated Identification and Classification of Stereochemistry: Chirality and Double Bond Stereoisomerism. Technical Report. LaSIGE, Department of Informatics, Faculty of Sciences, University of Lisbon (2013) doi: arXiv:1303.1724

Todeschini, R., Consonni, V.: Handbook of Molecular Descriptors, vol. I. Wiley-VCH Verlag, Darmstadt (2009)

Todeschini, R, Consonni, V, Xiang, H., Holliday, J., Buscema, M., Willett, P.: Similarity coefficients for binary chemoinformatics data: overview and extended comparison using simulated and real data sets. J. Chem. Inform. Model. 52, 2884–901 (2012)

van't Hoff, J.H.: Voorstel tot Uitbreiding der Tegenwoordige in de Scheikunde gebruikte Structuurformules in de Ruimte, benevens een daarmee samenhangende Opmerking omtrent het Verband tusschen Optisch Actief Vermogen en chemische Constitutie van Organische Verbindingen [Proposal for the Extension of Current Chemical Structural Formulas into Space, together with Related Observation on the Connection between Optically Active Power and the Chemical Constitution of Organic Compounds]. Greven, Utrecht (1874a)

van 't Hoff, J.H.: Sur les formules de structure dans l'espace. Archives neerlandaises des sciences exactes et naturelles 9, 445–454 (1874b)

van't Hoff, J.H.: La Chimie dans l'Espace. Bazendijk, Rotterdam (1875)

van't Hoff, J.H., Herrmann, F.: Die Lagerung der Atome im Raume. Vieweg, Braunschweig (1877)

Villani, G.: La chiave del mondo. Dalla filosofia alla scienza: l'onnipotenza delle molecole, CUEN, Napoli, Italy (2001)

Villani, G.: Theoretical investigation of hydrogen transfer mechanism in adenine-thymine base pair. Chem. Phys., 316, 1-8 (2005)

Villani, G.: Theoretical investigation of hydrogen transfer mechanism in guanine-cytosine base pair, Chem. Phys., 324, 438-446 (2006)

Villani, G.: Complesso e Organizzato. Sistemi strutturati in fisica, chimica, biologia ed oltre, Franco Angeli, Milano, Italy (2008)

Villani, G.: Theoretical investigation of the hydrogen atoms transfer in the adenine-thymine base pair and its coupling with the electronic rearrangement. Concerted vs stepwise mechanism. Phys. Chem. Chem. Phys. 12 2664-2669 (2010a)

Villani, G.: Theoretical investigation of the hydrogen atoms transfer in the cytosine-guanine base pair and its coupling with the electronic rearrangement. Concerted vs stepwise mechanism. J. Phys. Chem. B 114, 9653-9662 (2010b)

Villani, G.: Theoretical investigation of the hydrogen atom transfer in the hydrated A–T base pair. Chem. Phys. 394 9-16 (2012)

Villani, G.: Theoretical investigation of the hydrogen atom transfer in the hydrated C-G base pair. Mol. Phys. 111 201-214 (2013a)

Villani, G.: Theoretical investigation of the coupling between hydrogen-atom transfer and stacking interaction in adenine-thymine dimers. Chem. Phys. Chem 14, 1256-1263 (2013b)

Villani, G.: Theoretical investigation of the coupling between hydrogen atoms transfer and stacking interaction in guanine-cytosine dimers. Phys. Chem. Chem. Phys. 15, 19242-19252 (2013c)

Villani, G.: Chemistry: a systemic complexity science. Pisa University Press, Pisa (2017a)

Villani, G.: Quantum Mechanical Investigation of G-quartet Systems of DNA. *New Journal of Chemistry* 41 2574 – 2585 (2017b)

Watts, D.J., Strogatz, S.H.: Collective Dynamics of "Small-world" Networks. Nature 393, 440–442 (1998)

Wittgenstein, L.: Philosophical Investigation. Blackwell, Oxford (1968)

Wolley, R.G.: Must a molecule have a shape? J. Am. Chem. Soc. 100, 1073-1078 (1978)

Zeidler, P.: The Epistemological Status of Theoretical Models of Molecular Structure. HYLE 6, 17-34 (2000)