doi:10.2533/chimia.2023.646 Chimia 77 (2023) 646–654 © A. Togni

Do You Speak Chemistry?

Learning Chemistry Means Learning Its Language

Antonio Togni*

Abstract: This article is a broad personal reflection about the language of chemistry, deriving from decades of experience in both research and teaching. The three levels of thought as represented by the Johnstone triangle, a critical attitude towards systematic nomenclature, the concept of pure substance, and the central significance of structural formulas as the most prominent form of representation in chemistry, are the main topics discussed. The chemical transliteration of the Platonic solids is presented as a mean of expression unique to chemistry.

Keywords: Johnstone triangle · Platonic solids· Pure substance · Structural formula · Systematic nomenclature

Antonio Togni was born in the Italian part of Switzerland (Misox) and studied chemistry and obtained a PhD at ETH Zurich. After a postdoctoral stay at the California Institute of Technology (1983–1984), he joined the Central Research Laboratories of the former Ciba-Geigy Ltd. in Basel. In 1992, he took a risky move, leaving industry and returning to his alma mater as a non-tenure-track assistant professor. This choice paid out and he became a full professor in 1999. This is also about the time when he started slowly moving away from his original main research interests in organometallic chemistry and asymmetric catalysis towards fluorine chemistry. Since January 2021, he has been a happy retiree who quit research, but temporarily continues teaching. From 2016 until his retirement, he acted as the Vice Rector for the Doctorate at ETH and also served as a member of the Research Council of the Swiss National Science Foundation from 2012 to 2016.

1. Introduction

This article reflects the lecture I gave on August 24, 2023, at the Symposium Future of Chemical Education held in the frame of the Fall Meeting of the Swiss Chemical Society. Several aspects have been taken from a series of lectures I have been giving at ETH for students of the *Lehrdiplom*[1] in chemistry, now already since several years. However, the considerations about the language of chemistry to be offered in the present article are not only meant for ongoing high school teachers, but to everybody speaking, writing, or thinking about chemistry at any level of experience in this discipline. My thoughts cover several aspects – practical, technical, philosophical, educational – as I was able to discern and appreciate during more than forty years as a researcher and roughly thirty years as a university teacher in chemistry (note that I never conducted research in chemical education).

Many argue that the language of science is mathematics thereby often forgetting that chemistry has its own unique language that cannot be viewed as or even reduced to mere equations and quantitative relationships. Others may be naively tempted to think that systematic nomenclature as recommended by IUPAC *is* actually the language of chemistry because it is made up of words, even if often very long and seldom easy to remember as such. As I will argue, the language of chemistry is much more than this, not only in terms of forms of verbal and/or written communication, but much more as an instrument when it comes to theorizing about chemical phenomena. Additionally, if we broadly intend language as a mean of expression, chemistry expresses itself through the unique ability of making its own objects.

The multifaceted aspects of such a language makes it difficult to understand and learn. It raises the question concerning a sensible approach in teaching, not only at high school, but also at university level.

2. Talking Chemistry and the Johnstone Triangle

In order to illustrate some of the most important characteristics of ‘spoken chemistry’, I shall use an example stemming from my own experience. During my time as a researcher at the Central Research Laboratories of the former Ciba-Geigy Ltd. I had been working quite intensively with ethyl isocynoacetate as a starting material for the synthesis of a specific type of oxazolines.[2] To a colleague asking me about my research, I would make a series of specific key statements. My monologue would include information at three different levels, represented as the corners of a triangle in Fig. 1. Indeed, I would possibly first talk about *macroscopic* properties – a brown, smelly liquid – and immediately try to offer an *interpretation* of those properties by moving to the *molecular* level – an isocyano group attached to an enolizable moiety. Of course, I would also draw the structure on any available piece of paper, *e.g.* the back of an envelope, thereby accessing the *formal/symbolic* level, which would also give the opportunity to formulate *e.g.* a reaction equation.

The essential point now to be noted is that moving around that triangle, from one corner to another and in any direction, thereby touching upon the three different levels *simultaneously*, is the most natural thing to do for a chemist when talking, as in this example, about a single compound as a chemical entity. However, it is exactly this peculiarity of the chemical discourse that makes chemistry very difficult to understand for the non-expert. The spontaneous, almost instinctive switching between the macroscopic, molecular and formal levels of thought is very characteristic for and unique to chemistry. However, chemists often forget – this happens to me as well – that learning and getting used to this is not a seamless process.

**Correspondence:* Prof. em. Dr. A. Togni, E-mail: atogni@ethz.ch

Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, CH-8093 Zürich, Switzerland

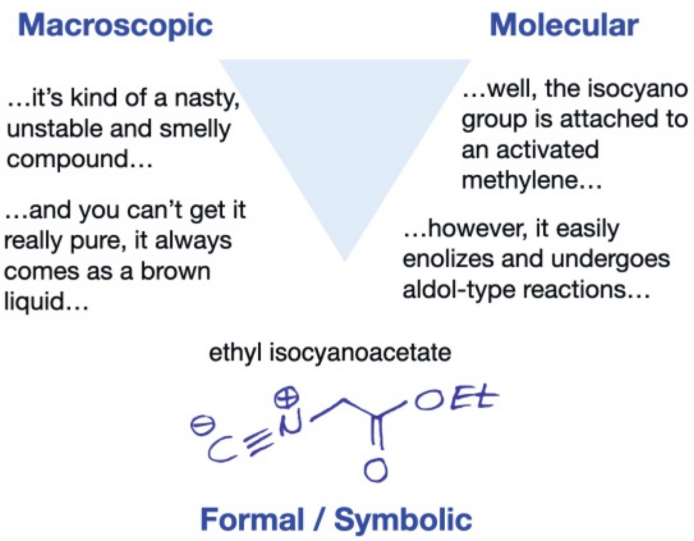

Fig. 1. Three levels of thought in a conversation about a single compound.

As far as teaching chemistry is concerned, this situation has been first analyzed by the late Alex Johnstone, a chemistry education researcher in Scotland. The so-called Johnstone triangle (see Fig. 2)[3] has been formulated in many ways and has constituted the basis for conceiving new chemistry curricula at high schools in the UK and other countries. Without going into the details of the Johnstone triangle for didactic purposes, it is sufficient to say that its main guiding idea is that a learner's working memory is not able to cope with the three levels of thought describing chemical objects and phenomena *simultaneously*. Hence, it is clear that a chemistry monologue as simple as the one described above, will turn out to be rather inscrutable and enigmatic, if not to say essentially incomprehensible to a person without specific training. This needs to be kept in mind also when teaching chemistry at first-year university level, when this specific training is not yet (or, in the worst case, will never be) effective.

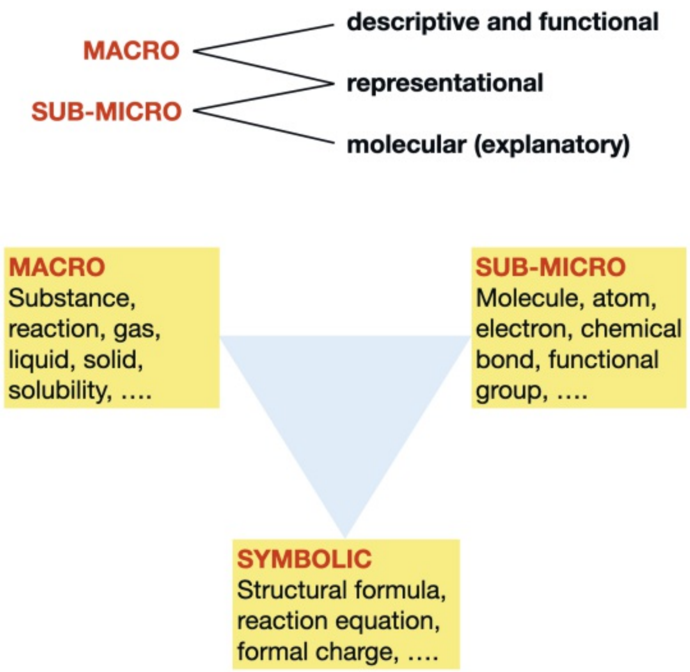

Fig. 2. Original representation by Johnstone of three levels of thought in chemistry (top) and the widely used illustration as a triangle.

3. Naming and Representing Molecules and Compounds

Is systematic nomenclature of molecular compounds important for understanding chemistry? I learned the nomenclature of organic chemistry as well as that of coordination chemistry when I was a student. Later I had to apply it as a researcher when describing new compounds in publications. However, since the advent and general use of molecule editors such as ChemDraw, formulating an unequivocal systematic name of even a relatively complex compound starting from the structural formula is no longer a problem and requires just one click with the function 'Convert structure to name'. What used to be and partly still is a task suited for exams in first-year organic chemistry courses (and partly even at high schools!) loses value and didactic significance. My critical and sceptical attitude towards investing significant teaching time for nomenclature also derives from the ascertainment that knowing the systematic name of a compound does not add any substantial and pertinent information about the compound itself. In terms of chemical knowledge, once I know the name, I know exactly what I knew before and nothing more!

Yet, as I will illustrate later, specific aspects of nomenclature, precisely as the act of classifying by naming, are eminently important when it comes to understanding reactivity in terms of relationships between compounds and compound classes.

However, let us first consider systematic names as one possible way of describing specific molecules and at the same time specific compounds. Since names are words, though very special ones, we should ask the simple question whether there is a similarity between such words and words in 'regular' language, specifically in terms of *immediate understanding* when reading them. By way of example, consider the names given below, generated by ChemDraw, of three compounds that have been made and studied in my research group.[4] They are given in the order of their length which also mirrors the slightly increasing structural complexity, from a substituted linear-chain compound, to heterocyclic derivatives having two and three fused rings, respectively. I take it for granted that they are correct.

1: (2*S*,3*R*,4*R*,*E*)-3-hydroxy-4-methyl-2-(methyl-amino)oct-6-enoic acid
2: 1-(4-azido-1,1,2,2-tetrafluorobutyl)-3,3-dimethyl-1,3-dihydro-1λ^3-benzo[*d*][1,2]iodaoxole
3: (1*R*,10*bR*)-3-benzoyl-8-bromo-1,2,3,5,6,10*b*-hexahydropyrazolo[5,1-*a*]isoquinoline-1-carbonitrile

These names count a total of 243 characters. A piece of regular text of about the same length from *e.g.* a book or a newspaper would require a reading time of just a few seconds for the reader to understand and capture its precise meaning. However, despite the fact that I have been confronted with these compounds directly, I will still not be able to remember all three names correctly after giving myself a generous reading time of 20 seconds. Given the same reading time, a student or even a chemist not familiar with these compounds would only be able to grasp minimal parts of them but could not really say with an acceptable accuracy what we are talking about.

By strong contrast, consider a little piece of classical world literature, such as the first six verses of Dante Alighieri's masterwork '*The Divine Comedy*':[5]

When half way through the journey of our life
I found that I was in a gloomy wood,
because the path which led aright was lost.
And ah, how hard it is to say just what
this wild and rough and stubborn woodland was,
the very thought of which renews my fear!

Even when reading this text for the very first time and never having heard about Dante Alighieri, a reader will immediately understand that it must be part of an old poem. More importantly, he/she will realize that a possibly unsettling message is being conveyed and shall be likely impressed, or motivated to think about, even though he/she might not be able to recite the verses by heart. The level and quality of understanding are totally different when compared to those deriving from quickly reading complex chemical names. While this is true essentially for any piece of common written text, I have not chosen Dante's verses randomly, but with a specific intention. The image of a 'gloomy wood, a dark forest in more prosaic simple terms, can be metaphorically tailored to chemistry. Hence, a sensible credo to be adopted by chemistry teachers could read:

Do not turn chemistry into a gloomy wood in which students may quickly lose their path and remember it later only with renewed fear!

Thus, it appears that systematic names are not suited to ensure immediate understanding when communicating about chemical objects, very simple molecules with short and easy-to-remember names – *e.g.* ethanol, acetone, sulfuric acid, *etc*. – being obvious exceptions. Instead of names, we need *visual representations*,[6] structural formulas being the most obvious ones. It is very important to emphasize the importance of structural formulas as a means of communication about molecular objects, as model representations, coding, and to a certain extent as metaphors, and as expression of theoretical thinking. When simply viewing a structural formula, chemists understand what they mean and what they are talking about. The perception about the object being described, as opposed to the case of systematic names, *i.e.* words, is immediate. Moreover, drawing a structural formula means (re) constructing that structure while drawing, which requires to anticipate and project knowledge about the very chemical object being drawn.

A structural formula is a two-dimensional *graphic* representation that can also transport some information about the three-dimensionality of molecules, though only to a limited extent. Another type of *pictographic* representations, for example in form of so-called space-filling models, is shown in Fig. 3 for molecules **1**–**3** together with their structural formulas. Space-filling models are specifically meant to convey shape and spacial extent of the molecules, *i.e.* their van der Waals volume. Such representations reflect either quantum-chemical calculations or they derive from physical measurements, typically X-ray diffraction of single crystals (see also ORTEP representations). However, in terms of primary communication about a chemical object, they play a different, auxiliary role and cannot replace structural formulas.

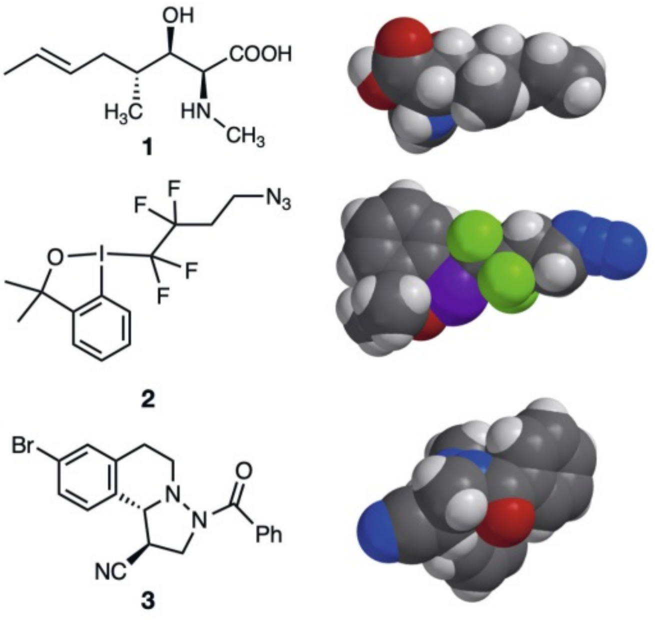

Fig. 3. Structural formulas and space-filling model representations of compounds **1**–**3**.

4. If the Periodic Table is the Alphabet, Are Molecules Words?

If taken as a metaphor, this analogy may sound somewhat simplistic. However, as pointed out by Pierre Laszlo in his erudite book '*La parole des choses ou le langage de la chimie*'[7] and in a more recent article,[8] if one considers groups of atoms (methyl, phenyl, hydroxyl, carbonyl, *etc.*) instead of just elements, the analogy makes much more sense. It becomes the starting point for looking at the molecular world and correspondingly at its teaching in a linguistic sense.

Without elaborating more in detail on this somewhat provoking point of view, I want to discuss a simple example on how to assemble five different atoms to a sensible molecule according to specified criteria. The example is adapted from a preparatory problem I formulated on the occasion of the 55th International Chemistry Olympiad that took place recently at ETH (https://www.icho2023.ch/):

Draw a complete Lewis structure of a compound with elemental composition CFNOS, in which:

- the sulfur atom is bonded to three of the remaining four atoms
- N and O display the lowest possible formal oxidation number
- there are two multiple bonds and no formal charges

CFNOS is an alphabetic sequence of five letters. I am not aware of any word in any of the languages I know of that could be composed using these five letters. Nonetheless, one could possibly think of CFNOS as an abbreviation or acronym, such as 'Canadian Force Naval Operations School', which indeed exists, or maybe 'Chemistry for non-ordinary scientists', which I freely invented. However, 'inventing' the molecule CFNOS requires to take into account the specific information given above, as if it were a little collection of spelling, syntax or grammar rules to be applied in order to find the right word.

Fig. 4 shows the unique solution of the problem, corresponding to an actually known compound,[9] presented in form of a Johnstone triangle.

CFNOS, sulfinyl cyanide fluoride, as it has been named in the original publication, should be more suggestively written as FS(O)CN in order to indicate the role of the sulfur atom as a central atom. Sulfur(IV) derivatives bearing CN substituents are interesting because of the possible interaction between the lone pair at sulfur and the π* orbital of the CN group. This would allow for a mesomeric contribution ($\mathrm{S{-}C{\equiv}N} \leftrightarrow {}^{+}\mathrm{S{=}C{=}N}^{-}$), which could be

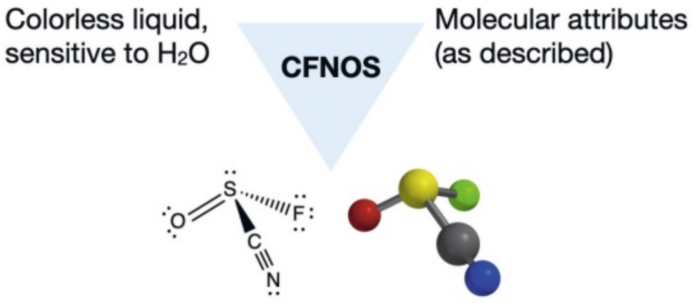

Fig. 4. The Johnstone triangle for the molecule CFNOS. Structure calculated by DFT ωB97X-D / 6-311+G** (Spartan'20).

evaluated by analyzing the bond lengths in the SCN unit. Therefore, this compound has been originally prepared exactly with the purpose of investigating this aspect by the use of spectroscopic methods. Moreover, CFNOS is an example of a compound for which the molecular and symbolic levels of thought have clearly preceded and inspired the macroscopic description. It is also a very well-suited example to be discussed in the general frame of the VSEPR model.

5. Terminology and Some Fundamental Concepts

Before tackling some of the most important aspects of the way of thinking of chemistry, as related in a very broad sense to its language, I will briefly argue about the relatively undemanding problem of naming elements existing as diatomic molecules. Consider the following reaction:

$$2\,H_2 + O_2 \rightarrow 2\,H_2O$$

A very simple verbal description of this reaction could be "hydrogen and oxygen react together forming water". Zealous teachers may possibly argue in favour of saying "*di*hydrogen and *di*oxygen..." instead, in order for learners to explicitly take into account the stoichiometric coefficients. For consistency and coherence one would therefore have to say for example tetraphosphorus for P_4, octasulfur for S_8, and hexacontacarbon for fullerene, C_{60}. Such more precise designations may be quite legitimate and are certainly not wrong. However, they are not strictly necessary and do not represent any rigorosity bonus, rather a risk of pedantry. The word "hydrogen" (or "oxygen", *etc*.) is intended to convey the information that element H exists in molecular form as H_2 under a broad range of conditions. Theoretical knowledge and knowledge about the substance hydrogen, very much in the sense of the German word *Stoffkenntnis*, is supposed to convey this information. In other words, talking about hydrogen is an exemplary opportunity for raising awareness about the three levels of thought as illustrated by the Johnstone triangle (Fig. 5). Not saying "dihydrogen" when talking about H_2 is not necessarily a sign of fuzzy language.

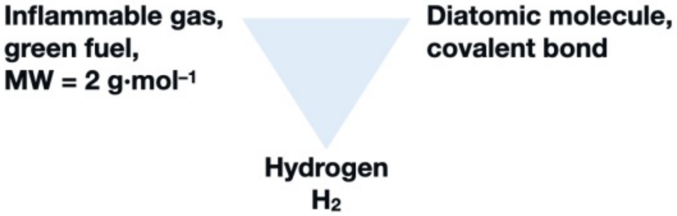

Fig. 5. Simplified Johnstone triangle for hydrogen.

Finally, the word "dihydrogen" is quite specifically used to portray H_2 acting as a side-on bonded ligand in transition-metal complexes, since their discovery in 1984.[10] Thus, the term can be viewed as being "occupied" as the designation of that specific form of hydrogen. Coordination chemistry uses by analogy the words "dioxygen" and "dinitrogen" to describe corresponding complexes (though not only for the side-on arrangement).

Let us turn our attention to the concept of *pure substance*, which is very central to chemistry and the way we speak chemistry.[11] A pure substance is usually defined according to a molecular approach. Somewhat simplistically, we say and teach that a pure substance is made up of a single type of molecules (or of atoms in the case of pure elements, or of anions and cations in the case of salts). When we consider water we would say, for example, that 18.0153 g of water (1 mole) is made up by or contain $6.022\cdot10^{23}$ H_2O molecules. However, the common essential material water contains not only H_2O molecules but also HDO, D_2O, $H_2{}^{17}O$, $HD^{17}O$, $D_2{}^{17}O$, $H_2{}^{18}O$, $HD^{18}O$, and $D_2{}^{18}O$ molecules because of the natural isotopic composition of the elements hydrogen and oxygen. Additionally, self-ionization generates H_3O^+, OH^-, and the corresponding isotopomers in equilibrium. Even an isotopically pure sample of water, whatever this is meant to be, is never going to be composed of H_2O molecules exclusively because self-ionization cannot be suppressed. It is furthermore important to realize that hydrogen bridges, proton-transfer processes,[12] and self-ionization obliterate the identity of an individual H_2O molecule in a sample of water on a very short time-scale. The apparently simple problem of knowing what we mean when we say "water" has even led to a philosophical debate![13] If we consider a sample of an isolated purified enzyme, as an example of a material constituted by much more complex molecules than H_2O, then it becomes very clear that the only-one-type-of-molecule criterion is not suited to define a pure substance – think of the innumerable isotopic combinations. Moreover, a single molecule cannot really represent a pure substance. So what is pure water and by extension a pure substance, given that a pure substance is not a natural category? In other words, the material world cannot be described and understood only in terms of pure substances. However, it can be *experimentally* subdivided into pure substances according to an *operationalized definition of purity*. For preparative chemists it is common and obvious to carry out purification operations on a product obtained by synthesis. Such operations usually involve *e.g.* phase transformations – distillation, sublimation, crystallization – or chromatography. A substance is then pure if further purification operations do not, or no longer alter the purity as ascertained by specific analytical criteria. Such criteria can be as simple as a reproducible sharp melting point, or otherwise the 'absence' of impurities as detected by NMR spectroscopy, a maximum deviation of 0.4% by element content in a conventional elemental analysis, or the presence of contaminants only below a certain threshold, often in the ppm range.

Despite the fact that 100% purity simply does not exist in reality, the artificial, idealized concept of the pure substance fulfills a fundamental role with regard to a systematic description of chemistry.[14]

When carrying out a reaction and formulating a corresponding equation, chemists implicitly make use of the concept of pure substances. Take the generic reaction shown in Fig. 6 as an example.

Pure starting materials A and B are brought together under certain specified conditions such that product C is formed. The equation relates pure substances – reactants and product(s) – to each other. Note that, as it is the case in most representations in synthetic chemistry, the equation does not necessarily need to be a balanced (stoichiometric) one. Think of, for example, reagents that are needed in excess, the excess being quenched upon workup, or a catalyst. These are not explicitly part of the equation but are taken into account by the conditions often given above and below the reaction arrow. Pure substances A, B, and C can obviously take part also in other reactions and each single reaction can be viewed as the simplest possible constituting unit of a *chemical network*. The crucial point is now that such a network actually represents the logical structure of systematic chemical knowledge.

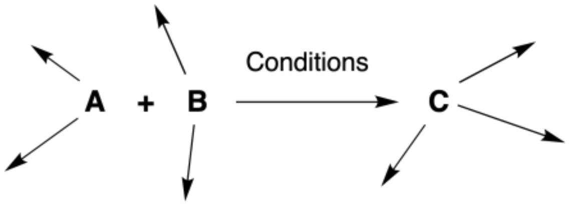

Fig. 6. A generic chemical reaction as the simplest form of a chemical network.

Chemical knowledge relies heavily on the *classification of pure substances* based on similarities in chemical behaviour and properties.[15] Hence, two substances are chemically similar if, under the same conditions, they react to form products also displaying chemical similarities. Think, for example, of the oxidation of a secondary alcohol to a ketone. Thus, alcohols and ketones form substance classes related to each other by the reaction conditions needed to transform one into the other. It is clear that the network relationship connecting a specific alcohol to the resultant ketone is valid for the substance classes as well.

A pure substance and the corresponding substance class will display a certain typical reactivity as conveyed by *functional groups*. Fig. 7 shows the previously mentioned molecule **1** and highlights its functional groups together with examples of their typical reactivities to be interpreted as network connections.

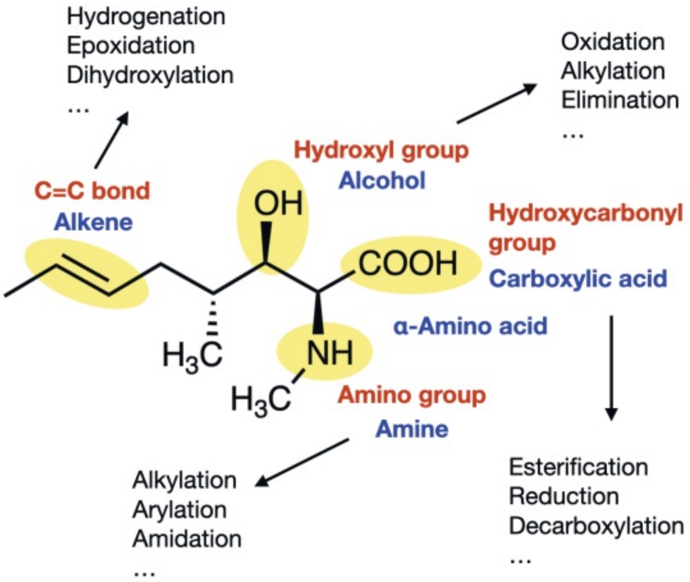

Fig. 7. Functional groups (red), substance classes (blue) and reactivities, as exemplified by compound **1**.

Being able to recognize and name functional groups and to assign their typical reactivities becomes fundamentally important in view of a systematic understanding of (molecular) chemistry. Examples of just a few typical functional groups together with their characteristic reactivity should be taught as early in a curriculum as possible. This would not require to deal necessarily with systematic nomenclature and would help setting the stage for a more quintessential understanding of chemistry and its language at large.

A structural formula unequivocally identifies a pure substance and conveys its position within the network of known, but also of *still unknown* substances. In other words, the attribution of a compound to a certain substance class makes it possible to *predict* which new compounds can be accessed by exploiting known functional-group transformations. In the eminently important context of *predictive power*, as entailed by structural formulas, I refer by way of a quintessential summary to a very appropriate and profound quote by Joachim Schummer:[14]

The chemical sign language is actually one of the most predictive theories of science at all !

6. Notes on Retrosynthesis and Reaction Mechanisms

Predicting the formation of a product by a targeted functional-group transformation is at the core of organic synthesis. Planning the total synthesis of a structurally complex product occurs by the use of a retrosynthetic analysis.[16] By way of example, Fig. 8 shows the most important steps in the retrosynthetic analysis that led to the first solid-phase total synthesis of epothilone A – a natural product with strong cytotoxicity against tumour cells – based on a crucial olefin metathesis step, as realized by K. C. Nicolaou and coworkers.[17] A *retro*synthetic analysis dissects the targeted product stepwise, thereby identifying sensible specific transformations that can be carried out in the opposite, *i.e.* the *forward* direction. The goal is efficiency, *i.e.* the minimization of the total number of steps and the use of readily accessible starting materials. From a practical point of view it is therefore an expression of the envisaged synthetic strategy.

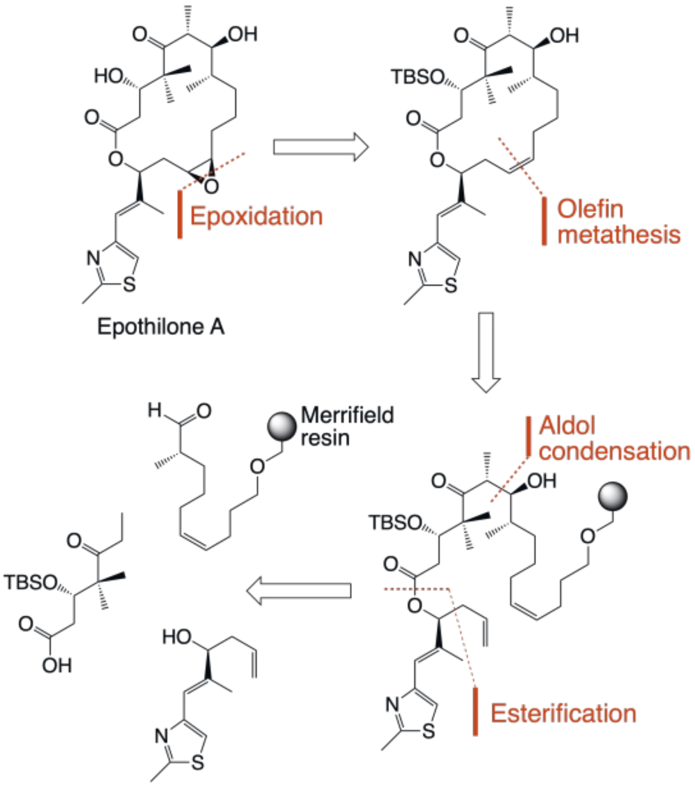

Fig. 8. Simplified retrosynthetic analysis for the solid-phase synthesis of epothilone A.

In the representation of a retrosynthetic scheme the final product of the synthesis, *i.e.* epothilone A in our example, appears as if it were to take up the role of the starting material. However, this does not mean that each step in the *retro* direction needs to be realized in the laboratory. Correspondingly, reaction conditions for such fictitious steps do not need to be specified. Thus, removing the epoxide oxygen atom from epothilone A, thereby restoring the alkene functionality, for example, is not a step that must be carried out, though this would be in principle possible. Only the *forward* direction, *i.e.* the epoxidation of the alkene, must be executed.

These are the most significant aspects of a retrosynthetic analysis, which constitutes one of the most characteristic and profound expressions of the language of formulas and their predictive power and can be considered as a most important theoretical instrument of chemistry.

Structural formulas are not only used to express the identity of a pure substance, as discussed, they also fulfill an important role in the description and interpretation of reaction mechanisms. A reaction mechanism is an explanation of how and why a certain transformation may take place in terms of forming new chemical bonds and cleaving others, *i.e.* on an atomistic level. Thus, the actual elucidation of a reaction mechanism corresponds to inferring events on a molecular level from a combination of experimental kinetic rate laws, the isolation and characterization of intermediates, isotope labelling studies, as well as from quantum-chemical calculations, just to name a few of the methods and techniques typically used. 'Knowing' the mechanism of a specific reaction means having done sufficient work to be able to exclude possible alternative mechanisms and to make it as plausible as possible. In cases where this work has not been done, a possible or putative mechanism can nevertheless be formulated, as we often see in the literature. A postulated mechanism is a simplified version based on analogies and reasonable assumptions of what the actual mechanism could possibly look like.

In the context of the language of chemistry, we speak about and represent mechanisms according to established qualitative models and conventions deriving from typical categories such as nucleophilic/electrophilic, acidic/basic, electron-donating/-withdrawing, polar/unpolar *etc.* and from principles of frontier-orbitals interactions. Thus, a description of a mechanism is implicitly loaded with theoretical considerations that are typical for chemistry. Nonetheless, the most popular representation of a reaction mechanism relies just on *pushing electron (pairs)* by using *curved arrows*.[18] In this sense, it reduces a potentially highly complex event at the molecular level to a highly symbolic and simplified illustration.

Fig. 9 depicts this kind of mechanistic description for the case of the crucial polyene cyclization step in the classical Johnson's biomimetic synthesis of progesterone which can be accessed in racemic form in two steps from the product shown.[19] I have chosen this particular example, though admittedly it is not necessarily the simplest one, for two reasons. Firstly, I vividly remember the excitement of the late Duilio Arigoni telling about this reaction mechanism in an advanced organic chemistry course in 1978, at the time the work was published and when I was a third-year student at ETH, and secondly because of the relatively high number of curved arrows. These account for the formation of the three new crucial C–C bonds and concomitantly three new rings in a single synthetic step, starting from a linear-chain unsaturated starting material. One is tempted to say that this is an elegant mechanism. However, the difficulties students may have in understanding this kind of mechanistic representations should not be underestimated because of the high level of abstraction they imply.

Pushing electrons is also used when formulating resonance structures. This is very useful when offering an explanation why, for example, 4-nitrophenol is more acidic than phenol itself. The better delocalization of the negative charge of the 4-nitrophenolate anion accounts for its stabilization and hence for the observed higher acidity of the protonated form.

Finally, when representing, discussing and interpreting mechanisms, we often also draw, hence visualize (molecular) orbitals. They may help understanding why a certain transformation occurs the way it does. Knowledge about orbitals obviously stems from MO-theory and from modern quantum-chemical calculations.

Fig. 10 shows for example the crucial single-electron-transfer (SET) process occurring in the Ti-catalyzed electrophilic fluorination of β-ketoesters.[20] Such a process involves the HOMO of the corresponding enolate and the LUMO of the N-F reagent (*e.g.* Selectfluor), which corresponds to the vacant σ* orbital of the N–F bond. Thus, the SET leads to a weakening and elongation of the N–F bond and the following radical recombination will involve the enolate C atom with the largest coefficient in the HOMO. Representing and talking about vacant orbitals is actually quite a peculiar thing to do, but we usually take them for granted as a 'normally' existing part of a molecule.

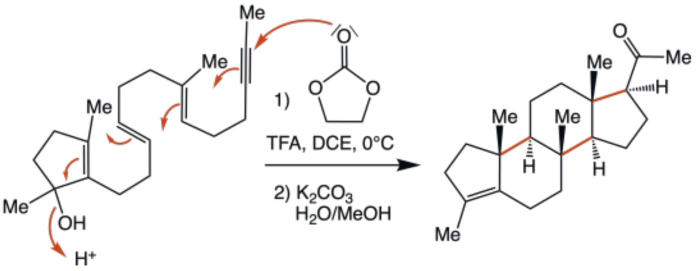

Fig. 9. Pushing electron pairs using curved arrows as an explanation of the polyene cyclization step in Johnson's biomimetic progesterone synthesis.

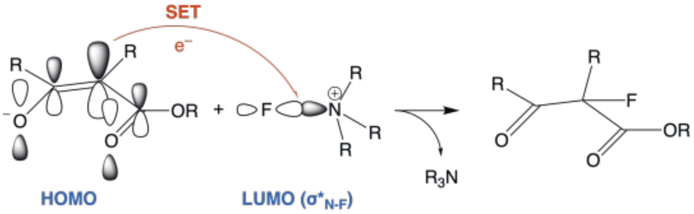

Fig. 10. Frontier orbitals involved in the SET process in the Ti-catalyzed fluorination of β-ketoesters.

7. The Chemical Transliteration of Three Platonic Solids

Creating new molecular objects is chemistry's most prominent and unique form of expression, beyond language.[21] In my view, one of the most significant and impressive examples of this unique capacity of chemistry is the so-called *chemical transliteration*[22] of the Platonic solids.

The three Platonic solids tetrahedron, cube and dodecahedron have in common that three edges meet at each vertex. If we imagine the three edges to be three equivalent chemical bonds involving a methine group (CH), the vertex, then drawing the three Platonic solids corresponds to drawing the structural formulas of three polycyclic alkanes: tetrahedrane (C_4H_4), cubane (C_8H_8) and dodecahedrane ($C_{20}H_{20}$), shown in Fig.11. In the case of the two remaining Platonic solids, octahedron and icosahedron, this is not possible because of four and five edges, respectively, meeting at one vertex. I can imagine that possible syntheses of these compounds must have enormously fascinated those who thought about and eventually realized them, for a variety of reasons. Certainly the high symmetry and intrinsic beauty of these convex polyhedra and the association with objects with a very long history and symbolic value are intuitively very understandable. In chemical terms, all three were not surprisingly unknown compounds, each and everyone with its specific synthetic challenges. Ring strain and therefore thermodynamic instability were very significant parameters, but difficult to predict for tetrahedrane and cubane and closing a cage-like molecule such as dodecahedrane would not be easy at all.

Without elaborating the synthetic details of any of the three compounds, it is important to ask what chemistry has gained, al-

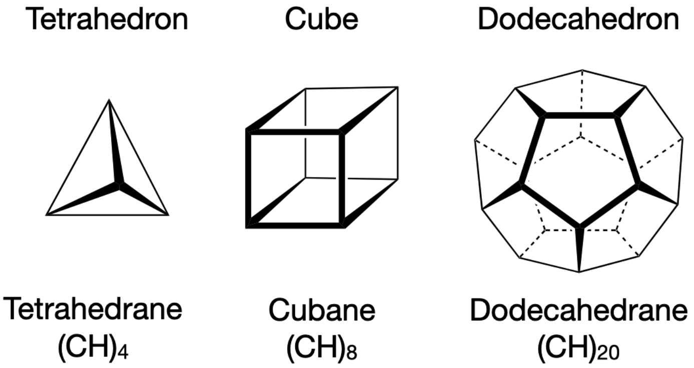

Fig. 11. The three Platonic solids that can be chemically transliterated to corresponding polycyclic alkanes.

so with respect to a possible educational value, from the effort to make such purportedly 'useless' molecules. As I will briefly argue, tetrahedrane is probably the most important in terms of learnings of general and far-reaching validity. Unsubstituted tetrahedrane is still unknown and probably will remain so. The main reason is its exorbitant ring strain lying in the calculated range 527–586 kJ/mol, a value that clearly exceeds the average bond dissociation energy of a C–C single bond.[23] However, thermodynamic instability does not necessarily mean that substituted tetrahedranes cannot exist. Indeed, the first derivative which was successfully prepared and reported by Günther Maier and coworkers in 1978 is tetra-*tert*-butyltetrahedrane[24] (see Fig. 12). This compound is astonishingly stable up to 135 °C, despite its ring strain of 541 kJ/mol, corresponding to 90.2 kJ/mol per C–C bond, the highest value ever determined. The key factor is here the kinetic stabilization due to the bulky *t*-butyl groups. The intramolecular repulsion of these groups forces the tetrahedral arrangement of the C_4 core of the molecule by what has been called a corset effect. Kinetic stabilization by sterics is a generally valid principle in chemistry, exemplarily demonstrated by the tetrahedranes. However, depending on the specific case, not just any bulky group is suited to achieve the desired stabilization. Symmetry, shape and spacial extent of the stabilizing groups need to be judiciously taken into account. The known tetrahedranes – the *t*-butyl group can be formally replaced by $SiMe_3$[25] – undergo a thermal or photochemical rearrangement to the corresponding cyclobutadienes, prototypical antiaromatic derivatives, hence compounds of fundamental theoretical significance.

Cubane has quite a long history and a rich derivative chemistry involving partial and full substitution patterns. Indeed, the first synthesis has been disclosed already in 1964 by Eaton and Cole[26] and octafluorocubane (C_8F_8), for example, was reported only very recently and found to readily form a stable radical anion.[27] Fig. 13 shows the structure of these two remarkable compounds. It is noteworthy that cubane dicarboxylic acid can be prepared in five steps from cyclopentenone in *ca*. 25% overall yield in multi-kg quantities. Syntheses of cubane typically involve photochemical [2+2] cycloadditions and Favorskii rearrangements.

As opposed to tetrahedrane, cubane is stable in its unsubstituted form and only slowly decomposes above 220 °C despite its ring strain of 676 kJ/mol (56.3 kJ/mol per C–C bond). Thus, the very cubic symmetry takes care of the corset effect, so to speak. Therefore, it is not hard to imagine that trying to pull apart any two carbon atoms along an edge of the cube will be opposed by the rest of the structure, thus keeping them together.

Finally, dodecahedrane resulted from a 21-step synthesis from cyclopentadiene, as realized by Paquette and coworkers in 1982.[28] As shown in Fig. 14, the second step of the synthesis, a domino Diels-Alder reaction, generates an intermediate already containing four of the twelve 5-membered rings and fourteen of the twenty carbon atoms. The following two main phases of the synthesis consisted in completing the number of carbon atoms and closing the molecule by removing functional groups, respectively.

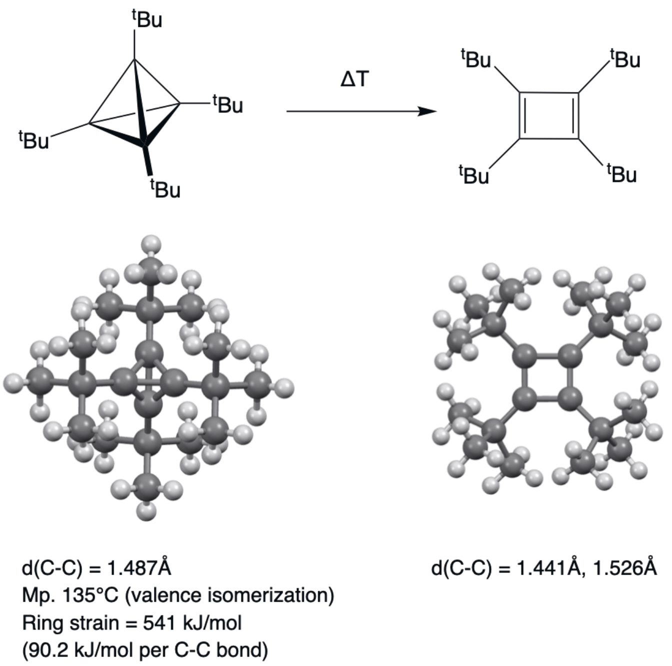

Fig. 12. Crystal structures of tetra-*tert*-butyltetrahedrane (CCDC 1132313) and its valence isomer tetra-*tert*-butylcyclobutadiene (CCDC 1267649).

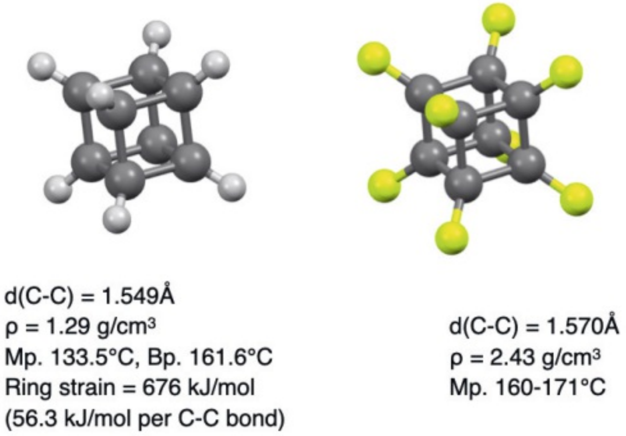

Fig. 13. Crystal structures of cubane (CCDC 1132082) and octafluorocubane (CCDC 2144167).

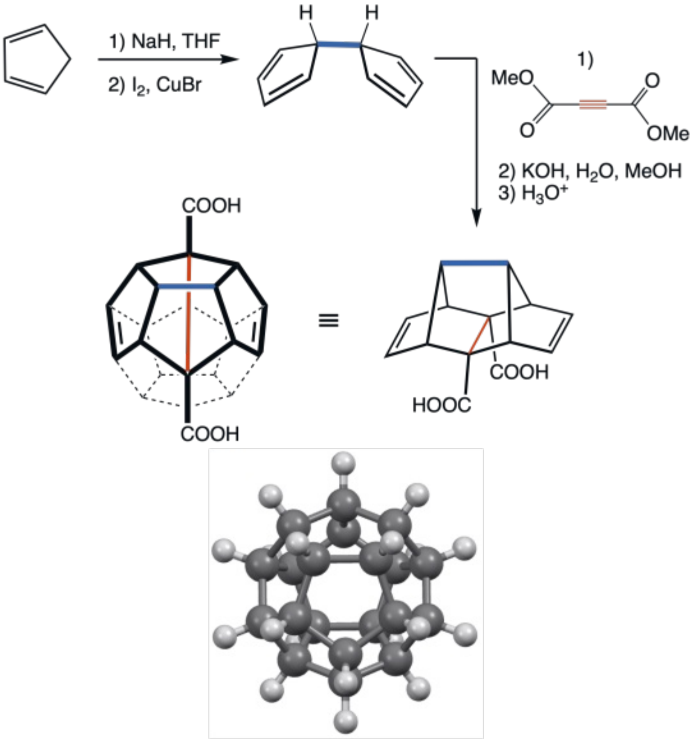

Fig. 14. First steps of the synthesis of dodecahedrane and its crystal structure (CCDC 104163).

8. Conclusion

I hope I was able to convince the readers, from those approaching chemistry as interested students in high schools to trained specialists like myself, that indeed chemistry has its own unique, diverse and multifaceted language, characterized in particular by structural formulas. These are symbolism, model, and coding at the same time and allow for immediate understanding, as opposed to verbal and written descriptions in regular language, including systematic names, of the same chemical objects. However, in order to become proficient in the language of formulas it needs a lot of practice, including drawing by hand, even if this may seem somehow outdated!

Nowadays, an article about the language of chemistry would not be complete without some concluding remarks with respect to language models and a possible corresponding digital revolution. Though I did not do the 'experiment', I doubt that ChatGPT could be a competent interlocutor in a discussion like the one exemplified at the beginning about ethyl isocyanoacetate, involving the three levels of thought symbolized by the Johnstone triangle. In a recent study,[29] Castro Nascimento and Pimentel come to the conclusion that the ability of ChatGPT to understand chemistry is not (yet) as reliable as one would hope. On the other hand, efforts by *e.g.* Laino, Schwaller and coworkers[30] seem to lead the way towards an effective digitalization of chemistry by the use of large language models.

In any case, chemistry can only be learned and appreciated as a cultural achievement through a very conscious and reflected approach to its language.

Received: August 17, 2023

---

[1] In Switzerland, the *Lehrdiplom* is the qualification necessary for teaching at a high school. It can be obtained by students having a MSc in chemistry and requires 60 ECTS. I have been teaching parts of one of the mandatory course called *Vertiefte Grundlagen der Chemie* (In-depth fundamentals of chemistry), in which one of the topics is the language of chemistry.

[2] See *e.g.*: a) A. Togni, S. D. Pastor, *J. Org. Chem.* **1990**, *55*, 1649, https://doi.org/10.1021/jo00292a046; b) H. Allgeier, C. Angst, G. Bold, R. Duthaler, R. Heckendorn, A. Togni, 'Preparation of phosphino- and phosphono-pentenoates as NMDA antagonists', EP 302826 A2 (Feb. **1989**), CH pat. appl. 1987.

[3] For the original publication, see: a) A. H. Johnstone, *School Science Review* **1982**, *64*, 377. See also: b) A. H. Johnstone, *J. Chem. Educ.* **1993**, *70*, 701, https://doi.org/10.1021/ed070p701; c) N. Reid, 'The Johnstone Triangle: The Key to Understanding Chemistry', Advances in Chemistry Education Series, RSC, **2021**, https://doi.org/10.1039/9781839163661; d) S. Rees, V. Kind, D. Newton, *Isr. J. Chem.* **2019**, *59*, 470, https://doi.org/10.1002/ijch.201800079; e) G. P. Thomas, *Chem. Educ. Res. Pract.* **2017**, *18*, 533, https://doi.org/10.1039/C6RP00227G; f) V. Talanquer, *Int. J. Sci. Educ.* **2011**, *33*, 179, https://doi.org/10.1080/09500690903386435; g) K. S. Taber, 'Foundations for Teaching Chemistry', Routledge, London, **2020**; h) 'Multiple Representations in Chemical Education', Eds. J. K. Gilbert, D. Treagust, Springer, **2009**.

[4] a) Compound **1** see: A. Togni, S. D. Pastor, G. Rihs, *Helv. Chim. Acta* **1989**, *72*, 1471, https://doi.org/10.1002/hlca.19890720707; b) compound **2** see: D. Abegg, M. Tomanik, N. Qiu, D. Pechalrieu, A. Shuster, B. Commare, A. Togni, S. B. Herzon, A. Adibekian, *J. Am. Chem. Soc.* **2021**, *143*, 20332, https://doi.org/10.1021/jacs.1c09724; c) compound **3** see: S. Milosevic, A. Togni, *J. Org. Chem.* **2013**, *78*, 9638, https://doi.org/10.1021/jo401247d.

[5] This version is taken from the Langdon translation: https://oll.libertyfund.org/title/langdon-the-divine-comedy-vol-1-inferno-english-trans. The original in Italian goes back to the 14th century and reads:
Nel mezzo del cammin di nostra vita
mi ritrovai per una selva oscura,
ché la diritta via era smarrita.
Ahi quanto a dir qual era è cosa dura
esta selva selvaggia e aspra e forte
che nel pensier rinova la paura!

[6] For a review article, see: R. Hoffmann, P. Laszlo, 'Representation in Chemistry', *Angew. Chem. Int. Ed.* **1991**, *30*, 1, https://doi.org/10.1002/anie.199100013.

[7] P. Laszlo, 'La parole des choses ou le langage de la chimie', Collection Savoir: Sciences, Hermann, Paris, **1993**.

[8] P. Laszlo, 'Towards Teaching Chemistry as a Language', *Sci. & Educ.* **2013**, *22*, 1669, https://doi.org/10.1007/s11191-011-9408-6.

[9] Synthesis and characterization of FS(O)CN: a) J. Jacobs, H. Willner, *Z. Anorg. Allg. Chem.* **1993**, *619*, 1221, https://doi.org/10.1002/zaac.19936190712. Structure and spectroscopy: b) H.-G. Mack, H. Oberhammer, J. Jacobs, M. Kronberg, H. Willner, *Inorg. Chem.* **1996**, *35*, 806, https://doi.org/10.1021/ic9511940.

[10] G. J. Kubas, R. R. Ryan, B. I. Swanson, P. J. Vergamini, H. J. Wasserman, *J. Am. Chem. Soc.* **1984**, *106*, 451, https://pubs.acs.org/doi/pdf/10.1021/ja00314a049. See also: G. J. Kubas, *Acc. Chem. Res.* **1988**, *21*, 120, https://pubs.acs.org/doi/pdf/10.1021/ar00147a005.

[11] For philosophical perspectives, see *e.g.*: a) 'Stuff – The Nature of Chemical Substances', Eds. K. Ruthenberg, J. van Brakel, Königshausen &Neumann, Würzburg, **2008**; b) M. Fernández-González, *Sci. & Educ.* **2013**, *22*, 1723, https://doi.org/10.1007/s11191-011-9428-2.

[12] See *e.g.*: C. Knight, G. A. Voth, *Acc. Chem. Res.* **2012**, *45*, 101, https://pubs.acs.org/doi/10.1021/ar200140h.

[13] For the view of philosophy of science, see: a) M. Weisberg, 'Water is *NOT* H_2O', in 'Philosophy of Chemistry', Eds. D. Baird, E. Scerri, L. McIntyre, Springer, Dordrecht, **2006**, pp. 337. See also: b) P. Needham, 'The Discovery that water is H_2O', *Int. Stud. Phil. Sci.* **2002**, *16*, 205, https://doi.org/10.1080/0269859022000013300; c) P. Needham, 'What is Water?', *Analysis* **2000**, *60*, 13, https://doi.org/10.1093/analys/60.1.13; d) H. Chang, 'Is Water H_2O ? Evidence, Realism and Pluralism', Boston Studies in the Philosophy of Science 293, **2012**.

[14] J. Schummer, 'The Chemical Core of Chemistry I: A Conceptual Approach', *HYLE* **1998**, *4*, 129 http://www.hyle.org/journal/issues/4/schumm.pdf.

[15] For a philosophical perspective, see *e.g.*: J. R. Bursten, 'Microstructure without Essentialism: A New Perspective on Chemical Classification', *Philosophy of Science* **2014**, *81*, 633, https://doi.org/10.1086/678043.

[16] For a general discussion of retrosynthesis, see: E. J. Corey, X.-M- Cheng, 'The Logic of Chemical Synthesis', Wiley, **1995**, (6th print.).

[17] K. C. Nicolaou, N. Winssinger, J. Pastor, S. Ninkovic, F. Sarabia, Y. He, D. Vourloumis, Z. Yang, T. Li, P. Giannakakou, E. Hamel, *Nature* **1997**, *387*, 268, https://doi.org/10.1038/387268a0.

[18] See *e.g.:* a) D. P. Weeks, 'Pushing Electrons. A Guide for Students of Organic Chemistry', Brooks/Cole, Belmont, **1998**; b) A. Ghosh, S. Berg, 'Arrow Pushing in Inorganic Chemistry', Wiley, **2014**; c) D. Klein, 'Organic chemistry as a second language', 5th ed., Wiley, **2020**.

[19] M. B. Gravestock, W. S. Johnson, B. E. McCarry, R. J. Parry, B. E. Ratcliffe, *J. Am. Chem. Soc.* **1978**, *100*, 4274, https://doi.org/10.1021/ja00481a044.

[20] S. Piana, I. Devillers, A. Togni, U. Röthlisberger, *Angew. Chem. Int. Ed.* **2002**, *41*, 979, https://doi.org/10.1002/1521-3773(20020315)41:6<979::AID-ANIE979>3.0.CO;2-E.

[21] Remember the famous quote by Marcellin Berthelot "*La chimie crée son objet"* (chemistry creates its objects): M. Berthelot, 'La Synthèse chimique', 2ème Ed., Librairie Germer Baillière, Paris, **1876**, p. 275.

[22] This term has been used for the first time by Paquette, see ref. [28d].

[23] G. Maier, *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 309, https://doi.org/10.1002/anie.198803093.

[24] Synthesis: a) G. Maier, S. Pfriem, U. Schäfer, R. Matusch, *Angew. Chem. Int. Ed. Engl.* **1978**, *17*, 520, https://doi.org/10.1002/anie.197805191. Structure: b) H.Irngartinger, A. Goldmann, R. Jahn, M. Nixdorf, H. Rodewald, G. Maier, K.-D. Malsch, R. Emrich, *Angew. Chem. Int. Ed.* **1984**, *23*, 993, https://doi.org/10.1002/anie.198409931.

[25] G. Maier, J. Neudert, O. Wolf, D. Pappusch, A. Sekiguchi, M. Tanaka, T. Matsuo, *J. Am. Chem. Soc.* **2002**, *124*, 13819, https://doi.org/10.1021/ja020863n. For a tris(trimethylsilyl)pentafluorophenyltetrahedrane, see: M. Nakamoto, Y. Inagaki, M. Nishina, A. Sakeguchi, *J. Am. Chem. Soc.* **2009**, *131*, 3172, https://doi.org/10.1021/ja810055w.

[26] Syntheses: a) P. E. Eaton, T. W. Cole, Jr., *J. Am. Chem. Soc.* **1964**, *86*, 3157, https://doi.org/10.1021/ja01069a041; b) P. E. Eaton, T. W. Cole, Jr., *J. Am. Chem. Soc.* **1964**, *86*, 962, https://doi.org/10.1021/ja01059a072; c) J. C. Barborak, L. Watts, R. Pettit, *J. Am. Chem. Soc.* **1966**, *88*, 1328, https://doi.org/10.1021/ja00958a050; d) P. E. Eaton, N. Nordari, J. Tsanaktsidis, S. P. Upadhyaya, *Synthesis* **1995**, 501, https://doi.org/10.1055/s-1995-3961. Review articles: e) P. E. Eaton, *Angew. Chem. Int. Ed.* **1992**, *31*, 1421, https://doi.org/10.1002/anie.199214211; f) G. W. Grifin, A. P. Marchand, *Chem. Rev.* **1989**, *89*, 997, https://doi.org/10.1021/cr00095a003; g) K. F. Biegasiewicz, J. R. Griffiths, G. P. Savage, J. Tsanaktsidis, R. Priefer, *Chem. Rev.* **2015**, *115*, 6719, https://doi.org/10.1021/cr500523x.

[27] M. Sugiyama, M. Akiyama, Y. Yonezawa, K. Komaguchi, M. Higashi, K. Nozaki, T. Okazoe, *Science* **2022**, *377*, 756, https://doi.org/10.1126/science.abq0516.

[28] Synthesis: a) R. J. Ternansky, D. W. Balogh, L. A. Paquette, *J. Am. Chem. Soc.* **1982**, *104*, 4503, https://doi.org/10.1021/ja00380a040. For a conclusive report, see: b) L. A. Paquette, J. C. Weber, T. Kobayashi, Y. Miyahara, *J. Am. Chem. Soc.* **1988**, *110*, 8591, https://doi.org/10.1021/ja00234a003. For structural aspects, see: c) J. C. Gallucci, R. T. Taylor, T. Kobayashi,

J. C. Weber, J. Krause, L. A. Paquette, *Acta Cryst.* **1989**, *C45*, 893, https://doi.org/10.1107/S0108270188013721. See also: d) L. A. Paquette, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 4495, https://doi.org/10.1073/pnas.79.14.4495.

[29] C. M. Castro Nascimento, A. S. Pimentel, 'Do Large Language Models Understand Chemistry? A Conversation with ChatGPT', *J. Chem. Inf. Model.* **2023**, *63*, 1649, https://doi.org/10.1021/acs.jcim.3c00285.

[30] See: A. Cardinale, A. Castrogiovanni, T. Gaudin, J. Geluykens, T. Laino, M. Manica, D. Probst, P. Schwaller, A. Sobczyk, A. Toniato, A.C. Vaucher, H. Wolf, F. Zipoli, *CHIMIA* **2023**, *77*, 484, https://doi.org/10.2533/chimia.2023.484, and references cited therein.

License and Terms