OF MINDS AND MOLECULES

New Philosophical Perspectives on Chemistry

Edited by
Nalini Bhushan
Stuart Rosenfeld

OXFORD
UNIVERSITY PRESS
2000

13

How Symbolic and Iconic Languages Bridge the Two Worlds of the Chemist

A Case Study from Contemporary Bioorganic Chemistry

EMILY R. GROSHOLZ
ROALD HOFFMANN

Chemists move habitually and with credible success—if sometimes unreflectively—between two worlds. One is the laboratory, with its macroscopic powders, crystals, solutions, and intractable sludge, as well as the things that are smelly or odorless, toxic or beneficial, pure or impure, colored, or white. The other is the invisible world of molecules, each with its characteristic composition and structure, its internal dynamics and its ways of reacting with the other molecules around it. Perhaps because they are so used to it, chemists rarely explain how they are able to hold two seemingly disparate worlds together in thought and practice. And contemporary philosophy of science has had little to say about how chemists are able to pose and solve problems, and, in particular, to posit and construct molecules, while simultaneously entertaining two apparently incompatible strata of reality. Yet chemistry continues to generate highly reliable knowledge, and indeed to add to the furniture of the universe, with a registry of over ten million well-characterized new compounds.

The philosophy of science has long been dominated by logical positivism, and the assumptions attendant on its use of predicate logic to examine science, as well as its choice of physics as the archetype of a science. Positivism thus tends to think of science in terms of an axiomatized theory describing an already given reality and cast in a uniform symbolic language, the language of predicate logic. (See especially the locus classicus of this position, Carnap, 1937.)

We here wish to question certain positivist assumptions about scientific rationality, based on an alternative view brought into focus by the reflective examination of a case study drawn from contemporary chemistry. Our reflections owe something to Leibniz (1686, 1695, 1714), Husserl (1922), Kuhn (1970), and Polanyi (1960, 1966), and draw on the earlier writings of both of us—Hoffmann (1995; Hoffmann & Laszlo, 1991) and Grosholz (1991; Grosholz & Yakira, 1998). We will offer a nonreductionist account of methods of analysis and synthesis in chemistry. In our view, reality is allowed to include different kinds of things existing in different kinds of ways, levels held in intelligible relation by both theory and experiment, and couched in a multiplicity of languages, both symbolic and iconic.

We argue that there is no single correct analysis of the complex entities of chemistry expressed in a single adequate language, as various reductionist scripts require; and yet the multiplicity and multivocality of the sciences, and their complex "horizontal" interrelations, do not preclude but in many ways enhance their reasonableness and success. Nor is this view at odds with our realism; we want to distinguish ourselves quite strongly from philosophers engaged in the social construction of reality (see, e.g., Pickering, 1992; Shapin, 1992; Fuller, 1994; for a balanced analysis of the problem, see Labinger, 1995). We understand the reality whose independence we honor as requiring scientific methods that are not univocal and reductionist precisely because reality is multifarious, surprising, and infinitely rich.

Formulating the Problem

The article drawn from the current literature in chemistry that we shall consider is "A Calixarene with Four Peptide Loops: An Antibody Mimic for Recognition of Protein Surfaces," authored by Yoshitomo Hamuro, with Andrew Hamilton, Mercedes Crego Calama, Hyung Soon Park, and published in December 1997 in the international journal *Angewandte Chemie* (Hamuro et al., 1997. We will refer to this article as Hamuro et al.) The subfield of the article could be called bioorganic chemistry. One way to look at biology is to examine its underlying chemistry, in a well-developed program that is both one of the most successful intellectual achievements of the twentieth-century, and a locus of dispute for biologists. For many years, organic chemists had let molecular and biochemistry "get away" from chemistry; recently, there has been a definite movement to break down the imagined fences and reintegrate modern organic chemistry and biology. The article we examine is part of such an enterprise.

We have learned something about the structure of the large, enigmatic, selectively potent molecules of biology. But describing their structure and measuring their functions do not really answer the question of how or why these molecules act as they do. Here, organic chemistry can play an important role by constructing and studying molecules smaller than the biological ones, but which model or mimic the activities of the speedy molecular behemoths of the biological world.

The article opens by stating one such problem of mimicry, important to medical science and any person who has ever caught a cold. The human immune system has flexible molecules called antibodies, proteins of some complexity that recognize a wide variety of molecules including other proteins.

> The design of synthetic hosts that can recognize protein surfaces and disrupt biologically important protein–protein interactions remains a major unsolved problem in bioorganic chemistry. In contrast, the immune system offers numerous antibodies that show high sequence and structural selectivity in binding to a wide range of protein surfaces. (Hamuro et al., 1997, p. 2680)

The problem is thus to mimic the structure and action of an antibody; but antibodies in general are very large and complicated. Hamuro et al. ask the question, Can we assemble a molecule with some of the structural features of an antibody, simplified

and scaled down, and, if so, will it act like an antibody? But what are the essential structural features in this case?

Prior investigation has revealed that an antibody at the microscopic level is a protein molecule that typically has a common central region with six "hypervariable" loops that exploit the flexibility and versatility of the amino acids that make up the loops to recognize (on the molecular level) the near infinity of molecules that wander about a human body. The article remarks that "this diversity of recognition is even more remarkable, because all antibody fragment antigen binding (FAB) regions share a common structural motif of six hypervariable loops held in place by the closely packed constant and variable regions of the light and heavy chains" (p. 2680).

What is recognition at the microscopic level? It is generally not the strong covalent bonding that makes molecules so persistent, but is rather a congeries of weak interactions between molecules that may include bonding types that chemists call hydrogen bonding, van der Waals or dispersion forces, electrostatic interactions (concatenations of regions of opposite charge attracting or like charge repelling), and hydrophobic interactions (concatenations of like regions attracting, as oil with oil, water with water). These bonding types are the subject of much dispute, for they are not as distinct as scientists would like them to be. (For an introduction to chemistry and molecular interactions, see Joesten et al., 1991). In any case, the interactions between molecules are weak and manifold. Recognition occurs as binding, but it is essentially more dynamic than static. At body temperature, recognition is the outcome of many thermodynamically reversible interactions: the antibody can pick up a molecule, assess it, and then perhaps let it go. In the dance of holding on and letting go, some things are held on to more dearly.

Whatever happens has sufficient cause, in the geometry of the molecule, and in the physics of the microscopic attractions and repulsions between atoms or regions of a molecule. The article remarks,

> Four of these loops . . . generally take up a hairpin conformation and the remaining two form more extended loops. X-ray analyses of protein–antibody complexes show that strong binding is achieved by the formation of a large and open interfacial surface (>600 Å) composed primarily of residues that are capable of mutual hydrophobic, electrostatic, and hydrogen bonding interactions.* The majority of antibody complementary determining regions (CDRs) contact the antigen with four to six of the hypervariable loops.* (pp. 2680–81)[1]

The foregoing passage is a theory about the structure and function of antibodies, but it is asserted with confidence and in precise detail. Standing in the background, linking the world of the laboratory—where small (but still tangible) samples of antibodies and proteins are purified, analyzed, combined, and measured—and the world of molecules are theories, instrumentation, and languages. There is no shortage of theories here; indeed, we are faced with an overlapping, interpenetrating network of theories backed up by instrumentation. These include the quantum mechanics of the atom along with a multitude of quantum mechanically defined spectroscopies, chemistry's highly refined means for destructively or nondestructively plucking the strings of molecules and letting the "sounds" tell us about their features (Hoffmann & Torrence, 1993, pp. 144–147). Three are equally ingenious techniques for separating and purifying molecules, that we will loosely term chromatographies. They proceed at a larger scale and, when traced, are also the outcome of a sequence of holding on and letting go, like antibody recognition.

Further, statistical mechanics and thermodynamics serve to relate the microscopic to the macroscopic. These theories are probabilistic, but they have no exceptions because of the immensity of the number of molecules—10^{23} in a sip of water—and the rapidity of molecular motion at ambient temperatures. Thus, the average speed of molecules "scales up" to temperature, their puny interactions with light waves into color, the resistance of their crystals to being squeezed to hardness, their multitudinous and frequent collisions into a reaction that is over in a second or a millennium (Atkins, 1984, 1987, 1991; Joesten et al., 1991; Hoffmann, 1995).

These theories are silent partners in the experiments described in the article, taken for granted and embodied, one might say, in the instruments. But a further dimension of the linkage between the two worlds is the languages employed by the chemists, and that is what we now propose to examine at length.

Solving the Problem

The construction of "a calixarene with four peptide loops" serves two functions in this article. It serves as a simplified substitute for an antibody, though we doubt that the intent of the authors is the design of potential therapeutic agents. More important, the calixarene serves to test the theory of antibody function sketched in the preceding discussion: Is this really the way that antibodies work? The authors note that earlier attempts to mimic antibodies have been unsuccessful, and they propose the alternative strategy, which is the heart of the article: The search for antibody mimics has not yet yielded compact and robust frameworks that reproduce the essential features of the CDRs.* Our strategy is to use a macrocyclic scaffold to which multiple peptide loops in stable hairpin-turn conformations can be attached (p. 2681).

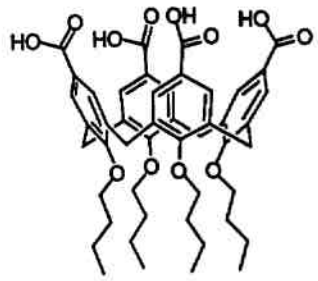
Stage 1a: The Core Scaffold

The experiment has two stages. The first is to build the antibody mimic, by adding peptide loops to the scaffolding of a calix[4]arene—a cone-shaped concatenation of four benzene rings, strengthened and locked into one orientation by the addition of small-length chains of carbon and hydrogen (an alkylation), with COOH groups on top to serve as "handles" for subsequent reaction. (The benzene ring of six carbons is a molecule with a venerable history, whose structure has proved especially problematic for the languages of chemistry, as we point out later in this chapter.) The authors write,

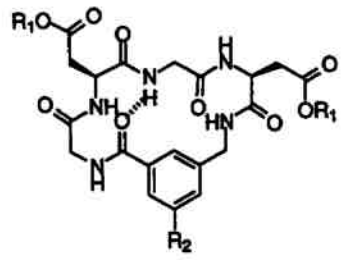
> In this paper we report the synthesis of an antibody mimic based on calix[4]arene linked to four constrained peptide loops. . . . Calix[4]arene was chosen as the core scaffold, as it is readily available* and can be locked into the semirigid cone conformation by alkylation of the phenol groups. This results in a projection of the para-substituents onto the same side of the ring to form a potential binding domain.* (p. 2681)

Diagram 1 is given to illustrate this description, as well as the following "recipe:" "The required tetracarboxylic acid 1 was prepared by alkylation of calix[4]arene* (*n*-butyl bromide, NaH) followed by formylation (Cl_2CHOCH_3, $TiCl_4$) and oxidation ($NaClO_2$, H_2NSO_3H)*" (p. 2681).

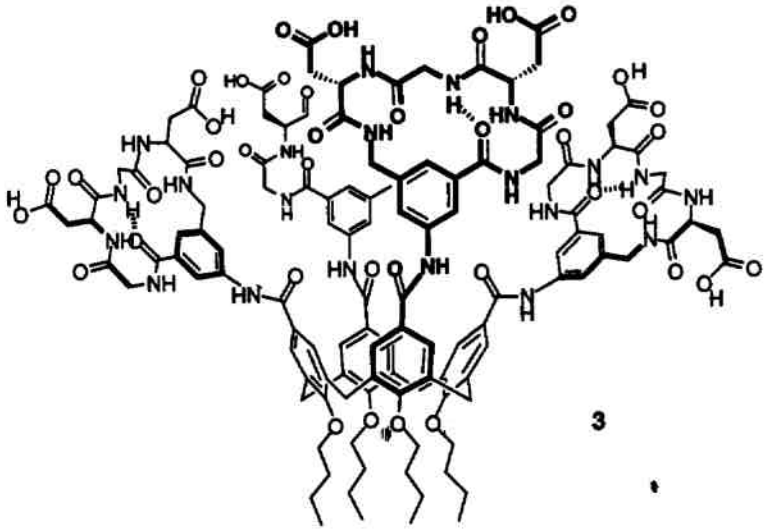
The iconic representation offered is of a microscopic molecule, but the language is all about macroscopic matter, and it is symbolic. The symbolic language of chemistry is the language of formulas employed in the laboratory recipe. It lends itself to the chemist's bridging of the macroscopic and the microscopic because it is thoroughly equivocal, at once a precise description of the ingredients of the experiment (e.g., *n*-butyl bromide is a colorless liquid, with a boiling point of 101.6°C, and is immiscible with water), and a description of the composition of the relevant molecules. For example, *n*-butyl bromide is construed by the chemist as $CH_3CH_2CH_2CH_2Br$; it has the formula C_4H_9Br, a determi-



1



2. $R_1 = Bu^t$, $R_2 = NH_2$,
4. $R_1 = H$, $R_2 = NO_2$,



3

nate mass relationship among the three atomic constituents, a preferred geometry, certain barriers to rotation around the carbon–carbon bonds it contains, certain angles at the carbons, and so forth (Atkins 1987; Joesten et al., 1991; Hoffmann, 1995).

The laboratory recipe is thus both the description of a process carried out by a scientist, and the description of a molecule under construction: a molecule generic in its significance, because the description is intended to apply to all similar molecules, but particular in its unity and reality. There are parallels in other fields of knowledge. Thus, in mathematics, the algebraic formula of a function applies equally to an infinite set of number pairs and to a geometric curve; its controlled and precise equivocity is the instrument that allows resources of number theory and of geometry to be combined in the service of problem solving (Grosholz, 1991, chaps. 1 and 2). Likewise, here the algebra of chemistry allows the wisdom of experience gained in the laboratory to be combined with the (classical and quantum) theory of the molecule, knowledge of its fine structure, energetics, and spectra.

But the symbolic language of chemistry is not complete, for there are many aspects of the chemical substance/molecule that it leaves unexpressed: (1) We cannot deduce from it how the molecule will react with the enormous variety of other molecules with which it may come in contact; (2) We cannot even deduce from it the internal statics, kinematics, and dynamics of the molecule in space.[2] Molecules identical in composition can differ from each other because they differ in constitution, the manner and sequence of bonding of atoms (tautomers), in spatial configuration (optical or geometrical isomers), and in conformation (conformers). (See Joesten et al., 1991; Zeidler & Hoffmann, 1995; and Sobczynska, 1995/6.) An adequate description of the molecule must invoke the background of an explanatory theory, but to do so it must also employ iconic languages. Thus, the very definition of the calixarene core scaffold involves a diagram. (It was also necessary for the authors to identify C_4H_9Br as *n*-butyl bromide, a nomenclature that implies a specific connectivity of atoms.)

This diagram of calixarene is worth careful inspection, as well as careful comparison with its counterparts in the more complex molecules (for which it serves as core scaffold) furnished to us, the readers, by means of computer-generated images. First, it leaves out most of the component hydrogens and carbons in the molecule; they are understood, a kind of tacit knowledge shared even by undergraduate chemistry majors. The hexagons are benzene rings, and the chemist knows that the valence of (the number of bonds formed by) carbon is typically four and so automatically supplies the missing hydrogens. But this omission points to an important feature of iconic languages: they must always leave something out, because they are only pictures, not the thing itself, and because the furnishing of too much information is actually an impoverishment. In a poor diagram, one cannot see the forest for the trees. Not only must some things remain tacit in diagrams, but also the wisdom of experience that lets the scientist know how much to put in and how much to leave out—wisdom gleaned by years of translating experimental results into diagrams for various kinds of audience—is itself often tacit. It can be articulated now and then but cannot be translated into a complete set of fixed rules.

Second, the diagram uses certain conventions for representing configurations in three-dimensional space on the two-dimensional page, like breaking the outlines of molecules which are supposed to be behind other molecules whose delineation is

unbroken. (In other diagrams, wedges are used to represent projection outward from the plane of the page, and heavy lines are used to represent molecules that stand in front of other molecules depicted by ordinary lines.) Sometimes, although not in the context of a journal article, chemists show three-dimensional representation, such as "ball-and-stick" models. But then, in addition, one may want to see more precise angles and interatomic distances in correct proportion and so resort to the images produced by X-ray crystallography. Or one may want some indication of the motion of the molecules, because all atoms vibrate and rotate. Arrows and other iconographies of dynamic motion are used in such diagrams. The cloudy, false-color, yet informative photographs of scanning tunneling microscopy come in here, as well as assorted computer images of the distribution of electrons in the molecule.[3]

Finally, the convention of a hexagon with a perimeter composed of three single lines alternating with three double lines to represent a benzene ring deserves a chapter in itself. This molecule has played a central role in the development of organic chemistry. No single classical valence structure was consistent with the stability of the molecule. Kekulé solved the problem by postulating the coexistence of two valence structures in one molecule. In time, practitioners of quantum mechanics took up the benzene problem, and to this day it has served them as an equally fecund source of inspiration and disagreement. The electrons in benzene are delocalized, that much people agree on; but the description of its electronic structure continues to be a problem for the languages of chemistry (Brush, 1998).

An Interlude on Symbolic and Iconic Languages

Philosophers of science working in the logical positivist tradition have had little to say about iconic languages. Symbolic languages typically lend themselves to logical regimentation, but pictures tend to be multiform and hard to codify; thus, if they proved to be indispensable to human knowledge, the logical positivist would be quite vexed. As any student of chemistry will tell you, conventions for producing "well-formed icons" of molecules exist and must be learned, or else your audience will misread them. But no single iconic language is the correct one or enjoys anything as precisely determined as a "wff" in logic. The symbolic language of chemistry is, to be sure, a precisely defined international nomenclature that specifies in impressive detail a written sequence of symbols so as to allow the unique specification of a molecule. But, significantly, the iconic representations of a molecule are governed only by widely accepted conventions, and a good bit of latitude is allowed in practice, especially a propos what may be omitted from such representations. Symbolic languages lend themselves to codification in a way that iconic languages don't.[4]

Symbolic languages, precisely because they are symbolic, lend themselves best to displaying relational structure. Like algebra, they are tolerant or relativistic in their ontological import: it doesn't so much matter what they pertain to, as long as their objects stand in the appropriate relations to each other. But iconic languages point, more or less directly, to objects; they are not ontologically neutral but, on the contrary, are ontologically insistent. They display the unity of objects, a unity that might metaphysically be called the unity of existence. But there is no way to give an exhaustive

summary of the ways of portraying the unity of existence; it is too infinitely rich, and thought has too many ways of engaging it. We should, therefore, not jump to the conclusion that knowledge via an iconic language is impossible or incoherent: iconic languages despite being multiform employ intelligible conventions, they are constrained by the object itself, and they are made orderly by their association with symbolic language. An inference cannot be constructed from icons alone, but icons may play an essential role in inference.[5]

How does the iconic form of the chemical structure expressed as a diagram that displays atom connectivities and suggests the three-dimensionality of the molecule, bridge the two worlds of the chemist? The most obvious answer is that it makes the invisible visible, and does so, within limits, reliably. But there is a deeper answer. It seems at first as if the chemical structure diagram refers only to the level of the microscopic since, after all, it depicts a molecule. But in conjunction with symbolic formulas, the diagram takes on an inherent ambiguity that gives it an important bridging function. In its display of unified existence, it stands for a single particular molecule. Yet we understand molecules of the same composition and structure to be equivalent to each other, internally indistinguishable. (In this, the objects of physics and chemistry are like the objects of mathematics.)

Thus, the icon (hexagonal benzene ring) also stands for all possible benzene rings, or for all the benzene rings (moles or millimoles of them!) in the experiment, depending on the way in which it is associated with the symbolic formula for benzene. The logical positivist in search of univocality might call this obfuscating ambiguity, a degeneracy in what ought to be a precise scientific language that carries with it undesirable ontological baggage. And yet, the iconic language is powerfully efficient and fertile in the hands of the chemist.

Now we can better understand why the kind of world-bridging involved in posing a problem/construction in chemistry requires both symbolic and iconic languages for its formulation. On the one hand, the symbolic language of chemistry captures the composition of molecules, but not their structure (constitution, configuration, and conformation), aspects that are dealt with better, though fragmentarily, by the many iconic idioms available to chemists. Moreover, the symbolic language of chemistry fails to convey the ontological import, the realism, intended by practitioners in the field. Hamuro et al. are not reporting on a social construction or a mere computation, but a useful reality: the diagram confidently posits its existence. On the other hand, icons are too manifold and singular to be the sole vehicle of scientific discourse. Their use along with symbolic language embeds them in demonstrations and gives to their particularity a representative and well-defined generality, sometimes even a universality.

We return to our reading of the Hamuro et al. article.

Stage 1b: The Peptide Loops

Hamuro et al. chose cyclic hexapeptides to mimic the "arms" of the antibody because they can be modified so as to link up easily with the core scaffold, and because they form hairpin loops: "The peptide loop was based on a cyclic hexapeptide in which two residues were replaced by a 3-aminomethylbenzoyl (3amb) dipeptide analogue* containing a 5-amino substituent for facile linkage to the scaffold" (p. 2681). The recipe for