The Metaphysics of Molecular Structure

Robin Findlay Hendry

Abstract In this paper I distinguish two kinds of structure that appear in chemical explanations: geometrical structure and bond structure. I examine structural descriptions of a range of chemical substances including sodium chloride, water, cyclohexane and proteins, arguing that neither kind of structure is more basic than the other. Such pluralism should not be surprising, however, because at least in chemistry, structure is a creature of classification, and therefore of comparison.

1 Introduction

Chemists appeal to structure at the molecular level—molecular structure—to explain the thermodynamic, chemical and spectroscopic behaviour of chemical substances. Structure is the sole basis of the systematic nomenclature by which substances are named (Thurlow 1998). But what is a molecular structure? In general, the structure of a thing is how its parts fit together. The parts of molecules are atoms and ions, which leaves how they fit together. In fact chemical explanations invoke two kinds of structure, which I will call geometrical structure (the relative positions of the atoms and ions) and bond structure (the framework of bonds between the atoms and ions). The two kinds of structure are perfectly reconcilable, and some substances have both. But they are quite distinct. In what follows I will describe them, and the relationships they bear to each other. In the final section I will argue for pluralism about structure, and that this should not be surprising, given that structure is primarily a classificatory notion.

R.F. Hendry (✉)
Department of Philosophy, Durham University, 50 Old Elvet, Durham, DH1 3HN UK
e-mail: r.f.hendry@durham.ac.uk

2 Geometrical Structure

When chemists describe the 'structure' of a substance, at least sometimes they mean something that can be specified fully in terms of the (average) relative positions of the constituent atoms and ions. Sodium chloride (NaCl) is—pretty much— positively-charged charged sodium ions and negatively-charged chloride ions in a one-to-one ratio. *Solid* NaCl is composed of 'two interpenetrating face-centred cubic sub-lattices' (Greenwood 1968, p. 48), in each of which each a sodium (or chloride) ion is surrounded by six chloride (or sodium) ions arranged octahedrally. So it may be considered as a (potentially infinite) array of unit cells, each cell containing four sodium ions and four chloride ions (see Fig. 1).

There are four of each kind of ion in a cell because the eight ions at the corners are each shared with seven other unit cells, so each counts only as one eighth; the 12 ions at the edges are each shared with three other cells, so each counts as one quarter; the six ions at the faces of the cube are shared with one other cell, so each counts as one half; and finally the ion at the centre falls entirely within the cell, so counts as 1.

The structure of an ionic solid arises from the way the constituent ions pack together so as to maximise interactions between ions of opposite charge, and minimise interactions between those of like charge, given the charges on the ions, the relative size of the cations and anions (M^+ and X^-, respectively), and the stoichiometry of the substance (i.e. whether it is of the form M_2X, MX, MX_2 or so on). Although the structure is characterised by the relative positions of the ions, as represented by distances between the ionic centres (which can be regarded as the sum of two 'ionic radii'), the ions are not entirely static: they vibrate around their equilibrium positions to an extent that is dependent on temperature, so the distances fluctuate. Since the structure survives such fluctuations, it must be characterised by small regions around *average* relative positions. At 801 °C, however, enough of the ions have enough energy to overcome the forces holding them in the lattice and the structure breaks down, forming a liquid consisting mostly of dissociated ions: since the ions are now free to move under electrical forces, the molten salt is an electrical conductor while the solid is an insulator. Clearly, the geometrical structure of solid
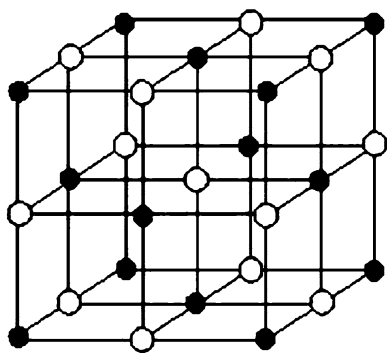

**Fig. 1** Solid sodium chloride (After Greenwood 1968, p. 48)

NaCl does not survive transition to the liquid phase: it is phase-specific. Molten NaCl, like other liquids, has its own structure, which can be characterised in terms of radial distribution functions describing the probability density of various molecular or atomic species as a function of their distance from a central atom. Once again, the structure is fully specified by geometrical relations between the constituent ions, and is phase-specific, in that it exists only within a particular state of aggregation. Water is similar. Depending on pressure, ice is described as displaying one of a number of different structures (see Eisenberg and Kauzmann 1969, Chap. 3; Finney 2004), in all of which hydrogen bonds play an essential role (Needham 2013), linking together the partial negative charges on oxygen atoms to the partial positive charges of protons on neighbouring H_2O molecules. As in NaCl, this structure breaks down on transition to the liquid phase. It is not that the H_2O molecules cease to form hydrogen bonds with each other, or that these bonds cease to constrain their relative positions and orientations: it is rather that, in this higher temperature range, the H_2O molecules are freer to move around them, and the hydrogen bonds themselves are constantly forming and reforming. So even though, at short range, the structure of liquid water is quite like that of ice, over longer ranges this breaks down, as displayed in the radial distribution functions used to describe its structure (see Fig. 2).

If a structure is constituted by the *average* relative positions of the atoms or ions, then structure in this sense must depend on the energy range and timescale over which that average is taken. The cell structure of solid NaCl, as we saw, breaks down above its melting point, and so if we choose a wide enough energy range, the long-range geometrical order of solid NaCl is lost. Similarly, once it is acknowledged that even in the solid state, atoms and ions are constantly in motion, 'structure' depends on timescale. Eisenberg and Kauzmann (1969, 150–152) point out that H_2O molecules in ice undergo vibrational, rotational and translational motions, the molecules vibrating much faster than they rotate or move through the lattice. At very short timescales (shorter than the period of vibration), the 'structure' is a snapshot of molecules caught in mid-vibration. It will be disordered because different molecules will be caught at slightly different stages of the vibration. As timescales get longer, the 'structure' averages over the vibrational motions, and then the rotational and translational motions, yielding successively more regular but diffuse structures. None of this should be surprising: different kinds of structural feature persist over different energy ranges and timescales, and so the energy ranges and timescales we focus on in some particular case will determine which structural features are part of 'the' structure (alongside, of course their relevance to the things we want to explain).

3   Bond Structure

The bond structure of a substance, the framework of bonds between its constituent atoms or ions, is quite different from its geometrical structure, which is constituted by geometrical relationships between them. Consider, for instance, cyclohexane, which is a cyclic alkane—a hydrocarbon involving only single bonds between

Fig. 2 Radial distribution functions at various temperatures for H$_2$O and D$_2$O, (Reproduced from Eisenberg and Kauzmann 1969, p. 157)
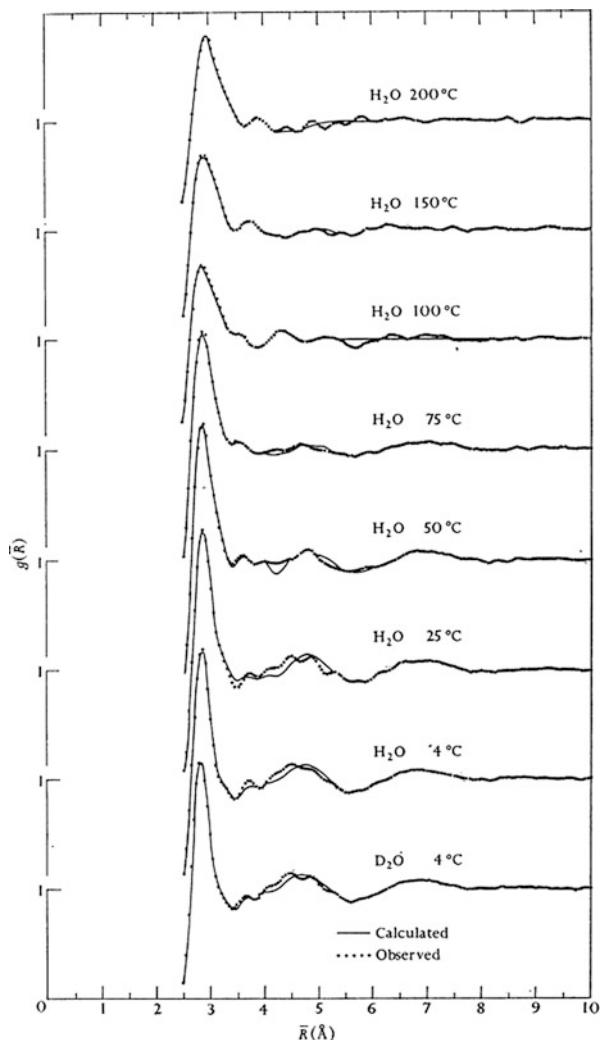
carbon atoms in a ring structure—with empirical formula C$_6$H$_{12}$. In cyclohexane
the six carbon atoms are bonded together in a ring, and to each is attached two
hydrogen atoms (see Fig. 3).

The bond structure of cyclohexane is easily distinguished from its geometrical
structure. Firstly, consider any pair of hydrogen atoms which are attached to the
same carbon atom. These two hydrogen atoms may be geometrically adjacent to
each other in the sense that they are not far apart, and no other atom is between
them (they are in each other's line of sight). But they are not bonded directly to
each other. Secondly, the bond structure is compatible with wide variation in the
relative positions of the atoms, and different geometrical structures. Cyclohexane

Fig. 3 Two representations of the bond structure of cyclohexane: in the image on the right, the hydrogen atoms are left out for clarity
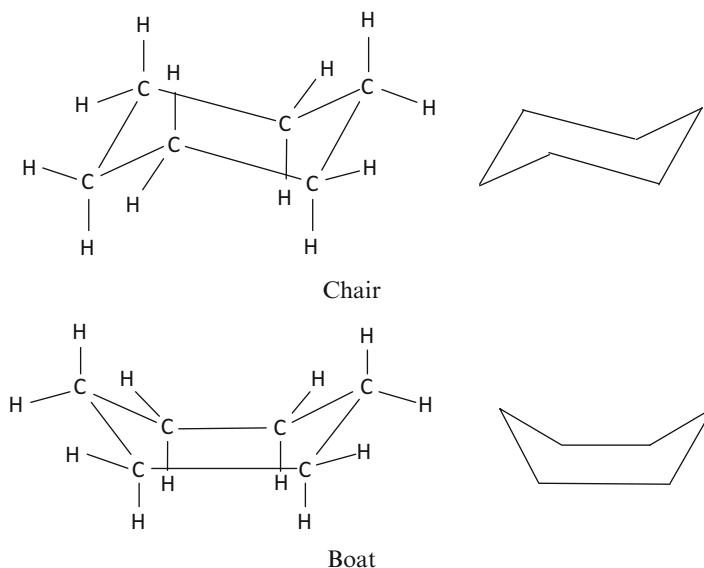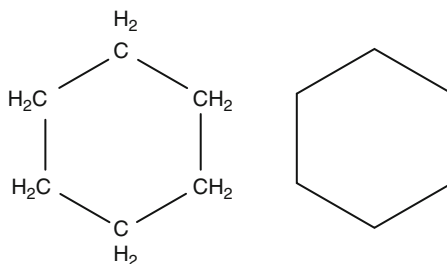

Chair

Boat

Fig. 4 Conformations of cyclohexane: in the images on the right, the hydrogen atoms are left out for clarity

exhibits a number of different *conformations*: that is, geometrical configurations of its atoms. Cyclohexane's lowest energy conformation is the chair (see Fig. 4), but individual cyclohexane molecules are constantly in motion. The energy difference between chair and boat is small, and molecules flip between them many thousands of times a second.

Across all the different conformations, however, one thing remains constant: the pattern of connections between the atoms. This is the bond structure. In the 1860s there appeared a number of different but equivalent ways of representing the bond structure of molecules (see Rocke 1984, 2010), which employed either diagrams on paper or three-dimensional models. They were equivalent in the sense that the structures they represented, attributed on the basis of chemical evidence, were topologically identical. They were constructed under rules of valence which determined, for each element, how many atoms of the various other types it could

be bonded to in a molecule. The topological nature of these bond structures was recognised explicitly in Arthur Cayley's discussion of these 'chemical graphs' as 'trees', his application of this to isomerism, and his formal proof of how many distinct aliphatic hydrocarbons there are with empirical formula C_nH_{n+2} (see Biggs et al. 1976, Chap. 4).

By the mid-1870s, graphical formulae came to be understood as embedded in three-dimensional space. The embedding made available new kinds of chemical evidence for distinguishing between structures. Jacobus van't Hoff explained why there are two isomers of compounds in which four different groups are attached to a single carbon atom by supposing that the valences are arranged tetrahedrally (the two isomers are conceived of as mirror images of each other). Adolf von Baeyer explained the instability and reactivity of some organic compounds by reference to strain in their molecules (Ramberg 2003, Chaps. 3 and 4), which meant distortion away from their preferred geometry. These stereochemical theories were intrinsically spatial, because their explanatory power depended precisely on their describing the arrangement of atoms in space. From the beginning of the twentieth century, bond structures became dynamic, as chemists and physicists began to develop models of how molecules vibrate and rotate, to explain their spectroscopic behaviour (Assmus 1992). This involved filling out structures with details, such as bond lengths, bond angles and force constants, which had previously been absent.

The valence rules have a curious status. They provided a reliable guide to the development of organic chemistry, successfully attributing structures to a vast number of chemical substances, and many of the structures attributed to substances in the 1860s are still accepted in modern chemistry. G.N. Lewis thought 'this group of ideas which we call structural theory' (1923, 20–21) to be one of the most successful in science, yet he recognised that chemical substances did not always behave (physically or chemically) in accordance with their structural formulae: this was a problem especially in inorganic chemistry (1923, p. 67). Moreover, there were always well-known anomalies: substances, like carbon monoxide, in which some atom does not display its usual valence. Whether they cover all of chemistry, or just some well-behaved fragment of it, the idea of these valence rules is worth exploring: they assign valences to atoms according to their elemental identity, and require that, in a valence structure, all of an atom's valences should be used up in single or multiple connections to other atoms. They govern the construction of graphs, and so they should be expressible in first-order logic: there must be an axiomatisation of what chemists call the 'classical theory of molecular structure,' even though that theory remained entirely implicit in the nineteenth century.[1]

---

[1]Note that the theory is not 'classical' in any way that connects it with 'classical mechanics.' In particular, it is not a dynamical theory, and so is not governed by Newtonian dynamics; it is not a statistical theory, and so does not assume that the atoms over which it quantifies are governed by Boltzmann (or any other) statistics.

4   Geometrical Structure Versus Bond Structure

Clearly, geometrical structure and bond structure are not the same thing. What is the relationship between them? Is one more basic or fundamental than the other?

Some substances may have a geometrical structure yet (arguably) lack a bond structure. G.N. Lewis established this when he argued that the structure of ionic substances like potassium chloride (KCl) can be represented without appeal to any bonds between atoms. Lewis (1913) considers a proposal to represent ionic bonding in potassium chloride with a directed arrow, as K→Cl, which would signify that an electron has passed from K to Cl. He argues that this would be misleading, because *even if* (*per impossibile*, given the qualitative identity of electrons), one could tell which electron had come from which potassium atom, the bonding that holds the substance together does not arise from that donation, but rather from the opposite charges that result from it. Furthermore, 'a positive charge does not attract one negative charge only, but all the negative charges in its neighborhood' (Lewis 1913, p. 1452). In potassium chloride, the bonding is electrostatic and therefore radially symmetrical. An individual ion bears no *special* relationship to any *one* of its neighbours, but the same relationship to each of them. This relationship is non-directional, and so cannot be represented by the lines connecting atoms that appear in classical structural formulae. Nor did Greenwood's description of NaCl mention bonds (see above): bonds are not indispensible to a description of its structure. So even though, in the representation of the structure of NaCl (Fig. 1, above), we can see lines between neighbouring ions, they are merely an aid to the eye in discerning the three-dimensional structure of a unit cell. The lines do not represent real physical features of NaCl's structure. If this is right, then some substances have a geometrical structure but no bonds, and therefore no bond structure. How is it possible to have bonding without bonds? There is certainly bonding, because the ions are held together in the lattice by something or other (to a large extent, electrostatic attraction). So although here is a 'bond' in one abstract sense of the word (as in the phrase 'the bond of sisterhood'), there is no bond in the sense which is important to Lewis' argument: the pairwise physical relationship between individual atoms or ions, which is represented by the lines between atoms in molecular structure diagrams. Lewis' argument is meant to establish that there is geometrical structure in NaCl, but no bonds in that second sense of 'bond,' and therefore no bond structure.

Furthermore, every molecule has a geometrical structure, in the sense that its parts are distributed somehow in space, and they bear spatial relations to each other. Given that not every substance has a bond structure, this seems to favour geometrical structure over bond structure for the leading role in the relationship between them: geometrical structure is a more general, and so more basic notion, because having a geometrical structure is necessary, even if not sufficient, for having a bond structure. But that would be misleading for two reasons. Firstly, it is not so clear that having a geometrical structure is necessary for having a bond structure, at least in any way that would make it more basic. From a mathematical point of view, a bond structure is a set-theoretic object: if we take the set of a molecule's

constituent atoms, a bond structure is some subset of the set of ordered pairs that can be formed from the members of this set. This set-theoretic structure is all that is needed to fulfil one important explanatory role for bond structure in chemistry: that of explaining how many structural isomers a particular substance may have. And from a purely logical point of view, something might have this set-theoretic structure without it (or its parts) being located in space at all. This is just how the explanatory role of structure was seen by the pioneers of structure in chemistry in the 1860s, such as Edward Frankland (see Hendry 2008b). Even though bond structures did eventually come to be regarded as embedded in space, that was an extension of the explanatory role of structure to account for optical isomerism. From a purely mathematical point of view, then, geometrical relationships do not determine bonding relationships. Perhaps bond structure is only contingently embedded in space. But the mathematical point of view is not all there is, and a bond structure is not just a graph: it is a graph generated by a particular *physical* relation (the bonding relation). Is a *bond* structure something that is necessarily embedded in space? To answer that question we need to know more about what a bond is. (That is not merely a rhetorical pointer to a later discussion: the answer is not clear: in Hendry 2008b, 2010 I discuss two opposed accounts.) Bonds clearly have geometrical constraints: distinct bonds do not overlap or cross, and it may well be that fixing the geometrical configuration *physically* (though not mathematically) determines the bond structure uniquely, in the following way.

In the 'Atoms in Molecules' (AIM) programme, Richard Bader and his co-workers have sought to recover the traditional bond structure of molecules as a topological feature of the electron-density distribution (see Bader 1990; Popelier 2000; Gillespie and Popelier 2001). From the electron-density distributions for many different molecules can be defined 'bond paths' between atoms that generate 'molecular graphs' which are strikingly close to the classical molecular structures of those molecules. As Bader puts it, 'The recovery of a chemical structure in terms of a property of the system's charge density is a most remarkable and important result' (1990, p. 33). Bader's elegant results are interesting and significant in a number of ways. Firstly, AIM offers a substantive answer to a longstanding question: what is a chemical bond? The answer (according to AIM) is that bonds are topological features of the electron density distribution (or rather the particular regions of electronic charge that bear these topological features). Secondly, although it recovers the classical bond topology, AIM seems to make geometrical structure prior to bond structure. The quantum-mechanical calculations that underlie AIM, like all tractable quantum-mechanical calculations concerning molecules, begin by making the Born-Oppenheimer approximation (see Hendry 1998, forthcoming), which involves separating nuclear and electronic variables, and fixing (or 'clamping') the nuclear positions. The electric field due to the nuclei is then used as a constraint on the calculation of a resultant electron density distribution. If the nuclear positions are well chosen (i.e. correspond to the nuclear positions in the molecule's equilibrium geometry), then from the resulting electron density distribution we have 'read off' the bond structure of a real molecule from its nuclear geometry. Physically, if not mathematically, it might seem that geometry

determines bond structure. But that is too quick: it's not so clear that we can simply 'read off' the bond structure from the geometry, because the whole calculation relies on minimising the energy of the system. (That, in fact, is taken to be a mark of how closely the Born-Oppenheimer calculation approximates the 'exact' energy.) By concentrating on the lowest-energy states, the whole procedure would seem simply to ignore higher-energy states that correspond to higher-energy geometrical configurations of different bond topologies. Perhaps we don't find *the* unique bond topology for the geometry, but rather the bond topology which has the lowest energy in that geometry (and, probably, has that geometry as its lowest-energy geometrical configuration).

Let us pursue this idea that a bond structure is something that can be displayed by a substance *in addition* to its geometrical structure. This is supported by the fact that bond structure may survive phase transitions which the geometrical structure cannot. Thus, for instance ice, liquid water and steam all display different geometrical structures, but the topological structure of its molecules, as represented in its structural formula (a central oxygen atom bonded to two hydrogen atoms) remains constant across the different states of aggregation. Secondly, in the substances that have it, bond structure is explanatorily prior, in the sense that a molecule's bond structure is compatible with a range of different geometrical arrangements of its parts, and in fact determines which arrangements it may have. Consider once again the conformations of cyclohexane. In that case, the bond structure is a constant while the molecule moves between quite different geometrical configurations. And it is the persistent bond structure which *explains* the energetic ordering of the various conformations. The chair is the lowest-energy conformation because in that geometry the bond structure experiences the least strain: that is, in that geometry the arrangement of bonds around individual carbon atoms is closest to the tetrahedral, and the hydrogen atoms are less crowded, reducing their (repulsive) interactions. These considerations allow us, I think, to resist the idea that geometrical structure is prior to bond structure.

5   Structure as Abstraction

If neither geometrical structure nor bond structure is prior to the other, how might one understand the role of these two notions in the classification of substances, and in explanations of their behaviour and characteristic properties as arising from their structure? Seen as a classificatory notion, structure is derived from a process of abstraction. Molecules that differ in the properties of, and relations between, their parts can share a structure. Identifying a class of molecules as sharing structure, we just ignore the differences between them. Different classes of molecule or substance may be alike in different ways: we might expect there to be different kinds and levels of structure. We get to the structure of a substance by abstracting away from the particular clusters of property- and relation-instances that its constituent atoms and ions bear to each other, to focus on some subset of them which is salient because

it survives across some range of (e.g. thermodynamic or energetic) conditions that demands our attention. Pluralism about structure should not then be surprising simply because in a reasonably complex thing there is always, in principle, more than one way to abstract away from its full particularity in this way.

Take water and proteins as examples.[2] Water, as we have seen, has both a bond structure and a geometrical structure, although the bearers of these structures are different (individual molecules as opposed to collections of such molecules). Considered as a substance which can exist in different states of aggregation, we focus on the covalent bond structure of its molecules that is shared by all those different states (solid, liquid, and gas, up to highly rarefied states). You cannot abstract away from that bond structure without abstracting away from water itself (or so I argue: see Hendry 2006, 2008a). But hydrogen bonding is 'structural' too: in water, it plays an important role in understanding the structure of the substance in its particular states of aggregation. Neither is *the* structure, in the sense that, if we need to know about the details of water's structure in one of its particular states of aggregation, the bond structure won't tell us enough. It is the interaction between the individual molecules that gives rise to the (large-scale) geometrical structure, via hydrogen bonding (on which see Needham 2013): the opposite partial charges on oxygen and hydrogen atoms give rise to chains and clusters of H_2O molecules whose existence has a major influence on the properties of the substance. But these chains and clusters are constantly forming and reforming.

Proteins also display different kinds and levels of structure. The primary structure of a protein is, roughly, the order of the connections between its constituent amino-acids, while secondary, tertiary and quaternary structure concern the different ways in which it is arranged in space. Interestingly, the very same physical interaction, hydrogen bonding, which gives rise to short-lived structure in water, maintains structure that is longer-lasting in proteins, at least in the narrow range of physical and chemical conditions within which cellular processes take place. But these higher levels of structure do not survive at higher temperatures, or in more hostile chemical environments. Concentrating on temperature, one might say that the higher levels of structure are conformations 'frozen in' below the temperature at which the hydrogen bonds that sustain them would begin to break and reform too quickly. Important phenomena (biology!) depend on them, but if you consider a wide enough range of conditions, the higher levels of structure disappear from view.

I conclude that, in identifying different kinds of structure and reasoning about them, we (implicitly) focus on relations which survive over the specific ranges of (chemical and physical) conditions in which the phenomena of interest can be given a unified explanation in terms of that kind of structure. Since there is a close relationship between structure and substance identity, these specific ranges of conditions are essentially those at which identifiable substances exist. Different substances are stable over different ranges of physical conditions, and

---

[2]See Tobin (2010) and Slater (2009) for discussion of classification and structure in proteins; Goodwin (2011) provides a reply.

it should be no surprise if structural explanations concerning substance X focus on physical interactions that underlie the particular structural relationships that survive across the conditions under which X exists, and structural explanations concerning substance Y focus on different physical interactions that underlie the different structural relationships that survive across the different conditions under which Y exists.

6 Conclusion

'Structure' sometimes invokes geometrical relations. Sometimes it invokes bond topology, which is understood always to be embedded in space. Furthermore neither kind of structure is more basic than the other. If these arguments support some form of pluralism, it is one that should take a robustly realist stance on structure and its role in classification. It is robustly realist for two reasons. Firstly, each kind of structure is constituted by real physical relations: spatial relations in the case of geometrical structure, bonds in the case of bond structure. Secondly, you can't ignore either kind of structure without significant loss of information about the substances that have them. If you ignore geometrical structure, you have no access to the explanations provided, for instance, by optical isomerism (van't Hoff), or stearic strain (von Beayer). If you ignore bond structure you ignore what, in many substances, is held constant over a range of different geometrical configurations (remember the conformations of cyclohexane), and explains which geometries are possible, and which are favoured energetically.

References

Assmus, A. (1992). The molecular tradition in early quantum theory. *Historical Studies in the Physical and Biological Sciences, 22*, 209–231.

Bader, R. (1990). *Atoms in molecules: A quantum theory*. Oxford: Oxford University Press.

Biggs, N. L., Lloyd, E. K., & Wilson, R. J. (1976). *Graph theory 1736–1936*. Oxford: Clarendon Press.

Eisenberg, D., & Kauzmann, W. (1969). *The structure and properties of water*. Oxford: Oxford University Press.

Finney, J. L. (2004). Water? What's so special about it? *Philosophical Transactions of the Royal Society B, 359*, 1145–1165.

Gillespie, R., & Popelier, P. (2001). *Chemical bonding and molecular geometry*. Oxford: Oxford University Press.

Goodwin, W. (2011). Structure, function, and protein taxonomy. *Biology and Philosophy, 26*, 533–545.

Greenwood, N. N. (1968). *Ionic crystals, lattice defects and nonstoichiometry*. London: Butterworths.

Hendry, R. F. (1998). Models and approximations in quantum chemistry. In N. Shanks (Ed.), *Idealization in contemporary physics* (pp. 123–142). Amsterdam: Rodopi.

Hendry, R. F. (2006). Elements, compounds and other chemical kinds. *Philosophy of Science, 73*, 864–875.

Hendry, R. F. (2008a). Microstructuralism: Problems and prospects. In K. Ruthenberg & J. van Brakel (Eds.), *Stuff: The nature of chemical substances* (pp. 107–120). Würzburg: Königshausen und von Neumann.

Hendry, R. F. (2008b). Two conceptions of the chemical bond. *Philosophy of Science, 75*, 909–920.

Hendry, R. F. (2010). The chemical bond: Structure, energy and explanation. In M. Dorato, M. Rèdei, & M. Suárez (Eds.), *EPSA Philosophical issues in the sciences: Launch of the European Philosophy of Science Association* (pp. 117–127). Berlin: Springer.

Hendry, R. F. (forthcoming). *The metaphysics of chemistry*. Oxford: Oxford University Press.

Lewis, G. N. (1913). Valence and tautomerism. *Journal of the American Chemical Society, 35*, 1448–1455.

Lewis, G. N. (1923). *Valence and the structure of atoms and molecules*. Washington: Chemical Catalogue Company.

Needham, P. (2013). Hydrogen bonding: Homing in on a tricky concept. *Studies in History and Philosophy of Science, 44*, 51–66.

Popelier, P. (2000). *Atoms in molecules: An introduction*. London: Pearson.

Ramberg, P. (2003). *Chemical structure, spatial arrangement: The early history of stereochemistry 1874–1914*. Aldershot: Ashgate.

Rocke, A. (1984). *Chemical atomism in the nineteenth century: From Dalton to Cannizzaro*. Columbus: Ohio State University Press.

Rocke, A. (2010). *Image and reality: Kekule, Kopp and the scientific imagination*. Chicago: Chicago University Press.

Slater, M. (2009). Macromolecular pluralism. *Philosophy of Science, 76*, 851–863.

Thurlow, K. J. (1998). IUPAC Nomenclature part 1, organic. In K. J. Thurlow (Ed.), *Chemical nomenclature* (pp. 103–126). Dordrecht: Kluwer.

Tobin, E. (2010). Microstructuralism and macromolecules: The case of moonlighting proteins. *Foundations of Chemistry, 12*, 41–54.