40. Models of Chemical Structure

William Goodwin

Models of chemical structure play dual crucial roles in organic chemistry. First, they allow for the discovery and application of *laws* to the complex phenomena that chemists hope to understand. Second, they are a source of novel concepts that allow for the continuing development of structure theory and theoretical organic chemistry. In chemistry, therefore, the centrality and significance of models to the scientific enterprise is manifest and furthermore chemistry is a relatively clear, useful, and interesting context in which to consider more general philosophical questions about the nature and role of models in science.

One of the most important and influential trends in the philosophy of science over the last 50 years has been the increase in both the attention paid to the concept of a model and the employment of this concept in philosophical reflection on the nature and dynamics of science. This trend has been usefully described and analyzed by many philosophers (see for instance: [40.1–3]). Without trying to be exhaustive, in this paper I plan to identify a few of the most significant philosophical insights that have emerged out of this increased interest in scientific models and then to reflect on these insights in the context of chemistry, which has been relatively neglected in the philosophical literature. I hope to show both that in chemistry the centrality and significance of models to the scientific enterprise is manifest, and that chemistry is a relatively clear, useful, and interesting context in which to consider more general philosophical questions about the nature and role of models in science.

Models have been characterized in many different ways in the philosophical literature, but for the purposes of this paper it will suffice to think of them as instruments for representation that are not primarily linguistic. The important contrast is with the linguistic statements of a theory (in the logical sense). So the double helix model of DNA represents DNA molecules not because it is a statement in a language that describes this molecule, but because it is a physical object with certain similarities to the objects that it is intended to represent. Likewise, the billiard ball model of a gas is an image of an interacting system (or the abstract idea of such a system) along with, perhaps, a narrative about how to understand this image, which can be used to represent a gas for certain purposes. Many types of objects, other than statements, have been thought of as models (including mathematical structures, abstract objects, and fictional objects). Given the diversity of things that models might be, it seems best to summarize the central insight behind the relatively recent philosophical interest in models as follows: Representational instruments that are not primarily linguistic are crucial to understanding the nature and development of science.

This central insight has been developed in a variety of ways, but I want to focus on two of them. First, models are crucial to the dynamics of science – how scientific representations, theories, experiments and concepts change over time in response to feedback from the world. And second, models are crucial to science's capacity to confront complexity – its ability to have useful things to say about real world and/or complex systems. Models, being nonlinguistic entities, help philosophers to get a grip on these aspects of science at least in part because they are representationally rich, that is, they are not limited in their representational capacities (as purely linguistic representations would be) by the arbitrary associations between their component

symbols and aspects of the world. In the case of the dynamics of science, the rich representational capacities of models both supply (or allow for) new features that can be exploited in the models' representational role and thereby act as an incubator for novel concepts. Similarly, models can act as intermediaries between theory and complex real world phenomena because their richer resources allow for representation of the more concrete and local features crucial to understanding such phenomena.

Chemistry, like any large and diverse field of scientific inquiry, is replete with models of many different sorts. Much of the modeling in chemistry is of the standard sort discussed in the philosophical literature – that is, a response to the problem of getting abstract mathematical theories to apply to complex real world phenomena. There has been interesting philosophical work on the form that this sort of modeling takes in chemical contexts [40.4–6] and [40.7] for example; however in this paper I want to focus on what is, I think, a distinctively central and important role for modeling in chemistry. Chemistry, at least large parts of it, is concerned with representing the structures of the substances it studies. For the most part, chemists do not use linguistic resources to represent structure; instead, they build models. Sometimes, like Watson and Crick and most students of organic chemistry, they build physical models, but most frequently they use diagrammatic representations like structural formulas. While there are often linguistic components to such formulas (letters for the atoms, for example), the representational resources of these diagrams are not limited to the arbitrary relations between their component signs and predicates or terms in the current language of the theory. Furthermore, much of importance of these models for both the dynamic development of chemistry and for facilitating the application of chemical theories to concrete, real world cases derives from these extra linguistic representational resources.

Structural formulas (Fig. 40.1), which were initially developed over the course of the nineteenth century, are the centerpiece of a research program that has been immensely successful ([40.8] for a summary of the development of structural formulas). The guiding strategy of this research program, articulated by Aleksandr Butlerov in 1861, was to have one structural formula for each chemical compound, and then to explain the chemical properties of these compounds by determining, "the general laws governing the dependence of chemical properties on chemical structure" [40.9, p. 256]. Structural explanations of chemical properties (and some physical properties as well) have been central to chemistry – particularly organic chemistry, on which I shall focus – ever since. A structural explanation of a chemical property proceeds by identifying the structural features of a compound that are responsible for some (usually contrastive) chemical fact. These structural features are typically general features of the compound as represented in its models, i. e., in its structural formula, which might also be realized by other structural formulas or models as well. In other words, structural patterns identifiable in the models are correlated with chemically significant facts (usually, these days, differences in energy or stability). It is these structural correlations that have ended up playing the role of the *laws* that Butlerov imagined. Thus the *laws* of structural chemistry are formulated in terms of something like chemically significant patterns in the models; in this sense, then, in structural chemistry the laws piggyback on the models. The models don't just interpret or concretize the laws, they make them possible in the first place. Understanding the complex chemical facts of organic chemistry depends, at the most basic level, on building structural models of chemical compounds.

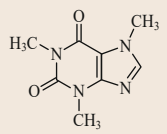

Fig. 40.1 A structural formula for caffeine

Structural chemistry has not been a static research program. Over the course of its 160 year development, there have been immense changes in chemists' understanding of structure. And not surprisingly, these changes have been reflected in the models used to represent that structure. In the course of this continuous refinement – the back and forth between conceptions of structure and representations of it – models of structure have played a crucial role as a source of new general structural features. In other words, features of the models not originally used in their representational role are reinterpreted as representationally significant in order to explain or account for new theories or experimental facts. Additionally, even after a new aspect of *chemical structure* has been recognized and represented in the models, there still remains the daunting task of making that structural feature experimentally and synthetically relevant (making it useful) and models have also supplied some of the very local structural concepts that allow for the experimental and theoretical development of whole new subfields. Thus, just as models are crucial for structural chemistry to discover and apply laws to the complex phenomena it studies, so too are these models crucial as a source of the novel concepts necessary for the continued development of structure theory and theoretical organic chemistry.

40.1 Models, Theory, and Explanations in Structural Organic Chemistry

Structural formulas are the central representational tool used in the explanations and predictions of the theory of organic chemistry. These formulas play many representational roles in organic chemistry, including the role of denoting expressions for chemical kinds. In this role, they serve as descriptive names for these kinds, individuating them according to their composition, bond connectivity and (aspects of their) stereochemistry. Because they can be put (roughly) into one-to-one correspondence with the kinds that they purport to denote, they are also able to act as *stand-ins* for these chemical compounds ([40.10–12] for more on the roles of structural formulas as both names and models in organic chemistry.). When acting as stand-ins for chemical kinds, structural formulas can be *manipulated to teach us things about themselves*, and the things learned about these models can, in many cases, "be transferred to the theory or to the world which the model represents" [40.2, p. 33] ([40.13] for examples of how models acted as paper tools in the early development of organic chemistry.). That is to say that chemists can learn from their models by exploring the implications of (or for) abstract theory in the concrete contexts of particular chemical kinds and often this exploration takes the form of the manipulation of structural models. To bring out some of the interconnections between models of structure, theory, and the explanations in structural chemistry I am going to briefly consider the theory of resonance, which is a modification of structure theory developed in the first half of the twentieth century that is still central to organic chemistry today.

Though it was developed prior to any generally accepted account of the nature of chemical bonding, structure theory has had to evolve in the face of changing theoretical accounts of the nature of the chemical bond. This is to be expected given Butlerov's aspiration for structural formulas; namely, that they explain the chemical and physical properties of the compounds that they depict. As the theoretical understanding of what a bond was changed, so to did the depictions of chemical bonds in structural formulas, and this was crucial to leveraging the revised understandings of bonding into new structural explanations of chemical or physical properties. The theory of resonance is one important way (and was historically the first broadly accepted way) that the quantum mechanical character of chemical bonding is recognized and applied within structural organic chemistry.

Even before the development of quantum mechanics, in response to the explanatory demands placed on chemical structures, several modifications of structure theory were suggested that [40.14, p. 2]:

> "considered it possible for the true state of a molecule to be not identical with that represented by a single classical valence-bond structure, but to be intermediate between those represented by two or more different valence-bond structures. "

These suggested modifications where motivated by cases where a chemical compound did not behave as it would have been expected to behave given its representation using a single structural formula but where, by thinking about the compound as a mixture intermediate between two or more such formulas, the behavior could again be accounted for in structural terms. *Pauling*, in his seminal *Nature of the Chemical Bond* [40.15], laid out the theory of resonance by employing quantum mechanics to rationalize and systematize (but not to deduce), the use of multiple structures to represent individual chemical kinds and then went on to demonstrate the broad usefulness of this theory in organic chemistry.

The theory of resonance is interesting, from the point of view of the role of structural models in chemistry, for two related reasons. First, it was the manipulation of valence bond structures (which are structural formulas that explicitly depict the bonding electrons of the constituent atoms) that first revealed the possibility of explaining recalcitrant chemical and physical phenomena by thinking of chemical kinds as appropriately represented by a combination of distinct structural formulas. Facts in the world of the model – that multiple different valence bond structures where plausible for a given chemical kind – were used to suggest modifications designed to improve the explanatory power of structure theory. Furthermore, these facts about the structural models (multiple possible valence bond structures for a given kind) were systematized and rationalized using the theory of quantum mechanics so that the delocalization of bonding electrons (which is the central implication of quantum mechanics for chemical bonding) could be recognized and exploited in organic chemistry. Manipulations of the model supplied the vehicle for making quantum mechanics first applicable to structural organic chemistry. In this sense, then, the structural models mediated between the theory (quantum mechanics) and the world. Secondly, and similarly, it is the actual exploration of the range of available valence bond structures that often proves crucial to the use of the theory of resonance in generating the structural explanations that are useful in organic chemistry. In a typical case of an explanation invoking resonance, none of the individual structural formulas making up a resonance hybrid allows for the explanation of all of the chemical or physical properties of interest. In-

stead, more standard structural analysis is applied to some of the individual component formulas of the resonance hybrid, and then the behavior of the chemical kind as a whole is understood as some proportional mixture of the structural prediction based on its component structures. In other words, without the more complex depiction of a chemical kind allowed by resonance theory, it would not be possible to successfully supply structural explanations of its chemical or physical behavior. Getting the structural *laws* to apply to some chemical kinds requires a more complex model of their structure – their depiction as a resonance hybrid. Explorations in the world of the model uncovering potential significant resonance structures mediate between the theory (structure theory) and the world by allowing for the successful application of this theory to complex cases to which it would otherwise not be useful.

In order to see how the theory of resonance allows structural formulas to mediate between more general structure theory and experimental facts, it will be useful to consider some of Pauling's work on the structure of proteins. Proteins are polymers of amino acids formed when the carboxyl group of one amino acid reacts with the amino group of another forming an amide linkage called a peptide bond. As a result, there is a recurring structural pattern in proteins: tetrahedral carbon atoms (bonded to the R-groups of the amino acids) joined by amide groups (-NH-CO-). Amide groups are therefore a fundamental structural component of proteins; however, in order to predict the behavior of these groups, and thus to outline the basic structural features of proteins, it is not sufficient to consider only one of the structural formulas that can be used to represent them. Instead amide groups, at least in Pauling's treatment, were thought of as a resonance hybrid of two component structural formulas, and their structural behavior was anticipated to be a sort of weighted average of the behavior predicted by the structure theory for these individual structural formulas.

The most important feature of the amide linkage from the point of view of predicting the structure of proteins is that the carbon and nitrogen of the amide linkage lie in a single plane with the two tetrahedral carbons that they connect. *Pauling* regarded the planarity of the amide group as, "a sound structural principle" concluding that a "structure in which the atoms of the amide group are not approximately coplanar should be regarded with skepticism" [40.16, p.19]. Though he provided substantial experimental confirmation of the planarity of the peptide bond, it was the theoretical arguments for this principle that invoked the theory of resonance. An amide linkage is typically represented by a structural formula in which there is a single bond between the nitrogen and the carbonyl carbon, which is itself double bonded to oxygen. However, another possible structural formula for the amide linkage has a double bond between the carbon and the nitrogen while there are three unshared electron pairs around the oxygen (resulting in a net formal charge of -1 on the oxygen) and no unshared pairs around the nitrogen (resulting in a formal charge of $+1$). The theory of resonance indicates that the first, and more typical, structure should be the most significant contributor to the overall structure of the amide linkage, but that the second structure might also be important to consider (Fig. 40.2).

The most salient difference between these two structures is where the double bond is located, either between carbon and oxygen or between carbon and nitrogen. Pauling argued based on experimental measurements of the bond lengths in some simple peptides (by x-ray crystallography), that in the actual peptide linkage (on average) the relative contribution of these two structures was 60% for the typical structure and 40% for the secondary structure. These numbers were based on the fact that the measured C-O bond length in the peptide bonds was longer than typical double bonds between carbon and oxygen (in cases where no alternative resonance structures were available) but also shorter than typical single bonds between C and O. Similarly, the measured C-N bond length was shorter than typical single bonds, but longer than typical double bonds between these atoms. If he supposed that the relative contributions of the two structures were 60% and 40% respectively, and thus that the C-O bond was 60% double and 40% single, while the C-N bond was 40% double and 60% single, then the predicted length of the bonds closely matched the measured values.

Once Pauling had argued that the second resonance structure with a double bond between carbon and nitrogen was an important contributor to the overall structure of the linkage, it followed from standard structural theory that the peptide linkage should be essentially planar. Double bonds do not allow free rotation; that is, you have to break the bond (costing a lot of energy) in order to rotate around the axis of the bond. The energetic cost of rotation around the double bond is the reason that double bonds lead to stereoisomerism (there are distinct chemical compounds, with different structural formulas, that reflect different arrangements of substituents around a double bond). Since the C-N bond in

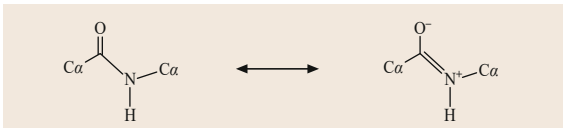

Fig. 40.2 Resonance structures of the peptide bond

the peptide linkage was 40% of a double bond, *Pauling* was able to estimate that the energetic cost of rotating around this bond would be about 40% of the bond energy of a typical double bond between these atoms. Furthermore, he was able to estimate the strain energy of deviations from planarity, concluding: "we can calculate strain energies of about 0.9 kcal/mole for 10° distortion of the amide group" [40.16, p. 14]. This effectively meant that large deviations from planarity in peptide linkages would be very energetically expensive and would therefore constitute a "highly unusual steric relationship" [40.16, p. 16].

Pauling's analysis of the structure of peptide linkages shows both what structural accounts of chemical phenomena look like, and how resonance theory allowed structural analysis to be applied to a broader range of chemical facts. The measured bond lengths of the amide linkages in peptides do not correspond to the average bond length that one would expect based on the typical representation of the amide linkage using structural formulas (with typical CO double bonds and typical CN single bonds). In order to provide a structural explanation for this deviation from expected values, Pauling needed to find some recurring structural feature of the amide linkage that could explain it. What he found, by manipulation of the formulas, is that such amide linkages could be represented by another structural formula, one obtained by redistributing the valence electrons (and thus moving around the bonds). According to the principles of resonance theory, this alternative structure was energetically plausible and should be regarded as a potential contributor to the overall structure of the amide linkage. However, a structural analysis of this second structure by itself would also not explain the experimental bond lengths. Instead, only by regarding the actual structure as intermediate between the two resonance structures, did an explanation of the observed bond lengths became possible. Exploration in the world of the model, then, was crucial to providing a structural account of the observed bond distances. Furthermore, the success of this explanation gave Pauling confidence that his resonance structures provided a reasonable representation of the peptide bond, and thus that he could apply a structural analysis to this representation in order to make a significant structural prediction about all proteins. Again this prediction (the planarity of the peptide bond) was based on giving a weighted analysis of the the various accessible resonance structures. Without the detour through the range of plausible resonance structures, explored by the chemist through manipulations of structural formulas, neither the original explanation, nor the extremely significant prediction that Pauling made about the structure of proteins would have been possible. (Now there are other ways of taking into account the implications of quantum mechanics on chemical bonding, and these can also presumably support the same explanations and predictions; but the historical fact is that Pauling, who basically initiated the study of protein structure, used resonance theory.)

40.2 Structures in the Applications of Chemistry

In addition to their role supporting the explanations and predictions of structural organic chemistry, structural formulas and/or models also play a crucial role in applying the theory of organic chemistry to the solution of synthesis problems. Synthesis problems are the guiding application of the theory of organic chemistry. Of course not all organic chemists are working to synthesize compounds, but this is the characteristic goal around which the field developed, and it is possible to understand the theoretical structure of the field as reflecting this goal. That is, an important reason that the explanations of organic chemistry take the form that they do – looking for structural accounts of chemical phenomena, for example – is because this approach facilitates the solution of synthesis problems [40.17, 18]. Synthesis problems have a common form, and the intellectual challenges that they present derive from this form. By understanding the basic form of synthesis problems, and the basic strategies developed for solving these problems, it is possible to appreciate the central importance of the sorts of structural analysis undertaken in the typical explanations and predictions of organic chemistry. As seen in the last section, explorations in the world of the model, and thus the role of structural formulas as models, can be crucial to providing the structural analyses of organic chemistry. However, the role of structural formulas and models in solving synthesis problems is not limited to their support of structural explanations or predictions. Instead they play additional roles in delimiting the array of possible synthetic approaches and evaluating the plausibility of those approaches.

A synthesis problem begins with a target molecule. The chemist's goal is to come up with a method for making this target molecule by a sequence of chemical reactions that begins with compounds that chemists already know how to make. Often, no chemist has ever synthesized the target molecule before, though it may

be a natural product that is synthesized by some biological systems. The first step in solving such a problem is to get a clear idea about the structure of the target molecule. Since organic molecules are individuated by their structures, this amounts to insisting that the synthesis problem is well defined – it has a clear goal. Once the structure of the target is settled, the synthetic chemist must come up with some way to leverage knowledge about the outcomes of lots of chemical reactions run on different (typically simpler) compounds into a strategy for making the target compound, which, presumably, no chemist has ever experimented with before. It is crucial, therefore, for the synthetic chemist to exploit some notion of structural similarity. Structural patterns identified in the target indicate which known reactions might be plausibly employed in its synthesis. Furthermore, because the structural patterns identified in the target are in a novel context, the chemist must have some way of accounting for, or anticipating, the way that the structural context influences the behavior of known reactions (characterized and understood in simpler structural contexts). This is what the structural accounts (explanations and predictions) of theoretical organic chemistry do.

One way to think about the process of coming up with a synthesis for a target compound is through the process of retrosynthetic analysis [40.19]. Retrosynthetic analysis works backwards from the target molecule, systematically investigating all of the ways that one might produce the target molecule by a known chemical reaction (these are characterized by the structural patterns on which they operate, for instance, by the functional groups that they begin with and that they produce). All of the reactants that might produce the target molecule by one of these know reactions are then subjected to the same process, generating their own lists of possible second-order reactants. This process is repeated until it generates a path terminating in compounds that can already be synthesized. Given that there are thousands of known reactions, many of which might apply to a complex target molecule, the branching array of possibilities generated by such a process – the retrosynthetic tree – is immense and must be systematically pruned into a plausible synthetic plan. (I have described this process in significantly more detail, with concrete examples, in [40.20, 21].)

The pruning of the retrosynthetic tree, following Corey's conception, takes place in stages. In the first stage, strategic pruning, the synthetic chemist analyzes the target compound in order to identify the sources of synthetic complexity in it. By identifying these sources of complexity, the chemist can focus on paths in the retrosynthetic tree that reduce synthetic complexity and that, therefore, are more likely to terminate in compounds that have already been synthesized or are easy to make. Assessing the sources of synthetic complexity in a target molecule amounts to using a set of heuristic principles, grounded in both the collective experience of synthetic chemists and the theory of organic chemistry, to identify particular bonds or atoms whose structural environment will make them particularly difficult to create. The relative difficulty of dealing with these sources of complexity can also often be estimated, giving the synthetic chemist, in the end, a clear focus on branches of the retrosynthetic tree that eliminate the largest source of complexity. Though this can result in a drastic narrowing of the possible synthetic paths that need to be explored, strategic pruning must be followed up by plausibility assessment, where the possible paths removing the largest source of complexity are evaluated for their relative structural plausibility. As I characterized the retrosynthetic tree, the possible reactions that might produce a structure were characterized based on the presence of some structural feature in the target. Any reactions that might produce that product were part of the tree. However, not all of these reactions are actually plausible because, for example, the target has other structural features that would interfere with the success of that particular reaction. And even among those that are plausible, the synthetic chemists will want to decide which path or paths are most likely to work and to generate the fewest complications downstream. These assessments again depend on analyzing how a reaction, understood and characterized in some other, simpler structural context, would perform in the complex local circumstances of the target molecule. After plausibility assessment comes the final stage of synthetic design, which is optimization, where precise ordering of synthetic steps is worked through and control steps are added. These control steps are added in order to eliminate complicating factors identified by a careful structural analysis of the synthetic route. They work by adding chemical groups to synthetic intermediates in the proposed path that either eliminate the influence of complicating structural factors or promote the formation of desired products. These control groups can then be removed after they have done their job. Often the precise ordering of the synthetic path can influence which control groups are needed, and vice-versa, so the overall optimization of the synthetic route must involve both of these considerations.

This brief sketch of the process of designing a chemical synthesis has made it clear, I hope, that close structural analysis of the both the target molecule and the potential intermediates is crucial to the process. The possible reactions resulting in the target molecule (or some intermediate) guided by structural similarity to the products of known reactions is what generates

the array of potential precursors at each stage in the generation of the retrosynthetic tree. This exploration of possible reactants and reactions, and the array that it generates, all take place in the world of the model – in fact, the molecules depicted by these structural formulas may never have existed. The possible reactants must be deduced based on the reaction being considered, and this can be done by exploring what reacting structures would, upon application of the considered reaction, result in the target molecule; the reactions must be worked through backwards in the world of the model to generate potential precursors. Similarly, the strategic pruning of the array of possible reactions depends on investigating the detailed local environments of particular atoms or bonds in the model of the target molecule. Rules of thumb about the relative difficulties of producing these atomic arrangements or bonds (based in part on what the reactant structures would have to be if they were generated by certain procedures) guide the synthetic chemist to particular routes in the retrosynthetic tree. Recurring structural features identified in the local environment of the model of the target molecule provide the basis for the application of these rules of thumb, and thus for the large-scale decisions about synthetic strategy. Both plausibility analysis and optimization depend on determining how generically characterized reactions would be likely to perform in the complex local environment of the target molecule or intermediates. Often the typical explanations (and/or predictions) of organic chemistry can be used to figure out how individual structural features would affect the reaction. But in complex environments there are often multiple relevant, and potentially competing, structural features at play. To make sensible choices about strategy in these cases, synthetic chemists can either attempt to theoretically discriminate the plausibility of potential pathways, or to modify the structure to make its behavior more predictable using control steps. All of this takes place in the world of the model, using whatever theoretical principles are applicable in that local environment, to analyze and make sensible decisions about what synthetic pathways might work in the lab. Synthetic design is thus a process that, from beginning to end, involves manipulating, exploring, deducing possible precursors and analyzing structural models. Theory can be brought to bear on the problem only through its application to, and analysis in the context of, particular structural models.

40.3 The Dynamics of Structure

In this last section, I want to describe two ways that structural models have contributed to the development of the research program of structural chemistry by supplying new structural concepts. In the first case, models of structure had features that were not initially recognized to be representationally significant but which, when interpreted as significant, could be used to explain anomalous results. Chemists did not abandon the structural research program when they encountered unexpected experimental results; instead, they modified their models of structure, taking features readily available in the model and attributing new representational significance to them. General features carried around in the models were appropriated in order to modify the conception of chemical structure in the face of new experimental results. In the second case, particular structures supplied foothold concepts that allowed for experimental results to be brought to bear on these newly representationally significant features of structural formulas. Particular structures are cognitively richer than general types of models or abstract theories; they have all sorts of features that might turn out to support important inferences about the target system. In this example, chemists isolated particular cases where the significance of this new structural element was clear, used very local concepts to explain and predict in those cases, and then generalized from there. Thus features identified in particular structures were appropriated to develop and articulate the experimental consequences of this new aspect of chemical structure. Models of structure do play an important role in the dynamics of science by supplying concepts or features that can be appropriated to modify or develop a research program. Visual representations of structure, and models of structure more generally, act as incubators for the concepts essential to modifying and teasing out the experimental consequences of chemical structure.

One of the most dramatic changes in chemists' conception of structure occurred during the middle third of the twentieth century with the gradual realization that the *conformations* of molecules, and not just their bond connectivity, had a crucial role to play in understanding their physical and chemical behavior. A conformation is, roughly, any of the three-dimensional arrangements of atoms in space resulting from rotations around single bonds in a molecule. The development of the theory of conformations (typically called conformational analysis) occurred when features of structural formulas that had originally not been thought to have any representational significance, the 3-D arrangement of bonds or

its 2-D depiction, was recognized to represent something about the compounds that the formulas denote. With these new features available, new concepts were crafted to organize the phenomena and then articulated throughout the domain. Though it is somewhat artificial, in order to relate the development of conformational analysis to the themes of this paper, it can be understood to have occurred in two phases. First, the prior understanding of structural formulas had to be found to be insufficient, and the three-dimensional arrangement of bonds recognized to accommodate those insufficiencies. Second, once the three-dimensional orientation of bonds had been seen to be significant, the consequences of the newly enhanced conception of structure had to be developed and articulated.

40.3.1 Recognizing the Importance of Conformations

At the end of the nineteenth century, structural formulas (roughly) allowed for the generation of one distinct formula for each known, distinguishable chemical compound. The formulas used at this time included not just single bonds between adjacent atoms, but also occasional double bonds. Double bonds allow for *geometrical isomers* in which the same groups are connected to the four available positions in a double bond in two different ways. Similarly, there are two distinct ways of orienting four distinct groups around a carbon. As a result, given the number of *asymmetric double bonds* and the number of *centers of asymmetry* one could compute, using a formula due to Van't Hoff, the number of stereoisomers to be expected. This formula worked because: "It was based on the concept of *restricted* rotation about double bonds and of *free* rotation about single bonds" [40.22, p. 299]. Rotation around single bonds had to be free because otherwise one would have expected many more distinguishable isomers. In order for structural formulas to accurately map onto the results of isomer counting experiments, certain features of the models of organic molecules had to be regarded as representationally significant. For example, structural formulas had to distinguish the two distinct ways that groups can be oriented about a double bond because these represented two distinguishable compounds. At the same time, however, the experimental facts demanded that other features of the formulas not be taken to be representationally significant. The fact that there were lots of ways to produce formulas with the same bonding and orientation (differing by what we would now think of as rotations around single bonds) was explicitly not taken to be significant in the resulting structural formulas. When it came to individuating chemical compounds, the various physical models or structural formulas that could be generated by rotations about single bonds were distinct without being different. The possibility of rotational variants was an incidental feature of the symbol system that needed to be ignored when deducing the experimental facts from the models. They were not taken to reflect significant features of the target system.

Chemical structures are not frozen in time, however, and chemists aspired to add to the array of chemical and physical features that could be explained in terms of them. Chemists knew (or thought they did) that the many distinct models producible by rotations about single bonds weren't important for the individuation of chemical compounds, and thus for isomer counting experiments, but that left it open whether these differences might be employed to explain other sorts of experimental results. In fact, given the rich array of distinctions available in the models as yet uncorrelated with differences in the compounds they depicted, these distinctions would seem to have been ripe for exploration should new experimental results force modifications of the models.

Eventually, new experimental results did force such modifications. There are at least two distinct sorts of evidence that put pressure on the idea of free rotation about single bonds. First were failed isomer counting experiments, beginning in 1922, in which chemists were able to distinguish optically active forms of (unusual) compounds where, if all rotation about single bonds had been free, there should not be any such forms. More precisely, so long as all rotational variants around single bonds were regarded as indistinguishable, structural formulas did not predict the existence of distinct optically active forms, but optically distinct forms there were. The second sort of evidence came from discrepancies between the observed and measured entropy of ethane. These discrepancies "could only be explained by a barrier to free rotation about the two methyl groups" [40.22, p. 299]. These new experimental results were accommodated, eventually, by changing the representational significance of the models. Most fundamentally, the fact that a model has lots of rotational variants was now regarded as an explanatorily significant fact. Many of those differences between models of structure that had been irrelevant became differences that could correspond to differences in the energy or stability of the represented compound. For example, the distances between the *atoms* in the model (or suggested by the structural formula) became a feature used to connect differences in structure to differences in the energy or stability of the represented compound. It is because of differences in the distance relationships between the depicted atoms that rotational variants have different energies.

By imagining the atoms of a molecular *model* or *structural formula* to be interacting (either by attraction or repulsion) in a manner that varied according to the distance between them, the chemists looking to revise earlier interpretations of *chemical structure* could explain both why the rotations of ethane would be restricted and why there might be optically distinct forms of some strategically bulky organic molecules. This required the idea of nonbonding interactions between the atoms in the compound and the addition of this idea was nontrivial, depending for its plausibility on the dawning awareness of the nature of the chemical bond. But once this idea was in place, not only could the new experimental results be explained, but the success of Van't Hoff's formula could be preserved as approximately true. Most of the time, the newly postulated nonbonding interactions would be insufficient to allow for distinct forms of chemical compounds to be isolated. Sometimes, however, such distinctions would show up in physical properties that Van't Hoff hadn't been concerned to explain (like the entropy of ethane). And occasionally, in structurally rationalizable exceptional cases, these distinctions would result in failed isomer counting experiments. Instead of there being free rotation about single bonds, now the rotation about single bonds was just substantially freer than rotation about double bonds, except in certain special circumstances.

The significance of nonbonded interactions in ethane and in dramatically rotation-restricted organic molecules suggested that such interactions would also be significant to the physical properties of organic molecules in general. Thinking in terms of such nonbonding interactions required interpreting chemists' representations of structure, including structural formulas, to be significant in new ways. However, perhaps because "there was no technique available to demonstrate the phenomenon experimentally" [40.22, p. 299] this more general significance was not systematically explored until after the Second World War. Still, by this point, the rotational variants of structural formulas or physical models had demonstrated their usefulness by explaining several different sorts of novel experimental results (entropy measurements and failed isomer counting experiments) and had therefore earned their place as representationally significant.

40.3.2 Using Conformations in Organic Chemistry

Though this newly significant feature of chemists' structural models had been used to explain unexpected experimental results, it had not yet been integrated into the mainstream practice of organic chemists and used to generate results of its own. This began to change when Odd Hassel published his systematic investigations of the conformations of cyclohexane and its derivatives. Cyclohexane is an ideal experimental system for investigating the significance of conformations because, as investigation of a model will quickly show, there are only three conformations of this system (what are now called the *chair*, *boat*, and *twist-boat*), out of the infinite number that are theoretically possible, that have no angle strain (Fig. 40.3). In an earlier application of structural formulas as models ([40.10, 20], for a description of this work), chemists had shown that angle strain (or deviations from the standard tetrahedral bonding angles) was an important factor in the stability of rings. This meant that when trying to understand the behavior of cyclohexane, it was principally these three conformations that needed to be considered because all others would be energetically unfavorable. Hassel not only showed that the *chair* conformation was the most stable, but was also able to establish that the relevant nonbonded interactions were repulsive, because the chair form maximizes the distances between atoms in the ring.

Exploring a careful drawing of a chair conformation, or better yet a physical model of it, quickly reveals that there are two distinct types of bonds emanating from the carbon ring. In an obvious case of using models to introduce new conceptual distinctions, these are now called axial and equatorial bonds, according to whether they are parallel to the axis of symmetry of the molecule or in an equatorial belt around it. Furthermore, it is also clear that substituents attached axially are closer to the other atoms in cyclohexane than are substituents attached equatorially. As a result, substituted cyclohexanes generally prefer to have their substituents equatorial since this minimizes the nonbonding repulsive interactions. Hassel's work showed how the conformational preferences of cyclohexane derivatives could be rationalized using repulsive nonbonding inter-

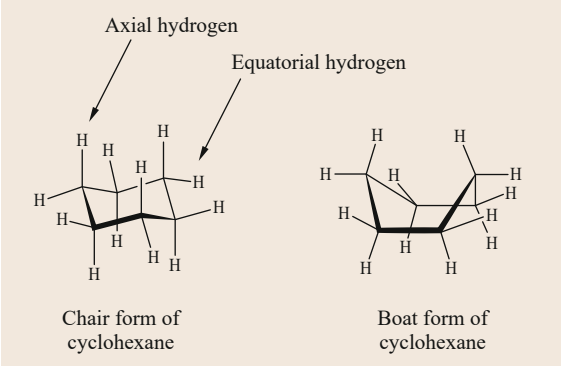

Fig. 40.3 **(a)** Chair form of cyclohexane. **(b)** Boat form of cyclohexane

actions in a way that had strong experimental support. Furthermore, he not only isolated a structural type in which the energetic implications of conformational differences were clear, but he also provided structural concepts (axial versus equatorial positions on the ring) useful in explaining the relative energies of structures of this type.

It was Barton who established the importance of conformational analysis in explaining and predicting the chemical behavior of synthetically important organic molecules. He did this by recognizing that steroids are instances of the structural type carefully studied by Hassel. The steroid nucleus consists of three cyclohexane rings fused to a five-membered ring. Because "the ring fusions of the steroid nucleus fix the conformation of the whole molecule" [40.22, p. 302] there are basically two different significant conformations of the steroid nucleus. In both of these conformations, all three of the cyclohexane rings are fixed in the *chair* form. So just as with cyclohexane itself, the significance of conformations for the behavior of steroids can be understood by considering just a few of the infinitely many possible conformations. Better still, following Hassel, the axial and equatorial substituents in the steroid nucleus can be distinguished and their relative stability rationalized in terms of repulsive nonbonded interactions.

Barton next showed how differences in the relative stability of steroids based on the conformational location of their substituents could be used to explain the chemical behavior of these molecules. Most simply, in a chemical reaction known for mechanistic reasons to result in the most stable product, one can often predict which of several candidate steroids will be preferred. Similarly, if one knows something about either the steric requirements or the geometry of the transition state, then one can often deduce which steroid will react more quickly or which product will be preferred in a reaction under kinetic control. What *Barton* did (originally in [40.23]) was to show that "an enormous literature of stereochemical fact" [40.22, p. 302] about steroids could be systematically and consistently interpreted using the conformational analysis of the steroid nucleus. He went through a variety of different results previously reported in the steroid literature and showed that the differences in rates or product distribution were what would be expected based on the conformational analysis of the steroid nucleus. This established by a sort of conscilience of inductions that, at least in the case of steroid chemistry, conformations had an important role to play in understanding chemical behavior.

Between Hassel and Barton, not only had conformations proved themselves to be useful in explaining significant chemical behavior, but also a set of structural circumstances (and concepts) had been articulated that allowed chemists to clearly discern the implications of conformation. With these resources in place, chemists were able to begin to apply these concepts in synthesis and experimental design. For example, once chemists understood why certain substitution patterns of the steroid nucleus were more stable then others, they could begin to exploit this knowledge in designing synthetic reactions. *Barton* describes how the tendency for adjacent diaxial substituents to rearrange into the more stable diequatorial form led to "a convenient route for shifting an oxygen function from one carbon atom to the adjacent carbon" [40.22, p. 304]. Similarly, because the conformation of the steroid nucleus was well known and restricted, it could be used to investigate the mechanisms of chemical reactions by effectively locking the substrate in a reaction into a particular geometry. For example, steroids were useful in establishing that "the phenomenon of neighboring group participation demands a conformational interpretation (diaxial participation)" [40.22, p. 304]. These, and other cases of application, depend on being able to recognize a set of structural circumstances in which conformational analysis is straightforward because it can be directly related back to cases that have already been successfully analyzed.

Of course, chemists were not content to apply conformational analysis just to cyclohexane and steroids. Instead, conformational analysis was articulated, from this base, along several different avenues. In the first place, it was applied to other molecules containing cyclohexane subunits, such as triterpenoids and oleanolic acid [40.22, p. 305]. Quantitative approaches were developed and this allowed for precise predictions of energy differences between conformations in these sorts of systems. Eventually, the structural limits of this approach were probed by identifying situations in which molecules with cyclohexane subunits did not behave as expected. New concepts, such as *conformational transmission* were then introduced to account for these deviations from expectation. These were refinements in the application of conformational analysis to the same basic type of system in which its clear consequences were originally discerned. Additionally, attempts were made to extrapolate the same basic approach used in analyzing cyclohexane to unsaturated six-membered rings and heterocyclic compounds. This is a case of pushing the approach into new territory. It required adapting the concepts used in the cyclohexane case to these structurally similar but importantly different new cases. New issues had to be confronted, such as how to account for the conformational implications of electron pairs. Eventually, conformations became one of the central tools

used to understand the behavior of biologically relevant molecules.

Once the rotational variants of structural models were recognized to be significant, chemists still faced the daunting task of organizing and sorting these infinite structural variations into categories that could be inferentially connected with experimental results, and eventually lead to new experiments. This was not done in a top-down way, by somehow deducing the implications of nonbonded interactions and conformations for chemical reactions. Instead, successfully doing this depended on finding a particular case where the conformational implications were clear and then generalizing and articulating from there. The concepts used to connect conformations with experiment came, initially, from considering cyclohexane. Models of cyclohexane played a crucial role in both the recognition of these concepts and their connection to experiment.

Cyclohexane was focused on because chemists already knew, from manipulation of models, that it had just a small number of strain-free conformations. This feature of cyclohexane is not shared with most other organic compounds, but it was crucial to its role in revealing the power of conformational analysis. Examination of these conformations showed that the chair form maximized interatomic distances, which led to the conclusion that the relevant nonbonded interactions were repulsive. Additionally, inspection of the chair form led to the important distinction between axial and equatorial positions about the ring, which was subsequently linked with important energy differences between structural variants of cyclohexane (including, ultimately, steroids). These conformational features of cyclohexane are also not features shared by most molecules. The concept of an axial or an equatorial substituent simply doesn't apply in most molecules, but this concept turned out to be crucial is deducing the chemical consequences of conformations. The distinctions between conformations that were actually used in order to connect this new aspect of chemical structure with experiment were available only in concrete representations of a particular structure. Models of cyclohexane are rich with discernible differences not previously identified as significant in chemical explanations. These previously neutral features supplied the concepts that eventually got connected with experimental results. It was then by generalizing, adapting, and articulating these foothold concepts that the broad applicability and novel applications of conformations were developed.

40.4 Conclusion

I hope to have established that models of structure, typically in the form of structural formulas, are essential tools for chemists. They mediate between theory and phenomena, providing the platform on which theoretical principles are both recognized and applied. They also facilitate application, as seen in the use of the theory of organic chemistry in solving synthesis problems by – in addition to its role in explanation and prediction – providing for the possible reaction pathways, strategic evaluations, plausibility assessments, and optimization that are crucial to synthetic design. Furthermore, structural models have also played important roles as sources of the concepts that chemists use to adapt their models to both theoretical and experimental developments. In sum, structural models are the keystone of the success of structural chemistry, not only because they are crucial to its theoretical content and application at any particular time, but also because of their contribution to its continued viability as a research program.

References

40.1 D.M. Bailer-Jones: Tracing the development of models in the philosophy of science. In: *Model-Based Reasoning in Scientific Discovery*, ed. by L. Magnani, N. Nersessian, P. Thagard (Kluwer, Dordrecht 1999) pp. 23–40

40.2 M. Morgan, M. Morrison: *Models as Mediators. Perspectives on Natural and Social Science* (Cambridge Univ. Press, Cambridge 1999)

40.3 R. Frigg, S. Hartmann: Models in science. In: *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta http://plato.stanford.edu/archives/fall2012/entries/models-science/ (Fall 2012 Edition)

40.4 M. Weisberg, P. Needham, R. Hendry: Philosophy of chemistry. In: *The Stanford Encyclopedia of Philosophy*, ed. by E. N. Zalta, http://plato.stanford.edu/archives/win2011/entries/chemistry/ (Winter 2011 Edition)

40.5 M. Weisberg: Qualitative theory and chemical explanation, Philos. Sci. **71**, 1071–1081 (2004)

40.6 A. Woody: Putting quantum mechanics to work in chemistry: The power of diagrammatic representation, Philos. Sci. **67**(Supp.), S612–S627 (2000)

40.7 E.F. Caldin: Theories and the development of chemistry, Br. J. Philos. Sci. **10**, 209–222 (1959)
40.8 O.T. Benfey: *From Vital Force to Structural formulas* (Houghton Mifflin, Boston 1964)
40.9 W.H. Brock: *The Chemical Tree: A History of Chemistry* (W. W. Norton, New York 2000)
40.10 W. Goodwin: Structural formulas and explanation in organic chemistry, Found. Chem. **10**, 117–127 (2008)
40.11 W. Goodwin: Visual representations in science, Philos. Sci. **76**, 372–390 (2009)
40.12 W. Goodwin: How do structural formulas embody the theory of organic chemistry?, Br. J. Philos. Sci. **61**(3), 621–633 (2010)
40.13 U. Klein: *Experiments, Models, Paper Tools* (Stanford Univ. Press, Stanford 2003)
40.14 G.W. Wheland: *The Theory of Resonance* (Wiley, New York 1944)
40.15 L. Pauling: *Nature of the Chemical Bond* (Cornell Univ. Press, Ithaca 1948)
40.16 L. Pauling, R.B. Corey: Fundamental dimensions of polypeptide chains, Proc. R. Soc. London. Ser. B Biol. Sci. **141**(902), 10–20 (1953)
40.17 W. Goodwin: Experiments and theory in the preparative sciences, Philos. Sci. **79**(4), 429–447 (2012)
40.18 W. Goodwin: Quantum chemistry and organic theory, Philos. Sci. **80**(5), 1159–1169 (2013)
40.19 E.J. Corey, X.M. Cheng: *The Logic of Chemical Synthesis* (Wiley, New York 1989)
40.20 W. Goodwin: Implementation and innovation in total synthesis, Found. Chem. **10**, 177–186 (2008)
40.21 W. Goodwin: Scientific understanding and synthetic design, Br. J. Philos. Sci. **60**, 271–301 (2009)
40.22 D.H.R. Barton: *The Principles of Conformational Analysis* (1969), Nobel Lecture, http://nobelprize.org/nobel_prizes/chemistry/laureates/1969/barton-lecture.html
40.23 D. Barton: The conformation of the steroid nucleus, Experientia **6**(8), 316–320 (1950)