

Bawden, D. (2015). Storing the wisdom: chemical concepts and chemoinformatics. *Informatics*, 2(4), pp. 50-67. doi: 10.3390/informatics2040050



**CITY UNIVERSITY  
LONDON**

[City Research Online](#)

**Original citation:** Bawden, D. (2015). Storing the wisdom: chemical concepts and chemoinformatics. *Informatics*, 2(4), pp. 50-67. doi: 10.3390/informatics2040050

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/12986/>

### Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

### Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

### Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).

*Review*

## Storing the Wisdom: Chemical Concepts and Chemoinformatics

**David Bawden**

Centre for Information Science, City University London, London EC1V 0HB, UK;

E-Mail: d.bawden@city.ac.uk; Tel.: +44-207-040-8390

Academic Editor: Peter Willett

*Received: 26 September 2015 / Accepted: 17 November 2015 / Published: 20 November 2015*

---

**Abstract:** The purpose of the paper is to examine the nature of chemical concepts, and the ways in which they are applied in chemoinformatics systems. An account of concepts in philosophy and in the information sciences leads to an analysis of chemical concepts, and their representation. The way in which concepts are applied in systems for information retrieval and for structure–property correlation are reviewed, and some issues noted. Attention is focused on the basic concepts of substance, reaction and property, on the organising concepts of chemical structure, structural similarity, periodicity, and on more specific concepts, including two- and three-dimensional structural patterns, reaction types, and property concepts. It is concluded that chemical concepts, despite (or perhaps because of) their vague and mutable nature, have considerable and continuing value in chemoinformatics, and that an increased formal treatment of concepts may have value in the future.

**Keywords:** chemistry; chemoinformatics; concepts; abstract objects; information systems; chemical structure; periodicity

---

### 1. Introduction

Chemistry has always been regarded as a particularly information-intensive science, which has had, since the establishment of modern chemical science in the nineteenth century, an unusually wide and diverse range of information sources. Over the past few decades, in common with information management generally, chemical information sources have moved from printed to digital form and a variety of associated chemoinformatics systems have been established.

One feature of chemical information and informatics systems, throughout the period of their existence, has been their use of a small number of chemical concepts as the basis for information organisation and retrieval. The purpose of this paper is to examine the nature of these concepts, with some specific examples, and the changing ways in which they have been applied for knowledge organisation and information handling. This paper takes a historical perspective, rather than focusing solely on the latest developments, as the issues involved can only be understood in the context of the development of chemical knowledge, and of the transition of chemical information systems from print to digital form. No attempt is made to identify and comment on all chemical concepts; rather, some major concepts are identified, with emphasis on those of greatest significance for chemoinformatics.

We first examine the nature of concepts in general usage, in philosophy, and in the information sciences, as a precursor to a consideration of specifically chemical concepts.

## 2. Concepts

Typical definitions of “concept” in general English language dictionaries include: “a thing thought of, a general notion, an idea” (Chambers); “an abstract idea” (Oxford); “an idea of what something is, and how it works, an abstract or generic idea generalized from particular instances” (Merriam–Webster). The commonly held idea, therefore, is that a concept is an idea of an abstract and general nature. Somewhat more definitely, a concept is often regarded as a general idea of a class of entities defined by some essential characteristic features. This leads to the more sophisticated philosophical understanding of a concept.

In philosophy, concepts have been the subject of analysis and study for many centuries [1]. There is some agreement about the general nature of concepts, but much disagreement about the details. “Concepts are the constituents of thought. However, the nature of concepts—the kinds of things concepts are—and the constraints that govern a theory of concepts have been the subject of much debate” ([2], p. 1). “Today there is no consensus about what concepts are, which theories of concepts are most important, or how theories of concepts should be classified” ([3], p. 1519) As to what concepts are, specifically their ontologies, philosophers today tend to take one of three positions: concepts as mental representations, concepts as abilities, and concepts as abstract objects [2,4,5]. The first of these propositions holds that a concept is a structured representation of some piece of reality in the mind of a person; a popular perspective in cognitive science. The second associates a concept with abilities possessed by a conscious agent. The concept of “cat” for example, could be equated with an ability to distinguish cats from other animals. The third associates concepts with abstract entities, the “senses” introduced by the philosopher Frege [6,7].

Considered from the viewpoint of information systems and information retrieval, the first of these positions poses problems, since concepts in this sense are necessarily personal and subjective. Similarly, the second perspective is associated with the abilities of particular people at particular moments in time. The third perspective, that concepts should be understood as abstract objects, seems most appropriate for our purposes. It allows for concepts to be objective entities, in as much as the same concept can be possessed by different people at the same time, and by the same person over time, and they may therefore be used for communication and the retrieval of information. This does not require the concept to be defined formally, or with great precision, which is in line with the common

sense understanding of a concept as a somewhat vague and general notion. This has evident significance for dealing with concepts in information systems, including chemical information systems, as will be discussed later.

Abstract objects have been of interest to philosophers, including philosophers of science, because of their ontological status, lacking both concrete substance and also a location in space and time [8,9]. They are of particular concern in the philosophy of mathematics, a subject arguably entirely devoted to the study of such objects [10]. As with concepts generally, there is active debate about the nature and status of abstract objects. However, regardless of such on-going debates, concepts, understood as abstract objects, and usually somewhat ill-defined and fuzzy, are an important part of the theory-base of all sciences, including chemistry.

### 3. Concepts in Information Science

The idea of concept is also important, theoretically and practically, for the information sciences, since concepts are often the basis of resource description, classification, indexing, and retrieval. As Hjørland puts it: “Concepts seem to be all-present and pervasive in library and information science (LIS). Concepts are what are behind users’ questions, in the understanding of intermediaries and in the information being sought and retrieved. In addition, the goal of information retrieval technology is to identify information corresponding to a certain concept, but which is often hidden under different labels and symbols that mix up different concepts and thus produce noise as well as a lack of recall. Most directly, concept theory is related to knowledge organisation, to the development of classification systems, taxonomies, thesauri, ontologies and so on. Different theories of concepts have implications for how LIS investigates its core topics and, therefore, the theoretical assumptions have to be examined” ([3], p. 1527).

Thus, a concept, understood as “an abstract notion or idea” is one of the sets of entities, which may be the subject of works in FRBR (functional requirements for bibliographic records), the basis for modern standards of cataloguing and resource description [11,12]. For classification, a concept is “the abstract notion of an entity, topic or class, as opposed to its name”, and “a classification is a concept-based system, since its classes are intended to reflect the abstract idea of the subjects they refer to” ([13], p. 30 and p. 379). Indexing (and, to an extent, abstracting), other than that which simply reuses terms in the text of the original document, is necessarily conceptual, since it involves an initial conceptual analysis, to identify concepts present, regardless of how they may have been named in the original [14,15]. A concept may have many names—synonyms, popular slang and expert terminology, different language variants—and there may be no neat link between a concept and a name. This accounts for many of the practical difficulties of the indexing and retrieval of conceptual material, compared with that of material able to be described in a complete and unambiguous manner, such as chemical substances, in the form of distinct elements or “small-molecule” compounds. (It is necessary to be specific about the nature of the substance, as an anonymous referee points out, since materials such as polymers, alloys, and mixtures cannot be described in this way.)

The idea of concept for the information sciences has been debated, generally drawing on the philosophical positions noted above, without any exact or formal definition having been agreed upon [16–19]. Hjørland makes a particularly strong plea for the analysis of different theories of

concepts in terms of their relevance for information science [3]. He emphasises that the way in which concepts are understood and applied in information science will be influenced by the overall philosophical and theoretical approaches and assumptions in the subject domain in which information systems are being applied. When we consider chemical information systems, therefore, the kind of concepts they must deal with will be strongly influenced by the nature of the knowledge and concepts in chemistry itself.

#### 4. Concepts in Chemistry

Chemistry is typically regarded as a “central science” between physics and biology, thus having both a quantitative and qualitative nature, a “classifying science”, and a conceptual science, with a small number of basic concepts, from which are built a larger number of secondary concepts [20–24].

The typical history of any scientific concept follows from the emergence of the concept as a vague idea, followed by a gradual clarification and a more detailed expression, within the development of a coherent theory of which the concept is a part [25]. It might therefore be thought that concepts are an unsatisfactory stopping point *en route* to a more satisfactory formal theory, particularly where the concept is a qualitative or semi-quantitative summary of a property, which might be calculated in exact quantitative terms, for example by denoting a chemical moiety as “electron withdrawing”. However, this is by no means always the case, and certainly not in chemistry where qualitative and semi-quantitative concepts have always played, and continue to play, an important role in understanding and communicating the subject, as will be discussed below. Nor do older concepts necessarily lose relevance; see, for instance, Needham’s philosophical analysis of the concept of chemical substance, which might have been thought to be rendered redundant by more modern concepts of chemical structure [26].

Rouvray emphasises that generic “chemical concepts are often vague and surprisingly ill-defined. In this respect, chemical concepts do not differ from concepts in general this fact, however, need not be regarded as detrimental to the future development of chemistry” ([27], p. 11). According to Rouvray, the only chemical concepts that are *not* to an extent vague, general, and expressed in terms of fuzzy thinking, are those that have a formal mathematical definition, such as the symmetry group or the density matrix. In this respect, chemical concepts are good exemplars of concepts in general, as discussed above. Such concepts are created by, and intuitively understood by subject experts, but may not be readily detected, or processed by computers; hence, there arise some of the issues of concepts in chemoinformatics systems discussed below.

Chemistry is typically understood as the science that deals with the study of matter, with the appearance and behaviour of different forms of matter, and with the transformations that which matter can undergo [23]. Its most basic concepts are therefore necessarily those dealing with matter (*element, atom, compound, molecule*), with transformations (*reaction*), and with the nature and effects of substances (*properties*). These are indeed the essential concepts for the teaching of chemistry [22]; as an example, see the recent textbook by Rice [24].

Two very general concepts, both stemming from the development of modern chemistry in the nineteenth century with its concepts of atoms and molecules, are central to modern chemistry, and the most powerful organising principles for the communication of chemical information. They are

*periodicity* and *chemical structure*, the latter itself based on the additional fundamental concept of the *chemical bond*.

The idea of periodicity, based initially on the observation of periodically repeating properties in a table of elements arranged by increasing atomic weight, and later rationalised in terms of atomic number and electron shells, forms the basis for the periodic table, the major organising principle for chemistry as a whole [28,29]. Despite its ubiquity, it is worth noting that the periodicity concept continues to evolve [30]. There have been numerous variants of the periodic table for different purposes, with new ones still being developed, and no variant captures perfectly all the relevant information and issues [31,32]. Despite this limitation, typical as we have seen, of all concept-based principles, the periodical table and periodicity in general, has been and remains one of the fundamental concepts in the organisation of chemical information.

The concept of chemical structure, that the nature and properties of substances are determined by the arrangement of their atoms and inter-atomic bonds, is the second major conceptual organising principle of modern chemistry [33,34]. Applicable in all aspects of the subject, it has been particularly powerful in understanding the complexities of organic chemistry. Although accounts of chemical structure with much greater physical and mathematical sophistication have been devised, the basic concept, denoted by simple graphical representation, is still sufficient for many purposes including the organisation and communication of chemical knowledge. As Lewis puts it “No generalization of science, even if we include those capable of exact mathematical statement, has ever achieved a greater success in assembling in simple form a multitude of heterogeneous observation than this group of ideas which we call structural theory” ([35], pp. 20–21).

An associated concept is that of *structural similarity*: the idea that substances whose structures have some similarity will have similar natures and properties. This applies whether the similarity is local due to the presence of the same important structural sub-unit, or global due to an overall structural similarity. This concept, originally due to Crum Brown [36], has proved very powerful in chemistry, and has had practical consequences for informatics systems.

Although these two concepts, periodicity and chemical structure, have proved remarkably powerful and fruitful, they share the properties of all concepts in that they are not formally defined so as to be entirely precise, nor are they complete in the sense of encapsulating and expressing all information about a chemical entity or transformation. Therefore, a wide variety of other concepts, qualitative and semi-quantitative, have sprung up around them. At the risk of over-simplification, we may categorise these as:

- Two-dimensional patterns of molecular structure
- Two-dimensional patterns of bonding
- Three-dimensional patterns of structure
- Types of reaction
- Properties of a substance or substructure

In the first category, we have concepts of functional groups and structural moiety, defined as patterns of atoms and bonds in a localised part of a molecule. These share the general features of concepts by virtue of being not always precisely defined, and being chosen for pragmatic reasons. Some groups are structure-based: amines, aldehydes, ketones, phenols, *etc.* Others are property-based,

and may encompass structurally diverse moieties, such as electron-withdrawing groups, nucleophiles, leaving groups, *etc.*

In the second category, we find concepts related to the nature of bonding within a general pattern of atoms. Examples of this are aromaticity, tautomerism, and non-classicality. The best known is certainly aromaticity, a concept that has stimulated much debate over the years [37–42]. The name was originally given to a group of organic compounds distinguished by their pleasant smell, later applied to those undergoing, or not undergoing, certain reactions, and later rationalised theoretically by invoking the idea of the delocalisation of electrons. Subsequent experimental study and analysis through a variety of theoretical approaches has given more sophisticated understanding of the concept, with a consequent and continuing broadening in the types of substances said to have “aromatic character”. Despite the interest in, and value of the concept, Ollis was able to write in 1967 “no satisfactory and generally acceptable definition of the term “aromaticity” has been given, but we doubt that this is necessary. We all know (or do we know?) what “aromaticity” means, but this has not inhibited the use of new terms such as pseudo-aromaticity, non-benzenoid aromaticity, homoaromaticity, antiaromaticity, and even antihomoaromaticity” ([43], p. 3). Similar remarks could be made today. Aromaticity shows particularly clearly the nature and value of concepts in chemistry, in that “the concept of aromaticity pervades many areas of organic chemistry ... [but] ... it can be, and has been, argued that the concept of aromaticity is so vague to serve no useful purpose ... the concept may be vague, but this is not necessarily a disadvantage” ([40], pp. 285, 293 and 296). An ill-defined and vague concept, surrounded by a penumbra of related concepts and terminology, which is nonetheless of great value for organising the subject’s knowledge base is likely to pose a particular problem for information systems, and the treatment of the aromaticity concept by such systems will be discussed later.

Tautomerism is a concept of similar nature. It is generally understood to be state of equilibrium that exists between two structural forms of a compound differing only in the position of a mobile grouping, typically a hydrogen atom, although as Sayle shows, it is a problematic concept because of the limitation of the valence approximations underlying the idea of the chemical structure [44]. Another similar concept is non-classicality, involving structures, typically reactive intermediates, which do not follow the norms of chemical bonding.

In the third category, dealing with the three-dimensional structure of substances, we have the concepts of chirality, stereoisomerism (geometrical and optical), stereoselectivity of reaction, *etc.* [45].

In the fourth category are the classifications of reaction type, for which a variety of intellectual and automated means have been devised [46–48]. These may be: generally indicative of the results of the reaction, e.g., rearrangement, substitution, oxidation/reduction; descriptive of a reaction mechanism, e.g., nucleophilic aliphatic substitution, halogen abstraction; or indicative of some particular characteristic of the reaction, e.g., stereoselective, metal-ion catalysed. They may be structure-based, e.g., oxidative cleavage of furans, or they may not.

In the last category are concepts relating to the properties of chemical substances as a whole, whether a bulk property of the whole substance, such as boiling point or partition coefficient, or a property induced by a structural pattern within it. A well-known example of the latter is the “pharmacophore” or “pharmacophoric pattern”, a part of a molecule causing the substance as a whole to exhibit a specific biological property. This concept has been a contested one over time, with its nature changing from a specific pattern of atoms and bonds, rather like the functional groups of the

first category, to one of a pattern of more abstract spatial features [49]. The “toxicophore” causing a specific toxic property is a closely related concept. Another example is the concept of the “synthon” or, in more recent usage, “retron”, typically a ring or a pattern of bonding, which makes it a suitable candidate for synthesis planning [48,50,51].

Concepts in all five of these categories have been, and remain, important in the organisation of chemical knowledge. Goodwin, for example, notes that the whole edifice of modern organic chemistry has been built since the nineteenth century on a small set of concepts, particularly chemical structure, functional groupings, and reaction types [52]. They have been equally important for the communication of chemical information, and we will see that all of them have been used within chemical information systems. It should be noted that this is not intended as a comprehensive list of concepts of importance in chemistry, or chemoinformatics. The focus of this article is on those concepts that are central and unique to the science of chemistry. This is not intended to minimise the importance of more general concepts, which are of value in chemistry. One example of these is complexity, a multifaceted and contested concept in itself [53,54], which is of increasing interest in chemistry. For an overview, see Bonchev and Seitz [55], and for examples, see [50,51,56]. Other general concepts of chemical relevance are symmetry and group theory [57]. There are also important concepts within chemoinformatics itself, such as the canonical numbering of atoms. Again these are not the focus of this paper.

## 5. Representations of Chemical Concepts

We now turn to the consideration of how chemical concepts of this kind—those uniquely associated with chemistry and chemical knowledge—have been dealt with in chemical information systems. As a prelude to doing so, it is worth noting that the development of concepts themselves, and of their representation in a format such as a diagram for purposes of communication, have often gone closely together [58]. Lewis’s “dot notation” for his fundamentally important concept of a chemical bond formed by a shared pair of electrons is one such example [59]. Arguably the best-known example is Crum Brown’s “structure diagrams”, conveying the concept of organic chemical structure [60,61], following the earlier graphical structure representations due to researchers such as Kekulé, Butlerov, and Couper [33,34]. Another example is the mesomeric symbolism devised by Ingold to explain the characteristics of molecules not captured by conventional structure diagrams [34].

The ubiquity and usefulness of such representations has led to their being regarded as the natural language of organic chemistry, or as Goodwin puts it, “In (organic chemistry) it is not plausible to regard (structure) diagrams as simply heuristic aids for expressing or applying what is essentially a linguistic theory. Instead, it is more plausible to think of linguistic representation as supplementing theories whose principal expression is diagrammatic” ([34], p. 621). Because the structure diagram representation has been developed and extended over time, “the evolving norm-governed practice of producing and employing structural (diagrams) “stores the wisdom” of organic chemists because these norms have been continually refined structural (diagrams) and the norms constraining their use embody the theory of organic chemistry” ([34], p. 622).

Nor is chemistry, at least for the most part, governed by mathematical theories. A great deal of explanation in chemistry is based on qualitative, or semi-quantitative theorising, very often based on



information implicit in structure diagrams [62,63]. This is as true for explanations about transformation and reactions as it is about the substances themselves.

The same is true for inorganic chemistry, whose theories are again neither primarily linguistic nor mathematical. Here, the periodic table, embodying and representing the periodicity concept “summarizes relationships between the elements and plays a crucial role in organizing information about them” ([23], p. 13).

The relation between concepts, representations, and information handling is a crucial one for chemistry. Concepts of this sort are, as Goodwin puts it, both descriptions and models, and therefore play a distinctive part both in the practice of the science, and in the systems that make its knowledge available [34]. We now turn to the representation of concepts within chemical information systems, and to their use in the organisation and retrieval of information.

## 6. Concepts in Chemical Information Systems

The organisation and communication of chemical information began as a reaction to the establishment of chemistry as a recognised science in the nineteenth century [64]. A wide range of printed information resources—journals, abstracts, indexes, monographs, textbooks, and data compilations—were founded, some of which exist to this day. From the 1960s onwards, the digital computer began to be applied to chemical information systems, resulting in an equally wide range of digital systems; some the equivalents and successors of the major printed information resources, some entirely novel in nature, offering processing and access facilities not conceivable in the print context. For a review of these developments, see Willett [65], and for accounts of numerous aspects of the state-of-the-art at various stages to the present day, see the reviews by Lynch, Harrison, Town, and Ash [66]; Bottle [67]; Ash and Hyde [68]; Ash, Chubb, Ward, Welford, and Willett [69]; Ash, Warr, and Willett [70]; Leach and Gillet [71]; and Currano and Roth [72]. These texts also describe the development of the major chemical information sources mentioned below, and others, some of which have been converted to digital form after many decades as major printed reference sources [73,74].

To a very large extent, the kinds of chemical concepts referred to above have been applied as organising principles for chemical information and informatics systems of all kinds. We will consider how this has been achieved, using the rough typology of chemical concepts derived above:

### *Basic concepts*

- Substance
- Reaction
- Property

### *Organising concepts*

- Chemical structure
- Periodicity

### *Detailed concepts*

- Two-dimensional patterns of structure
- Two-dimensional patterns of bonding

- Three-dimensional patterns of structure
- Reaction types
- Property concepts

We may note that a number of processes and techniques, which have been applied within chemical informatics systems, have been named so as to indicate some formal handling of concepts. Examples are: formal concept analysis, a data mining technique based on information theory [75,76]; concept-based ontologies [77,78]; and search outputs described as *ad hoc* concepts [79]. Although these methods and approaches may make some use of the kinds of concepts described in this paper, they do not fall within the scope of this discussion. In particular, what might seem to be the advantages of the more formal ways of defining concepts do not seem to have gained traction in face of the familiar, though sometimes vague, chemical conceptions?

### 6.1. Basic Concepts

The three basic concepts noted above have been used in two main ways within chemical information systems.

First, they have commonly determined the nature of particular information sources. Many sources, both printed and digital, have been created for the sole purpose of storage and retrieval of information on chemical substances *per se*, on chemical reactions, and on properties of substances. The content and arrangement of such systems are primarily or wholly determined by which of the three basic concepts they are oriented towards. Well-known examples are the Chemical Abstracts Service substance databases, the Beilstein, Landolt–Börnstein and Gmelin compilations, the CASREACT and Reaxys reaction databases, and the Cambridge Structural Database of crystal structures. For discussion and examples, see [69,70,80–86].

Second, printed or digital systems, arranged around these concepts are likely to be geared for specific look-up of a known item: information about a specific chemical substance; about a specific reaction, defined by name or by the nature of the reactants; or about a particular property of a particular substance. Indeed, in pre-digital times, this was often the only means of accessing information in such sources. The advent of digital files, and the application of substructure searching in this kind of source, has meant that substances may be identified by virtue of their containing a particular partial atom-bond structure; but, nonetheless, what is returned is a set of defined substances. Similarly, digital property compilations lend themselves to more sophisticated data manipulations than simply looking up a property value for a substance, but the basic conceptual structure remains unaltered.

### 6.2. Organising Concepts

While the basic concepts have generally determined the content and arrangement of chemical information sources, the two organising concepts have generally determined the information organisation within them, and the access provided to their user.

The importance of organisation by, and access through, chemical structure has already been noted, particularly for organic chemistry. In pre-digital times, this was provided through indexes of

systematic nomenclature, molecular formula, structure notations, and the like. Digital systems complement this, through graphical searching of structures and substructures, making tangible the idea of chemical structure as language [87]. The associated concept of structural similarity has been applied with considerable success to systems for information retrieval and for structure-property correlation [88–93].

The great success of structure and substructure searching by graphical means has, ironically, led to organic structure-based searching being regarded as something different from “topic” or “concept” searching, which is assumed to rely on text terms, a view which has persisted over a long period [94,95]. This is despite the fact that structure-based search relies for its success on what we have seen is one of the primary concepts of chemistry.

Similarly, the periodic table concept has been the main intellectual tool for organising and providing access within collections of information in inorganic chemistry. A well-known example is Gmelin’s Handbook, established in 1819, which, ever since the general acceptance of the usefulness of the table, has used it as the basis for its classification.

### 6.3. Detailed Concepts

These more detailed chemical concepts have all played significant parts in chemical information handling. In some cases this has brought to light issues with the concepts themselves. Both these points are illustrated by examples below. It is worth noting that there is often no sharp distinction between systems in the use of the different categories. For example, as noted below, functional groups and reaction types are often used together in organising organic reaction information.

#### 6.3.1. Two-Dimensional Patterns of Structure

Patterns of this nature, typically denoted as functional groups or ring systems, have been one of the principle tools by which chemical knowledge has been organised, communicated, and retrieved, in textbooks, information compilations, and databases. A well-known example of such a compilation, originally printed and now a digital database, is Patai’s Chemistry of Functional Groups (now part of the Wiley Online Library).

This sort of pattern was adopted in the first generations of cheminformatics systems. Many fragment codes, an early and ambiguous form of digital structure representation, were based on functional groups as being natural chemically meaningful and significant sub-units used for indexing and retrieval, and for structure-property correlation, and still retaining some value particularly for patents’ information systems [66,69,70]. Similarly, early chemical notations, most particularly Wiswesser Line Notation (WLN), were designed to incorporate these intuitively natural structural groupings as discrete elements within the notation, by allocating a single symbol to stand for a multi-atom/bond group, for example R for a benzene ring in WLN [96]. This, however, came at the cost of an inefficiency in processing such notations, a need for expert human intervention in their use, and a lack of flexibility in specifying searches. For these reasons they have been superseded by representations in which the structural concepts are not explicit. This has many advantages, but one may muse on whether something conceptually valuable has been lost, or perhaps this is another illustration of the way in which chemical concepts and their representations change with changing contexts.

The functional group concept is also applied in reaction information systems, where the reaction concept is expressed as the reactant and/or product functionality. For example, the CASREACT system is searchable by functional groups in reactant and/or product [97], examples of the groups used being alcohols, carboxy derivatives, hydrazines, and unsaturated aldehydes, while the Science of Synthesis (formerly Houben Weyl) compendium is organised by the functionality, which is the end point of synthesis [81].

It may be noted that, in all the examples given above, the functionality concepts are expressed as an alphabetical lists of terms. While this is convenient for most users who will be familiar with the terminology, it fails to offer the assistance of any classificatory structure. However, it is likely that the vagueness of many of these familiar structural concepts would cause problems in setting up any hierarchical arrangement.

Going beyond functional groups simply defined by their atom-bond structure, chemoinformatics systems have used a variety of qualitative conceptual categorisations for groups. For example, early synthesis planning systems, having identified structural groups then defined them in terms such as “electron withdrawing” or “electron donating”, or assigned a measure of their activity as a base, nucleophile, *etc.* [98]. Later generations of these systems have further developed such categorisations [50,51,99]

### 6.3.2. Two-Dimensional Patterns of Bonding

The much-debated concept of aromaticity has led to particular problems for chemoinformatics systems, and to a number of alternative approaches to deal with it [100]. Some systems have dealt with this by declaring a special “alternating” bond type, or a somewhat more all-embracing “aromatic” bond type, both based on perception of alternating bond patterns. Others base the definition on a count of pi-electrons; some following the Hückel  $4n + 2$  rule, others including the “anti-aromatic”  $4n$  count. Some recognise aromaticity only in 6-membered rings, or in even-membered rings, and others make no such distinction.

Tautomerism poses similar problems for chemoinformatics systems [44], and again is dealt with in different ways. Some systems store only one tautomer, other store both. Still others store a single structure with a “tautomer” bond type. Warr gives a detailed comparison of the treatment of tautomerism by 27 different system and database providers [101].

These difficulties and the consequent problems for consistency of the representation of the same substance in different systems are a particularly clear example of the practical problems posed for chemoinformatics systems by the nature of some chemical concepts, as discussed above.

### 6.3.3. Three-Dimensional Patterns of Structure

Chemoinformatics systems have developed considerable capability in calculating and displaying the three-dimensional structures of molecules [71]. However, this detail is not necessarily of help in representing the concepts of three-dimensional structures, particularly for organisation and retrieval. For these purposes, systems typically rely on representations and notations developed for the communication of information in printed products, *i.e.*, representing three dimensions on a two-dimensional surface, in such a way as to capture the significant concepts of chirality and

stereoisomerism, whether this be by ball-and-stick or space-filling models, or by representation of shape or electrostatic field [71]. The *cis-trans*, R-S, and E-Z notations are most commonly used. However, as with the previous category, these representations are far from fully satisfactory, see, for example, [102], and again, informatics systems have handled these concepts inconsistently—as an early example, see [103], and, as a later example, see [104]—while others ignore them entirely. Here is another case where lack of formality in concept description, convenient though this is for general chemical purposes, causes problems for informatics systems.

#### 6.3.4. Reaction Types

Reaction information systems typically arrange their contents by reaction classification, by reaction type, defined by structural change (of atoms and/or bonds), or by reaction mechanism [46,48,81]. Identification of structural changes and of mechanism may overlap in some reaction descriptions, and functional group specification of reactant and product are often included in the descriptions. For example, typical sections in the Organic Reaction Mechanisms compilation (now part of the Wiley Online Library) are: oxidation and reduction, nucleophilic aromatic substitution, reactions of aldehydes and ketones and their derivatives, and molecular rearrangements-pericyclic reactions. Examples of the “reaction type” access in the Organic Reactions database are: addition, acylation, elimination, peroxidation, and thioacetal cleavage [81].

These are plainly based on the chemically intuitive concept of reaction type, and are generally presented in the form of an alphabetical list. As with the functionalities in Section 6.3.1, this will be convenient for those familiar with the terminology of the concept, but lacks the facility of a more formal, and perhaps hierarchical, classification. Numerous formal classifications of chemical reactions have been devised for different purposes [46–48] but none have achieved wide use in systems of this type. This may be seen as another reflection of the nature of chemical concepts, and the possible advantage of their pragmatic and general nature.

#### 6.3.5. Property Concepts

A good example of this, as noted above, is that of the pharmacophore, or pharmacophoric pattern, and much effort has gone into devising systems to identify such patterns in substance databases [71,105]. The abstract concept is here converted to a concrete and usable set of molecular descriptors, by analysis of the pattern of three-dimensional structural features (the first of the five categories above), such as acid, base, aromatic ring, hydrogen-bond donor, hydrophobic group, *etc.* [106]. Similarly, the identification of synthetically significant structural elements is an important aspect of systems to aid organic synthesis [74,98,107].

## 7. Conclusions

It is evident that concepts have been, and remain, vital to the progress of chemistry, particularly in the organisation and communication of chemical knowledge, as well as the development of chemistry's theory-base. This is despite the fact that these concepts are, at least in some cases, acknowledged to be vague, and to lack a consensus as to their exact meaning. However, it seems that

this very vagueness and mutability may be a strength in enabling chemical concepts to remain relevant, to show analogy and similarity, and to support creative insights.

Given this, it seems essential for the explicit inclusion of such concepts within modern information systems, and indeed concepts play significant roles in chemoinformatics systems. However, their representation has caused problems and inconsistencies. The rapid developments in these systems have stemmed almost entirely from innovations in the algorithmic handling of structure representations, for structure and substructure searches, similarity and dissimilarity/diversity analysis, and structure-property correlation. There has been no equivalent advance in the nature, handling, or concepts within such systems. It may be that current developments in the use of the semantic web for chemical information, particularly with chemically relevant ontologies, may mark the beginning of such a development [108,99].

We may reflect that many of the major advances in chemoinformatics systems have resulted from the application of formal methods, from set theory, graph theory, topology, information theory, probability, and other areas. Although some formal methods of conceptual analysis have been attempted in the chemoinformatics context, they have not yet led to widely applicable improvements in system capabilities. Perhaps a combination of the relatively new philosophical interest in concepts as abstract objects, a renewed focus on concepts in the information sciences, and the interest of philosophers of science in chemical concepts, may find fruitful application in chemoinformatics in the future.

### Acknowledgments

I thank two anonymous referees for helpful comments, and in particular for raising the issue of the variety of kinds of chemical concepts.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Weitz, M. *Theories of Concepts: A History of the Major Philosophical Tradition*; Routledge: London, UK, 1988.
2. Margolis, E.; Laurence, S. Concepts. In *The Stanford Encyclopedia of Philosophy*, Spring 2014 ed.; Zalta, E.N., Ed.; Available online: <http://plato.stanford.edu/archives/spr2014/entries/concepts> (accessed on 17 September 2015).
3. Hjørland, B. Concept theory. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 1519–1536.
4. Margolis, E.; Laurence, S. The ontology of concepts-abstract objects or mental representations? *Noûs* **2007**, *41*, 561–593.
5. Van Loocke, P. *The Nature of Concepts: Evolution, Structure and Representation*; Routledge: London, UK, 1999.
6. Peacocke, C. *A Study of Concepts*; MIT Press: Cambridge, MA, USA, 1992.

7. Zalta, E. Fregean senses, modes of presentation, and concepts. *Philos. Perspect.* **2001**, *15*, 335–359.
8. Hale, B. *Abstract Objects*; Blackwell: Oxford, UK, 1987.
9. Ney, A. *Metaphysics: An Introduction*; Routledge: Abingdon, UK, 2014.
10. Parsons, C. *Mathematical Thought and Its Objects*; Cambridge University Press: Cambridge, UK, 2008.
11. Zhang, Y.; Salaba, A. *Implementing FRBR in Libraries*; Neal-Schuman: New York, NY, USA, 2009.
12. Maxwell, R.L. *Maxwell's Handbook for RDA*; Facet: London, UK, 2014.
13. Broughton, V. *Essential Classification*, 2nd ed.; Facet: London, UK, 2015.
14. Foskett, A.C. *The Subject Approach to Information*, 5th ed.; Facet: London, UK, 1996.
15. Lancaster, F.W. *Indexing and Abstracting in Theory and Practice*, 3rd ed.; Facet: London, UK, 2003.
16. Marradi, A. The concept of concept: Concepts and terms. *Knowl. Org.* **2012**, *39*, 29–54.
17. Szostak, R. Complex concepts into basic concepts. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 2247–2265.
18. Friedman, A.; Thellefsen, M. Concept theory and semiotics in knowledge organization. *J. Doc.* **2011**, *67*, 644–674.
19. Smiraglia, R.P.; van den Heuvel, C. Classifications and concepts: Towards an elementary theory of knowledge interaction. *J. Doc.* **2013**, *69*, 360–383.
20. Knight, D. *Ideas in Chemistry: A History of the Science*; Athlone Press: London, UK, 1992.
21. Rouvray, D.H. *Concepts in Chemistry: A Contemporary Challenge*; Research Studies Press: Taunton, UK, 1997.
22. Taber, K. *Chemical Misconceptions: Prevention, Diagnosis and Cure-Theoretical Background*; Royal Society of Chemistry: Cambridge, UK, 2002; Volume 1.
23. Atkins, P. *Chemistry: A Very Short Introduction*; Oxford University Press: Oxford, UK, 2015.
24. Rice, J.E. *Organic Chemistry Concepts and Applications for Medicinal Chemistry*; Academic Press: San Diego, CA, USA, 2014.
25. Chalmers, A. *What is This Thing Called Science?* 4th ed.; Open University Press: Buckingham, UK, 2013.
26. Needham, P. Chemical substances and intensive properties. *Ann. N. Y. Acad. Sci.* **2003**, *988*, 99–113.
27. Rouvray, D.H. Are the concepts of chemistry all fuzzy? In *Concepts in Chemistry: A Contemporary Challenge*; Research Studies Press: Taunton, UK, 1997; pp. 1–15.
28. Scerri, E.R. *The Periodic Table: Its Story and Significance*; Oxford University Press: Oxford, UK, 2007.
29. Marchese, F.T. Periodicity, visualization and design. *Found. Chem.* **2013**, *15*, 31–55.
30. Babaev, E.V.; Hefferlin, R. The concepts of periodicity and hyperperiodicity: From atoms to molecules. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D.H., Ed.; Research Studies Press: Taunton, UK, 1997; pp. 41–100.
31. Hjørland, B. The periodic table and the philosophy of classification. *Knowl. Org.* **2011**, *38*, 9–21
32. Scerri, E. Trouble in the periodic table. *Educ. Chem.* **2012**, 13–17. Available online: [http://www.rsc.org/images/Scerri%20Trouble%20PT\\_EiC\\_January2012\\_tcm18-212413.pdf](http://www.rsc.org/images/Scerri%20Trouble%20PT_EiC_January2012_tcm18-212413.pdf) (accessed on 19 November 2015).

33. Villaveces, J.L.; Daza, E.E. The concept of chemical structure. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D.H., Ed.; Research Studies Press: Taunton, UK, 1997; pp. 101–132.
34. Goodwin, W. How do structural formulas embody the theory of organic chemistry? *Br. J. Philos. Sci.* **2010**, *61*, 621–633.
35. Lewis, G.N. *Valence and the Structure of Atoms and Molecules*; The Chemical Catalog Company: New York, NY, USA, 1923.
36. Crum Brown, A. On chemical constitution and its relation to physical and physiological properties. *Philos. Mag.* **1869**, *37*, 395–400.
37. Chemical Society. *Aromaticity: An International Symposium*, full proceedings; Chemical Society: London, UK, 1967.
38. Badger, G.M. *Aromatic Character and Aromaticity*; Cambridge University Press: Cambridge, UK, 1969.
39. Randic, M. Aromaticity and conjugation. *J. Am. Chem. Soc.* **1977**, *99*, 444–450.
40. Garratt, P.J. *Aromaticity*. Wiley: New York, NY, UK, 1986.
41. King, R.B. Aromaticity from planar organic molecules to polyhedral inorganic species. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D.H., Ed.; Research Studies Press: Taunton, UK, 1997; pp. 152–203.
42. Krygowski, T.M.; Cyranski, M.K. *Aromaticity in Heterocyclic Compounds*; Springer: Berlin, Germany, 2009.
43. Ollis, W.D. Introduction. In *Aromaticity: An International Symposium*; Chemical Society: London, UK, 1967; pp. 1–3.
44. Sayle, R.A. So you think you understand tautomerism? *J. Comput. Aided Mol. Des.* **2010**, *24*, 485–496.
45. Gilat, G. The concept of structural chirality. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D.H., Ed.; Research Studies Press: Taunton, UK, 1997; pp. 325–351.
46. Bawden, D. Classification of chemical reactions: Potential, possibilities and continuing relevance. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 212–216.
47. Kraut, J.; Eiblmaier, J.; Grethe, G.; Löw, P.; Matuszczyk, H.; Saller, H. Algorithm for reaction classification. *J. Chem. Inf. Model.* **2013**, *53*, 2884–2895.
48. Warr, W.A. A short review of chemical reaction database systems, computer-aided synthesis design, reaction prediction and synthetic feasibility. *Mol. Inform.* **2014**, *33*, 469–476.
49. Güner, O.F.; Bowen, J.P. Setting the record straight: The origin of the pharmacophore concept. *J. Chem. Inf. Model.* **2014**, *54*, 1269–1283.
50. Cook, A.; Johnson, P.A.; Law, J.; Mirzazadeh, M.; Ravitz, O.; Simon, A. Computer-aided synthesis design: 40 years on. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 79–107.
51. Bøgevig, A.; Federsel, H.-J.; Huerta, F.; Hutchings, M.G.; Kraut, H.; Langer, T.; Löw, P.; Oppawsky, C.; Rein, T.; Saller, H. Route design in the 21st century: The ICSYNTH software tool as an idea generator for synthesis prediction. *Org. Process Dev.* **2015**, *19*, 357–368.
52. Goodwin, W. Scientific understanding and synthetic design. *Br. J. Philos. Sci.* **2009**, *60*, 271–301.
53. Mitchell, M. *Complexity: A Guided Tour*; Oxford University Press: New York, NY, USA, 2009.



54. Bawden, D.; Robinson, L. "Waiting for Carnot": Information and complexity. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2177–2186.
55. Bonchev, D.; Seitz, W.A. The concept of complexity in chemistry. In *Concepts in Chemistry: A Contemporary Challenge*; Rouvray, D.H., Ed.; Research Studies Press: Taunton, UK, 1997; pp. 353–381.
56. Krivovichev, S.V. Which inorganic structures are the most complex? *Angew. Chem. Int. Ed.* **2014**, *53*, 654–661.
57. Vincent, A. *Molecular Symmetry and Group Theory*, 2nd ed.; Wiley: Chichester, UK, 2013.
58. Rouvray, D.H. The changing role of the symbol in the evolution of chemical notation. *Endeavour* **1977**, *1*, 23–31.
59. Lewis, G.N. The atom and the molecule. *J. Am. Chem. Soc.* **1916**, *38*, 762–785.
60. Crum Brown, A. On the theory of isomeric compounds. *Trans. R. Soc. Edinb.* **1864**, *23*, 707–719.
61. Ritter, C. An early history of Alexander Crum Brown's graphical formulas. In *Tools and Modes of Representation in the Laboratory Sciences*; Klein, U., Ed.; Kluwer: Dordrecht, The Netherlands, 2001; pp. 35–46.
62. Goodwin, W. Explanation in organic chemistry. *Ann. N. Y. Acad. Sci.* **2003**, *988*, 141–153.
63. Goodwin, W. Sustaining a controversy: The non-classical ion debate. *Br. J. Philos. Sci.* **2013**, *64*, 787–816.
64. Meadows, J. *The Victorian Scientist: The Growth of a Profession*; The British Library: London, UK, 2004.
65. Willett, P. From chemical documentation to chemoinformatics: 50 years of chemical information science. *J. Inf. Sci.* **2008**, *34*, 477–500.
66. Lynch, M.F.; Harrison, J.M.; Town, W.G.; Ash, J.E. *Computer Handling of Chemical Structure Information*; MacDonald: London, UK, 1971.
67. Bottle, R.T. *The Use of Chemical Literature*, 2nd ed. revised; Butterworths: London, UK, 1971.
68. Ash, J.E.; Hyde, E. *Chemical Information Systems*; Ellis Horwood: Chichester, UK, 1975.
69. Ash, J.; Chubb, P.; Ward, S.; Welford, S.; Willett, P. *Communication, Storage and Retrieval of Chemical Information*; Ellis Horwood: Chichester, UK, 1985.
70. Ash, J.E.; Warr, W.A.; Willett, P. *Chemical Structure Systems*; Ellis Horwood: Chichester, UK, 1991.
71. Leach, A.R.; Gillet, V.J. *An Introduction to Chemoinformatics*, 2nd ed.; Springer: Dordrecht, UK, 2007.
72. Currano, J.N.; Roth, D.L. *Chemical Information for Chemists: A Primer*; RSC Publishing: Cambridge, UK, 2014.
73. Buntrock, R.E. Beilstein and Gmelin: Classical chemical information for people who hate classics. *Database* **1992**, *15*, 104–106.
74. Lawson, A.J.; Swienty-Busch, J.; Géoul, T.; Evans, D. The making of Reaxys-towards unobstructed access to relevant chemical information. In *The Future of the History of Chemical Information, American Chemical Society Symposium Series*; McEwan, L.R., Buntrock, R.E., Eds.; Oxford University Press: New York, NY, USA, 2014; Volume 1164, Chapter 8, pp. 127–148.
75. Loukine, E.; Stumpfe, D.; Bajorath, J. Molecular formal concept analysis for compound selectivity profiling in biologically annotated databases. *J. Chem. Inf. Model.* **2009**, *49*, 1359–1368.

76. Gardiner, E.J.; Gillet, V.J. Perspectives on knowledge discovery algorithms recently introduced in chemoinformatics: Rough set theory, association rule mining, emerging patterns, and formal concept analysis. *J. Chem. Inf. Model.* **2015**, *55*, 1781–1803.
77. Fugmann, R.; Nickelsen, H.; Nickelsen, I.; Winter, J.H. TOSAR—a system for the structural formula-like representation of concept connections in chemical publications. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 52–55.
78. Van der Vet, P.E.; Mars, N.J.I. Structured system of concepts for storing, retrieving, and manipulating chemical information. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 564–568.
79. Hicks, M.G. Similarity and the Beilstein information system: Searching for concepts with current facts. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 631–638.
80. Bawden, D. Chemical reaction information. In *Chemical Structure Systems*; Ash, J.E., Warr, W.A., Willett, P., Eds.; Ellis Horwood: Chichester, UK, 1991; pp. 57–87.
81. Currano, J.N. Reaction searching. In *Chemical Information for Chemists: A Primer*; Currano, J.N., Roth, D.L., Eds.; RSC Publishing: Cambridge, UK, 2014; pp. 224–254.
82. Ash, J.E.; Warr, W.A. Databanks. In *Chemical Structure Systems*; Ash, J.E., Warr, W.A., Willett, P., Eds.; Ellis Horwood: Chichester, UK, 1991; pp. 154–191.
83. Allen, F.H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr. Sec. B Struct. Sci.* **2002**, *58*, 380–388.
84. Wagner, A.B. Physical properties and spectra. In *Chemical Information for Chemists: A Primer*; Currano, J.N., Roth, D.L., Eds.; RSC Publishing: Cambridge, UK, 2014; pp. 146–183.
85. Groom, C.R.; Allen, F.H. The Cambridge Structural Database in retrospect and prospect. *Angew. Chem. Int. Ed.* **2014**, *53*, 662–671.
86. Bruno, I.J.; Groom, C.R. A crystallographic perspective on sharing data and knowledge. *J. Comput. Aided Mol. De.* **2014**, *28*, 1015–1022.
87. Currano, J.N. Searching by structure and substructure. In *Chemical Information for Chemists: A Primer*; Currano, J.N., Roth, D.L., Eds.; RSC Publishing: Cambridge, UK, 2014; pp. 107–145.
88. Bender, A.; Glen, R.C. Molecular similarity: A key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
89. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **2006**, *11*, 1046–1053.
90. Stumpfe, D.; Bajorath, J. Similarity searching. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 260–282.
91. Lewis, R.A.; Wood, D. Modern 2D QSAR for drug discovery. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 505–522.
92. Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular similarity in medicinal chemistry. *J. Med. Chem.* **2014**, *57*, 3186–3204.
93. Willett, P. The calculation of molecular structural similarity: Principles and practice. *Mol. Inform.* **2014**, *33*, 403–413.
94. Charton, B. Searching the literature for concepts. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 45–46.
95. Twiss-Brooks, A. Searching using text: Beyond web search engines. In *Chemical Information for Chemists: A Primer*; Currano, J.N., Roth, D.L., Eds.; RSC Publishing: Cambridge, UK, 2014; pp. 93–108.

96. Smith, E.G. *The Wiswesser Line-Formula Chemical Notation*; McGraw-Hill: New York, NY, USA, 1968.
97. Blower, P.E.; Myatt, G.J.; Petras, M.W. Exploring functional group transformations on CASREACT. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 54–58.
98. Loftus, F. Computer-aided synthesis design. In *Chemical Structure Systems*; Ash, J.E., Warr, W.A., Willett, P., Eds.; Ellis Horwood: Chichester, UK, 1991; pp. 222–262.
99. Batchelor, C. Chemistry ontologies. In *The Future of the History of Chemical Information*, American Chemical Society Symposium Series; McEwan, L.R., Buntrock, R.E., Eds.; Oxford University Press: New York, NY, USA, 2014; Volume 1164, Chapter 13, pp. 219–235.
100. Barnard, J.M. Structure representation and searching. In *Chemical Structure Systems*; Ash, J.E., Warr, W.A., Willett, P., Eds.; Ellis Horwood: Chichester, UK, 1991; pp. 9–56.
101. Warr, W.A. Tautomerism in chemical information management systems. *J. Comput. Aided Mol. Des.* **2010**, *24*, 497–520.
102. Fujita, S. The stereoisogram approach for remedying discontents of stereochemical terminology. *Tetrahedron Asymmetry* **2014**, *25*, 1612–1623.
103. Mills, J.E.; Baughman, B. REACCS in the chemical development environment. 3. Graphically nonequivalent representation of molecules and reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 431–435.
104. Gakh, A.A.; Burmett, M.N.; Trepalin, S.V.; Yarkov, A.V. Modular Chemical Descriptor Language (MCDL): Stereochemical modules. *J. Cheminform.* **2011**, doi:10.1186/1758-2946-3-5.
105. Jiang, L.; Rizzo, R.C. Pharmacophore-based similarity scoring for DOCK. *J. Phys. Chem. B* **2015**, *119*, 1083–1102.
106. Pickett, S.D.; Luttmann, C.; Guerin, V.; Laoui, A.; James, E. DIVSEL and COMPLIB-strategies for the design and comparison of combinatorial libraries using pharmacophoric descriptors. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 144–150.
107. Todd, M.H. Computer-aided organic synthesis. *Chem. Soc. Rev.* **2005**, *34*, 247–266.
108. Walker, M.A. Public chemical databases and the Semantic Web. In *The Future of the History of Chemical Information*, American Chemical Society Symposium Series; McEwan, L.R., Buntrock, R.E., Eds.; Oxford University Press: New York, NY, 2014; Volume 1164, Chapter 12, pp. 197–217.

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).