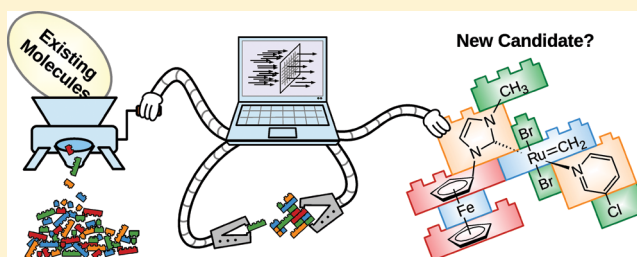


Automated Design of Realistic Organometallic Molecules from Fragments

Marco Foscato,[†] Giovanni Occhipinti,[†] Vishwesh Venkatraman,[‡] Bjørn K. Alsberg,[‡] and Vidar R. Jensen^{*,†}[†]Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway[‡]Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

S Supporting Information

ABSTRACT: A method for the automated generation of realistic, synthetically accessible transition metal and organometallic complexes is described. Computational tools were designed to generate molecular fragments, preferably harvested from libraries of existing, stable compounds, to be used as building blocks for the construction of new molecules. These fragments are enriched with information about the number and type of possible connections to other fragments and are stored in library files. When connecting fragments in the subsequent building process, compatibility matrices, which define the connection rules between fragments, are used to delineate organometallic fragment spaces from which molecules can be generated in an automated fashion. The approach is flexible and allows ample structural variation at the same time as the combination of known fragments is easily restrained to avoid generation of exotic and unrealistic substructures and molecules. The method was tested in the generation of ruthenium complexes, with a given coordination environment, which can serve as candidates in catalyst development. The results demonstrate that molecules generated with the described method do not contain exotic arrangements of atoms and are by far more realistic than those obtained by the application of valence rules alone.



1. INTRODUCTION

Methods for the automated generation of molecular models, either as complete structures with three-dimensional coordinates or containing constitutional information only, have mainly been developed for organic drug-like molecules in the context of *in silico* screening and *de novo* drug design.¹ Although these methods are widely used in organic drug design, their impact in inorganic and organometallic chemistry has so far been limited.^{2–6} The main reason for this difference has to do with the increased complexity of inorganic and organometallic species, i.e., the diversity of chemical bonds and flexibility with respect to both electronic and molecular structure.^{7–9} In addition, the most common chemical representations are often not able to handle features particular to organometallic compounds, such as dative bonds and π -complexes.¹⁰

Considering molecules to be arrangements of atoms, the automated generation of all the combinations of various numbers of atoms of different elements is the most general approach to defining the chemical space to be searched for new chemicals.^{11–15} It has been demonstrated that a huge and largely unexplored chemical universe of small organic molecules can be generated by the combination of a limited set of elements (H, C, N, O, S, and halogens).^{16,17} Extending the list of elements to include transition metals and other atoms commonly used in organometallic chemistry (e.g., P, B, and Si) quickly makes the combinatorial problem intractable. And the mere combinatorial problem is compounded by the higher and

more varied coordination numbers and the less restrictive valence rules typically seen for these elements.

To limit the number of combinations and focus on a more pragmatic portion of the chemical space, molecular fragments can be used as building blocks in lieu of single atoms. Such a fragment-based strategy has been implemented in a number of tools for building organic molecules,¹⁸ and a few pioneering studies on automated generation of organometallic compounds from fragments have also appeared.^{2,19,20}

Hay and Firman¹⁹ developed a method for *de novo* structure-based design of molecular receptors with multidentate ligands capable of forming metal complexes. The method makes use of hydrocarbon-based linkers that connect “complex fragments,” which include both the coordinating group (host component) and the metal atoms (guest), to build host–guest type transition-metal complexes and has been demonstrated in the design of receptors for metal ions or nonmetallic anions.^{21–25} However, the authors reported the need for filtering the generated structures to identify the most synthetic accessible ones.^{5,21} The same problem was noted in the analysis of the outcome of an evolutionary algorithm applied to *de novo* optimization of transition metal catalysts.² In the latter method,² as proposed earlier by Rothenberg and co-workers,²⁰ fragments belonging to different classes were used to build organometallic compounds. Three libraries of fragments,

Received: December 18, 2013

Published: February 17, 2014

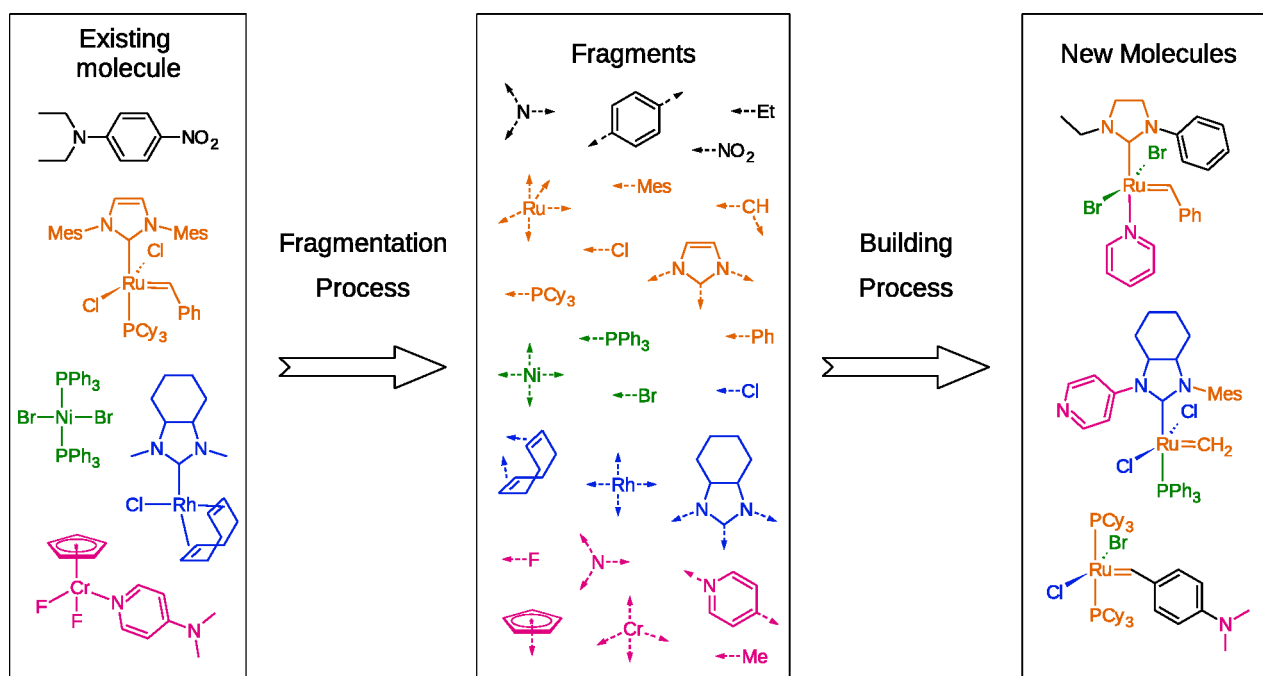


Figure 1. Schematic representation of the two main steps, fragmentation and building, involved in the fragment-based construction of molecules.

corresponding to three different classes of fragments, were created on the basis of the role of each fragment in the final structure: the metal-containing core part, so-called trial parts (metal-coordinating groups), and free parts (freely varying substituents on the core, on trial parts, and on other free parts). Despite the precise control of the metal-coordinating environment, many of the complexes were found to contain undesirable functional groups on the ligands or were synthetically inaccessible for other reasons.²

The poor synthetic accessibility of automatically constructed molecules is a well-known issue in *de novo* design of organic drugs.^{1,2,6,27} Different strategies have been proposed either to limit the generation of unfeasible molecules by controlled assembly of selected building blocks^{18,28–31} or to reject poorly scoring molecules after an *in silico* evaluation of their synthetic accessibility.^{32–38} Following the first strategy, the concept of “chemical fragment space” has been proposed as the combination of fragments and rules controlling the connection of such fragments.^{39,40} It has been demonstrated that when connection rules are based on retrosynthetic reasoning, the chemical fragment space provides effective control of the synthetic accessibility of the automatically generated molecules.^{41,42}

Although the three above-mentioned automated builders^{2,19,20} used fragments, the connection rules were based on classification of the fragments according to their position with respect to the metal (i.e., metal-containing core, ligands, linkers, and substituents) rather than retrosynthetic reasoning. The use of layers of fragments permits a precise and accurate selection of the coordinating moieties surrounding the metal atom (the inner layer), preserving its main properties, which is certainly a crucial point in the design of a catalyst.^{2–4,20,43–46} Nevertheless, it has been suggested^{2,5,21} that the use of layer-classified fragments can be improved, in particular with respect to the synthetic accessibility of the automatically generated ligands.

To improve the performance of fragment-based tools for building organometallic compounds, we apply the concept of

chemical fragment space to organometallic chemistry and devise a special treatment of bonds involving metal atoms. Hence, the concept of the “organometallic fragment space” is introduced as the combination of annotated molecular fragments and connection rules for organometallic compounds. Here, the implementation of this concept is presented and discussed along with its application in selected test cases aiming to achieve automated generation of realistic organometallic compounds.

The focus of the present contribution is to explore the possibilities of the organometallic fragment space, a concept which inherently deals only with the molecular constitution, that is, the identity and connectivity of atoms in molecules. Although the approach can be implemented both for two-dimensional (2D) and three-dimensional (3D) structures, only 2D representations will be built and discussed here. Adaptation of the approach to generate 3D molecular models will be discussed in a future contribution.

2. METHODS

The construction of molecules with a fragment-based strategy involves two main steps (Figure 1): (i) the generation of the fragments from known/existing molecules (fragmentation) and (ii) the construction of new molecules by connection of some of these fragments (building process). Before describing in detail the two processes, it is necessary to define the concept of fragments in the context of the present method.

2.1. Fragments and Attachment Points. Fragments are molecular building blocks annotated with information on the number, position, and type of all the possible connections with other building blocks (Figure 2). Each fragment contains a portion of a molecular system that can be of any size. Each of the possible connections with other fragments is represented by an attachment point (AP), in the figures depicted as a dashed arrow originating from the “source atom,” i.e., the atom that can be used to make a new fragment–fragment connection by formation of a formal bond. An AP contains all of the

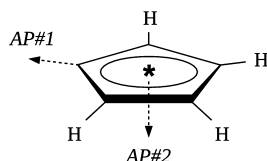


Figure 2. A typical fragment with two attachment points (AP#1, AP#2). Each AP is shown as a dashed arrow originating from the source atom. The source atom can either be a regular atom (AP#1) or a dummy atom (AP#2).

information needed to identify the source atom and define the chemical context characterizing the source atom (i.e., portion of functional group and type of substitution). Such characterization with respect to the chemical context is further described below in the Attachment Point Class subsection.

Since neither fragments nor APs are required to respect valence rules, there are no limitations with respect to arrangements of atoms within fragments. The only requirement is that every atom of a polyatomic fragment must be connected to at least one other atom of the same fragment. The neglect of valence rules means that, although the connectivity is used to define formal bonds, these bonds are not associated with a precise bond type. The use of formal bonds rather than a strict bond formalism reflecting electron count opens for treatment of any kind of interatomic interaction as if it were a regular covalent bond. This means that any kind of chemical object can be represented by the fragments and their connections. For instance, three-center two-electron bonds, noncovalent interactions, or even partially broken/formed bonds can be treated as formal bonds both within building blocks, via intrafragment formal bonds, and in interfragment formal bonds with the corresponding definitions in APs. To illustrate, contact ion pairs⁴⁷ such as those involved in catalyzed olefin polymerization⁴⁸ can either be represented by a single fragment or by two fragments (Figure 3). In return, since this formalism differs

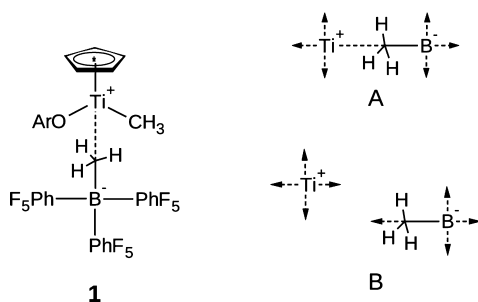


Figure 3. The contact ion pair half-titanocene compound (1), which can be constructed either by using the fragment (A), which contains a formal bond in lieu of the ionic interaction, or by connection of two fragments (B) with the formation of a formal interfragment bond.

from the usual connectivity table formalisms and thus is not compatible with existing tools for implicit hydrogen assignment, all formal charges and all atoms, including hydrogens, pertaining to a fragment must be explicitly reported in a fragment.

2.2. The Fragmentation Process. Fragmentation is the process of generating fragments from libraries of existing molecules (Figure 4). Molecules to be fragmented are harvested from databases of compounds, such as the Cambridge Structural Database⁴⁹ or PubChem.⁵⁰ Alternatively, in-house generated molecular models can also be used as long as

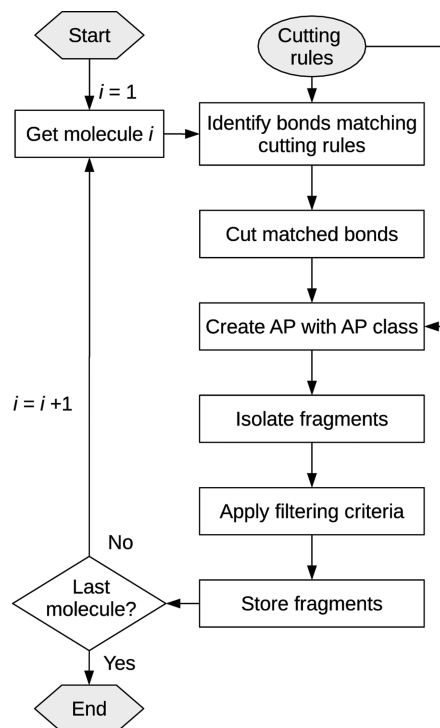


Figure 4. The main steps of the fragmentation.

all atoms and their connectivity are provided in the specified file format (currently SDF⁵¹).

For each molecule to be fragmented, all of the bonds matching a given list of bond types (termed “cutting rules,” explained in the next subsection) are removed, and the molecule is broken into as many fragments as possible. Fragments that are identified as isolated, yet internally continuously connected portions of the initial molecule, are then treated as independent units. Moreover, some fragments are removed using filtering criteria, which can be based on the fragment’s empirical formula, substructure, properties of APs, and rules for detection and handling of duplicates. Finally, after the filtering, the remaining fragments are stored in the fragment library file.

In general, for each bond matching at least one cutting rule, two APs are generated, one on each side of the broken bond (see Figure 5). However, in the case of ligands capable of multihapto binding, i.e., ligands with two or more contiguous atoms bonded to the same metal atom,^{53,54} more than one organometallic bond is broken but only two APs are generated, one on the metal center and one on the multihapto ligand. Coordinating but not contiguous atoms (not bonded to each other) of a polydentate ligand do not match the definition of a multihapto ligand,⁵⁴ which means that each such atom is treated independently in the fragmentation (Figure 6). However, for multidentate ligands characterized by one or more coordination sites with hapticity larger than one, each of the isolated groups of contiguous atoms that are also connected to the same metal is treated independently as a single multihapto ligand. This more local treatment is, formally, in disagreement with the original definition of hapticity,⁵⁴ according to which the bidentate coordinating olefin 1,5-cyclooctadiene is a η^4 ligand (hapticity = 4) but allows for efficient construction of the correct set of metal–ligand bonds by splitting the global role of a polydentate, multihapto ligand

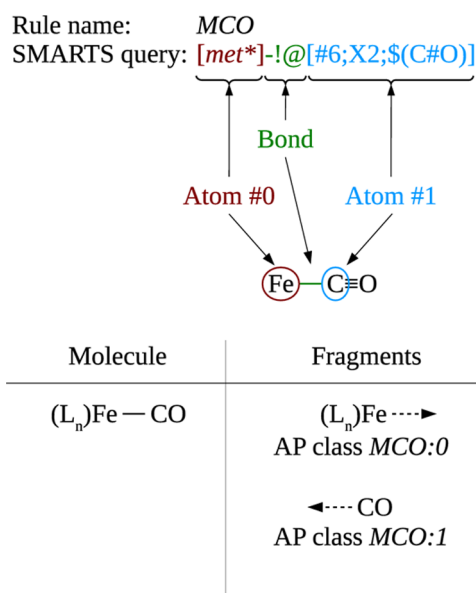


Figure 5. Application of a cutting rule on a target bond matching the given SMARTS query (*met** = list of transition metal symbols). The class of the APs on the resulting fragments is derived from the name of the cutting rule (e.g., *MCO*) and the unambiguous identification of Atom#0 (in red) and Atom#1 (in cyan) by means of the two SMARTS subqueries highlighted in red and cyan, respectively.

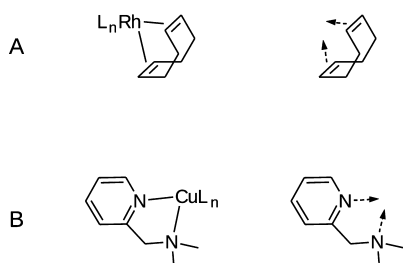


Figure 6. Example of fragmentation of a bidentate ligand with multihapto coordinating sites (A) and a η^1 -coordinating bidentate ligand (B).

in local contributions. Therefore, the metal-(η^4 -1,5-cyclooctadiene) system is represented as a pair of isolated η^2 ligands, and two APs are generated, one placed on each of the coordinating π systems of 1,5-cyclooctadiene (Figure 6).

2.2.1. Cutting Rules. Cutting rules define the type of bonds to be broken during the fragmentation process. Daylight SMARTS⁵⁵ were used to build substructure queries, with one SMARTS query per cutting rule. Each SMARTS query consists of three parts: two atom subqueries, one for each of the atoms connected by the bond to be matched (Atom #0 and Atom #1, red and cyan in Figure 5), and a bond subquery specifying the properties of the bond (green in Figure 5). The two atom subqueries are used independently to unambiguously identify the atoms on either sides of a bond to be broken and to assign properties to the resulting APs (see the Attachment Point Class subsection). Therefore, cutting rules for multihapto ligands have to match all the atoms participating in the same metal-multihapto ligand system. On one side, the metal is matched by one of the two atom subqueries (red in Figure 7), and on the other side, all the atoms of the multihapto ligand that are connected to the metal fit the other atom subquery (cyan in Figure 7). It should be noted that, thanks to the possibility of

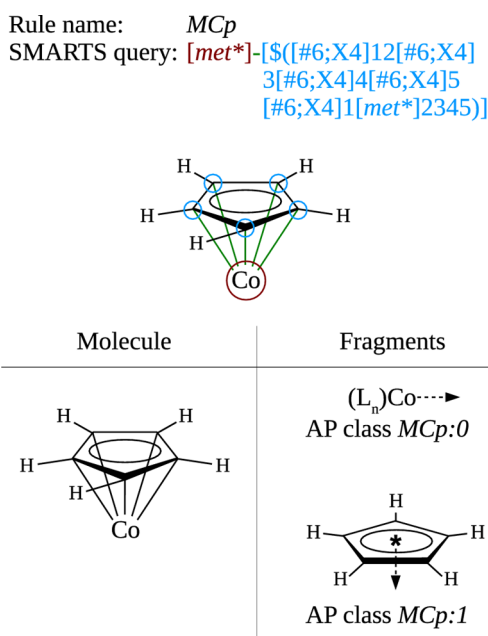


Figure 7. The class of the AP placed on the dummy atom at the center of the cyclopentadienyl (Cp) ring determined by the subquery highlighted in cyan. This subquery matches all five carbon atoms (cyan circles) of the Cp (*met** = list of transition metal symbols).

using logical operators (AND, OR) within SMARTS queries, multihapto ligands containing various elements can be matched by a single atom subquery. Further details regarding the use of cutting rules are given in the Supporting Information, section S1, together with the full list of rules used in this contribution.

2.2.2. Attachment Point Class. The attachment point class (AP class) is an encoded representation of the chemical moiety of the source atom in the molecule subjected to fragmenting. In fact, the AP class derives directly from the cutting rule applied to break a bond and generate a pair of APs (see Figures 5 and 7).

The AP class is a string including the name of the cutting rule representing the moiety from which the AP was generated and an index (0 or 1, see Figures 5 and 7) to identify which of the two atom queries was matched by the source atom of the AP. For nonsymmetric cutting rules ("non-symmetric" means that the two atom subqueries are not identical), the two APs generated by breaking a bond each bear the complementary class of the other AP. For example, from the cutting rule matching the metal-carbonyl bond in $(L_n)Fe-CO$, assuming that the given name of such a rule is *MCO*, the fragmentation produces a fragment $(L_n)Fe$ with an attachment point of class *MCO:0* and a fragment CO with an attachment point of class *MCO:1* (Figure 5).

When cutting symmetric bonds, the two atom queries are identical, and both APs are annotated with the same class (convention: atom index = 0) since it is not possible to distinguish the two sides of the bond. In the case of multihapto ligands, the AP is placed on a dummy atom representing that ligand, or that contiguous part of the ligand, with its hapticity (which is given by the cutting rule). Since the corresponding cutting rule has to match all the contiguous atoms defining the multihapto ligand, the AP class of the dummy atom is defined by the atom subquery matching all the contiguous atoms of (that part of) the multihapto ligand (Figure 7).

AP class on fragments to be attached






		AP class on fragments to be attached								
		$\leftarrow \cdots \text{CO}$		$\leftarrow \cdots \text{NR}_2$						
		$\leftarrow \cdots \text{Fe}$		$\leftarrow \cdots \text{OR}$						
		<i>MCO:0</i>	<i>MCO:1</i>		<i>alcohol:0</i>	<i>amine:0</i>		<i>ketone:0</i>		
AP class on the growing molecule	$\text{Fe} \cdots$	<i>MCO:0</i>	<div>T/F Fe-Fe</div>	<div>T/F Fe-CO</div>		<div>T/F Fe-OR</div>	<div>T/F Fe-NR₂</div>		<div>T/F </div>	
	$\text{OC} \cdots$	<i>MCO:1</i>	<div>T/F OC-Fe</div>	<div>T/F OC-CO</div>		<div>T/F OC-OR</div>	<div>T/F OC-NR₂</div>		<div>T/F </div>	
	$\text{RO} \cdots$	<i>alcohol:0</i>	<div>T/F RO-Fe</div>	<div>T/F RO-CO</div>		<div>T/F RO-OR</div>	<div>T/F RO-NR₂</div>		<div>T/F </div>	
	$\text{R}_2\text{N} \cdots$	<i>amine:0</i>	<div>T/F R₂N-Fe</div>	<div>T/F R₂N-CO</div>		<div>T/F R₂N-OR</div>	<div>T/F R₂N-NR₂</div>		<div>T/F </div>	

Figure 8. Compatibility matrix. The compatibility matrix uses the ordered list of all AP classes as row and column indices and thus covers all the combinations of AP classes. For clarity, only a handful of examples of AP classes are indicated in the figure. Each element of the matrix is a Boolean (True/False) defining the compatibility between that particular combination of AP classes. The compatibility between two AP classes (defined by a True entry) allows for the generation of a chemical moiety resulting from the formation of a bond between the two atoms bearing APs of the given classes. An example of the resulting moieties is shown in each of the elements of the matrix.

2.3. Building Process. To exploit the organometallic fragment space, we implemented an automated builder in which molecules are built as tree-like structures. Branches of fragments are grown around a selected seed fragment in a three-step process: (i) selection of the seed fragment, (ii) growth, and (iii) finalization of the molecular system.

Seed fragments are simply fragments that are chosen to be at the core of the molecule being constructed. Usually, the seed fragment contains features important for the field of application. A typical fragment may, for example, be a single metal atom (Figure 3B) or a metal atom along with some of its bonded neighboring atoms (Figure 3A). Still, there is no distinction between seed fragments and other fragments.

The growth of a tree-like molecule proceeds by the attachment of new fragments on APs of the growing molecule. Which fragments that can be attached is determined by the compatibility of that AP with the APs of the fragments in the fragment library. New fragment–fragment bonds can only be generated between pairs of atoms bearing APs with compatible AP classes. Thus, the chemical traits of the newly generated interfragment connections are explicitly imposed by the defined organometallic fragment space, i.e., by the fragments, the APs, and the classes of the APs. How compatibility between different APs is controlled is described in detail in the next subsection (Compatibility Matrix).

In addition to the AP class compatibility, two parameters are used to control the molecular growth: (i) the probability of appending one more fragment via one of the still available and

compatible APs of the growing molecule (P_{grow}) and (ii) the probability of symmetric substitution (P_{symm}) that is defined as the addition of N copies of the same fragment on N symmetric APs, where AP classes and the constitutional environment of the source atoms (i.e., the identity and connectivity of neighboring atoms) are used to define the constitutional symmetry. For example, the addition of three identical substituents on a central P atom generates a symmetric phosphine such as PPh_3 . While P_{symm} has a constant value, P_{grow} decreases with the size of the molecule according to eq 1:

$$P_{\text{grow}}(x) = 1 - \frac{1}{1 + e^{-s(x-m)}} \quad (1)$$

where x is the distance from the seed fragment (determined as the number of fragment–fragment connections), m is the value of x corresponding to 50% probability, $P_{\text{grow}}(m) = 0.5$, and s represents the slope of the curve when $x = m$. The influence of s and m is illustrated in Figure S1 of the Supporting Information. In addition to P_{grow} , which only influences the probability of building molecular branches with a given number of subsequent fragment–fragments connections, a size threshold is used to impose a maximum number of non-hydrogen atoms in the final molecules.

At each level x , the builder loops through all the APs of the growing molecule and appends new fragments with a probability given by the local value of P_{grow} . Thus, when all APs have been processed, the growing molecule can still bear a number of available APs not connected to any fragment. To

complete the building process, fragments with only one AP, termed “capping groups,” are used to saturate such unused APs. Typical capping groups are fragments such as $-H$ (single hydrogen atom), $-X$ (single halogen atom), $-CH_3$ (methyl group), or even complex groups depending on the characteristics of the building job. In general, the list of capping groups should include fragments for all the AP classes present on the seed, such as $-PR_3$ for L-type AP classes on a metal-atom seed fragment, and all the classes of APs of the library of fragments that require saturation. Alternatively, if the AP cannot remain unused and the capping procedure is not suitable, the generated molecule is rejected and the building process starts over again from scratch.⁵⁶

2.4. Compatibility Matrix. The compatibility matrix is the operator that turns cutting rules into connection rules and thus defines the compatibility between attachment points classes. It is a square, nonsymmetric matrix of *True/False* entries (see Figure 8 and section S3 of the Supporting Information). The complete list of AP classes constitutes the index in both dimensions. The row index represents the class of the AP on the growing molecule and the column index, the corresponding class of the AP on a candidate fragment to be attached. Moreover, since the AP class is an encoded representation of the chemical moiety surrounding the AP, each matrix element represents the arrangement of atoms resulting from the connection (formal bond formation) of the two chemical moieties represented by the two AP classes (Figure 8). Therefore, whether or not the two moieties should be allowed to join can be controlled by setting the corresponding matrix element to *True* or *False*.

The combination of APs having two complementary AP classes, such as *MCO:0* and *MCO:1*, leads to the original arrangement of atoms from which the two APs were generated. Next, the combination of noncomplementary AP classes, that is, AP classes with different rule names or with the same rule name and same atom index (AP classes from nonsymmetric cutting rules are not self-complementary), leads to chemical moieties that differ substantially from the molecule from which the fragments originated. Thus, the variability of the generated molecules can be enlarged by connections involving noncomplementary AP classes (cross-complementarity connections). For example, the combination of AP classes derived from cutting rules acting on alcohols and amines, namely *alcohol:0* and *amine:0*, respectively, with APs derived from ketones may be used to introduce esters by means of element (*alcohol:0;ketone:0*) and amides via (*amine:0;ketone:0*), see Figure 8. This kind of compatibility is justified by the likeliness that derivatives of carboxylic acids can be obtained for a skeleton bearing a ketone. The same kind of synthesis-related reasoning can justify many other cross-complementarity connections, including AP classes involving metals. For example, by introducing compatibility between *MCO:0* and all AP classes corresponding to donor ligands (L-ligand), i.e., not only *MCO:1*, the original carbonyl ligand may be exchanged with all other kinds of L-ligands. However, allowing for broad cross-complementarity compatibility has the potential drawback of producing unwanted or exotic moieties. For example, setting elements such as (*MCO:1;alcohol:0*), (*MCO:1;amine:0*), or (*MCO:1;ketone:0*) to *True*, allows for the generation of exotic arrangements of atoms, with wrong formal charge, that every chemist would usually try to avoid. Instead, combinations such as (*MCO:0;alcohol:0*) and (*MCO:0;amine:0*), allow for the construction of alkoxy and

amido complexes, but the use of such ligands in lieu of CO leads, if all other ligands remain constant, to a formal change in the oxidation state of the metal atom; see Figure 8.

It should be kept in mind that AP classes, as defined here, are not universally compatible. As a consequence, the large majority of the elements of the compatibility matrix that represent cross-compatibility connections should by default be set to *False*. Such matrix elements should only be set to *True* after careful evaluation. The importance of this point is illustrated by examples in the Results and Discussion section.

Even if there are always two entries that represent the combination of the same two AP classes, e.g., (*MCO:0;MCO:1*) and (*MCO:1;MCO:0*), and these two entries are located in symmetric positions with respect to the diagonal of the matrix, the compatibility matrix is not symmetric. In fact, the two indices of the matrix refer to two different objects (the growing molecule as a row index and the incoming fragment as a column index). Therefore, the two symmetric entries refer to two different bond formation events, involving the same pair of AP classes but in different directions. In particular, (*MCO:0;MCO:1*) means “from” *MCO:0* on the growing molecule “to” *MCO:1* on the approaching fragment, whereas (*MCO:1;MCO:0*) means “from” *MCO:1* on the growing molecule “to” *MCO:0* on the approaching fragment. The capability to distinguish between the two directions of growth allows for precise control of the molecular constitution, in particular with respect to the introduction of certain moieties (e.g., metal centers) in the periphery of the growing molecule. To illustrate, setting the entry (*MCO:0;ketone:0*) to *True* (Figure 8) allows the use of an acyl group as a covalent ligand on a metal center, but as long as the entry (*ketone:0;MCO:0*) is set to *False*, available acyl groups of the growing molecule will not bind any metal-containing fragment via *MCO:0* APs.

2.5. Limitations. The fragment-based builder generates compounds by assembling available fragments using compatible APs. Apart from all the fragment–fragment connections, for which chemical traits are controlled by the compatibility matrix, the builder is not aware of the chemistry embedded within each of the fragments used to build a molecule. On the one hand, this allows for the treatment of any kind of chemical object; i.e., the builder simply does not return any error regardless of the content of a fragment, but on the other hand this can lead to the coexistence of incompatible moieties, such as oxidants and reductants, in the same molecule if such groups belong to different fragments in the final molecule. Moreover, the protonation state of acid/base species is not taken into account in the building process and fragments with strongly acidic groups or organic oxidants should be eliminated during the filtration of the fragments (Figure 4). The filtering criteria, should be evaluated in relation to the chemistry of the particular case under study.

The current implementation does not allow for generation of intramolecular connections between APs on fragments belonging to the same molecule. Even if this is only a limit of the current implementation and not of the AP-compatibility scheme in general, the consequence is that ring closure is not permitted. As already presented by others,¹⁹ fragments can be used to construct rings and polycyclic systems. Moreover, since the AP class is an intrinsic property, the compatibility between a pair of APs can also be determined at the intramolecular level as long as the directionality, as defined by elements in the upper or lower triangle of the compatibility matrix, is handled

properly. Therefore, the AP-compatibility scheme can also be applied to generate cyclic or polycyclic molecules and chelates.

3. RESULTS AND DISCUSSION

The use of an organometallic fragment space was evaluated in test cases that aimed to build lists of molecules matching the general formula $L_2X_2Ru=CR_2$, where L are neutral ligands (formally two-electron donors), X are anionic ligands (also known as covalent ligands, formally contributing one electron to a covalent bond), and R may be H or an aromatic ring. Such ruthenium carbenes include Grubbs catalysts for olefin metathesis (Figure 9)^{57–60} and are the subject of substantial ongoing exploration and optimization of catalytic properties.^{2,61–68}

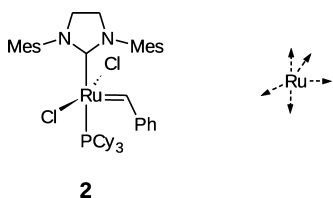


Figure 9. Grubbs second generation catalyst for olefin metathesis (2) and the seed metal fragment used to generate related Ru complexes.

The automated generation of such complexes is not only needed for the virtual screening of potential catalysts but could also inspire chemists' intuition by providing new combinations of potentially less known ligands developed for different fields of applications. Thus, the fragmentation process may be seen as a "data mining tool" capable of extracting a number of alternative ligands from databases of existing molecules. This opportunity was exploited by collecting fragments from the Cambridge Structural Database (CSD)⁴⁹ and combining them to build complexes respecting the general formula $L_2X_2Ru=CR_2$. See the Computational Details section for details on the fragmentation and building processes. Two case studies are presented to highlight and discuss different adaptations and uses of the organometallic fragment space.

3.1. Case Study I: Variability on All of the Ligands. The aim of this case study is to generate realistic compounds matching the general formula $L_2X_2Ru=CR_2$ with precisely defined variability with respect to the three types of ligands and substituents, L, X, and R; see Figure 10: (i) multiple kinds of L-ligands with a high degree of variability and various substitution patterns, (ii) a selected list of predefined, not further substituted X-ligands, and (iii) restrained variability for a selected list of substituents on the R groups.

Below, we start out by first producing molecules using only valence rules to define the AP compatibility and then compare with a procedure in which the organometallic fragment space is carefully controlled via the compatibility matrix.

3.1.1. Case Study IA: The Valence-Only Scheme. To achieve valence rules-only compatibility, the compatibility matrix was constructed to permit any type of fragment–fragment connection respecting valence rules for all of the organic elements. In other words, the compatibility matrix contained as many *True* entries as possible without being in conflict with the valence rules, leading to a matrix with 84.4% of the elements set to *True*. This means that most of the machinery and control mechanisms presented in the Methods section are ignored in order to carry out this particular test. The

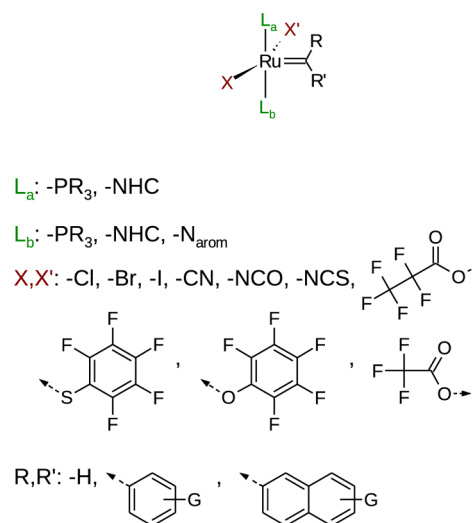


Figure 10. Schematic representation of the variability imposed on the different types of ligands by means of the compatibility matrix settings. The pairs X, X' and R, R' corresponding to atoms/groups that fulfill the requirements for symmetric APs detection (same AP class and bound neighborhood). G: any atom.

valence-rule only approach is not capable of imposing the variability scheme specified in Figure 10, the aim of this experiment is thus only to show what kinds of structures are generated without explicit control on the automatic building process.

A representative selection of the 3000 molecules generated in this experiment is given in Figure 11. Since the majority of the APs in the library of fragments (see Computational Details) are derived from common organic bonds and can be interpreted as produced by homolytic bond cleavage, most of the resulting organometallic ligands are classified as one-electron donors (X-ligands). Consequently, the analysis of the coordination environments of the generated complexes (Figure 12)^{69,70} shows that only a small fraction (7.3%) of the 3000 molecules matches the desired formula $L_2X_2Ru=CR_2$. In addition, several molecules bear exotic or too reactive arrangements of atoms that lead to unrealistic ligands (see 6, 7, 8, 11, 12, and 15 in Figure 11) even if the valence rules were followed in their construction. In conclusion, generation of fragment–fragment connections solely based on the valence-rules scheme does not lead to realistic organometallic structures, at least not with appreciable yield.

3.1.2. Case Study IB: Chemically Meaningful Compatibility Scheme. In this experiment, the organometallic fragment space is fully exploited thanks to a compatibility matrix derived from careful chemical reasoning. In fact, the compatibilities between AP classes have been selected so as to produce molecules with a given coordination environment and given a variability of the ligands. It should be kept in mind that the only difference with respect to the case study described above is the selection of the *True* entries of the compatibility matrix. The compatibility matrix in the present experiment permits only a small fraction (0.3%) of all the possible combinations of AP classes, the actual matrix being graphically depicted in the Supporting Information (section S3).

Figure 13 lists a representative subset of the 3000 molecules produced automatically (see also the Supporting Information, section S7.1). This time around, all generated complexes respect the formula $L_2X_2Ru=CR_2$, as depicted in Figure 12.

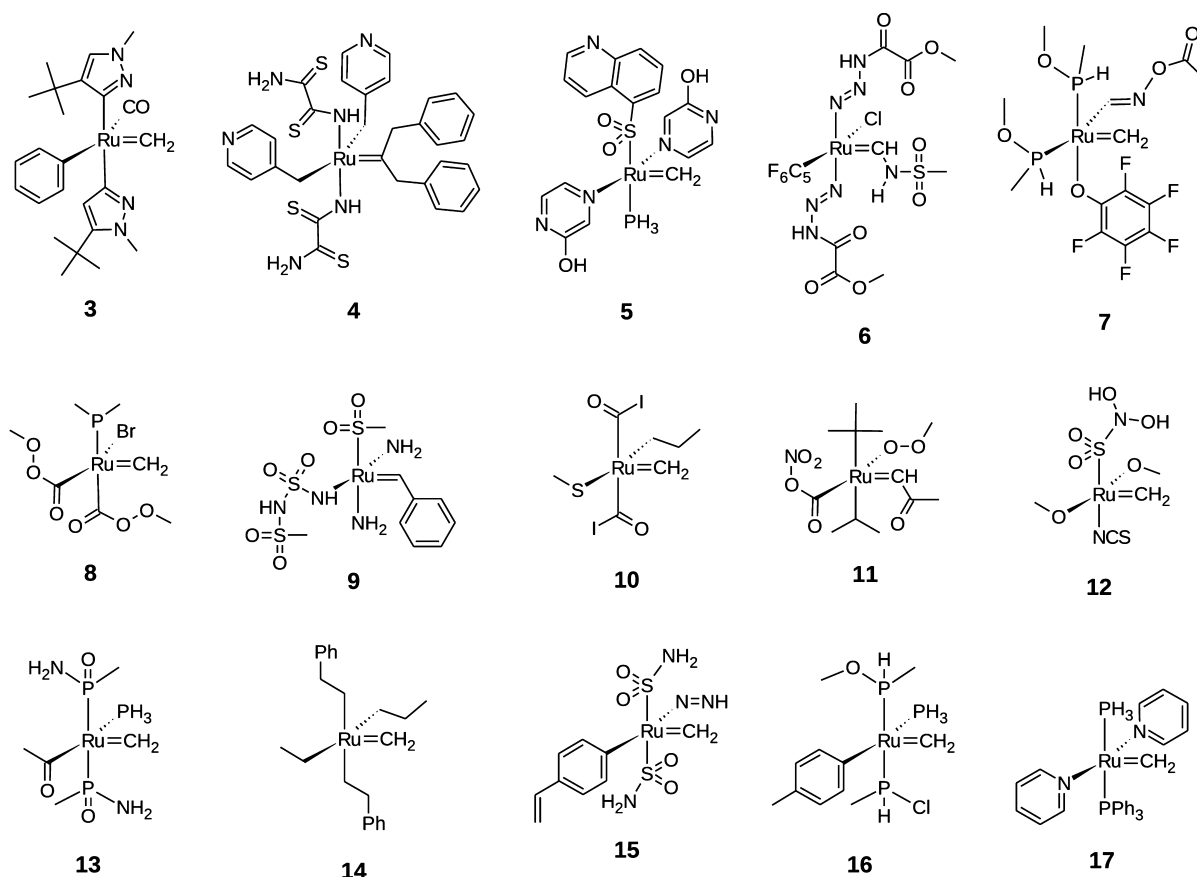


Figure 11. Examples of molecules generated with the valence-only AP-compatibility scheme.

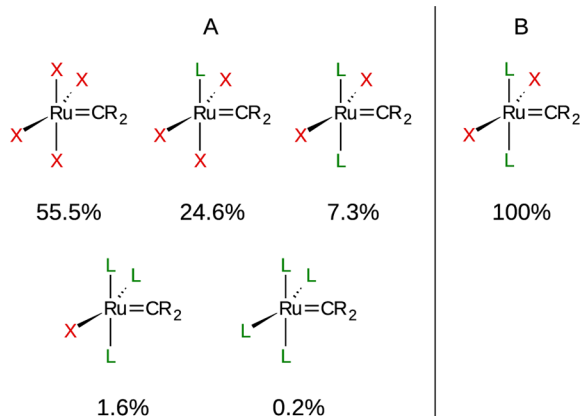


Figure 12. Distribution of the Ru formal electron count of the automatically generated molecules using the valence-only compatibility scheme (A) and chemically meaningful settings of the compatibility matrix (B).

Moreover, none of the fragment–fragment connections resulted in exotic chemical moieties. In fact, all of the combinations of AP classes that led to exotic moieties when using the above valence-only compatibility scheme were not permitted in the present experiment (*False* entries in the compatibility matrix), meaning that the problematic arrangements of atoms reported in Figure 11 could not be produced.

Visual inspection of the molecules generated using the AP class-based compatibility scheme (cf., Figure 13) reveals that many of the ligands thus produced are very similar, or even identical, to molecules already synthesized, for example, all the

phosphines and the imidazoles **20** and **21** in Figure 13, even if not part of the original library of molecules to be fragmented. Finally, it can be concluded that the vast majority of ligands produced by this approach appear to be synthetically accessible, and presumably, only relatively few cases would require substantial synthetic effort.

The evaluation of the feasibility of the complete organometallic molecule, in contrast to that of the isolated ligands alone, is much more challenging. While computational tools are capable of evaluating the synthetic accessibility of organic drug-like molecules,^{33–37,71,72} the application of such tools in organometallic chemistry has, to our knowledge, still to be investigated and validated. Indeed, the experienced organometallic chemist can, in particular when assisted by quantum chemical calculations, provide an opinion as to the feasibility of a given compound, but often only actual synthesis and experimental investigation can offer clarification. Since it is beyond the scope of the present work to offer such prediction for large collections of molecules, we only emphasize here the most common aspects that may play a role in preventing the formation of some of the designed complexes: (i) excessive steric hindrance between the ligands surrounding the metal center, (ii) combination of electronic and steric effects making a dative ligand not suitable for coordination to the metal center, and (iii) change in substituents affecting linkage isomerism; that is, some ligands can coordinate the metal via more than one atom, such as dimethyl sulfoxide, which can bind metal atoms via both O and S. These aspects are related to the makeup of a particular metal complex, and the organometallic fragment space provides a means to increase the control of the

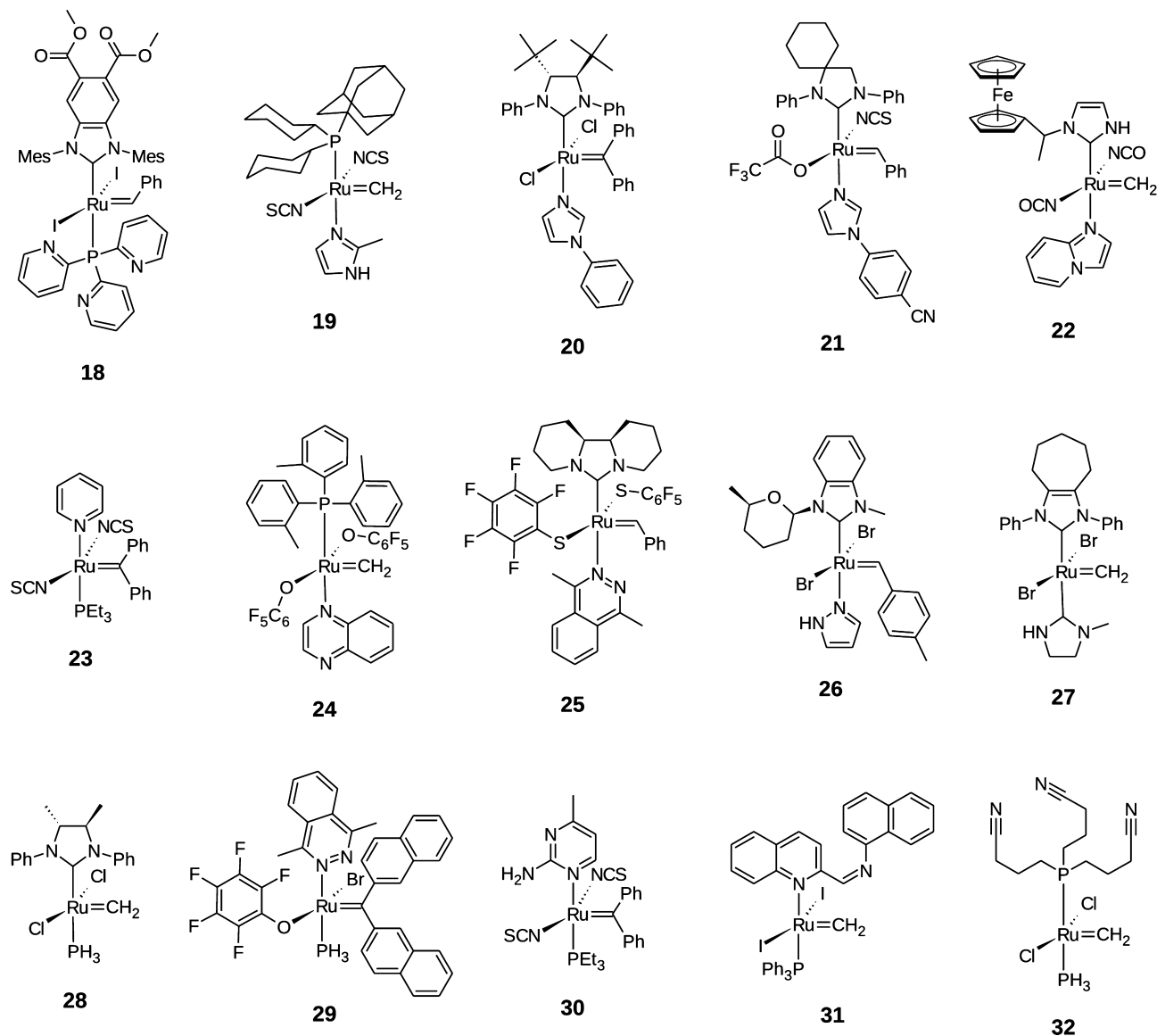


Figure 13. Examples of molecules automatically generated from the organometallic fragment space in case study IB.

automatic generation of molecules by adaptation of the cutting rules, filtering of the fragment library, and tuning of the AP class compatibilities. However, these details are strongly dependent on the actual application of the organometallic fragment space (i.e., type of metal involved and application field of the generated complexes) and should be seen as a means to impose more of the experience and wishes of the chemist into the organometallic fragment space and, hence, into the automated generation of molecules.

A final consideration is more related to the tree-like structure of the generated molecules, and thus to the present implementation of the building process. In fact, some of the constructed ligands, after proper conformational adaptation, can coordinate the metal in a polydentate fashion. In particular, **18**, **26**, and **31** (Figure 13) bear ligands with auxiliary sites that could potentially coordinate the metal, for example, the tri-2-pyridylphosphine ligand in **18**, the tetrahydropyranyl substituent of **26**, and the imine group of **31**. To evaluate this possibility, conformational flexibility of the complex must be determined. The present contribution is only concerned with the constitutional definition of the molecules (only 2D

structures are produced). Aspects such as ring closure and chelate formation are left for future improvements of the method.

3.2. Case Study II: Combinations of Existing Substituents. When modifying the characteristics of organometallic compounds, the selection of proper substituents on a predesigned scaffold is a common combinatorial strategy for fine-tuning the properties of well-known classes of ligands.^{20,73} Following this philosophy, this case study aims to demonstrate that organometallic fragment spaces can also be used to focus on the narrow variability resulting from the combination of existing substituents on a subset of scaffolds.

Ru catalysts bearing imidazolidin-2-ylidene (A in Figure 14) and 2-alkoxybenzylidene (B in Figure 14) ligands are known as Hoveyda–Grubbs catalysts for olefin metathesis and are of great significance because of their high activity and stability.^{74,75} Both the NHC ligand and the alkoxyarylmethylidene chelating group allow for a varied set of structural modifications that have been partially exploited to tune the properties of the catalysts while keeping the main features of the coordination environment.^{76–87} To automatically generate new candidate catalysts,

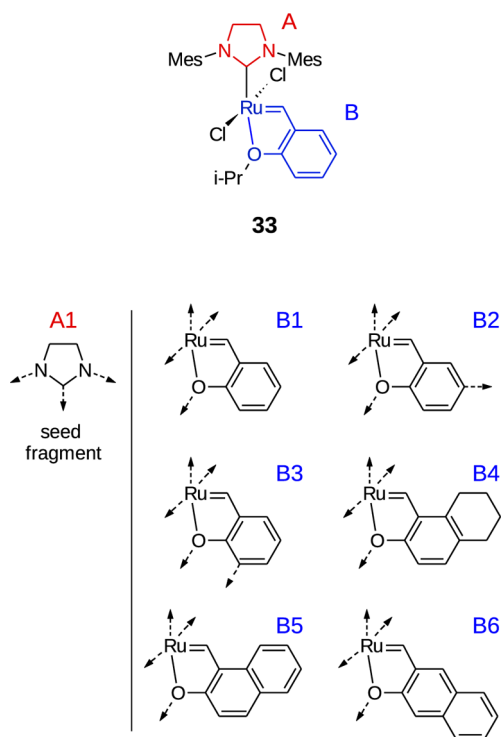


Figure 14. Hoveyda–Grubbs second-generation catalyst (top) and related fragments (bottom) used to build a list of alternative catalysts by combination of existing substituents in case study II.

an organometallic fragment space was designed by harvesting existing substituents from the literature and combining them with a list of scaffolds that share both the imidazolidin-2-ylidene, symmetrically *N*-substituted with aromatic groups, and the oxyarylmethylidene chelating ligands (Figure 14).⁸⁸

The fragments used included three main groups: (i) the imidazolidin-2-ylidene ligand, which was selected as a seed fragment since it participates in all of the structures to be generated (A1 in Figure 14), (ii) six Ru-oxyarylmethylidene fragments (B1–B6 in Figure 14), and (iii) substituents on both A1 and B1–B6. Since the aim of this test case is to use only existing substituents in a combinatorial fashion without any modification, the corresponding fragments have only one AP each and are generally complex in the sense that they could easily be further fragmented to smaller units. The details on the generation and constitution of the library of fragments are reported below in the Computational Details section.

A selection of the 3000 molecules generated is reported in Figure 15; see also section S7.2 of the Supporting Information.

Due to the use of carefully chosen fragments and a compatibility matrix constructed using chemical reasoning, all generated ligands represent sensible and synthetically accessible molecules. In fact, all of the substituents have been harvested from existing or commercial compounds. In addition, the limited variability with respect to metal-coordinating groups makes it likely that most of the complexes represent realistic organometallic compounds that can be prepared in laboratory. Besides, the limitation with respect to the size of the entire molecules (maximum 90 non-hydrogen atoms; see the Computational Details section) reduces the actual number of permitted combinations of fragments and avoids some of the problems mentioned above with respect to overcrowding of the metal center.

4. CONCLUSIONS

An automated fragment-based method for virtual construction of organic and inorganic molecules has been developed. The method introduces the concept of the organometallic fragment space, which involves a combination of annotated fragments and connection rules that control the chemical traits of the generated molecules. A compatibility matrix explicitly defines the kind of chemical moieties that can be produced by the building process, providing the user with complete control of the automated process. The performance of the method could be evaluated by examination of the results from two test cases, and comparison with molecules generated via a building process based only on valence rules. The results demonstrate that, when using chemically meaningful settings of the compatibility matrix, most of the automatically generated molecules are realistic. In comparison, the valence-only approach generates mostly exotic and unstable moieties. The method is very flexible with respect to the kind of chemical systems to be treated and should have considerable potential in particular for the construction of organometallic species with multihapto ligands and other bonds and interactions not found in organic chemistry alone.

5. COMPUTATIONAL DETAILS

5.1. Case Study I. The Cambridge Structural Database (CSD: version 5.34, November 2012)⁴⁹ was analyzed with ConQuest (1.15)⁸⁹ to extract molecules containing the following ligands: *N*-heterocyclic carbenes based on both imidazol-2-ylidene and imidazolidin-2-ylidene skeletons; acyclic phosphines; as well as aromatic nitrogen-containing heterocycles such as pyridines, pyridazines, pyrimidines, imidazoles, and pyrazoles. ConQuest's substructure queries are reported in the Supporting Information, section S6. The search returned an initial data set of 31 603 entries (data set termed **SM1**). Each member of **SM1** was submitted to a prefiltering step in order to detect, and possibly fix, potential incompatibilities with the fragmentation tool. Next, the fragmentation was performed using the cutting rules listed in the Supporting Information, section S1.1. During filtering and fragmentation, 7819 molecules were rejected, most of them due to missing atoms (5720),⁹⁰ others due to failures in converting CSD's notation for aromatic bonds (1242), and some due to other problems in the automated detection of the constitutional features of the molecule (857).⁹¹ Thus, the fragmentation successfully processed 23 784 molecules of **SM1**. Duplicate fragments⁹² were removed and filtering criteria were applied for each fragment (see the Supporting Information for details, section S4), resulting in a library of 2054 fragments. Five other fragments were added to complete the library according to the target chemistry (Figure 10): a selection of four X-ligands (CF_3COO^- , $\text{C}_2\text{F}_5\text{COO}^-$, $\text{C}_6\text{F}_5\text{O}^-$, $\text{C}_6\text{F}_5\text{S}^-$) and a monatomic fragment (Fe with 2 APs) used to build ferrocene-containing substituents.⁹³ These fragments were generated by a simplified fragmentation of selected CSD entries (CSD reference codes: EGIBUS, DACCUG, AJEFAX, BADDOA) applying only the cutting rules acting on the suitable metal–ligand bonds. This particular procedure is justified by the fact that these fragments would be further fragmented if all the cutting rules were to be applied, but, in the current contribution, one of the tasks was to impose the use of precisely defined fragments, namely CF_3COO^- , $\text{C}_2\text{F}_5\text{COO}^-$, $\text{C}_6\text{F}_5\text{O}^-$, and $\text{C}_6\text{F}_5\text{S}^-$, on defined APs (see Figure 10) rather than building various carboxylate,

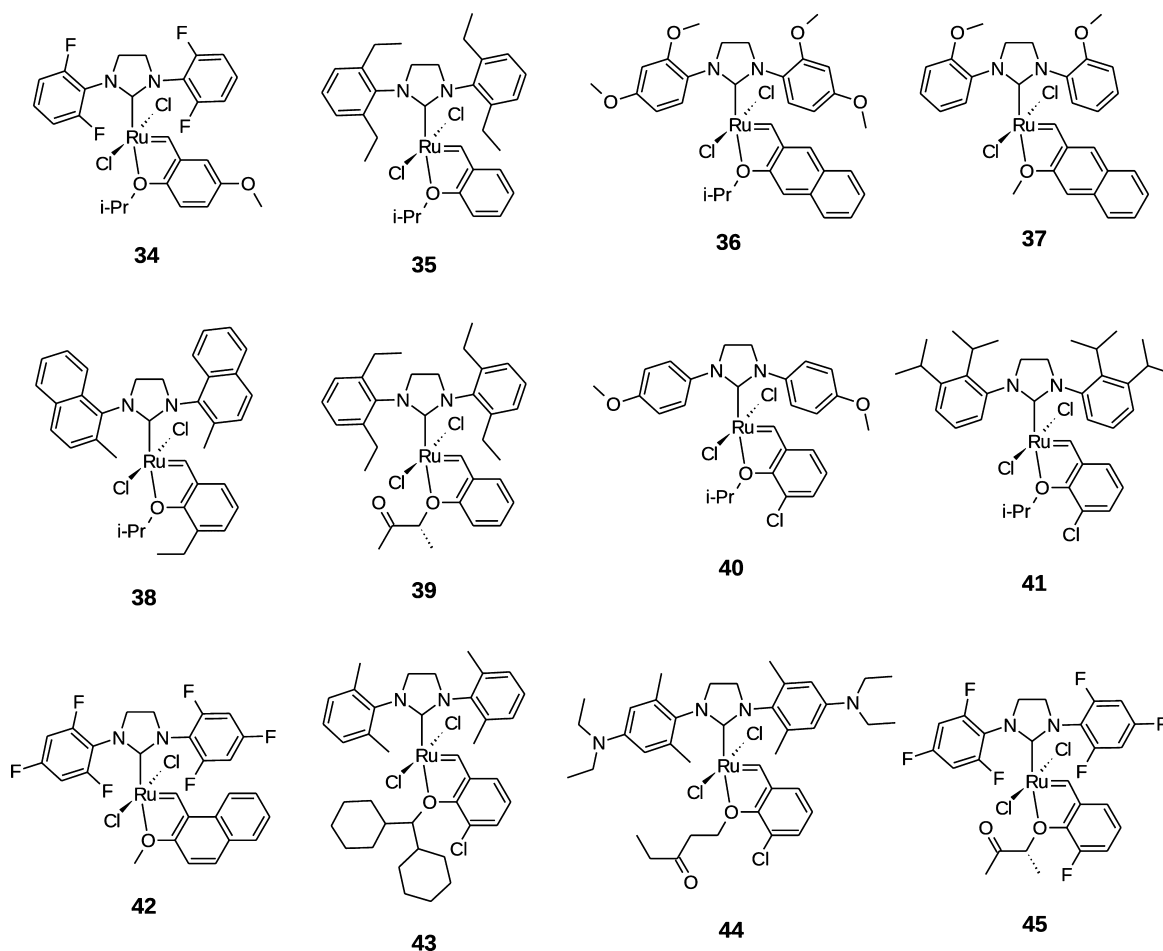


Figure 15. Representative molecules automatically generated from the combination of existing substituents used in Hoveyda–Grubbs catalysts.

alcoholate, and thiolate ligands on those APs. Capping groups were generated using the same approach, i.e., simplified fragmentation of selected CSD entries and selection of the proper fragments. Nine capping groups were collected: single hydrogen atom (H–), methylene (=CH₂), methyl (–CH₃), methoxy (–OCH₃), chloride (–Cl), phosphine (–PH₃), cyclopentadienyl (–Cp), Fe-cyclopentadienyl (–FeCp), and phenyl (–Ph).

The seed fragment for the building process is shown in Figure 9. The maximum size allowed for a molecule was set to 60 non-hydrogen atoms. P_{symm} was set to 95%, and the parameters controlling the behavior of P_{grow} were $s = 2.6$ and $m = 1.2$ (see eq 1). The same parameters were applied for both experiments, i.e., the application of the valence-only AP-compatibility scheme (case study IA) and the use of the strictly designed compatibility matrix (case study IB). A pictorial representation of the compatibility matrix used in the last case is reported in the Supporting Information, section S3.

5.2. Case Study II. To generate realistic Hoveyda–Grubbs-type catalysts, previously synthesized ligands were collected from molecules reported in CSD and commercial catalysts not included in CSD. From CSD, all molecules bearing aromatic substituents on nitrogen atoms of imidazolidin-2-ylidene ligands or the alkoxyarylmethylidene scaffold were harvested, with the corresponding substructure queries reported in the Supporting Information, section S6. Of the 338 molecules obtained, 20 were rejected as they could not be properly treated by the fragmentation process.⁹⁴ The remaining 318 molecules

were collected in a library in addition to 18 commercial Hoveyda–Grubbs catalysts not included in the list of molecules retrieved from CSD, see Supporting Information, section S5. The resulting data set contained 336 molecules (**SM2**). The fragmentation of the molecules of **SM2** was performed with a limited list of cutting rules, see the Supporting Information, section S1.2, and proper filtering criteria (see section S4) designed to produce Ru–alkoxyarylmethylidene scaffolds (Figure 14), fragments representing the substituents on the nitrogen atoms of imidazolidin-2-ylidene ligands, and fragments representing substituents of Ru–alkoxyarylmethylidene scaffolds. The resulting library, after removal of duplicates, contained 59 fragments. The composition of the library can be summarized as follows: six Ru–alkoxyarylmethylidene scaffolds, 36 aromatic substituents for imidazolidin-2-ylidene, five substituents for the oxygen atom of Ru–alkoxyarylmethylidene scaffolds (fragments B1–B6 in Figure 14), 11 substituents for the aromatic carbon in fragment B2, and one substituent for the aromatic carbon in fragment B3. Given the limited number of substituents for the ring positions of the Ru–alkoxyarylmethylidene scaffolds, 14 new fragments were added to the library, covering different functional groups commonly found as a decoration of aromatic rings (–CH₃, –NH₂, –OH, –F, –Cl, –Br, –I, –CN, –CH₂CH₃, –OCHH₃, –SH, –iPr, –NO₂, –CF₃). All 14 fragments were taken from the fragmentation of related CSD entries (CSD reference codes: IGOHOC, IKIJES, KUVNEU, MEDBUT, PAMKUL, PANMEY, QAZGAZ, QILKIF, RIPWES, RIXJET, RIRDAX,

RIRWAR, UJAKIZ, and YERZUQ). Thus, the library used for the building step contained a total of 73 fragments. The seed fragment was imidazolidin-2-ylidene, fragment A1 in Figure 14. Capping groups ($-\text{H}$, $-\text{CH}_3$, $-\text{iPr}$, $-\text{Cl}$, $-\text{Ph}$, $-\text{Ru}(\text{Cl}_2)(2\text{-isopropoxybenzylidene})$) were collected as needed from the previous library of capping groups (see section 5.1) or from simplified fragmentation of a selected molecule (CSD reference code JOFREX). The use of a Ru-containing capping group is necessary to build the smallest Hoveyda–Grubbs core scaffold via a capping procedure.

The parameters guiding the building process were chosen to generate molecules with a maximum of 90 non-hydrogen atoms. By setting $P_{\text{symm}} = 1.0$, the builder was forced to produce only NHC ligands with symmetric substitutions on the two nitrogen atoms. Finally, $s = 2.0$ and $m = 1.0$ were chosen to define the decay of P_{grow} (see eq 1). The compatibility matrix was defined so as to include all combinations originating from complementary AP classes along with the combinations necessary to use the 14 fragments as decoration of the Ru–alkoxyarylmethylidene scaffolds.

5.3. Software. In-house software written in Java has been developed for performing the analysis and filtering of the initial set of molecules and fragments, the fragmentation, and the building. The code makes use of functionality available in the CDK Chemistry Development Kit (CDK)^{95,96} library.

■ ASSOCIATED CONTENT

■ Supporting Information

Cutting rules, compatibility matrices, a list of filtering criteria for fragments, a list of commercial Hoveyda–Grubbs catalysts treated in this study, queries for surveying CSD, and two sets of Ru complexes generated automatically. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: vidar.jensen@kj.uib.no.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the Research Council of Norway (NFR) for financial support via the eVITA (grant number 205273/V30) and GASSMAKS (208335/E30) programs and for CPU and storage resources granted through the NOTUR (NN2506K) and NORSTORE (NS2506K) supercomputing programs.

■ ABBREVIATIONS

AP, attachment point; NHC, N-heterocyclic carbene; Cp, cyclopentadienyl; Cy, cyclohexyl; CSD, Cambridge Structural Database; $n\text{D}$, n -dimensional

■ REFERENCES

- (1) Schneider, G.; Fechner, U. Computer-Based de Novo Design of Drug-like Molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649–663.
- (2) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg, B. K. An Evolutionary Algorithm for de Novo Optimization of Functional Transition Metal Compounds. *J. Am. Chem. Soc.* **2012**, *134*, 8885–8895.
- (3) Burello, E.; Rothenberg, G. In Silico Design in Homogeneous Catalysis Using Descriptor Modelling. *Int. J. Mol. Sci.* **2006**, *7*, 375–404.

- (4) Maldonado, A. G.; Rothenberg, G. Predictive Modeling in Homogeneous Catalysis: A Tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891–1902.
- (5) Vukovic, S.; Hay, B. P. De Novo Structure-Based Design of Bis-Amidoxime Uranophiles. *Inorg. Chem.* **2013**, *52*, 7805–7810.
- (6) Drummond, M. L.; Sumpter, B. G. Use of Drug Discovery Tools in Rational Organometallic Catalyst Design. *Inorg. Chem.* **2007**, *46*, 8613–8624.
- (7) Ball, D. M.; Buda, C.; Gillespie, A. M.; White, D. P.; Cundari, T. R. Can Semiempirical Quantum Mechanics Be Used To Predict the Spin State of Transition Metal Complexes? An Application of De Novo Prediction. *Inorg. Chem.* **2002**, *41*, 152–156.
- (8) Buda, C.; Burt, S. K.; Cundari, T. R.; Shenkin, P. S. De Novo Structural Prediction of Transition Metal Complexes: Application to Technetium. *Inorg. Chem.* **2002**, *41*, 2060–2069.
- (9) Buda, C.; Flores, A.; Cundari, T. R. De Novo Prediction of the Ground State Structure of Transition Metal Complexes Using Semiempirical and Ab Initio Quantum Mechanics. Coordination Isomerism. *J. Coord. Chem.* **2005**, *58*, 575–585.
- (10) Bauerschmidt, S.; Gasteiger, J. Overcoming the Limitations of a Connection Table Description: A Universal Representation of Chemical Species. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 705–714.
- (11) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3–50.
- (12) Dobson, C. M. Chemical Space and Biology. *Nature* **2004**, *432*, 824–828.
- (13) Faulon, J.-L.; Visco, D. P.; Roe, D. Enumerating Molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Larter, R., Cundari, T. R., Eds.; John Wiley & Sons, Inc.: New York, 2005; pp 209–286.
- (14) Yu, M. J. Natural Product-Like Virtual Libraries: Recursive Atom-Based Enumeration. *J. Chem. Inf. Model.* **2011**, *51*, 541–557.
- (15) Reymond, J.-L.; Ruddigkeit, L.; Blum, L.; van Deursen, R. The Enumeration of Chemical Space. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 717–733.
- (16) Reymond, J.-L.; Awale, M. Exploring Chemical Space for Drug Discovery Using the Chemical Universe Database. *ACS Chem. Neurosci.* **2012**, *3*, 649–657.
- (17) Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.
- (18) Kutchukian, P. S.; Shakhnovich, E. I. De Novo Design: Balancing Novelty and Confined Chemical Space. *Expert Opin. Drug Discovery* **2010**, *5*, 789–812.
- (19) Hay, B. P.; Firman, T. K. HostDesigner: A Program for the de Novo Structure-Based Design of Molecular Receptors with Binding Sites That Complement Metal Ion Guests. *Inorg. Chem.* **2002**, *41*, 5502–5512.
- (20) Hageman, J. A.; Westerhuis, J. A.; Frühauf, H.-W.; Rothenberg, G. Design and Assembly of Virtual Homogeneous Catalyst Libraries – Towards in Silico Catalyst Optimisation. *Adv. Synth. Catal.* **2006**, *348*, 361–369.
- (21) Hay, B. P. De Novo Structure-Based Design of Anion Receptors. *Chem. Soc. Rev.* **2010**, *39*, 3700.
- (22) Hay, B. P.; Oliferenko, A. A.; Uddin, J.; Zhang, C.; Firman, T. K. Search for Improved Host Architectures: Application of de Novo Structure-Based Design and High-Throughput Screening Methods To Identify Optimal Building Blocks for Multidentate Ethers. *J. Am. Chem. Soc.* **2005**, *127*, 17043–17053.
- (23) Hay, B. P.; Firman, T. K.; Lumetta, G. J.; Rapko, B. M.; Garza, P. A.; Sinkov, S. I.; Hutchison, J. E.; Parks, B. W.; Gilbertson, R. D.; Weakley, T. J. Toward the Computer-Aided Design of Metal Ion Sequestering Agents. *J. Alloys Compd.* **2004**, *374*, 416–419.
- (24) Reyheller, C.; Hay, B. P.; Kubik, S. Influence of Linker Structure on the Anion Binding Affinity of Biscyclopeptides. *New J. Chem.* **2007**, *31*, 2095–2102.
- (25) Lumetta, G. J.; Rapko, B. M.; Garza, P. A.; Hay, B. P.; Gilbertson, R. D.; Weakley, T. J. R.; Hutchison, J. E. Deliberate Design

of Ligand Architecture Yields Dramatic Enhancement of Metal Ion Affinity. *J. Am. Chem. Soc.* **2002**, *124*, 5644–5645.

(26) Dey, F.; Caflisch, A. Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf. Model.* **2008**, *48*, 679–690.

(27) Böhm, H.-J.; Banner, D. W.; Weber, L. Combinatorial Docking and Combinatorial Chemistry: Design of Potent Non-Peptide Thrombin Inhibitors. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 51–56.

(28) Kutchukian, P. S.; Lou, D.; Shakhnovich, E. I. FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space. *J. Chem. Inf. Model.* **2009**, *49*, 1630–1642.

(29) Beccari, A. R.; Cavazzoni, C.; Beato, C.; Costantino, G. iGen: A High Performance Workflow for Chemistry Driven de Novo Design. *J. Chem. Inf. Model.* **2013**, *53*, 1518–1527.

(30) Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. DOGS: Reaction-Driven de Novo Design of Bioactive Compounds. *PLoS Comput. Biol.* **2012**, *8*, e1002380.

(31) Fechner, U.; Schneider, G. Flux (1): A Virtual Synthesis Scheme for Fragment-Based de Novo Design. *J. Chem. Inf. Model.* **2006**, *46*, 699–707.

(32) Baber, J.; Feher, M. Predicting Synthetic Accessibility: Application in Drug Discovery and Development. *Mini-Rev. Med. Chem.* **2004**, *4*, 681–692.

(33) Podolyan, Y.; Walters, M. A.; Karypis, G. Assessing Synthetic Accessibility of Chemical Compounds Using Machine Learning Methods. *J. Chem. Inf. Model.* **2010**, *50*, 979–991.

(34) Bonnet, P. Is Chemical Synthetic Accessibility Computationally Predictable for Drug and Lead-like Molecules? A Comparative Assessment between Medicinal and Computational Chemists. *Eur. J. Med. Chem.* **2012**, *54*, 679–689.

(35) Boda, K.; Seidel, T.; Gasteiger, J. Structure and Reaction Based Evaluation of Synthetic Accessibility. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 311–325.

(36) Huang, Q.; Li, L.-L.; Yang, S.-Y. RASA: A Rapid Retrosynthesis-Based Scoring Method for the Assessment of Synthetic Accessibility of Drug-like Molecules. *J. Chem. Inf. Model.* **2011**, *51*, 2768–2777.

(37) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminformatics* **2009**, *1*, 8.

(38) Boda, K.; Johnson, A. P. Molecular Complexity Analysis of de Novo Designed Ligands. *J. Med. Chem.* **2006**, *49*, 5869–5879.

(39) Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the Art of Compiling and Using “Drug-Like” Chemical Fragment Spaces. *ChemMedChem* **2008**, *3*, 1503–1507.

(40) Mauser, H.; Stahl, M. Chemical Fragment Spaces for de Novo Design. *J. Chem. Inf. Model.* **2007**, *47*, 318–324.

(41) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.

(42) Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K.-H.; Schneider, G.; Jacoby, E.; Renner, S. Probing the Bioactivity-Relevant Chemical Space of Robust Reactions and Common Molecular Building Blocks. *J. Chem. Inf. Model.* **2012**, *52*, 1167–1178.

(43) Jover, J.; Fey, N. Screening Substituent and Backbone Effects on the Properties of Bidentate P,P-Donor Ligands (LKB-PPscreen). *Dalton Trans* **2013**, *42*, 172–181.

(44) Fey, N. The Contribution of Computational Studies to Organometallic Catalysis: Descriptors, Mechanisms and Models. *Dalton Trans.* **2010**, 296–310.

(45) Fey, N.; Orpen, A. G.; Harvey, J. N. Building Ligand Knowledge Bases for Organometallic Chemistry: Computational Description of phosphorus(III)-Donor Ligands and the Metal–phosphorus Bond. *Coord. Chem. Rev.* **2009**, *253*, 704–722.

(46) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, *29*, 6245–6258.

(47) Macchioni, A. Ion Pairing in Transition-Metal Organometallic Chemistry. *Chem. Rev.* **2005**, *105*, 2039–2074.

(48) Nomura, K.; Liu, J.; Padmanabhan, S.; Kitiyanan, B. Nonbridged Half-Metallocenes Containing Anionic Ancillary Donor Ligands: New Promising Candidates as Catalysts for Precise Olefin Polymerization. *J. Mol. Catal. Chem.* **2007**, *267*, 1–29.

(49) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380–388.

(50) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Spellmeyer, D. C., Ed.; Elsevier: Amsterdam, 2008; Vol. 4, pp 217–241.

(51) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.

(52) For the machinery to work correctly, a preliminary evaluation of the input molecules is performed automatically to detect, and possibly fix, issues related to the notation and the connectivity matrix.

(53) This follows from the common practice, i.e., in the Cambridge Structural Database, of representing multihapto ligands specifying all metal–ligand atom bonds explicitly.

(54) Cotton, F. A. Proposed Nomenclature for Olefin-Metal and Other Organometallic Complexes. *J. Am. Chem. Soc.* **1968**, *90*, 6230–6232.

(55) *Daylight Theory Manual*; Daylight Chemical Information System, Inc.: Laguna Niguel, CA. <http://www.daylight.com/dayhtml/doc/theory/index.html> (accessed Mar 21, 2013).

(56) This may happen if the use of a suitable capping group would lead to undesired molecules. For example, to cap the AP of a carbene ligand would require a group including a metal center, but the designer might be interested neither in polymetallic systems (resulting from the use of such a capping procedure) nor in producing molecules having free carbenes (the result if no capping is performed).

(57) Hoveyda, A. H.; Zhugralin, A. R. The Remarkable Metal-Catalysed Olefin Metathesis Reaction. *Nature* **2007**, *450*, 243–251.

(58) Scholl, M.; Ding, S.; Lee, C. W.; Grubbs, R. H. Synthesis and Activity of a New Generation of Ruthenium-Based Olefin Metathesis Catalysts Coordinated with 1,3-Dimesityl-4,5-Dihydroimidazol-2-Ylidene Ligands. *Org. Lett.* **1999**, *1*, 953–956.

(59) Huang, J.; Stevens, E. D.; Nolan, S. P.; Petersen, J. L. Olefin Metathesis-Active Ruthenium Complexes Bearing a Nucleophilic Carbene Ligand. *J. Am. Chem. Soc.* **1999**, *121*, 2674–2678.

(60) Trnka, T. M.; Grubbs, R. H. The Development of L_2X_2RuCHR Olefin Metathesis Catalysts: An Organometallic Success Story. *Acc. Chem. Res.* **2001**, *34*, 18–29.

(61) Occhipinti, G.; Bjørsvik, H.-R.; Jensen, V. R. Quantitative Structure-Activity Relationships of Ruthenium Catalysts for Olefin Metathesis. *J. Am. Chem. Soc.* **2006**, *128*, 6952–6964.

(62) Occhipinti, G.; Hansen, F. R.; Törnroos, K. W.; Jensen, V. R. Simple and Highly Z-Selective Ruthenium-Based Olefin Metathesis Catalyst. *J. Am. Chem. Soc.* **2013**, *135*, 3331–3334.

(63) Fürstner, A. Teaching Metathesis “Simple” Stereochemistry. *Science* **2013**, *341*, 1229713.

(64) Kadyrov, R.; Wolf, D.; Azap, C.; Ostgard, D. J. The Enhancement of Metathesis Catalysts via Alkylidene and N-Heterocyclic Carbene Ligand Optimization. *Top. Catal.* **2010**, *53*, 1066–1072.

(65) Bieniek, M.; Samojłowicz, C.; Sashuk, V.; Bujok, R.; Śledź, P.; Luga, N.; Lavigne, G.; Arlt, D.; Grell, K. Rational Design and Evaluation of Upgraded Grubbs/Hoveyda Olefin Metathesis Catalysts:

Polyfunctional Benzyldiene Ethers on the Test Bench. *Organometallics* **2011**, *30*, 4144–4158.

(66) Poater, A.; Bahri-Laleh, N.; Cavallo, L. Rationalizing Current Strategies to Protect N-Heterocyclic Carbene-Based Ruthenium Catalysts Active in Olefin Metathesis from C–H (de)activation. *Chem. Commun.* **2011**, *47*, 6674–6676.

(67) Torker, S.; Müller, A.; Sigrist, R.; Chen, P. Tuning the Steric Properties of a Metathesis Catalyst for Copolymerization of Norbornene and Cyclooctene toward Complete Alternation. *Organometallics* **2010**, *29*, 2735–2751.

(68) Fogg, D. E. Inside the Black Box — Perspectives on Transformations in Catalysis. *Can. J. Chem.* **2008**, *86*, 931–941.

(69) Green, M. L. H. A New Approach to the Formal Classification of Covalent Compounds of the Elements. *J. Organomet. Chem.* **1995**, *500*, 127–148.

(70) In this analysis, ligands were classified according to the formal number of electrons donated to the metal (i.e., one electron for X ligands, two electrons for L ligands). In cases where several formal electron counts are possible, only the lowest number of electrons was taken into account. For example, X ligands with available lone pairs, such as alkoxido and dialkylamido, that may be considered to be three electron donors (XL ligands) depending on the chemical context, were counted as X ligands. For a significant fraction of the molecules (10.8%), the formal number of electrons could not be counted because the ligands did not resemble any common organometallic ligand.

(71) Ihlenfeldt, W.-D.; Gasteiger, J. Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs. *Angew. Chem., Int. Ed. Engl.* **1996**, *34*, 2613–2633.

(72) Gillet, V. J.; Myatt, G.; Zsoldos, Z.; Johnson, A. P. SPROUT, HIPPO and CAESA: Tools for de Novo Structure Generation and Estimation of Synthetic Accessibility. *Perspect. Drug Discovery Des.* **1995**, *3*, 34–50.

(73) Strassberger, Z.; Mooijman, M.; Ruijter, E.; Alberts, A. H.; Maldonado, A. G.; Orru, R. V. A.; Rothenberg, G. Finding Furfural Hydrogenation Catalysts via Predictive Modelling. *Adv. Synth. Catal.* **2010**, *352*, 2201–2210.

(74) Kingsbury, J. S.; Harrity, J. P. A.; Bonitatebus, P. J.; Hoveyda, A. H. A Recyclable Ru-Based Metathesis Catalyst. *J. Am. Chem. Soc.* **1999**, *121*, 791–799.

(75) Garber, S. B.; Kingsbury, J. S.; Gray, B. L.; Hoveyda, A. H. Efficient and Recyclable Monomeric and Dendritic Ru-Based Metathesis Catalysts. *J. Am. Chem. Soc.* **2000**, *122*, 8168–8179.

(76) Borré, E.; Rouen, M.; Laurent, I.; Magrez, M.; Caijo, F.; Crévisy, C.; Solodenko, W.; Toupet, L.; Frankfurter, R.; Vogt, C.; et al. A Fast-Initiating Ionically Tagged Ruthenium Complex: A Robust Supported Pre-Catalyst for Batch-Process and Continuous-Flow Olefin Metathesis. *Chem.—Eur. J.* **2012**, *18*, 16369–16382.

(77) Wakamatsu, H.; Blechert, S. A New Highly Efficient Ruthenium Metathesis Catalyst. *Angew. Chem., Int. Ed.* **2002**, *41*, 2403–2405.

(78) Grela, K.; Harutyunyan, S.; Michrowska, A. A Highly Efficient Ruthenium Catalyst for Metathesis Reactions. *Angew. Chem.* **2002**, *114*, 4210–4212.

(79) Kos, P.; Savka, R.; Plenio, H. Fast Olefin Metathesis: Synthesis of 2-Aryloxy-Substituted Hoveyda-Type Complexes and Application in Ring-Closing Metathesis. *Adv. Synth. Catal.* **2013**, *355*, 439–447.

(80) Berlin, J. M.; Campbell, K.; Ritter, T.; Funk, T. W.; Chlenov, A.; Grubbs, R. H. Ruthenium-Catalyzed Ring-Closing Metathesis to Form Tetrasubstituted Olefins. *Org. Lett.* **2007**, *9*, 1339–1342.

(81) Bieniek, M.; Bujok, R.; Cabaj, M.; Lugan, N.; Lavigne, G.; Arlt, D.; Grela, K. Advanced Fine-Tuning of Grubbs/Hoveyda Olefin Metathesis Catalysts: A Further Step toward an Optimum Balance between Antinomic Properties. *J. Am. Chem. Soc.* **2006**, *128*, 13652–13653.

(82) Ritter, T.; Day, M. W.; Grubbs, R. H. Rate Acceleration in Olefin Metathesis through a Fluorine–Ruthenium Interaction. *J. Am. Chem. Soc.* **2006**, *128*, 11768–11769.

(83) Rix, D.; Caijo, F.; Laurent, I.; Boeda, F.; Clavier, H.; Nolan, S. P.; Mauduit, M. Aminocarbonyl Group Containing Hoveyda–Grubbs-

Type Complexes: Synthesis and Activity in Olefin Metathesis Transformations. *J. Org. Chem.* **2008**, *73*, 4225–4228.

(84) Fujihara, T.; Tomike, Y.; Ohtake, T.; Terao, J.; Tsuji, Y. Ruthenium-Catalyzed Ring-Closing Metathesis Accelerated by Long-Range Steric Effect. *Chem. Commun.* **2011**, *47*, 9699–9701.

(85) Stewart, I. C.; Ung, T.; Pletnev, A. A.; Berlin, J. M.; Grubbs, R. H.; Schrodi, Y. Highly Efficient Ruthenium Catalysts for the Formation of Tetrasubstituted Olefins via Ring-Closing Metathesis. *Org. Lett.* **2007**, *9*, 1589–1592.

(86) Thomas, R. M.; Fedorov, A.; Keitz, B. K.; Grubbs, R. H. Thermally Stable, Latent Olefin Metathesis Catalysts. *Organometallics* **2011**, *30*, 6713–6717.

(87) Vieille-Petit, L.; Clavier, H.; Linden, A.; Blumentritt, S.; Nolan, S. P.; Dorta, R. Ruthenium Olefin Metathesis Catalysts with N-Heterocyclic Carbene Ligands Bearing N-Naphthyl Side Chains. *Organometallics* **2010**, *29*, 775–788.

(88) Although it is clear that the six scaffolds resulting from the combination of A1 and B1–B6 can easily be generated manually, part of the scope here is to perform the whole process automatically, meaning that the method should not be biased by an initial manual step. Besides, this approach also shows that even purely organic fragments can be used as seed fragments in the construction of organometallic compounds.

(89) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New Software for Searching the Cambridge Structural Database and Visualizing Crystal Structures. *Acta Crystallogr., Sect. B* **2002**, *58*, 389–397.

(90) Missing atoms were detected by comparing the molecular formula declared in CSD with the actual elemental analysis of the SD file provided by CSD.

(91) Apart from cases where the molecule did not contain any of the target bonds (because of inconsistencies in the notation), we experienced some problems in the application of the otherwise excellent SMARTS query tools of the CDK library.

(92) Two fragments are identical only if they have the same constitution (atoms and bonds) and the same pattern of APs, i.e., the same number of APs, reference atoms, and AP classes.

(93) Even if the fragment Fe with two APs for ferrocene-like systems was actually generated by the fragmentation of **SM1**, we preferred to delete all Fe-containing fragments during the fragmentation process and reintroduce only one such fragment afterwards.

(94) Missing atoms were detected for nine molecules, while the aromaticity of eleven molecules was not properly converted.

(95) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493–500.

(96) Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. Recent Developments of the Chemistry Development Kit (CDK) - An Open-Source Java Library for Chemo- and Bioinformatics. *Curr. Pharm. Des.* **2006**, *12*, 2111–2120.