JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Automated Building of Organometallic Complexes from 3D Fragments

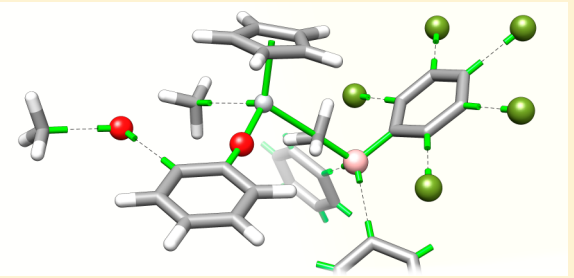
Marco Foscato,[†] Vishwesh Venkatraman,[‡] Giovanni Occhipinti,[†] Bjørn K. Alsberg,[‡] and Vidar R. Jensen*[,†]

[†]Department of Chemistry, University of Bergen, Allégaten 41, N-5007 Bergen, Norway

[‡]Department of Chemistry, Norwegian University of Science and Technology, Høgskoleringen 1, N-7491 Trondheim, Norway

Supporting Information

ABSTRACT: A method for the automated construction of three-dimensional (3D) molecular models of organometallic species in design studies is described. Molecular structure fragments derived from crystallographic structures and accurate molecular-level calculations are used as 3D building blocks in the construction of multiple molecular models of analogous compounds. The method allows for precise control of stereochemistry and geometrical features that may otherwise be very challenging, or even impossible, to achieve with commonly available generators of 3D chemical structures. The new method was tested in the construction of three sets of active or metastable organometallic species of catalytic reactions in the homogeneous phase. The performance of the method was compared with those of commonly available methods for automated generation of 3D models, demonstrating higher accuracy of the prepared 3D models in general, and, in particular, a much wider range with respect to the kind of chemical structures that can be built automatically, with capabilities far beyond standard organic and main-group chemistry.



INTRODUCTION

Techniques such as *de novo* design and virtual screening are powerful molecular design methods that only to a limited extent have been exploited in organometallic and transition-metal chemistry.[1−3] Efficient application of these techniques requires extensive automation of all steps of the molecular modeling procedures: from the building of new, realistic candidate compounds to the identification of the best promising candidates via estimation of molecular properties. In a recent contribution, we presented a method for the first of these steps, that is, the automated assembly of candidate organometallic compounds from fragments.[4] While this operation can be performed with various chemical representations, that may or may not include atomic coordinates, the evaluation of generated candidates and the identification of the most promising candidates may require the complete three-dimensional (3D) molecular structure. In fact, while analysis of two-dimensional (2D) structures, or even string-like chemical representations,[5,6] may provide good estimates of some molecular properties, high accuracy of calculated molecular properties in general can only be obtained based on 3D structures.[1,7,8] In addition, since automated design tools are likely to arrive at new chemical compounds, for which no experimental structural characterization exists, the 3D models have to be generated on-the-fly.

A number of methods and standalone tools have been developed in the context of organic drug design for the automated generation of 3D models from representations of lower dimensionality.[9−12] Both freely available and commercial tools have been systematically tested and validated against selected data sets of organic drug-like molecules.[12−20] In contrast, the complexity of organometallic and transition-metal chemistry has hampered the application of tools for automated generation of 3D models to this class of compounds.[21] Alternative approaches for the preparation and evaluation of 3D models have included development of specific force field methods, for instance capable of accounting for the ligand field stabilization energy,[22−24] or combination of sequences of empirical and semiempirical methods before ab initio calculations.[21,25−27] In addition, virtual construction of new 3D models has been achieved also by direct connection of structural fragments in 3D space and used in *de novo* design of host−guest systems[28] and in virtual fragment-based drug design,[29,30] which is also integrated in popular software packages.[31,32] While few of the automated methods originally developed for the generation of 3D structures of organic molecules have been demonstrated to work also for specific organometallic systems,[11,14] systematic analyses of their performance with respect to organometallic and transition-metal compounds have been reported only in a couple of recent contributions.[33,34] Baldi and co-workers evaluated the performance of widely available 3D generation tools pointing out issues of accuracy and coverage for the existing methods with respect to the metal-containing molecules.[33] To overcome these problems, they developed a new algorithm for the prediction

Received: May 28, 2014
Published: July 5, 2014

of 3D molecular geometries, termed COSMOS, with improved performances for organometallic compounds.[34]

In general, the aim of 3D generating methods is to give low-energy conformations that should be ready to use without further improvement and a simplified force field-based refinement is usually integrated within the algorithm for 3D generation.[10,33] However, in the context of automated molecular design, the role of the 3D generating methods may rather be to prepare reasonable initial 3D models that serve as input for further structure refinement, for example via geometry optimization using molecular mechanics, semiempirical methods, or density functional theory (DFT), or combinations thereof.[15,21] Still, the accuracy of the whole modeling process is dependent on the quality of the very first 3D representation. This aspect is of primary importance for applications within organometallic and transition-metal chemistry, where there are many limitations with respect to chemical representations[35,36] and few 3D generating tools are available. These problems are aggravated by the fact that most general-purpose force fields do not describe metal-containing molecules well.[37]

Here, we present an alternative method for automated generation of 3D molecular structures. The method is suited for the construction of 3D models for various analogues of a compound for which the molecular structure is known experimentally or is available from accurate molecular-level calculations. In fact, the homology between the geometry of the conserved region of a list of analogous compounds can be exploited to build the 3D models of many such analogues by using a 3D template.[33] Thus, in this method, 3D structural fragments are first generated and stored in libraries as 3D objects, and next used in the automated building of molecules by direct assembly of 3D fragments. Conversions between different chemical representations are avoided by the direct use of 3D building blocks. Moreover, since the fragments can be generated according to the specific needs of the user and do not have to represent stable molecular entities, any atomic arrangement can be modeled, thus providing the user with a flexible tool for automatic preparation of 3D models. The performance of the new method is evaluated and compared against common, existing 3D-generating methods.

METHODS

The construction of 3D models presented in this contribution is based on a general method for the automated design of organometallic compounds from fragments.[4] Whereas aspects of this method that are independent of the dimensionality of the chemical representation (i.e., string-like vs 2D vs 3D structural representation) have been described earlier,[4] the current contribution focuses on the particular challenges and solutions for direct construction of 3D molecular models from 3D structural fragments, an approach which here is generalized and implemented for the first time for the automated generation and design of organometallic and transition-metal compounds, including challenging species such as reaction intermediates and unconventionally bonded systems. The method described herein is referred to as the "full-3D" method.

In our previous contribution,[4] fragments were defined as "molecular building blocks", where so-called "attachment points" (AP) were used to store information as to the number, position and type of all the possible connections with other building blocks. For the implementation of the full-3D method, two types of information were included to define a 3D fragment: (i) the atomic coordinates of all atoms belonging to the fragment, and (ii) geometrical information defining the relative spatial position of the neighboring fragments connected by each AP of the fragment, to give what in the following will be referred to as the "AP vectors". Since fragments are mainly derived from pre-existing molecular structures by means of a fragmentation process (details reported elsewhere),[4] both kinds of geometrical information (i and ii) are derived from the complete 3D structures from which 3D fragments are generated. In particular, each AP vector is defined by the coordinates of the atom (or dummy atom) that is disconnected during the fragmentation step. In practice, the fragmentation process converts each dissociated bond into a pair of AP vectors (Figure 1, dashed arrows on the left). The geometrical

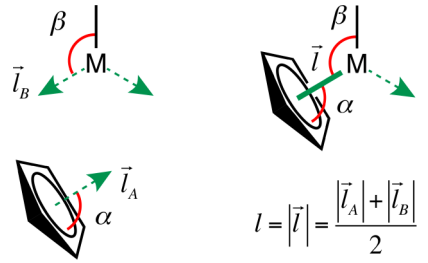


Figure 1. Two 3D fragments with AP vectors depicted as dashed arrows (left), and the structure resulting from connection of the two fragments. Vectors ($\vec{l}_A$, $\vec{l}_B$, $\vec{l}$,) are highlighted in green and angles (α, β) in red.

information implicitly included in AP vectors is then used to connect 3D fragments during the construction of a 3D model (Figure 1, right). The bond angles defining the fragment–fragment connections are simply those defined by the pair of AP vectors involved in the connection. The length of the new bond is taken to be the average of the lengths of the two AP vectors (see Figure 1, right). It is not possible to derive a value for the torsional angle of the fragment–fragment bond based on the atomic coordinates and the AP vectors of the participating fragments alone. Therefore, an arbitrary conformation, to be adjusted at a later stage (see below), is initially set for the new bond.

When this technique is applied to all fragment–fragment bonds in the construction of a molecule, the result is a 3D model with fairly realistic bond angles and distances. In fact, many such values are simply taken from crystallographic or previously modeled 3D structures used for the generation of the 3D fragments. Notably, the 3D fragments may even contain unusual geometrical features, such as those of transition states and unstable species that would otherwise be difficult to reproduce. Fragments of this kind may readily be introduced by the user. Next, the torsional angles of the fragment–fragment bonds, for which arbitrary values are used initially (see above), are optimized using the potential smoothing and search algorithm for the rotational space as implemented in the PSSROT program of the Tinker[38] package.[39−42] Two crucial aspects of the PSSROT calculation are defined in the following: the energy model and the definition of the rotational space.

The potential energy of the PSSROT calculations includes only the van der Waals term of the universal force field (UFF).[43] This simplification, although drastic, reduces the number of atom types and force fields parameters to a minimum, ensuring that also this part of the full-3D method is generally applicable, regardless of the nature of the chemical system. This means that, as long as the user provides the 3D

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

fragments needed for the construction of the target molecules, these building blocks will be assembled independently of their nature and they may contain exotic protonation states, oxidation state with peculiar geometries, or unusual chemical species. In spite of this geometrical flexibility, depending on the needs of the specific case and the availability of force field parameters, the PSSROT energy model can be extended to include the full variety of terms implemented in Tinker.

The rotational space is defined as the combination of all fragment−fragment connections and intrafragment rotatable bonds. While the fragment−fragment connections derive directly from the definition of the organometallic fragment space via the use of user-defined fragmentation rules (also known as "cutting rules", described elsewhere),[4] intrafragment rotatable bonds have to be identified on-the-fly from the molecular structure. To this end, an additional, predefined list of SMARTS[44] queries is used to identify the most common intrafragment rotatable bonds. The inclusion of such bonds in the definition of the rotational space allows for increased flexibility of the otherwise rigid fragments, permitting conformational adaptation during the PSSROT calculation.

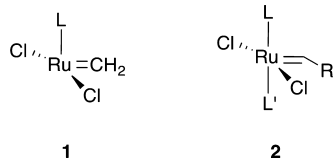
The current implementation of the molecular building algorithm only assembles tree-like structures,[4] which means that every cyclic substructure, independent of size and complexity, must be provided as part of a single 3D fragment. For instance, multidentate ligands and chelates are treated by including the whole metal−organic cycle in a single 3D fragment. The possibility of explicit ring-closure during the building process, for example by following strategies implemented in methods specifically aimed at the design of chelates,[28,45−47] will be the focus of future work.

RESULTS AND DISCUSSION

The performance of the full-3D approach in the generation of 3D models is evaluated in three different experiments. 3D models were also refined with molecular modeling methods, providing indication as to the actual usefulness of the generated 3D models as input for further calculations. In the first case study the full-3D approach is compared with a series of approaches based on chemical representations that do not include atomic coordinates, namely SMILES[36,44] strings. This experiment serves to demonstrate to what extent the full-3D method is able to overcome issues originating from loss of structural information during the automated SMILES-based treatment, not recovered by the SMILES to 3D converters. In the second case study, a varied set of 3D models is treated, providing insight as to the general applicability of the full-3D method. Finally, in the third case study, the full-3D method is tested on particularly challenging compounds with geometries and interactions not handled by tools mainly trained on standard organic chemistry.

In all three case studies the target molecules are active or metastable species that, because of the limited experimental structural information available, represent challenging tests for empirical tools. Since none of the molecules in this study have been characterized by X-ray or neutron diffraction, the reference structures have been obtained from high-accuracy computational studies.

Case Study 1. The comparison of the full-3D method against SMILES-based approaches was performed using a data set (termed **DS-1**) containing 82 organometallic ruthenium catalysts[48] for olefin metathesis.[49−52] Each molecule represents the active species (**1** in Figure 2) generated by activation of the corresponding precatalyst (**2** in Figure 2).[53,54] Consequently, all 82 molecules obey the general formula $(L)(Cl)_2Ru{=}CH_2$, where L is a dative ligand.



1 2

Figure 2. The 14-electron active species of **DS-1** (**1**) and the corresponding precatalysts (**2**). L and L′ are dative ligands. R is alkyl or aryl.

The list of dative ligands (L) includes ethers, thioethers, amines, phosphines, phosphites, arsines, carbonyl, thiocarbonyl, selenocarbonyl, isocyanides, N- and P-heterocyclic carbenes, and aromatic heterocyclic compounds (the complete list is reported elsewhere).[48] Predictive models derived from the analysis of this data set have already been used for the design of new catalysts,[3,48,55] demonstrating the chemical relevance of **DS-1** in the study of Ru-catalysts for olefin metathesis. Moreover, **DS-1** poses challenges that make this data set a suitable test ground for the generation of 3D models: (i) the diverse set of dative ligands allows for identification of classes of ligands erroneously treated by some of the evaluated approaches, (ii) the ruthenium coordination number (CN = 4) is often found for tetrahedral or square planar coordination geometries, but both such geometries are actually misleading for the complexes of **DS-1** (see below), and, as mentioned before, (iii) all molecules in **DS-1** are active species, and there is very little experimental information that can be used to train tools for prediction of such structures. In fact, only a handful of crystallographic structures of molecules that are related but not equal to those of **DS-1** exist. For this reason, DFT-optimized structures, previously published for all compounds in **DS-1**,[48] were taken as reference structures for this case study.

Next, 11 methods based on the use of the isomeric SMILES[44] representation (see Table 1) were evaluated for the generation of 3D models for **DS-1** and compared with the full-3D method. Each of these 11 methods is a particular combination of tools included in the Marvin suite,[56] Open Babel,[57] and COSMOS.[33,34] Each of these SMILES-based protocols involved three steps: (i) generation of isomeric SMILES from 3D structures, (ii) conversion to 3D, and, if needed, (iii) addition of hydrogen atoms, also known as "protonation" (see the Supporting Information, sections S1.1.1 and S2.1.1).

Since methods depending on the protonation step may return molecules with too few or too many hydrogen atoms (see Supporting Information, section S2.1.1), the following analysis focuses only on 3D models reflecting the correct molecular formula (those counted in the rightmost column of Table 1). Only the three-dimensional molecular structures of the latter, correct 3D models, are evaluated with respect to the reference structures. Whereas it is common practice to compare pairs of molecular geometries by means of root-mean-square deviation (RMSD) of atomic positions, this measure assumes precise identification and identical ordering of all atoms in the two structures being compared. Since this condition could not be always satisfied,[58] two different kinds of geometrical properties are used here for the comparison of the structures: the molecular shape, and the geometry of the metal center. The

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

Table 1. Definition of Protocols Used in Case Study 1 for Generation of 3D Models[a]

ID	3D-to-SMILES	SMILES-to-3D	protonation	wrong number of H	3D models produced
Generation of 3D models *via* SMILES					
1	Marvin	Marvin	none[b]	0	82
2	Marvin	COSMOS	Marvin	8	74
3	Marvin	COSMOS	Open Babel	11	71
4	Open Babel	Marvin	Marvin	2	80
5	Open Babel	Marvin	Open Babel	15	67
6	Open Babel	COSMOS	Marvin	10	70[c]
7	Open Babel	COSMOS	Open Babel	13	67[c]
8	Open Babel[d]	Marvin	Marvin	2	80
9	Open Babel[d]	Marvin	Open Babel	15	67
10	Open Babel[d]	COSMOS	Marvin	10	70[c]
11	Open Babel[d]	COSMOS	Open Babel	13	67[c]
Generation of 3D models from 3D fragments					
12		full-3D[e]		0	82

[a]The total number of output models, reproduced with correct molecular formula, is reported together with the number of molecules generated with wrong number of hydrogen atoms. [b]Protonation step not needed. [c]For two molecules COSMOS did not recognize the string generated by Open Babel as a correct SMILES string. [d]Canonical smiles. [e]This work, which does not involve SMILES strings or protonation steps; see the Methods section for details.

shape similarity index (T_S, see the Supporting Information, section S1.1.4) provides a normalized measure of the degree of similarity between the molecular shape of the generated 3D model and that of the reference structure. While low values for T_S in general can be attributed to a combination of effects that include different conformations as well as configurational isomerism and constitutional differences, the latter are not relevant for the present analysis since molecules with wrong number of hydrogen atoms were rejected.

The distributions of T_S values for the 3D models generated by the 12 methods are reported in Figure 3. In all cases a widespread of the T_S values demonstrates that a number of generated molecules have shapes significantly different from those of the respective reference molecules. Visual inspection of the 3D models generated by the full-3D method (series 12 in Figure 3) shows that, for this series, a low T_S value is always due to conformational differences between the generated molecule and the reference structure. This is an expected consequence of the simple van der Waals-based solution of the conformational problem described in the Methods section. In comparison, molecules with low T_S in the series 1−11 also suffer from conformational issues, but these conformational problems are compounded by significant geometrical distortions, some of which are highlighted by the analysis of the coordination geometry (see below). The models generated by the full-3D method suffer to a lesser extent from geometrical errors mainly because, in this method, all critical geometrical features, such as bond angles, most bond distances, and even entire portions of the system (e.g., cycles), are derived directly from the building blocks, namely the 3D fragments taken from experimentally determined molecular structures (e.g., from crystallography) of similar molecules or high-accuracy modeling. As a result, the distribution of T_S values for the full-3D method is shifted to higher values than those seen for all other methods. While the conformational problem could be handled by a number of specialized conformational search methods, which is beyond the scope of the present contribution, the presence of geometrical distortions is potentially fatal for any further modeling steps, including that of conformational search.

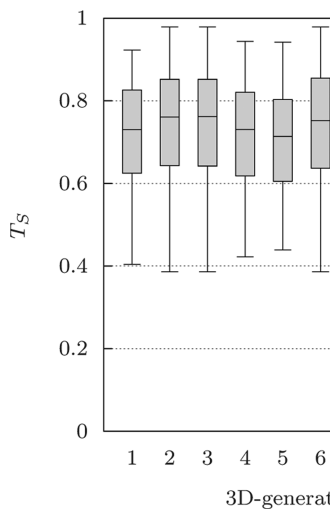


Figure 3. Boxplots[59] representing the distribution of the molecular shape similarity index (T_S) of the 3D models of data set **DS-1** generated by the 12 3D-generating protocols defined in Table 1.

As anticipated, the geometry around the ruthenium atoms in **DS-1** represents a challenge for the 3D generators. The Ru coordination number (4) suggests that a tetrahedral configuration might be a suitable starting, or guess, geometry for this atom. However, crystallographic structures of related compounds[60−62] as well as computational studies[48] show distorted geometries (*Ru 14-el* in Figure 4) that are more disphenoidal (*OC-4*) than tetrahedral (*T-4*).

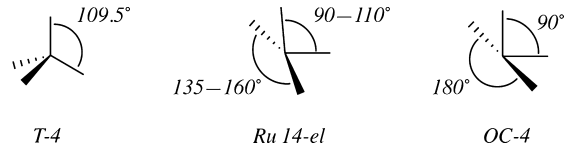


Figure 4. Schematic representation of relevant coordination geometries.

The mean angle difference (MAD)[33] of the metal center is used to provide a quantitative evaluation of the deviation from the corresponding reference geometry for Ru atoms in each of the generated 3D models. The challenge of predicting the geometry of these Ru centers is evident from the distributions of MAD values reported in Figure 5.

Owing to the use of a properly selected 3D fragment for the Ru atom (see the Supporting Information, section S1.1.2), the full-3D approach provided geometries with low MADs (series 12 in Figure 5) that are in general closer to the corresponding reference structures than the other methods. In fact, the MADs of series 1, 4, 5, 8, and 9 (involving Marvin SMILES-to-3D conversion) are only occasionally comparable to those of the full-3D method, and series 2, 3, 6, 7, 10, and 11 (involving COSMOS SMILES-to-3D conversion) give MADs that are always higher than those of the full-3D method. These results are explained by the fact that COSMOS and often also Marvin predict a tetrahedral geometry for these metal centers (see the Supporting Information, section S2.1.2). As is shown below, the geometrical distortions highlighted by both the shape similarity

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

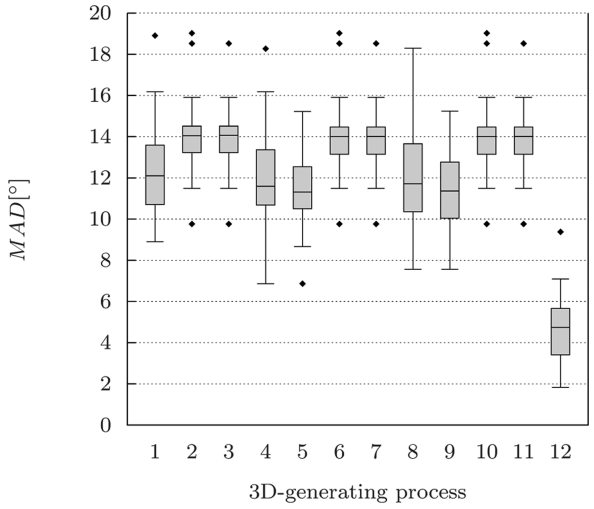


Figure 5. Boxplots[59] representing the distributions of mean angle differences (MADs) of the 3D models of data set **DS-1** generated by the 12 3D-generating protocols defined in Table 1.

index and the analysis of the geometry around the Ru atoms can seriously affect the outcome of any further calculation that makes use of these 3D models as input.

To evaluate the impact of the generated 3D geometries in subsequent molecular modeling, all 3D structures were subjected to three different refinement protocols, consisting of direct geometry optimization with (i) the universal force field molecular mechanics (labeled "UFF"), (ii) the semiempirical PM6 method (labeled "PM6"), and (iii) the OLYP density functional (labeled "$OLYP_B$", see the Supporting Information, section S1.1.3). The quality of the prerefinement 3D models, both in terms of correctness of the atomic arrangement and the number of atom clashes, affects the behavior of the calculations. Additionally, in some cases, manual correction of the input geometry, e.g., removal of atom clashes, would have been necessary in order for the refinement calculation to complete successfully. In fact, near-overlapping atoms were observed for some of the 3D models generated by COSMOS in series 2, 3, 6, 7, 10, and 11 (see section S2.1.3 of the Supporting Information).

The refinements based on density functional theory (DFT) illustrate some of the challenges of low-accuracy input geometries. As for the other refinement methods, efforts were made to identify a computational protocol for which a high number of calculations would converge without manual intervention. In particular, we initially attempted to converge the wave function directly in the geometry optimization using double-ζ basis sets (protocol labeled "$OLYP_A$", see the Supporting Information, section S1.1.3). Unfortunately, while the geometries prepared by the full-3D method could be refined with protocol $OLYP_A$, this approach turned out to be impossible for series 1−11 due to the huge number of cases where an initial, stable wave function could not be obtained. The convergence problems probably originate from low-quality input geometries, featuring near-degeneracies caused by certain internal coordinates, chiefly bond distances and angles, being far from their equilibrium values. Therefore, a preliminary geometry optimization step with minimum basis sets was introduced with the goal to improve the input geometry before applying the more demanding double-ζ basis sets (protocol labeled "$OLYP_B$", see the Supporting Information, section S1.1.3).

To analyze the accuracy of the structures obtained after refinement, a subset of **DS-1**, labeled **DS-2**, was identified to ensure that comparison of all 12 methods could be made for a common set of molecules. In fact, as explained in the Supporting Information (section S2.1.3), the number of successfully refined molecules differ for each of the methods 1−12. Therefore, **DS-2** includes only 35 molecules for which all series 1−12 returned a refined 3D model with (i) a proper molecular formula, (ii) a geometry successfully optimized by the three automated protocols (UFF, PM6, $OLYP_B$), and (iii) a preserved molecular constitution, that is, no substantial change in the atom connectivity. The Supporting Information (sections S2.1.4, S2.1.5, and S2.1.6) contains the complete analysis of all molecules satisfying the three criteria described above for each of the 12 3D-generating processes. In the following, the results for **DS-2** are presented and discussed in terms of the average shape similarity index ($\overline{T}_S$) and the overall mean angle deviation ($\overline{MAD}$).[63]

The averages of the shape similarity index suggest that UFF refinement is not always beneficial for the shape of 3D models, while, as expected, PM6 and $OLYP_B$ lead to significant improvements for methods 1−11 (Figure 6). In fact, PM6

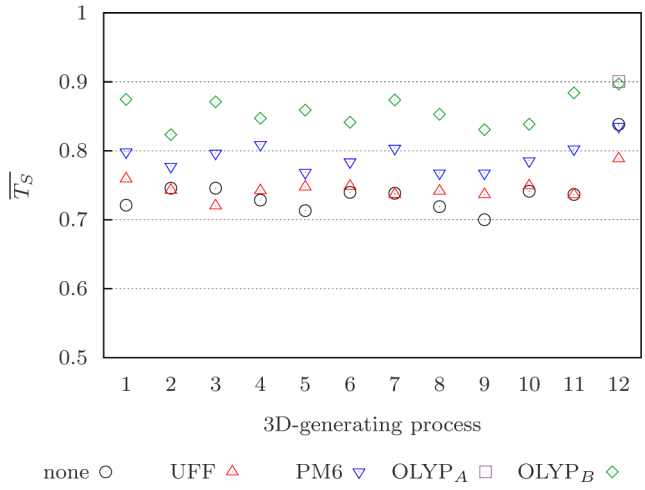


Figure 6. Average shape similarity index ($\overline{T}_S$) calculated for each of the 12 3D-generating processes, using subset **DS-2** before any refinement of the geometry (series "none"), and after molecular mechanics (series "UFF"), semiempirical (series "PM6"), and DFT (series "$OLYP_A$" and "$OLYP_B$") refinement processes.

and $OLYP_B$ are able to rectify geometrical distortions in the prerefinement 3D models, thus increasing the average values of T_S. This effect is much less visible in the case of the full-3D method (series 12), because of fewer geometrical distortions in the 3D models submitted to refinement. In particular, it can be observed that, while UFF refinement worsens the shape similarity index of the 3D models constructed with the full-3D method, PM6 provides no improvement of the average T_S with respect to the prerefinement 3D models (label "none" in Figure 6).

The fact that refinement by UFF and PM6 does not improve upon the quality of geometries generated by the full-3D method is also reflected in the analysis of the geometry around Ru atoms (Figure 7). The overall MAD values ($\overline{MAD}$) for UFF-refined structures are always very high. In fact, UFF

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

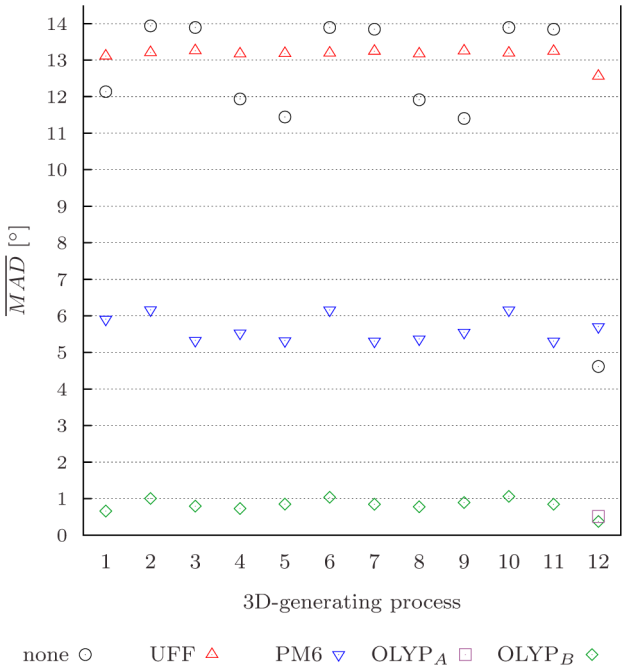


Figure 7. Overall mean angle difference ($\overline{MAD}$) calculated for each of the 12 3D-generating processes using subset **DS-2** before any refinement of the geometry (series "none"), and after molecular mechanics (series "UFF"), semiempirical (series "PM6"), and DFT (series "OLYP$_A$" and "OLYP$_B$") refinement processes.

refinement returned metal centers with disphenoidal geometry where two of the ligands are in a colinear spatial arrangement with the Ru atom (see *OC-4* in Figure 4). This occurs also for series 12 where the high quality of the local geometry of Ru atoms in the prerefinement 3D models (series 12 label "none" in Figure 7) is partly lost in the UFF and PM6 refinements. In fact, the full-3D approach, by way of using a proper 3D fragment for Ru atoms, is capable of preparing 3D models for which both the geometry of the metal center and the shape of the molecule are usually better than those obtained after UFF and PM6 refinements. In contrast, the average MAD for protocols 1−11 is always improved significantly by PM6 refinement. This improvement is related to the fact that, as described above, COSMOS (series 2, 3, 6, 7, 10, and 11) produces 3D models with tetrahedral Ru atoms. Similarly, the models generated by the Marvin suite (series 1, 4, 5, 8, and 9), despite showing a wider geometrical diversity than the other series, are much closer to a tetrahedral geometry than the reference structure (see the Supporting Information, section S2.1.2). Consequently, the overall MAD values before refinement are above 11° for series 1−11 (labeled "none" in Figure 7). It should be noted that high MAD values, which although represent deviation from the reference geometries, do not specify the direction of the deviation. In fact, the similar MAD values for the UFF and prerefinement models of series 1−11 reflect to opposite deviations with respect to the geometry of the reference structures. Whereas UFF-refined geometries are on average closer to the disphenoidal geometry (Supporting Information, Figure S9), the prerefined models of methods 1−11 are rather tetrahedral (Supporting Information, Figure S2).

The more accurate and computationally demanding refinement protocols manage to reduce the structural problems in the 3D models, in particular the coordination geometry of the metal center. However, the evaluation of the coordination geometry presented in Figure 7 considered only the relative position of the metal−ligand bonds around the central atom (Ru), neglecting the nature of the ligands involved.[33] In contrast, to address the stereochemistry of the complex, the relative positions of the different ligands should be used. Since the stereochemistry of the complex constitutes a fundamental property defining the identity of organometallic species, it is crucial for the whole 3D-generating process to produce refined structures with precisely controlled stereochemistry.

For the type of coordination geometry of the Ru compounds in **DS-1** (see Figure 4), six stereoisomers can be drawn (Figure 8). In the following, we aim to generate the same stereoisomer

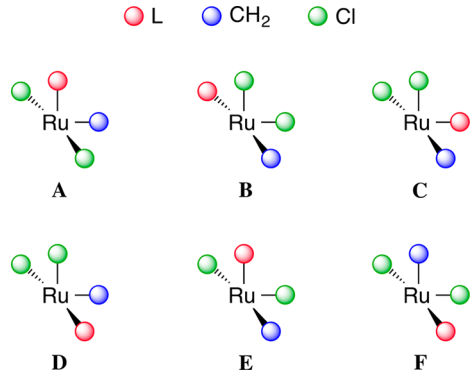


Figure 8. Stereoisomers of Ru(Cl)$_2$(L)(=CH$_2$). The geometry of the central atom is discussed in the text and presented in Figures 1 and 4. L represents a dative ligand.

for all the structures, corresponding to stereoisomer **A** in Figure 8. In other words, we do not explore the possibility that, for a given ligand, stereoisomers other than **A** might be more stable than **A**. We only aim to build models with controlled stereochemistry.

The identified stereochemistry of the metal centers produced after refinement is reported in Figure 9. The combinations of methods 1−11 with UFF refinement provide, at best, 30 molecules (37%) with proper stereochemistry. In contrast, the combination of UFF refinement and the full-3D approach performs well, returning the correct stereoisomer for 79 of the 82 molecules (96%). Of the remaining three molecules, one involved a dissociated ligand (tetrahydrofuran) and two showed distorted or formally pentacoordinate Ru centers. The latter distortions resulted from exceeding steric strain in the input structure caused by a bulky L (*tert*-butyl-substituted N-heterocyclic carbene). As expected, the performance of protocols 1−11 improved using PM6 refinement, and, even more so, with DFT. Both PM6 and OLYP$_B$ refinements manage to fix some of the input geometries, increasing the number of correct stereoisomers produced. Still, the performance of the full-3D method is better than that of approaches 1−11 for which there are always residual molecules generated with wrong stereochemistry.

The main explanation for the lack of stereochemical control in the 3D-generating protocols 1−11 is the tendency to produce tetra-coordinate Ru atoms with tetrahedral geometry. When generating a tetrahedral 3D model in which two of the four ligands are identical (in this case the two chlorides), only one isomer is possible. When such a tetrahedral model is refined, however, the geometry around the Ru atom changes toward the equilibrium geometry, which is closer to disphenoidal than tetrahedral (see *Ru 14-el* geometry in Figure

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

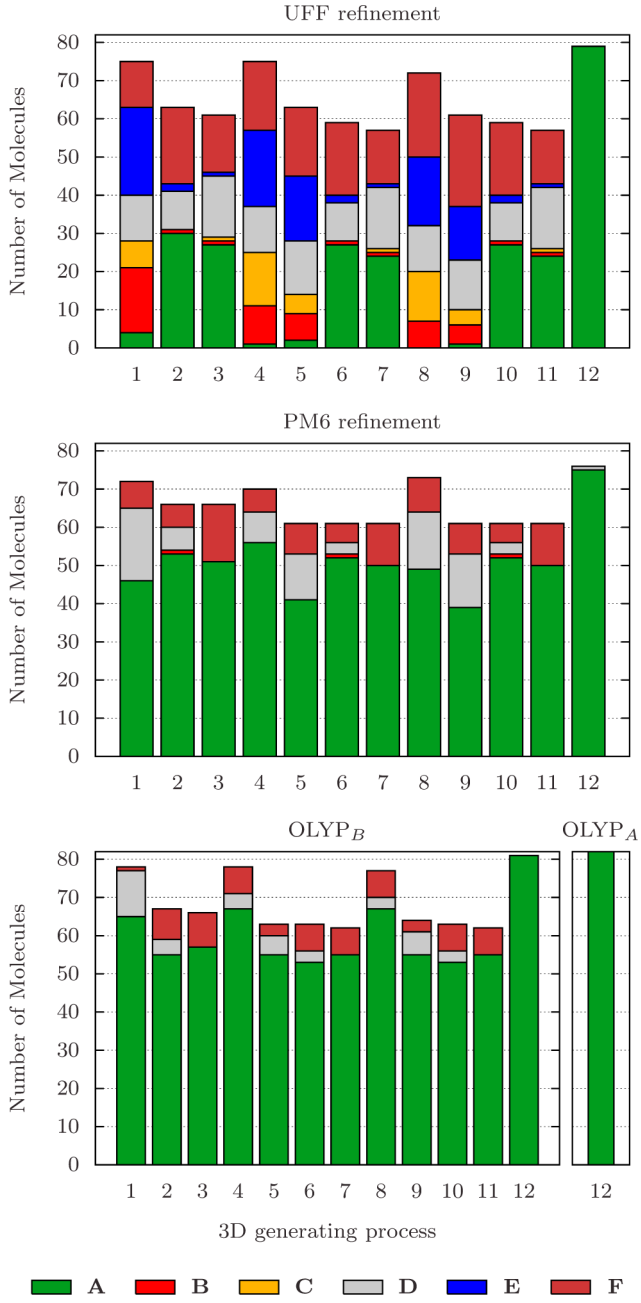


Figure 9. Distribution of stereoisomers for UFF-, PM6-, and OLYP-refined 3D models of data set **DS-1** obtained with the 12 3D-generating processes defined in Table 1. All molecules with correct molecular formula and connectivity are included. Stereoisomers **A**–**F** are defined in Figure 8.

4), thereby reducing the symmetry of the system. The resulting geometry usually resembles one of the stereoisomers depicted in Figure 8 (further details in the Supporting Information, section S2.1.3). The gradual displacements of the ligands are determined by the combination of the input geometry and the modeling method. When the model describing the potential energy surface is approximate and the input geometries are significantly distorted, e.g., tetrahedral or, even worse, slightly closer to a wrong stereoisomer, the number of molecules reproduced with correct stereochemistry drops drastically.

For the full-3D method, the number of properly refined structures is lower for PM6 than for UFF. The reason is that PM6 refinement sometimes leads to dissociation of the dative ligand L, presumably as a result of underestimation of the Ru–L interaction for some dative ligands. In addition, for L = tetrahydrofuran the PM6-refined model has the wrong stereochemistry due to serious distortions occurring during the geometry optimization with this particular ligand.[3]

Considering the increasing number of correct stereoisomers generated in series 1–11 upon going from UFF, via PM6, to OLYP$_B$, it is clear that DFT corrects distorted structures more effectively than UFF and PM6. As mentioned above, this is partly due to the introduction of a preliminary geometry optimization step using minimum basis sets (see the Supporting Information, section S1.1.3). Although this practice made it possible to obtain DFT-refinement also for significantly distorted input structures, the full set of 82 Ru compounds could be obtained with exact connectivity, geometry, and stereochemistry only through the use of the refinement protocol termed "OLYP$_A$", which does not include preoptimization of the geometry using minimum basis sets (see the Supporting Information, section S1.1.3). In fact, for one of the 3D models generated by the full-3D approach, preoptimization introduced a new bonding interaction, between the methylidene group and a branch of the dative ligand, altering the identity of the system. Since this event did not occur when the same 3D model was refined by the OLYP$_A$ method, we must conclude that the preoptimization step included in the OLYP$_B$ refinement protocol is often useful and removes geometrical distortions, but may also introduce undesirable modifications in some cases. However, analysis of the geometry of the metal center (see the Supporting Information, Figure S8) indicates that the two protocols OLYP$_A$ and OLYP$_B$ sometimes produce different geometries, with OLYP$_B$, in fact, overall being slightly closer to the reference (see Figure 7). The most likely explanation for this difference is that OLYP$_A$ tends to converge to local minima closer to the input geometry. In fact, minimal basis set preoptimization to some extent allows for basin hopping when considering the final outcome of the double-ζ basis set optimization level.

To conclude, the ensemble of data in this section demonstrates that the full-3D method is better at building these molecules than methods based on SMILES strings coupled with SMILES-to-3D conversion. In particular, two problems of the latter approaches are overcome by the full-3D method, the first of which is the lack of coherence in the treatment of hydrogen atoms, leading to the generation of hydrogen-depleted structures. This problem hampers the application of such tools to computational organometallic chemistry. Second, the low accuracy of the SMILES-to-3D conversion with respect to the particular coordination geometry of the Ru atoms compromises the control of the stereochemistry of the generated molecules. Although this may be partly fixed by quantum chemical refinement, the performance of the full-3D method remains superior to those of all the SMILES-based methods tested in this case study.

Case Study 2. In contrast to the previous case study, where the homogeneity of the metal centers was exploited to focus on the challenge of controlling the stereochemistry and the geometries of a particular set of Ru complexes, the present case study involves a more varied data set of organometallic species of Cu, Rh, and Pd, allowing for a more general comparison of the performance of the methods used to construct this set, which contains 58 molecules and covers various oxidation states, coordination numbers and geometries

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919–1931

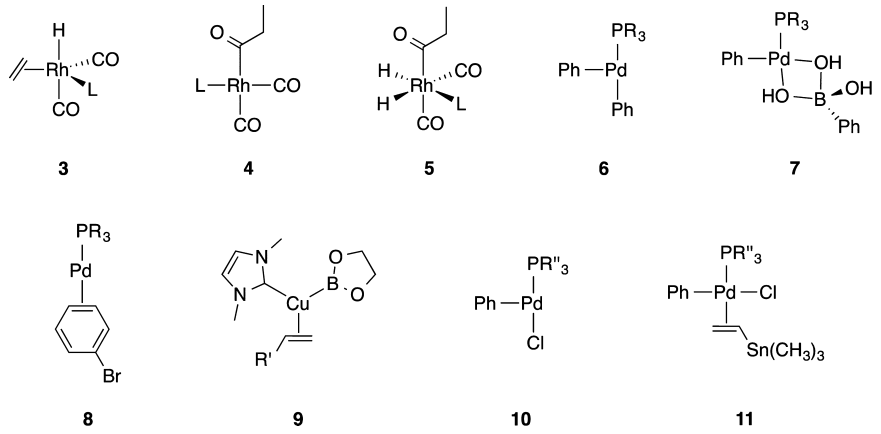


Figure 10. Schematic representation of the compounds included in **DS-3**. L: phosphines, phosphite esters, or N-heterocyclic carbenes; R: $-CF_3$, −Ph, or −*tert*-butyl; R′: −CN or −Ph; R″: −Me, −Ph, or −*tert*-butyl.

encountered in organometallic catalysts.[64−67] Line drawings covering the molecules of this data set (termed **DS-3**) are given in Figure 10.

One of the challenges of the species belonging to **DS-3** is that an explicit hydrogen atom formalism must be used. In particular, metal hydrides, such as the intermediates of rhodium-catalyzed hydroformylation (**1** and **5** in Figure 10)[66,68,69] and the boronate ligands (**7**) involved in the transmetalation step of palladium-catalyzed Suzuki−Miyaura coupling[64,70] require explicit hydrogen atoms. Thus, the present case study covers only those methods that demonstrated correct handling of explicit hydrogen atoms in the first case study above, namely protocols 1 (Marvin) and 12 (the full-3D method of the present work).

Data set **DS-3** also contains complexes with multihapto ligands (**3**, **8**, **9**, and **11** in Figure 10). The stereochemistry of asymmetric multihapto ligands, such as those in molecules **9** and **11**, represents a challenge for the full-3D method. Multihapto ligands are represented as a unique coordinating entity by means of a formal bond involving a dummy atom placed at the centroid of the multihapto system (see Figure 1).[4] Therefore, the bond between the metal and the multihapto ligand is usually free to rotate during the PSSROT calculation (see the Methods section). In reality, the energy barrier for such rotations, also known as "propeller rotation" in case of coordinated olefins, is significantly modulated by the substituents on the multihapto ligand and by the properties of the metal center.[71−74] The PSSROT calculation uses a van der Waals potential only and is not expected to reproduce the subtle stereoelectronic effects that may determine the properties of a bond connecting a multihapto ligand. Therefore, when the structure to be generated by the full-3D method has to reproduce a well-defined stereochemistry for a metal−multihapto ligand system, a single 3D fragment, including the metal and the multihapto ligand already connected with the proper orientation, should be defined by the user and used as a single building block. This strategy was applied for compound **9** for which the central metal-containing fragment included both the Cu atom and the two carbon of the coordinated double bond.

To give an overview of the results obtained with the two methods capable of coherent handling of explicit hydrogen atoms (1 and 12 from Table 1), the structures generated were evaluated in terms of molecular shape (Figure 11) and geometry of the metal center (Figure 12), both before and after DFT-based refinement. Since the compounds in **DS-3** have been assembled from different publications in which a range of different computational models were used, all

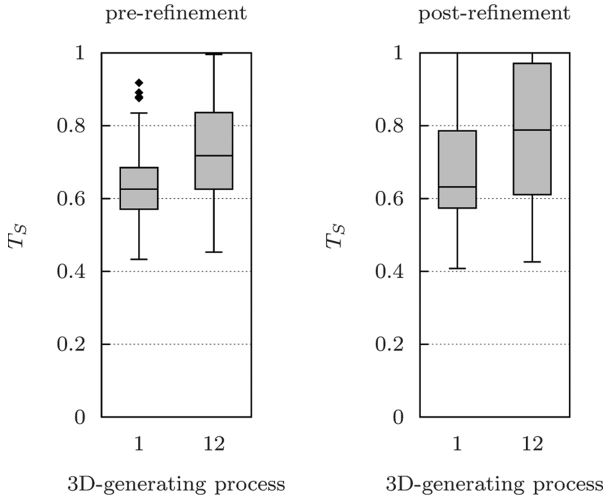


Figure 11. Boxplots[59] representing the distribution of the molecular shape similarity index (T_S) for 3D models of data set **DS-3** before (left) and after (right) DFT-based refinement of the geometry.

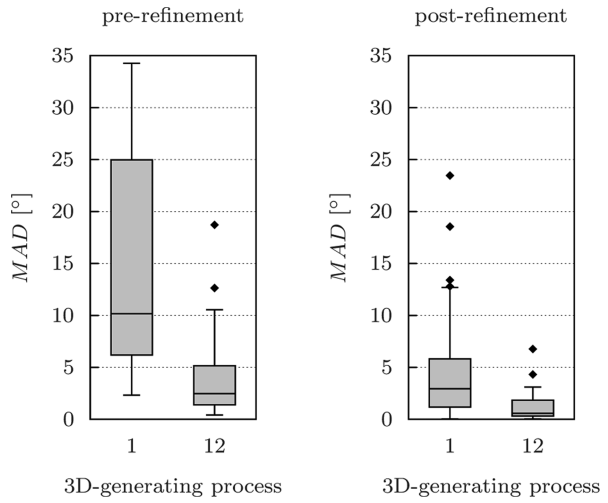


Figure 12. Boxplots[59] representing the distribution of mean angle differences (MADs) for 3D models of data set **DS-3** before (left) and after (right) DFT-based refinement of the geometry.

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

geometries were reoptimized at the same level, i.e., the protocol for DFT-refinement utilized in this case study (see the Supporting Information, section S1.2, for details). The thus optimized molecules are the reference structures used for the comparison of the 3D generating methods 1 and 12 in this case study.

For both 3D-generating processes, the molecular shape similarity index (T_S) of the unrefined molecules (Figure 11, left) is shifted toward lower values than those obtained in Case Study 1. Moreover, the minimum value obtained for T_S is lower for the DFT-refined models than prior to refinement. Both the lower minimum values and the general tendency toward lower T_S values are mainly caused by different conformations of the 3D models with respect to the reference structures. In fact, refinement of some 3D models with altered conformations leads to conformations even more different from the reference structures, thus reducing T_S. Further improvements of the full-3D approach should involve a more thorough handling of the conformational search problem, with a potential including also bonded interactions, or the introduction of an additional conformational search step.

Despite the good performance seen for the full-3D approach in general for this data set, the refinement of the 3D model of molecule **5** with L = $P(OCF_3)_3$ was not successful due to dissociation of L during geometry optimization. Since the most significant deviations between the reference structure and the prerefinement 3D model are due to conformational differences of the ligands (Figure 13), we must conclude that even this

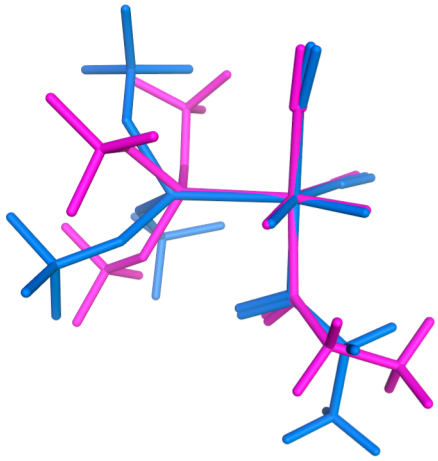


Figure 13. Comparison of the reference structure (blue) and the prerefinement 3D model (magenta) of molecule **5** with L = $P(OCF_3)_3$. The two structures were superposed considering only the metal atom and the six atoms directly connected to it.

apparently harmless conformational problem, combined with the underestimated metal−phosphorus bond energy resulting from the lack of dispersion terms in the functional used in these DFT calculations (OLYP),[75] may prevent the generation of the desired structure even after DFT refinement. Fortunately this seems to be a singular problematic case in **DS-3**.

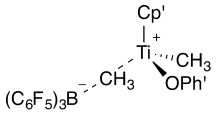
As reported for Case Study 1, the 3D models generated by method 1 also demonstrate significant distortion of the metal center. This is again shown by the angular analysis of the metal center (Figure 12). High MAD values for unrefined 3D models highlight molecules with significant geometrical deviations in the coordination of the metal atom, not only for the 3D models generated by Marvin (series 1) but also for a handful of models from the full-3D method (series 12). The distribution of MADs for series 1 reflects the wrong geometry prepared by Marvin for many three- and tetra-coordinate centers of structures **4**, **6**, **7**, **9**, and **10** in Figure 10. In particular, generation of trigonal rather that T-shaped geometries for compounds **6** and **10** and tetrahedral rather than square planar geometries for **4** returns very high MAD values that, during the subsequent refinement, results in stereochemical changes in the surroundings of the metal center by a mechanism similar to the one described above in Case Study 1. As a consequence, while the distribution of MADs after refinement is significantly improved, only 15 models (26%) generated by Marvin (series 1) have correct stereochemistry (Supporting Information, section S2.2.1). In contrast, the metal centers generated by the full-3D approach (series 12) demonstrate a low-shifted distribution of MAD values. Nevertheless, the few cases of high MADs are symptoms that deserve some discussion. In fact, these values correspond to compounds having reference structures with geometries that differ significantly from those of the other analogues of the same principal structure **3**−**11** in Figure 10. This effect is particularly evident for the analogues of molecule **4** for which the geometry of the metal spans from almost perfect square planar when L is a N-heterocyclic carbene to a geometry that is between tetrahedral and disphenoidal for L = $P(OCF_3)_3$. This variation within the same class is a challenge for the full-3D method that, de facto, exploits the possibility of using the same geometrical template for constructing the metal center of multiple analogues. Still, analysis of the MADs after refinement (Figure 12) demonstrates that DFT refinement significantly reduces the distance to the reference structures, in all cases giving proper stereochemistry (Supporting Information, section S2.2.1). For example, the largest MAD of the unrefined 3D models, 18.7° for molecule **4** with L = $P(OCF_3)_3$, drops to 3.0° after refinement. The few high MAD values remaining after DFT refinement are mostly due to analogues of **4** with weak electron donor ligands, with the exception of the highest MAD, which originates from another analogue of **4** sporting a very different conformation for its triisopropylphosphine ligand.

The overall experience of this case study is that the full-3D method performs well also on a varied data set such as **DS-3**. Nevertheless, the limits of the full-3D strategy were also exposed. Since the full-3D approach is based on the use of template structures for the construction of the 3D models of multiple analogues, potential strong influence from particular ligands is neglected. Whereas a possible solution could be to use multiple fragments for the same structural features depending on the properties of the surrounding chemical environment (e.g., strong electron-donating/-withdrawing substituents), it is not straightforward to tell, a priori, when this would be needed.

Case Study 3. In this final case study we demonstrate the capability of the full-3D method to construct chemical species with interactions and bonds normally not handled by tools primarily developed for organic compounds, including those of series 1−11 (Table 1). Due to these challenges, only the full-3D method was used in this case study. The full-3D approach does not evaluate, fix, or alter the chemistry of the 3D fragments provided as building blocks, and the connections between fragments are simply formal bonds with no relation to the electron count.[4] Therefore, any formally connected arrangement of atoms can be treated, irrespective of the chemical nature of the interactions between the atoms. As a challenging example, this case study tests the construction of

dx.doi.org/10.1021/ci5003153 | J. Chem. Inf. Model. 2014, 54, 1919−1931

3D models of 22 contact ion pairs that are active species in the titanium-catalyzed polymerization of olefins (data set **DS-4**, Figure 14).[76−80]



12

Figure 14. General structure of the contact ion pairs of data set **DS-4**. Cp′: cyclopentadienyl, pentamethylcyclopentadienyl, or indenyl; Ph′: substituted phenyl groups.

The main challenge of these species is to handle the relative position of the cation ($[TiCp'(OPh')CH_3]^+$) and the anion ($[B(C_6F_5)_3CH_3]^-$). The full-3D method is capable of building the ion pair as a single unit via an explicit formal bond between the two ions. Since this formal bond is nothing else than a connection between two 3D fragments, the full-3D approach can define the relative spatial position of the two species during the building of the 3D model. In fact, there is no limitation with respect to the kind of bonds that can be defined between or within 3D fragments.[4] Ionic interactions as well as the breaking/forming bonds of transition states can be represented as formal bonds.

Again, the quality of the generated 3D models, before and after DFT refinement, was analyzed in terms of similarity of the molecular shape and geometry of the metal center (Figure 15). The reference structures were obtained from reoptimization of the structures reported by Manz and co-workers (see Supporting Information, section S1.2, for details).[78,79]

The molecular shape similarity index highlights significant geometrical differences that, as pointed out earlier, correspond to conformational differences between the 3D models and the reference geometries. Rotations around the Ti−OPh′ bond and the formal bond connecting the two ions control the position

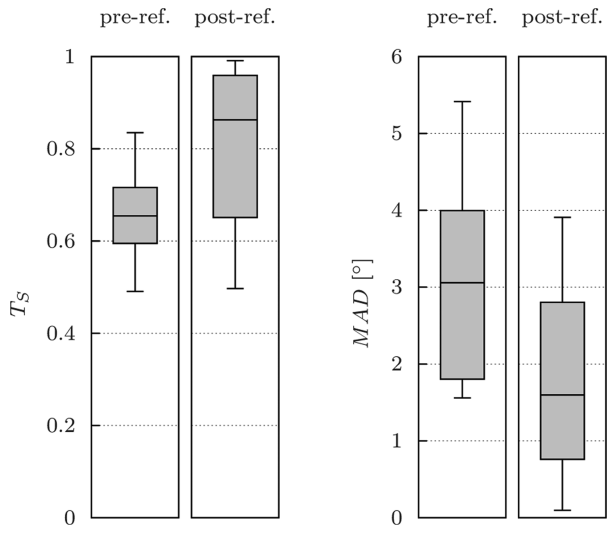


Figure 15. Boxplots[59] representing the distribution of the molecular shape similarity index (T_S, left) and the distribution of mean angle differences (MADs, right) before (left) and after (right) DFT-based refinement of the geometry obtained using the full-3D method on data set **DS-4**.

of large molecular branches (e.g., with many atoms), and thus even small differences in the conformations of such bonds may give large fluctuations in the shape similarity index. For instance, five molecules bear aryloxide ligands with asymmetric substitution, but the energy model of the full-3D conformational search (van der Waals interactions only) cannot distinguish which of the two main orientations, substituent distal or proximal to $Ti-CH_3$,[78,80] leads to the lowest-energy conformation. In addition, two other conformational effects contribute to lowering T_S: (i) propeller isomerism originating from different conformations of the pentafluorophenyl groups of the boron atom (Figure 14), and (ii) rotations around bonds to multihapto ligands. The last point in particular also determines the largest MAD values (Figure 15, right-hand side). In fact, the different conformations of the metal-to-multihapto ligand bonds translate into angular deviations with respect to the reference structures via the mechanism adopted for the calculation of MAD for multihapto ligands: all five metal−ligand formal bonds of η^5-ligands contribute to the MAD. Therefore, a conformational change around the bond between the metal and the centroid dummy atom representing the multihapto ligand quickly translates into higher MAD.

Finally, it should be stressed that the full-3D method returned 3D models for all the contact ion pairs of the data set (see the Supporting Information, section S2.3.1), demonstrating that this method is not limited by standard organic and main-group chemistry.

■ CONCLUSIONS

An alternative method for the automated generation of 3D chemical structures has been described and evaluated. This approach, termed the full-3D method, makes use of 3D fragments as building blocks in the construction of 3D models of chemical species that can be of any kind provided that all the structural features are made available in a library of 3D fragments that may also contain user-defined fragments. The performance of the full-3D approach has been compared with a range of 3D-generating methods based on the use of the SMILES notation and SMILES-to-3D conversion tools. The quality of the 3D models, both before and after geometry optimization (refinement), was seen to be superior to those of the SMILES-based approaches. In particular, while most of the latter approaches demonstrated problems such as inconsistent treatment of hydrogen atoms, generation of wrong coordination geometries, and lack of stereochemical control, the full-3D approach was able to overcome the vast majority of these issues. In addition, we have shown that the full-3D approach can be used for chemical species with bonds and interactions very different from those normally handled by standard cheminformatics tools mostly trained on organic chemistry. Thus, the full-3D approach represents an automated tool for the generation of 3D models of chemical species of any kind and should be useful in *de novo* design, virtual screening, and in simple generation of small libraries of analogues. Conformational differences were seen to be the single main source of deviation between the molecules generated by the full-3D approach and the reference structures. Given the challenge of combining accuracy with generality in a force field for inorganic compounds, it will not be a simple task to improve the conformational search within the full-3D method, although the use of external, more advanced conformational search routines is straightforward in cases where lack of parameters is not an issue. Nevertheless, the prospect of increasing the scope and

applicability of the method strongly suggests that improvement of the conformational search step is a worthwhile target of a future project.

■ ASSOCIATED CONTENT

Supporting Information

Computational details, detailed analysis of molecular shape and metal geometry for Case Study 1, counts of errors and successful refinements for all case studies, cutting rules used for fragmentation, examples of Gaussian input files. 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 Research Council of Norway (RCN) is acknowledged for financial support via the eVITA (grant no. 205273/V30) and GASSMAKS (208335/E30) programs and for CPU and storage resources granted through the NOTUR (NN2506K) and NORSTORE (NS2506K) supercomputing programs. ChemAxon (http://www.chemaxon.com) is thanked for free academic use of the Marvin package.

■ ABBREVIATIONS

3D, three-dimensional; 2D, two-dimensional; UFF, universal force field; DFT, density functional theory; CN, coordination number; DS, data set; MAD, mean angle difference; PSSROT, potential smoothing and search algorithm in rotational space

■ REFERENCES

(1) Burello, E.; Rothenberg, G. *In Silico* Design in Homogeneous Catalysis Using Descriptor Modelling. *Int. J. Mol. Sci.* **2006**, *7*, 375−404.

(2) Hay, B. P. *De Novo* Structure-Based Design of Anion Receptors. *Chem. Soc. Rev.* **2010**, *39*, 3700−3708.

(3) 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.

(4) Foscato, M.; Occhipinti, G.; Venkatraman, V.; Alsberg, B. K.; Jensen, V. R. Automated Design of Realistic Organometallic Molecules from Fragments. *J. Chem. Inf. Model.* **2014**, *54*, 767−780.

(5) Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Manganaro, A. QSPR Modeling of Enthalpies of Formation for Organometallic Compounds by SMART-Based Optimal Descriptors. *J. Comput. Chem.* **2009**, *30*, 2576−2582.

(6) Toropov, A. A.; Toropova, A. P.; Benfenati, E. QSAR-Modeling of Toxicity of Organometallic Compounds by Means of the Balance of Correlations for InChI-Based Optimal Descriptors. *Mol. Diversity* **2010**, *14*, 183−192.

(7) Fey, N. The Contribution of Computational Studies to Organometallic Catalysis: Descriptors, Mechanisms and Models. *Dalton Trans.* **2009**, *39*, 296−310.

(8) Maldonado, A. G.; Rothenberg, G. Predictive Modeling in Homogeneous Catalysis: A Tutorial. *Chem. Soc. Rev.* **2010**, *39*, 1891−1902.

(9) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.

(10) Sadowski, J.; Gasteiger, J. From Atoms and Bonds to Three-Dimensional Atomic Coordinates: Automatic Model Builders. *Chem. Rev.* **1993**, *93*, 2567−2581.

(11) Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Chemical Information in 3D Space. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030−1037.

(12) Ebejer, J.-P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52*, 1146−1158.

(13) Rusinko, A.; Sheridan, R. P.; Nilakantan, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. Using CONCORD to Construct a Large Database of Three-Dimensional Coordinates from Connection Tables. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 251−255.

(14) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-Ray Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.

(15) Mizutani, M. Y.; Nakamura, K.; Ichinose, T.; Itai, A. Starting Point to Molecular Design: Efficient Automated 3D Model Builder Key3D. *Chem. Pharm. Bull. (Tokyo)* **2006**, *54*, 1680−1685.

(16) Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tufféry, P. Frog: A FRee Online druG 3D Conformation Generator. *Nucleic Acids Res.* **2007**, *35*, W568−572.

(17) Lagorce, D.; Pencheva, T.; Villoutreix, B. O.; Miteva, M. A. DG-AMMOS: A New Tool to Generate 3D Conformation of Small Molecules Using Distance Geometry and Automated Molecular Mechanics Optimization for *in Silico* Screening. *BMC Chem. Biol. [Online]* **2009**, *9*, 6. http://www.biomedcentral.com/1472-6769/9/6 (accessed May 5, 2014).

(18) Miteva, M. A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D Conformation Ensemble Generator for Small Compounds. *Nucleic Acids Res.* **2010**, *38*, W622−W627.

(19) Lagorce, D.; Villoutreix, B. O.; Miteva, M. A. Three-Dimensional Structure Generators of Drug-like Compounds: DG-AMMOS, an Open-Source Package. *Expert Opin. Drug Discovery* **2011**, *6*, 339−351.

(20) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.

(21) 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.

(22) Deeth, R. J. The Ligand Field Molecular Mechanics Model and the Stereoelectronic Effects of d and s Electrons. *Coord. Chem. Rev.* **2001**, *212*, 11−34.

(23) Deeth, R. J.; Fey, N.; Williams−Hubbard, B. DommiMOE: An Implementation of Ligand Field Molecular Mechanics in the Molecular Operating Environment. *J. Comput. Chem.* **2005**, *26*, 123−130.

(24) Deeth, R. J.; Anastasi, A. E.; Wilcockson, M. J. An *in Silico* Design Tool for Fe(II) Spin Crossover and Light-Induced Excited Spin State-Trapped Complexes. *J. Am. Chem. Soc.* **2010**, *132*, 6876−6877.

(25) 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.

(26) Buda, C.; Cundari, T. R. *De Novo* Prediction of Ground State Multiplicity and Structural Isomerism for Transition Metal Complexes. *J. Mol. Struct. THEOCHEM* **2004**, *686*, 137−145.

(27) 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.

(28) 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.

(29) Loving, K.; Alberts, I.; Sherman, W. Computational Approaches for Fragment-Based and *de Novo* Design. *Curr. Top. Med. Chem.* **2010**, *10*, 14−32.
(30) Mortier, J.; Rakers, C.; Frederick, R.; Wolber, G. Computational Tools for *in Silico* Fragment-Based Drug Design. *Curr. Top. Med. Chem.* **2012**, *12*, 1935−1943.
(31) *Schrödinger Release 2014−2: Maestro*, version 9.8; Schrödinger, LLC: New York, NY, 2014.
(32) *Molecular Operating Environment (MOE), 2013.08*; Chemical Computing Group Inc.: Montreal, QC, Canada, 2013.
(33) Andronico, A.; Randall, A.; Benz, R. W.; Baldi, P. Data-Driven High-Throughput Prediction of the 3-D Structure of Small Molecules: Review and Progress. *J. Chem. Inf. Model.* **2011**, *51*, 760−776.
(34) Sadowski, P.; Baldi, P. Small-Molecule 3D Structure Prediction Using Open Crystallography Data. *J. Chem. Inf. Model.* **2013**, *53*, 3127−3130.
(35) 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.
(36) O'Boyle, N. M. Towards a Universal SMILES Representation - A Standard Method to Generate Canonical SMILES Based on the InChI. *J. Cheminf. [Online]* **2012**, *4*, 22. http://www.jcheminf.com/content/4/1/22 (accessed June 26, 2013).
(37) Comba, P.; Kerscher, M. Computation of Structures and Properties of Transition Metal Compounds. *Coord. Chem. Rev.* **2009**, *253*, 564−574.
(38) Ponder, J. W. *TINKER: Software Tools for Molecular Design*, 6.2 ed.; Washington University School of Medicine: Saint Louis, MO, 2013.
(39) Kostrowicki, J.; Scheraga, H. A. Application of the Diffusion Equation Method for Global Optimization to Oligopeptides. *J. Phys. Chem.* **1992**, *96*, 7442−7449.
(40) Nakamura, S.; Hirose, H.; Ikeguchi, M.; Doi, J. Conformational Energy Minimization Using a Two-Stage Method. *J. Phys. Chem.* **1995**, *99*, 8374−8378.
(41) Pappu, R. V.; Hart, R. K.; Ponder, J. W. Analysis and Application of Potential Energy Smoothing and Search Methods for Global Optimization. *J. Phys. Chem. B* **1998**, *102*, 9725−9742.
(42) Hart, R. K.; Pappu, R. V.; Ponder, J. W. Exploring the Similarities between Potential Smoothing and Simulated Annealing. *J. Comput. Chem.* **2000**, *21*, 531−552.
(43) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(44) *Daylight Theory Manual*; Daylight Chemical Information System, Inc.: Laguna Niguel, CA; http://www.daylight.com/dayhtml/doc/theory/index.html (accessed March 21, 2013).
(45) 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. R. Toward the Computer-Aided Design of Metal Ion Sequestering Agents. *J. Alloys Compd.* **2004**, *374*, 416−419.
(46) 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.
(47) 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.
(48) 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.
(49) Hoveyda, A. H.; Zhugralin, A. R. The Remarkable Metal-Catalysed Olefin Metathesis Reaction. *Nature* **2007**, *450*, 243−251.
(50) 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.
(51) 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.
(52) 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.
(53) Schrock, R. R. Recent Advances in Olefin Metathesis by Molybdenum and Tungsten Imido Alkylidene Complexes. *J. Mol. Catal. Chem.* **2004**, *213*, 21−30.
(54) Nuñez-Zarur, F.; Solans-Monfort, X.; Rodríguez-Santiago, L.; Sodupe, M. Differences in the Activation Processes of Phosphine-Containing and Grubbs−Hoveyda-Type Alkene Metathesis Catalysts. *Organometallics* **2012**, *31*, 4203−4215.
(55) Occhipinti, G.; Bjørsvik, H.-R.; Törnroos, K. W.; Jensen, V. R. Ruthenium Alkylidene Complexes of Chelating Amine Ligands. *Organometallics* **2007**, *26*, 5803−5814.
(56) *Marvin*, version 5.11.1; ChemAxon: Budapest, Hungary, 2012; http://www.chemaxon.com.
(57) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf. [Online]* **2011**, *3*, 33. http://www.jcheminf.com/content/3/1/33 (accessed May 29, 2013).
(58) The atom ordering of molecules generated from fragments is inherently different from that of the reference structures. Moreover, unambiguous atom−atom correspondence (atom mapping) could not be achieved a posteriori, for instance, when dealing with symmetrically equivalent atoms belonging to significantly distorted structures.
(59) Boxplots represent the first quartile, the median, and the third quartile. Whiskers extend up to the last values within 1.5 times the interquartile range both above and below the box. Values outside the range covered by the whiskers are depicted as single points (outliers).
(60) Romero, P. E.; Piers, W. E.; McDonald, R. Rapidly Initiating Ruthenium Olefin-Metathesis Catalysts. *Angew. Chem., Int. Ed.* **2004**, *43*, 6161−6165.
(61) Dubberley, S. R.; Romero, P. E.; Piers, W. E.; McDonald, R.; Parvez, M. Synthesis, Characterization and Olefin Metathesis Studies of a Family of Ruthenium Phosphonium Alkylidene Complexes. *Inorg. Chim. Acta* **2006**, *359*, 2658−2664.
(62) Wenzel, A. G.; Grubbs, R. H. Ruthenium Metallacycles Derived from 14-Electron Complexes. New Insights into Olefin Metathesis Intermediates. *J. Am. Chem. Soc.* **2006**, *128*, 16048−16049.
(63) Since all metal centers of the **DS-2** molecules are tetra-coordinate, the MADs are always calculated using the same number of angle differences. Therefore, the average of the MAD values corresponds to the overall mean angle deviation for this data set.
(64) Jover, J.; Fey, N.; Purdie, M.; Lloyd-Jones, G. C.; Harvey, J. N. A Computational Study of Phosphine Ligand Effects in Suzuki−Miyaura Coupling. *J. Mol. Catal. Chem.* **2010**, *324*, 39−47.
(65) Ariafard, A.; Yates, B. F. Subtle Balance of Ligand Steric Effects in Stille Transmetalation. *J. Am. Chem. Soc.* **2009**, *131*, 13981−13991.
(66) Sparta, M.; Børve, K. J.; Jensen, V. R. Activity of Rhodium-Catalyzed Hydroformylation: Added Insight and Predictions from Theory. *J. Am. Chem. Soc.* **2007**, *129*, 8487−8499.
(67) Dang, L.; Zhao, H.; Lin, Z.; Marder, T. B. DFT Studies of Alkene Insertions into Cu−B Bonds in Copper(I) Boryl Complexes. *Organometallics* **2007**, *26*, 2824−2832.
(68) Evans, D.; Osborn, J. A.; Wilkinson, G. Hydroformylation of Alkenes by Use of Rhodium Complex Catalysts. *J. Chem. Soc. A* **1968**, 3133−3142.
(69) Beller, M.; Cornils, B.; Frohning, C. D.; Kohlpaintner, C. W. Progress in Hydroformylation and Carbonylation. *J. Mol. Catal. Chem.* **1995**, *104*, 17−85.
(70) Miyaura, N.; Suzuki, A. Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem. Rev.* **1995**, *95*, 2457−2483.

(71) Albright, T. A.; Hoffmann, R.; Thibeault, J. C.; Thorn, D. L. Ethylene Complexes. Bonding, Rotational Barriers, and Conformational Preferences. *J. Am. Chem. Soc.* **1979**, *101*, 3801−3812.
(72) Albright, T. A.; Hoffmann, R.; Tse, Y.-C.; D'Ottavio, T. Polyene-ML_2 and -ML_4 Complexes. Conformational Preferences and Barriers of Rotation. *J. Am. Chem. Soc.* **1979**, *101*, 3812−3821.
(73) Albright, T. A. Rotational Barriers and Conformations in Transition Metal Complexes. *Acc. Chem. Res.* **1982**, *15*, 149−155.
(74) Del Klerk-Engels, B.; Delis, J. G. P.; Ernsting, J.-M.; Elsevier, C. J.; Frühauf, H.-W.; Stufkens, D. J.; Vrieze, K.; Goubitz, K.; Fraanje, J. Alkene Rotation in [Ru(η^5-C_5H_5) (L_2) (η^2-alkene][CF_3SO_3] with L_2 = iPr- or pTol-Diazabutadiene. X-Ray Crystal Structure of [Ru(η^5-C_5H_5(pTol-DAB) (η^2-ethene][CF_3SO_3]. *Inorg. Chim. Acta* **1995**, *240*, 273−284.
(75) Minenkov, Y.; Occhipinti, G.; Jensen, V. R. Metal−Phosphine Bond Strengths of the Transition Metals: A Challenge for DFT. *J. Phys. Chem. A* **2009**, *113*, 11833−11844.
(76) Macchioni, A. Ion Pairing in Transition-Metal Organometallic Chemistry. *Chem. Rev.* **2005**, *105*, 2039−2074.
(77) 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.
(78) Manz, T. A.; Phomphrai, K.; Medvedev, G.; Krishnamurthy, B. B.; Sharma, S.; Haq, J.; Novstrup, K. A.; Thomson, K. T.; Delgass, W. N.; Caruthers, J. M.; Abu-Omar, M. M. Structure−Activity Correlation in Titanium Single-Site Olefin Polymerization Catalysts Containing Mixed Cyclopentadienyl/Aryloxide Ligation. *J. Am. Chem. Soc.* **2007**, *129*, 3776−3777.
(79) Manz, T. A.; Sharma, S.; Phomphrai, K.; Novstrup, K. A.; Fenwick, A. E.; Fanwick, P. E.; Medvedev, G. A.; Abu-Omar, M. M.; Delgass, W. N.; Thomson, K. T.; Caruthers, J. M. Quantitative Effects of Ion Pairing and Sterics on Chain Propagation Kinetics for 1-Hexene Polymerization Catalyzed by Mixed Cp′/ArO Complexes. *Organometallics* **2008**, *27*, 5504−5520.
(80) Manz, T. A.; Caruthers, J. M.; Sharma, S.; Phomphrai, K.; Thomson, K. T.; Delgass, W. N.; Abu-Omar, M. M. Structure−Activity Correlation for Relative Chain Initiation to Propagation Rates in Single-Site Olefin Polymerization Catalysis. *Organometallics* **2012**, *31*, 602−618.