JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

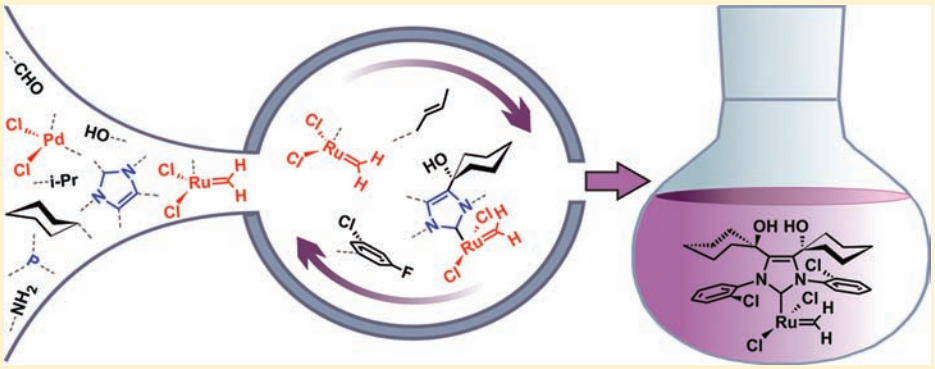
An Evolutionary Algorithm for *de Novo* Optimization of Functional Transition Metal Compounds

Yunhan Chu,[†] Wouter Heyndrickx,[‡] Giovanni Occhipinti,[‡] Vidar R. Jensen,*[,‡] and Bjørn K. Alsberg*[,†]

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

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

Supporting Information



ABSTRACT: Development of functional inorganic and transition metal compounds is usually based on *ad hoc* qualified guesses, with computational methods playing a lesser role than in drug discovery. A *de novo* evolutionary algorithm (EA) is presented that automatically generates transition metal complexes using a search space constrained around chemically meaningful structures assembled from three kinds of fragments: a part shared by all structures and typically containing the metal center itself, one or several parts consisting of ligand skeletons, and unconstrained parts that may grow and vary freely. In EA optimizations, using a cost-efficient fitness function based on a linear quantitative structure−activity relationship model for catalytic activity, we demonstrate the capabilities of the method by retracing the transition from the first-generation, phosphine-based Grubbs olefin metathesis catalysts to second-generation catalysts containing N-heterocyclic carbene ligands instead of phosphines. Moreover, DFT calculations on selected high-fitness, last-generation structures from these evolutionary experiments suggest that, in terms of catalytic activity, the structures arrived at by virtual evolution alone compare favorably with existing, highly active catalysts. The structures from the evolution experiments are, however, complex and probably difficult to synthesize, but a set of manually simplified variations thereof might form the leads for a new generation of Grubbs catalysts.

INTRODUCTION

Whereas a plethora of methods for virtual screening[1−5] and *de novo* design[5−7] are used routinely in the development of pharmaceutical compounds, such methods are only beginning to penetrate through to organometallic and transition metal chemistry, where discoveries of new and interesting compounds to a large extent continue to draw upon chemical knowledge, intuition, and serendipity.[8]

Existing *de novo* design methods are based on automatic computer generation of candidate molecular structures in order to optimize their biological activity. The latter, normally incorporated (approximately) in a "scoring" or "fitness" function, is either related to how well the structure of the candidate fits (e.g., in terms of the interaction energy) to a 3D model of the active site of a protein (termed receptor- or structure-based design) or to the degree to which candidates are similar to a known active compound (termed ligand-based design).[7] The two main classes of *de novo* drug design are illustrated in Figure 1 by actual examples from design of ligands (i.e., inhibitors) for the serine protease thrombin.[9−11] Drawing on protein engineering techniques, *de novo* design has in recent years also been demonstrated for construction of artificial enzymes that catalyze reactions not necessarily found in nature.[12−15]

Even if the *in silico* methods for drug design seldomly get the main credit for the development of a given drug and are almost always used in combination with other advanced methods such as parallel synthesis, they play an important role by discriminating roughly between active and inactive compounds. These methods are often less good at further optimizing a marginally active compound, a problem that to a large extent can be attributed to the crudeness of the fitness functions used, that is, to inadequate description of the candidate−protein interaction in case of structure-based design.[5]

Received: January 26, 2012
Published: April 23, 2012

   dx.doi.org/10.1021/ja300865u | *J. Am. Chem. Soc.* 2012, 134, 8885−8895

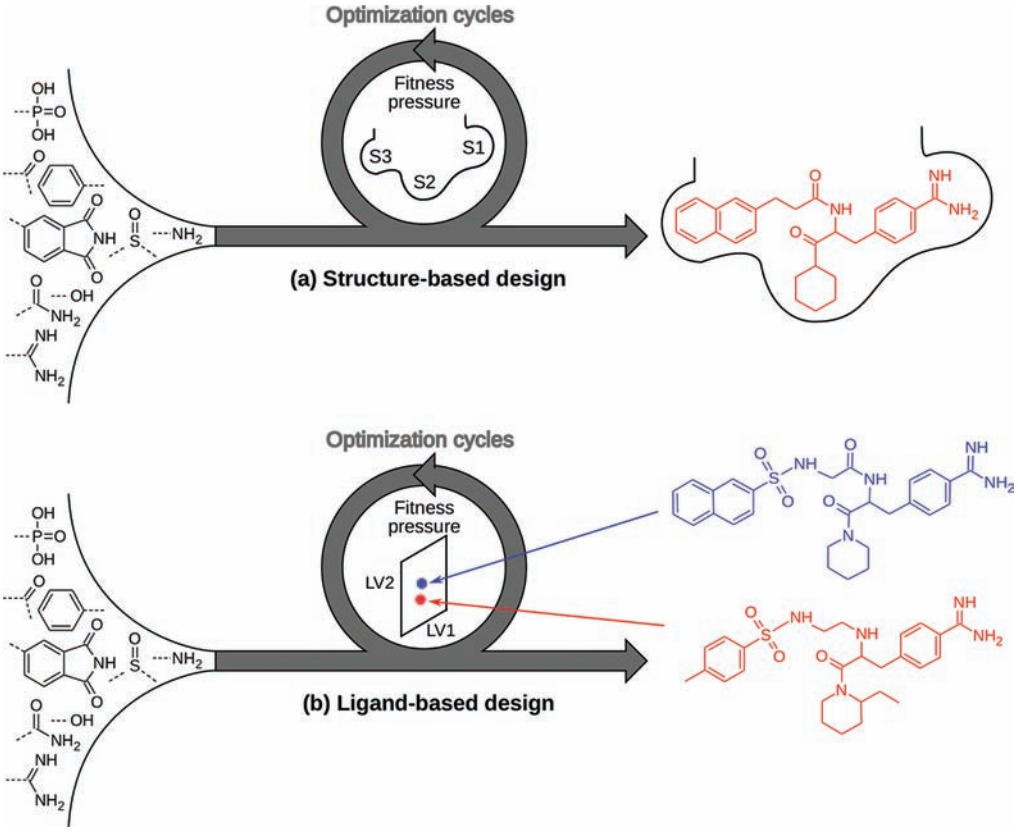


Figure 1. (a) Example of receptor-based design involving the binding pocket of thrombin as determined by the protein structure (PDB entry 1DWD) of its complex with a known inhibitor, NAPAP. The pocket contains three main interaction sites (here schematically indicated as S1, S2, and S3). Drawing upon a library of organic fragments (indicated to the left), candidate ligands (inhibitors) have been assembled and subjected to fitness pressure calculations in the form of empirically estimated binding affinities. A designed ligand (rendered in red) with a binding affinity predicted to be close to that of NAPAP is shown to the right.[9] (b) Example of ligand-based design. Drawing upon a library of organic fragments (indicated to the left), candidate ligands (inhibitors) are assembled and subjected to fitness pressure calculations in the form of a degree of similarity (see ref 10 for a review of similarity measures) with a template molecule. Structural similarity is here indicated by the closeness of the designed ligand (rendered in red) to the template (NAPAP, rendered in blue) in a 2D-projection of the structural space spanned by some latent variables (which are linear combinations of the original variables) LV1 and LV2 (which could be principal components, for example). In the present example,[11] the Tanimoto index was used as similarity-based fitness and *de novo* optimization produced the ligand shown, which subsequently essentially replicated the binding mode of NAPAP with thrombin in docking experiments.

In the current contribution, we explore whether an efficient and general *de novo* design method for transition metal compounds, with the fitness function expressing a property (such as catalytic activity) of the compound itself, may be achieved.

In general, *de novo* methods involve three main tasks: (i) construction of molecular structures; (ii) evaluation of the quality of each structure (i.e., computation of the fitness); and (iii) sampling of and optimization in structure space. The basic building blocks for the assembly of new structures can be based on either single atoms or fragments. Since atom-based assembly methods tend to generate intractable structural diversities, fragment-based methods are preferred in most cases.

The molecular structure space is vast, and selecting a suitable global optimization method is essential. Evolutionary methods such as genetic algorithms (GA),[16] genetic programming (GP),[17] evolutionary strategy (ES),[18] and evolutionary programming (EP)[19] are among the methods best suited to locate the global optimum when tailored genetic operators for structure generation are used to search the structure space; see refs 7 and 20 for key reviews. Therefore the most powerful methods for *de novo* molecular design are based on some form of an evolutionary algorithm (EA) optimizer.[21−26] The evolutionary process can be seen as a "swarm" of individuals (here molecular structures) traveling through a "fitness landscape" in search for the global optimum. Figure 2 shows a 2D version of such a landscape where the optimization (here assumed to be a maximization) is performed on two variables (represented here as "Gene 1" and "Gene 2") where a fitness (the height of the 2D surface) is associated with each coordinate position. The spread of the population ("swarm") of molecules together with the application of the evolutionary operators, in particular the mutation operator, ensure that the "swarm" is less likely to be trapped in a local maximum.

Whereas high-throughput approaches and genetic algorithms have indeed been adopted in optimization of catalysts (e.g., see refs 27−33), the coupling of such methods with automated molecular builders to allow for incremental *in silico* construction and optimization of general transition metal compounds against a computed scoring (fitness) function (a hallmark of *de novo* design methods) has, to our knowledge, not been achieved. A *de novo* approach to predict the ground state coordination isomers of a subclass (trigonal bipyramidal) of transition metal

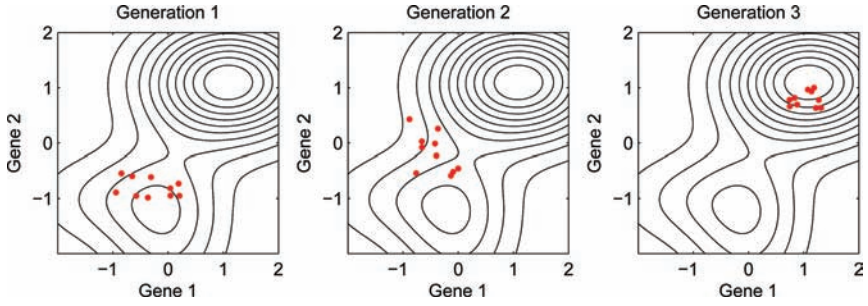


Figure 2. Three generations from an artificial evolutionary optimization in which a population of molecules (red dots) spread over an area in the 2D fitness landscape is seen to leave a local maximum and move to the region surrounding the global maximum in the upper right corner.

complexes using a hierarchy of molecular-level computational methods has been reported.[34] Similarly, automated molecular builders handling subclasses of metal chelate complexes also exist.[35,36] It is also encouraging that much progress has been made in the use of molecular-level computational methods for prediction of homogeneous catalysts in recent years, with a clear tendency toward the adoption of powerful quantitative structure–activity relationships (QSAR) and other computerized techniques to get the most out of the data.[36−44]All of these advances together suggest that the field is ready to explore the potential of a fully integrated drug-design-like *de novo* approach to design of catalysts and other molecular inorganic compounds.

Three main obstacles may explain why *de novo* methods so far have found little use in inorganic chemistry: (i) the increased diversity of chemical bonds and flexibility with respect to electronic and molecular structures; (ii) the need for more sophisticated calculations to obtain molecular properties (fitness) and the thus associated (increased) computational resources; and (iii) the lack of open source software among the most powerful existing *de novo* programs.

To overcome the first obstacle the rules governing how atoms and fragments combine should be extended to include more elements in the periodic table as well as ways to narrow the search space to chemically interesting regions. Introducing such extensions in existing *de novo* programs and methods would indeed be possible but is *de facto* hampered by the third obstacle, that is, the lack of software available for modification. Similarly, even if obstacles i and iii could be overcome, selection of fit over unfit structures would be hampered by the lack of molecular descriptors suitable for transition metal compounds and effective ways to explore the chemical structure space. The above-mentioned electronic and geometric diversity of transition metal compounds indicates that ranking and selection often will have to be based on electronic methods, that is, methods such as density functional theory (DFT), *ab initio*, and semiempirical theory which take into account the electronic degrees of freedom. Such methods are, for example, necessary for describing intermediates and transition states of reactions involving bond rupture and formation and may have to be applied when accurate fitness functions are needed in optimizations of catalysts. Alternatively, it may sometimes be possible to resort to computationally less demanding force-fields-based methods such as ligand-field molecular mechanics (LFMM[45]) and reactive force fields (e.g., ReaxFF[46]). The downside, however, is that such methods must be thoroughly parametrized to handle the chemistry in question prior to application; that is, they must be trained to mimic the behavior of more accurate electronic methods. At any rate, the methods that can be expected to be useful in *de novo* design of transition metal compounds are both more advanced and more computationally demanding than those frequently used in drug design. However, except for potential applications to metalloproteins, the systems to be treated are much smaller, and many of the fitness functions may still be easier to address than protein–ligand interaction energies.

Recent advances in both computational methods and hardware suggest that the time is ripe for applying evolutionary *de novo* methods to transition metal chemistry. Here we present such a method, which constrains the structural search space by utilizing the user's knowledge of the property to be optimized and the kind of compounds that are of interest. Search space constraints are introduced by assembling structures from three kinds of building blocks (termed parts), each with a different ability to undergo modification by tailor-made structural and evolutionary operators.

The capabilities of the EA method are demonstrated by sample optimizations of ruthenium-based catalysts for olefin metathesis[47] using a computationally obtained approximate measure[43] of catalytic activity as fitness.

RESULTS

The Overall Work-Flow of the EA. Using the principles of natural evolution, our *de novo* EA seeks to produce highly functional transition metal compounds. Its application, however, is not limited to this domain of chemistry only. Each optimization is initiated from a seed population of starting structures. A structure is represented as a chromosome in the form of a 2D graph. A fragment growth operation (see Supporting Information) is repeatedly used to construct the seed population. All such 2D structures are by default saturated with hydrogen atoms and subjected to conformational search upon transformation to 3D. Next, the fitness is calculated for the lowest-energy conformation identified in each case. Once a seed population has been prepared and the associated fitness values calculated, a four-step optimization cycle starts (see Figure 3 and Supporting Information): (i) ranking and identification of competitive parent structures; (ii) production of new offspring structures by structural operations (genetic crossover, mutation, or fragment growth); (iii) conformational search of offspring structures; and (iv) fitness calculation of offspring structures. The latter will typically be a property or relative energy (e.g., a rate-determining barrier[48]) involving the 3D structure, obtained in calculations using a molecular-level computational chemistry program.

As seen in Figure 3, the conformational searches and fitness calculations are spread on individual processor cores and are thus made to run in parallel. A tournament procedure is adopted to identify the more fit parent structures by pairwise comparisons of randomly picked individuals. Offspring structures generated

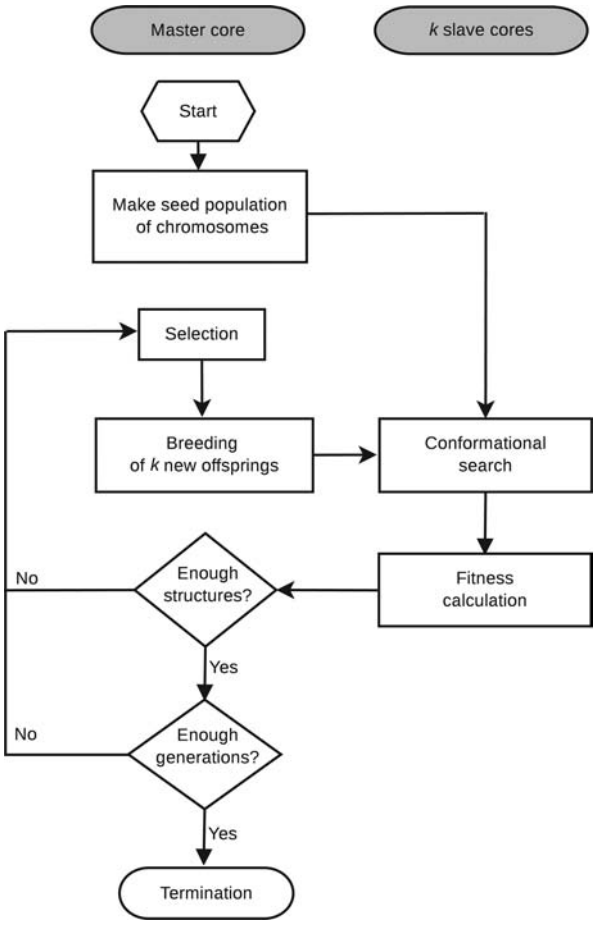


Figure 3. The overall work-flow of the *de novo* evolutionary algorithm. On a cluster-type architecture the user allocates n nodes, each of which contributes m processor cores to the overall job. Core number one on the master node generates k structures, where $k = nm - 1$, and transfers them simultaneously to the k slave cores to achieve parallel processing of the conformational search and fitness evaluation.

from genetic crossover or mutation or from fragment growth (see Supporting Information) are subjected to conformational search and fitness evaluation and serve to replace, gradually, the least fit structures in the current population. The optimization cycle continues until a predefined number of offspring structures have been produced and the offspring, together with the survivors of the current population, establish a new generation. The population thereby evolves over generations until the maximum number of generations is reached or until no fitter offspring structure occurs within a certain number of successive EA optimization cycles.

Structure Representation and Assembly. Transition metal compounds are characterized by one or more central metal atoms, to which a variable number of covalent (ionic) or dative (neutral) groups (ligands) are bound. When one optimizes such compounds in the context of *de novo* design, it will, in general, be possible to identify a minimal part of the structure that is known to be necessary for the intended property, such as the mediation of a transformation and that thus is left unmodified by structural operations (the geometry may be allowed to change) throughout the optimization. In the current EA scheme, this central portion of the structure is defined as the "core" part. The core is not subjected to 2D structural modifications, may have any number of substitution points (i.e., positions at which the structure may be extended and modified), and may be bound to any of the other two classes of parts (see below), as illustrated in Figure 4.

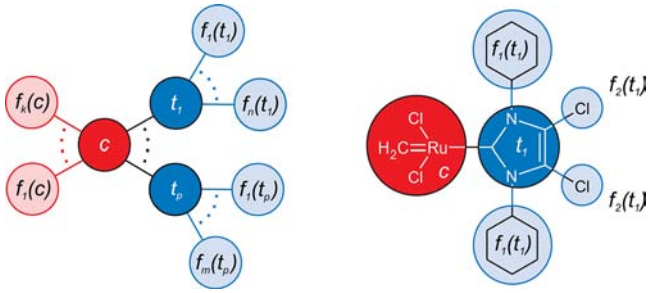


Figure 4. Construction of transition metal compounds from core (c), trial (t_1–t_p) and unconstrained, that is, free parts (attached to the core: $f_1(c)$–$f_k(c)$; attached to trial parts: from $f_1(t_1)$–$f_n(t_1)$ to $f_1(t_p)$–$f_m(t_p)$) (left). Example of how a 14-electron ruthenium alkylidene olefin metathesis catalyst may be assembled from a core (c), a trial (t_1), and four free parts. The latter consist of two copies each of $f_1(t_1)$ and $f_2(t_1)$, and this example thus illustrates the use of symmetry to specify that certain free parts remain identical throughout an EA optimization (right).

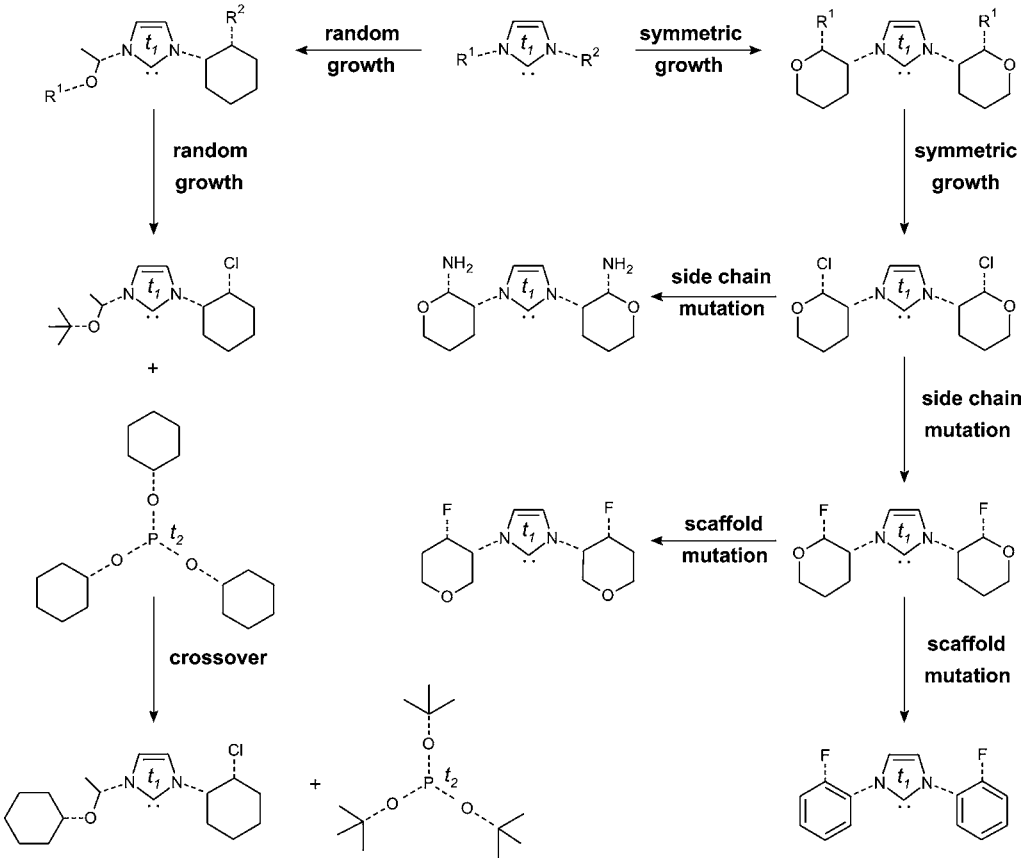
In a typical application, the metal center or centers, with or without portions of the ligand environment, will constitute the constant core part.

Our goal is to achieve a method that automatically can combine covalent and dative ligands to produce transition metal and organometallic compounds with given oxidation and coordination numbers for the central metal atoms. To our knowledge, the existing general automated molecular builders assemble radical fragments via electron-pair bonds to give organic compounds following standard valence rules. Whereas high, "non-organic" coordination numbers may be achieved via the definition of the core part (see above), attaching neutral donor ligands to a core part using such builders would require modification of the builder or the associated fragment libraries or both. Most builders would probably attach, say, an amine fragment as an amide group (M–NR_2) not as a donor ligand (M–NR_3). To solve this problem, we suggest to introduce the "trial" part, which encompasses the portion of the dative ligand that should not be modified by structural operations (see Scheme 1), that is, the ligand skeleton. However, unlike the core, the trial skeletons are allowed to be selected from a list, or a library, of ligand scaffolds, meaning that competition among various ligand classes may be included in evolution experiments (see Supporting Information). Each trial part will thus be subjected to modifications but in a restricted manner in which focus is on the chemically meaningful portion of the structural variation space. The introduction of trial parts thus circumvents the above-mentioned problems of automated building of structures involving donor–acceptor bonds and also offers a handle with which to restrict the structural variation in the optimization.

The remaining portions of the overall structure, termed "free parts", are varied freely following a strategy similar to that of existing fragment-based EA methods[49] and may be attached to trial parts or to the core. The individual fragments assembled to give free parts are generated by splitting of chemical structures from common large libraries of organic compounds (see the Methodological and Computational Details section as well as the Supporting Information). This ensures that the free parts are searched from a diverse chemical structure space. The division of molecular structures into parts that are varied according to

dx.doi.org/10.1021/ja300865u | J. Am. Chem. Soc. 2012, 134, 8885–8895

Scheme 1. An Overview of the Different Structural Operations Applied to the Trial and Free Parts: Growth (Random and Symmetric), Crossover, and Mutation (Side Chain and Scaffold)[a]



[a]Substitution points are indicated by R^n (enumeration by n) and dashed lines of attachment to the rest of the structure. Trial parts are indicated and enumerated as t_n.

different rules not only allows for construction and optimization of compounds other than those of standard organic chemistry but also offers powerful tools for constraining the searchable structure space in order to facilitate the location of more relevant and realistic molecules. An illustration of how a population of molecules constructed from core, trial, and free parts may evolve over generations is given in Figure 5.

An appropriate representation of molecular structures is needed for their generation and manipulation. Among the many different ways in which to represent molecular structures,[7] we selected, after careful comparison, a graph-based representation[50] available in the Chemistry Development Kit (CDK),[51] and the molecular structures are represented as combined graphs of core, trial, and free parts. The representation is object-oriented and enables fragments, atoms, and bonds of structures to be labeled and tracked throughout the optimization.

Case Study: Ruthenium Olefin Metathesis Catalysts. As an initial test of our *de novo* evolutionary method, we have chosen ruthenium alkylidene olefin metathesis catalysts.[47,52] The precursors of these catalysts, $L'LCl_2Ru{=}CHR$, are coordinated by two dative ligands, L′ and L, and the activation step consists of dissociation of one of these ligands, L′, to form the active 14-electron complex, $LCl_2Ru{=}CHR$. Moreover, the initial alkylidene group, usually benzylidene, is replaced in the first catalytic cycle, for example, by methylidene in the case of the generic ethylene metathesis reaction. The 14-electron $LCl_2Ru{=}CH_2$ complex is thus a reasonable model catalyst for optimization of general catalytic activity.

Three factors make the structural subspace generated by the dative ligand L appear tempting for test optimizations using the EA: (i) the drastic influence wielded by this ligand on catalyst activity and stability[53−55] making it an interesting target for further catalyst development, (ii) the considerable body of experimental and theoretical data against which such test optimizations can be validated, and (iii) the fact that a computationally inexpensive fitness function can be constructed based on an available QSAR model.[43] With respect to factors i and ii, a good portion of the efforts at improving the performance of the ruthenium-based catalysts are aimed at the dative ligand,[56,57] and optimization thereof has led to the development of two main classes of catalysts, the first[58,59] and second[60,61] generation Grubbs catalysts, based on phosphine and N-heterocyclic carbene (NHC, imidazol-2-ylidene (unsaturated) or dihydroimidazol-2-ylidene (saturated) ring) ligands, respectively, in addition to a considerable hierarchy of subclasses with known relative catalytic activities in particular among the second-generation catalysts.

Regarding the fitness (factor iii above), a clear correlation between certain molecular descriptors derived from DFT calculations (e.g., bond distances, angles, and partial charges) of the active 14-electron complexes $LCl_2Ru{=}CH_2$ and a theoretical, approximate measure of the catalytic olefin metathesis activity, termed "productivity", also obtained using DFT, has been demonstrated.[43] Even if the fitness could, for the purpose

dx.doi.org/10.1021/ja300865u | J. Am. Chem. Soc. 2012, 134, 8885−8895

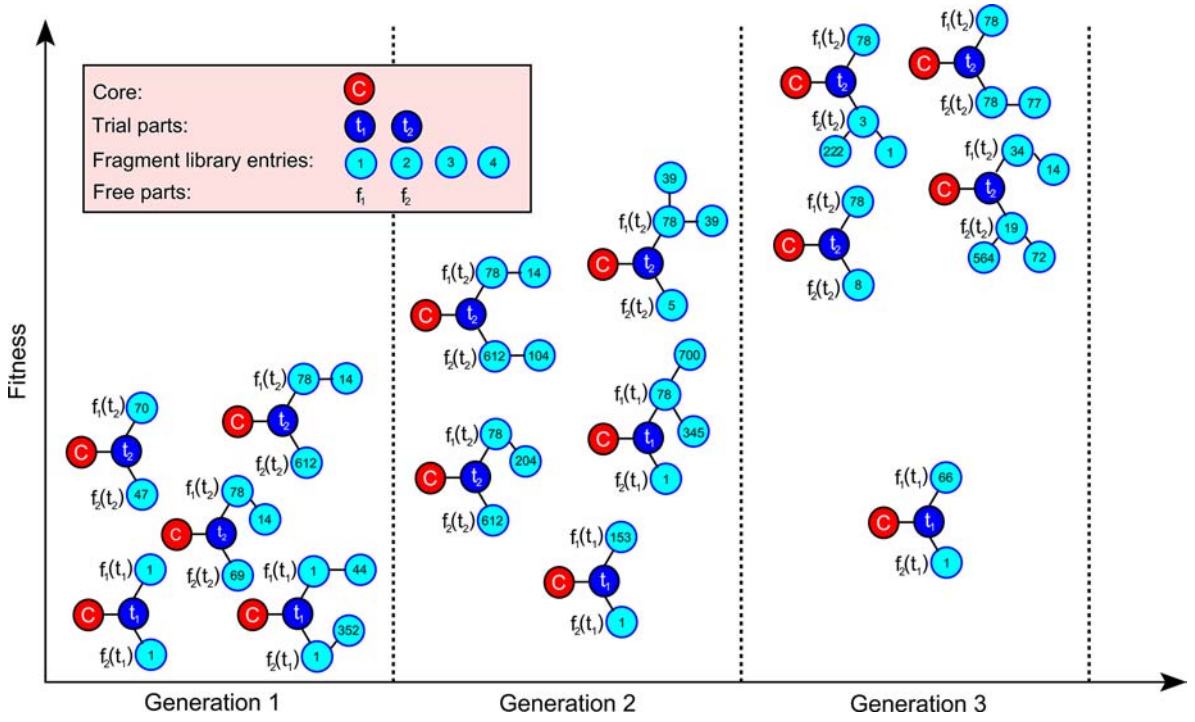


Figure 5. Schematic illustration of artificial evolution of molecules constructed from core, trial, and free parts.

of limited EA test optimizations, be based on DFT calculations, larger and more routine applications of the EA will require a computationally less expensive fitness. Thus, to ensure realistic testing, we constructed a QSAR model based on partial least-squares regression (PLSR)[62,63] for the DFT-based productivity of ref 43 using six readily interpretable molecular descriptors pertaining to the core part, obtained in PM6[64] geometry optimizations; see the Methodological and Computational Details section for more information. It is important to realize that this QSAR model is based on *comparable* descriptors from the part (core) shared by all the generated structures, eliminating the need for alignment of the molecular structures. An alignment step, such as used in the comparative molecular field analysis (CoMFA),[65] would make the optimization less efficient. However, there are alignment-free 3D QSAR methods that might be suitable for this purpose, such as grid-independent descriptors (GRIND),[66,67] VolSurf,[68] and those based on inductive logic programming.[69−71]

As explained above, the possibility to handle metal centers bearing dative ligands in *de novo* design is an important part of the motivation behind the development of the current method. This is reflected in the following test optimizations, which involve different competing ligand skeletons (trial parts) in combination with a constant, core part, $Cl_2Ru{=}CH_2$. In order to ensure equal chances to all ligands tested, the starting populations contained (nearly) equal numbers of the different trial parts. All ligand substituents have been subjected, in one or more evolution experiments, to variation of free parts using the 2238-entry fragment library described in the Methodological and Computational Details section. This section also contains the detailed settings and options of the EA experiments. All the EA experiments have been reproduced, with very similar developments of the populations and the productivities, and some of these additional experiments are given in the Supporting Information.

In the first evolution experiment, two types of trial parts were employed, namely, a general three-coordinate phosphorus scaffold (termed PR_3) and a general, unsaturated imidazol-2-ylidene scaffold ($I{-}R_2$) with available substitution points on the C=C backbone, see Figure 6. The two trial parts correspond to the first- and second-generation Grubbs catalysts, respectively, with the second-generation compounds known to display higher catalytic activities in general. Hence, if our fitness function reflects the observed activity of the catalyst, the second generation structures should tend to out-compete the first generation. This is confirmed in Figure 6, which shows the occurrence of the trial skeletons in the different generations. The rate of disappearance of the first-generation Grubbs catalysts (up to five per generation) is fast but would have been even faster if evolution were not operating also among the phosphorus ligands. It should be noted that even though the number of phosphorus ligands drops to zero in the eighth generation, new ligands of this kind are still being generated by fragment growth and could rebound if they would be fit to compete with the carbenes. This is not the case, however, and the number remains at zero until the end of the evolution run. The average predicted productivity is seen to grow smoothly and steadily throughout the optimization, driven by increases in both the maximal and minimal fitness values in the population, see Figure 6b. Particularly unfit complexes are eliminated early on in the optimization.

In the second evolution run, four trial parts compete against each other, namely, para-substituted aryl (termed PAr_3) and alkyl ($P(CH_2R)_3$) phosphine together with an imidazol-2-ylidene skeleton with hydrogen atoms (I) and chlorine atoms ($I{-}Cl_2$) on the NHC backbone; see Figure 7. Again, the less fit first-generation ligands ($P(CH_2R)_3$ + PAr_3) decline right from the start, leading to extinction of PAr_3 already in the fourth generation, followed by $P(CH_2R)_3$ in the seventh generation. From the moment that all phosphines have disappeared, the

dx.doi.org/10.1021/ja300865u | J. Am. Chem. Soc. 2012, 134, 8885−8895

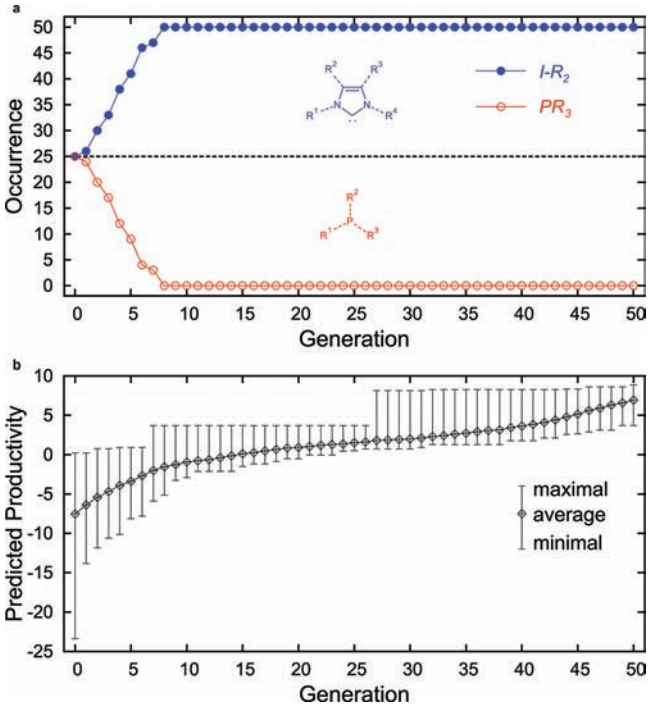


Figure 6. (a) The first EA test optimization, involving competition between two trial parts, an imidazol-2-ylidene scaffold with four substitution points (termed I−R_2) and a phosphorus atom with three substitution points (PR_3). Substitution points are indicated by R^n (enumeration by n) and dashed lines of attachment to the rest of the structure. (b) Average, minimal and maximal predicted productivities (i.e., fitness) throughout the first evolution experiment.

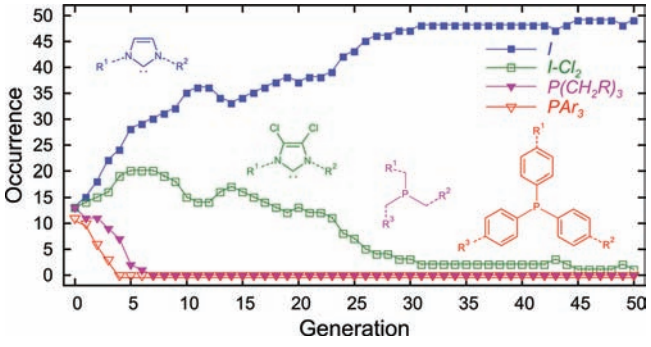


Figure 7. The second EA experiment, involving four trial parts, an imidazol-2-ylidene scaffold with two substitution points and two hydrogen atoms (termed I) or two chlorine atoms (I−Cl_2), as well as a triphenyl phosphine skeleton with substitution points at the phenyl para positions (PAr_3) and a trialkyl phosphine skeleton with three substitution points ($P(CH_2R)_3$). Substitution points are indicated by R^n (enumeration by n) and dashed lines of attachment to the rest of the structure.

chlorine-substituted NHCs begin to decline in numbers. The decline is rather slow and irregular, and this trial structure partly resists extinction, with one or two individuals still left in the population throughout almost 20 generations before the maximum number of generations is reached. The maximal fitness values of the initial generation may suggest probabilities for long-term survival or extinction and thus give hints, for example, as to the mechanism behind the slow decline of the chlorine-substituted carbenes I−Cl_2. The difference between the two kinds of carbene trial skeletons is only slightly in favor of I (0.47 kcal/mol predicted by the PM6-based QSAR model, which compares well with the 0.6 kcal/mol obtained from explicit DFT calculations[43]). The small difference in fitness is reflected in instances during evolution in which the intrinsically lesser fit I−Cl_2 is actually increasing in numbers at the expense of I.

The fact that the hydrogen-substituted imidazol-2-ylidene skeleton prevails is probably due to increased electron donation compared with the chlorine-substituted counterpart. Electron-donating substituents in the backbone positions are known to increase catalytic activity.[43] The drop in calculated productivity upon replacing backbone hydrogen by chlorine is thus consistent with a somewhat lower observed catalytic performance.[56]

As a further test of the quality of the structures with high predicted productivities in the last generations of the EA experiments, three such structures were subjected to explicit calculation of the productivity at the DFT level of theory, see the Supporting Information for computational details. With a fitness function defined to target a single property, a fair fraction of the structures generated by the method will necessarily be very difficult or impossible to synthesize. In the present case, the three complexes selected for explicit fitness calculation, **I1**, **I2**, and **I3**, were chosen among the high-productivity, final-generation structures due to their size (avoiding the largest) and their assumed ease of synthesis compared with the other final-generation complexes, see Figure 8. Gratifyingly, the explicitly DFT-calculated productivities of the three complexes compare favorably with those of existing commercial catalysts.[43] The DFT-calculated productivities of catalysts based on **I1** and **I2** are higher, for example, than that of the commonly used

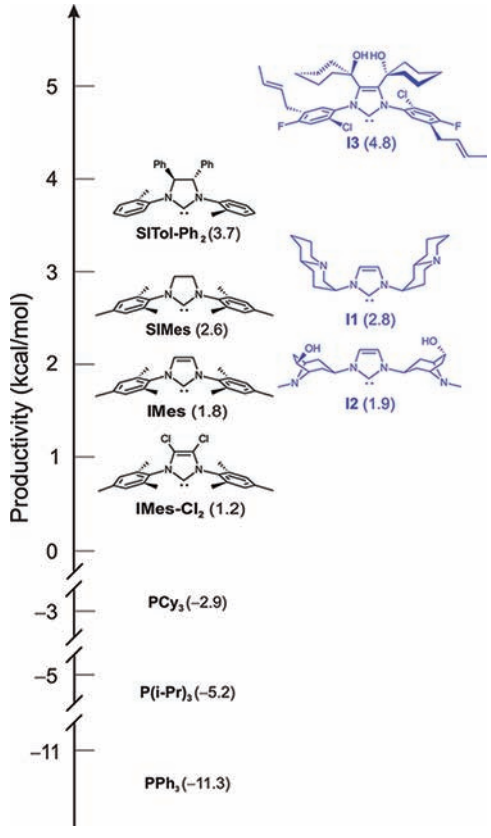


Figure 8. Explicitly DFT-calculated productivities[43] of NHC ligands **I1**, **I2**, and **I3** selected from the last generations of the EA experiments compared with those of known ligands L of catalysts L′LCl_2Ru=CHR.

dx.doi.org/10.1021/ja300865u | J. Am. Chem. Soc. 2012, 134, 8885−8895

mesityl-substituted unsaturated imidazol-2-ylidene (termed **IMes**). Particularly high productivities were calculated[43] for catalysts based on ligands with electron-donating substituents on the two carbon atoms of the N-heterocyclic carbene (NHC) backbone, and this prediction has subsequently been confirmed in a series of experimental studies, for example, see refs 72−74. It is thus remarkable that the productivity obtained for **I3** is, in fact, higher than that calculated for the best previous predictions,[43] including those of saturated NHCs such as **SIMes**.

Unfortunately, **I3**, albeit C_2-symmetric, contains structural features that would make the synthesis challenging, expensive, and time-consuming, if possible at all. The structure of the aryl substituent on the nitrogen atom is unusual and complex, and the C═C backbone is substituted with a sterically demanding 1-hydroxy-cyclohexyl group. Thus, it is tempting to try to reduce the chemical complexity of this ligand at the same time as preserving the predicted catalytic activity. To this end, we have progressively simplified its molecular structure and calculated the productivity of each simplified ligand explicitly (using DFT, see the Supporting Information for details); see Figure 9. It is immediately

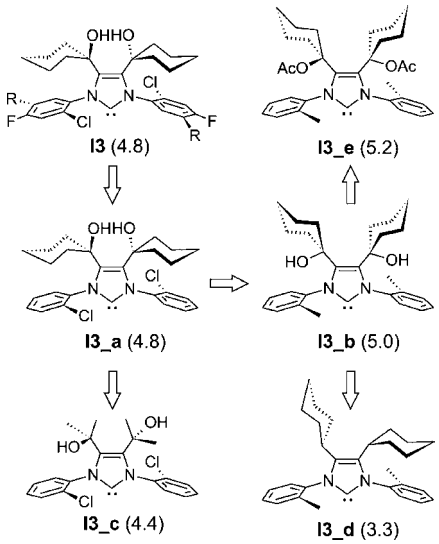


Figure 9. Explicitly DFT-calculated productivities[43] of catalysts $L'LCl_2Ru{=}CH_2$ for which L = **I3** and structural variations of **I3** (R = (E)-2-butenyl; Ac = acetyl).

evident that the chemical complexity of the aryl substituent is not essential for catalytic activity. For example, the calculated productivity of **I3_a** and **I3_b**, which both contain a monosubstituted N-aryl substituent, is at the same level as, or even slightly better than, that of **I3**. In contrast, attempts to simplify the molecular structure of the 1-hydroxy-cyclohexyl substituent of the C═C backbone leads to lower productivity.

The protection of the free hydroxyl function as acetate is beneficial for the calculated productivity. The higher productivity of **I3_e** compared with **I3_b** is most probably due to steric rather electronic effects. The larger acetoxy group increases the steric requirements of **I3_e**, which is expected to boost catalytic activity,[43] whereas its electron-withdrawing properties are expected to hamper the overall ligand-to-metal donation and thereby also the catalytic activity.[43] The explicitly calculated productivities of the ruthenium alkylidene compounds containing the ligands **I3_a**, **I3_b**, or **I3_e** imply that these complexes should possess exceptional catalytic activities.[43] Thus, if these ligands and their ruthenium alkylidene complexes can be prepared and are sufficiently stable, they could form a new generation of Grubbs-type catalysts.

The DFT-optimized geometries of the ruthenium complexes reveal that the structural parameters of the ruthenium center are similar to those of existing ruthenium-based catalysts. However, in contrast to other imidazol-2-ylidene ligands, the region around the C═C backbone in **I3_a**, **I3_b**, and **I3_e** is not planar and forms torsional angles N−C═C−N in the range 6−8°. These results suggest that, whereas the preparation of the ligands is probably not trivial, the subsequent synthesis of the ruthenium complexes should be straightforward. A retrosynthetic analysis of the ligands **I3_a**, **I3_b**, and **I3_e** or, better, of the imidazolium salt precursors suggests that the last step (the ring closing reaction) is the only truly challenging step. The synthesis of the acyclic precursor appears to be relatively unproblematic. The only starting material that is not commercially available is the fragment containing the substituted C═C backbone, but this can be prepared in a few steps starting from acetylene, cyclohexanone, and KOH, for example, see ref 75. Glorius and co-workers have recently discovered an efficient strategy for preparation of highly substituted imidazolium salts, which has been successfully applied in the synthesis of several 4,5-dialkyl-substituted unsaturated imidazolium salts.[76] One of the latter compounds involves a combination of substituents with an overall steric demand comparable to our target molecules. This compound has disubstituted alkyl substituents on the C═C backbone and a mesityl group on the nitrogen atoms, while our target compounds have trisubstituted groups on the C═C backbone and less sterically demanding substituents on the nitrogen atoms (2-tolyl or 2-chlorophenyl). Thus, it is conceivable that one or more of **I3_a**, **I3_b**, and **I3_e** can be prepared using the methodology disclosed by Glorius and co-workers[76] or a slightly modified version thereof.

DISCUSSION

The Case Study. It should be kept in mind that the above-predicted structures **I1**, **I2**, and **I3** have been generated automatically in an optimization of a single property (catalytic activity) only, with factors such as stability neglected. This case study thus illustrates a general problem of the present and similar EA methods: there is no guarantee that the predicted compounds are stable and may be obtained. Often one must therefore resort to extracting the essential pieces of information inherent in the high-fitness structures of the last generations. For example, as we have seen above (Figure 9 and accompanying text) the useful message from **I3** is not its exact molecular structure but rather the oxy-cyclohexyl-substituents on the C═C backbone (conserved in **I3_a**, **I3_b**, and **I3_e**), whereas the unusual and complex substituents on the N-bound groups could be discarded without loss of predicted catalytic activity. To our knowledge, no ruthenium-based catalysts for olefin metathesis involving NHC ligands with such cyclic substituents have been reported to date.

General Considerations. The particular challenges of transition metal and organometallic chemistry have in the present EA method been met by possibilities for restricting the search space around sets of combinations of metal centers and ligand skeletons and, in the present test optimizations, by careful preparation of a computationally inexpensive, yet apparently sufficiently accurate, fitness measure. The initial tests of the method indicate that it works as intended and reproduces known trends in activity among the class of compounds chosen for validation, the

ruthenium-based olefin metathesis catalysts. The method is likely to be broadly applicable in molecular inorganic chemistry. The degree to which it is useful, however, is to a large extent determined by the quality and computational cost of the fitness. Application of the method should thus in general be preceded by careful selection and preparation of a suitable fitness function, which also implies that as much mechanistic information as possible should be available. A relatively simple fitness function, describing the stability of a reaction intermediate with respect to the resting state, was used in the above case study. With sufficient knowledge of the reaction mechanism, the fitness function may take into account competition between different pathways and barrier heights. For example, if a given catalytic reaction involves two potentially rate-determining barriers and the goal is to optimize a more active catalyst, both barriers should be monitored throughout the EA optimization, with the fitness function always taken to be the highest of the two. With all the relevant mechanistic aspects known, the power of the method is, in principle, only limited by the available computing capacity. For example, realistic drug-design-style EA optimizations may require millions of individual fitness evaluations to complete, meaning such evaluations on average must finish within minutes, not hours, on a single processor core for the overall optimization to be tractable. In many cases, it will probably be necessary to prepare an indirect fitness measure such as in the present test runs, which rely on a QSAR model to offer an approximate measure of catalytic activity. Similarly, if sufficient attention is paid to parametrization, it is likely that efficient fitness functions may be obtained from the application of computationally inexpensive methods such as ligand-field molecular mechanics (LFMM,[45] which has already been used successfully in design[77]) and reactive force-fields (e.g., ReaxFF[46]) and steps to integrate such fitness providers will be taken. Introduction of structural operators that take synthetic accessibility[78] into consideration, removal of undesirable structures prior to fitness calculation,[79] and dynamical updating of QSAR models by on-the-fly calculation of true fitness during EA optimizations constitute other desirable targets for future work.

CONCLUSIONS

We have described a new fragment-based evolutionary algorithm (EA) for *de novo* optimization. The method is applicable to molecular systems in general but has been specifically developed to handle the particular challenges posed by organometallic and transition metal compounds.

With the exception of special and well-parametrized cases, the calculation of fitness can be expected to require quantum chemical methods. In the initial phase of the development, we have chosen the semiempirical method PM6,[64] as implemented in the MOPAC program,[80] as the basis for calculating the fitness. The possibility to address large EA optimization problems is enabled by simultaneously allocating the individual fitness evaluations to different processor cores.

In a series of EA experiments involving optimizations of ruthenium alkylidene complexes for olefin metathesis, we have demonstrated the capabilities of the EA method. The approach was not only able to retrace the transition of Grubbs catalysts from the so-called first-generation (phosphine-coordinated active compounds) to the second-generation (NHC-coordinated active compounds) catalysts, but also to discriminate well between less and more active catalysts within the same ligand classes. In addition, a small set of compounds were selected among the high-fitness, last-generation structures, and DFT calculations indicate that these compounds may indeed lead to new and highly active olefin metathesis catalysts. These predictions thus indicate that the EA method, even in combination with a computationally inexpensive fitness, has significant potential for *in silico* development of catalysts and other functional transition metal compounds.

METHODOLOGICAL AND COMPUTATIONAL DETAILS

The Library of Free Fragments. In order to obtain an initial version of a library of fragments to build free parts, a set of 10005 molecules from the KEGG LIGAND database provided by Ligand.Info[81] was used. The LIGAND database contains mostly organic compounds and has been reported to offer hit rates in high-throughput screenings of natural products superior to those of several other common libraries.[82,83] While organometallic compounds distinguish themselves from organic-only compounds in that they contain metal centers, they are bound to organic parts (ligands), which may be built from organic fragment libraries.

We have developed two programs to automatically build a library of fragments to form free parts: a splitting tool and a screening tool. The splitting tool cuts a hydrogen-depleted structure into fragments at rotatable and nonterminal bonds (i.e., single bonds that are not part of a ring and do not include atoms connected to only one other atom) and stores the generated fragments along with their substitution point information using the structure-data file format (SDF) of MDL. Next, the screening tool filters all the stored fragments according to user-defined rules. Any duplicates or undesirable fragments are discarded. We thus applied the KEGG data set as a source when constructing the free parts, with the following constraints: First, molecules that contain elements other than C, N, O, S, P, F, Cl, Br, I, B, As, Si, or Sn were not processed by the splitting tool. Similarly, molecules containing non-naturally occurring isotopes were also discarded, leaving 8868 for further processing. Fragments containing more than 14 non-hydrogen atoms or more than three rings were not entered into the library, nor those with an overall charge. We finally obtained a fragment library consisting of 2238 unique entries, including 1155 side chains with a single substitution. Of the 1083 scaffolds, 755 contain two substitution points, 251 have three, 66 have four, 10 have five, and 1 such scaffold has six substitution points.

The Fitness Pressure. The fitness function used in the present EA experiments is a QSAR model that predicts catalytic "productivity" as defined in ref 43. The productivity represents the stability of the metallacyclobutane intermediate relative to a series of 16-electron precursor-like, inactivated complexes and correlates very well ($R^2 > 0.98$) with DFT-calculated barrier heights,[84] as well as with the observed catalytic activity of the catalysts.[43]

The QSAR model ($Q^2 = 0.85$, RMSECV = 1.48 kcal/mol) correlates the productivity of 27 catalysts with molecular descriptors obtained in geometry optimizations using the semiempirical method PM6[64] as implemented in the MOPAC program.[80] With more than half (19) of these catalysts containing NHC ligands, and three bearing phosphorus-based ligands, the QSAR model is trained mainly at describing these common classes of ligands and should be used with care in case of more "exotic ligands". The descriptors were the average Ru−Cl bond distance, the Cl−Ru−Cl bond angle, the average Cl−Ru═C bond angle, the absolute value of X−Ru═C−H_{syn} torsion angle (H_{syn} is the hydrogen atom in syn position with respect to the X atom), the average partial charge on Cl, and the average partial charge on H. Here, Ru, C, Cl, and H are atoms of the core part, and X is the donor atom of the trial part

formally bound to Ru. Higher predicted productivity value for the generated structure means a higher fitness. The PLSR coefficients of the linear combination are presented in the Supporting Information.

As noted above, when we use regression models to define a simplified fitness measure, it is important to ensure that the EA optimization stays within the applicability domain (AD) of the model. In the present method, this is achieved mainly by constraining the search space by requiring that parts of the overall structure are kept fixed (the core parts) or are to be found among a finite set of (ligand) scaffolds (the trial parts). In addition, the present structural search space has also been limited by introducing a few upper bounds to selected key values, such as the maximum number of non-hydrogen atoms (60), and other structural requirements (see Table S1 in the Supporting Information).

Parameters of the EA Test Runs. The parameters pertaining to the structural operations, fitness applicability domain, and other relevant EA options are listed in Table S1 in the Supporting Information.

Two important settings, the maximum number of generations and the individual size of the population, are both set reasonably high (to 50) so as to ensure sufficient resolution and sampling. A relatively low number of offspring structures (five) generated per generation is commensurate with a moderate evolution pressure and smoothly evolving generations.

An optimization is terminated when no candidate is found fitter than the least well adapted structure in the current population within 100 successive searches (exhaustion size). Moreover, the maximum number of conformers visited in the conformational searches using the calculator (cxcalc) of the Marvin package[85] was set reasonably high (250). Tests showed that allowing more conformers to be included did not lead to significant improvement in the resulting 3D geometries.

■ ASSOCIATED CONTENT

Supporting Information

Details of the *de novo* evolutionary algorithm, complete computational details of the case study, results from additional evolution experiments, and details of the explicitly DFT-calculated productivity of catalysts based on predicted ligands **I1**, **I2**, and **I3**, as well as of the structural variation of **I3** (**I3_a**, **I3_b**, **I3_c**, **I3_d**, and **I3_e**).This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

vidar.jensen@kj.uib.no; bjorn.alsberg@chem.ntnu.no

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Y.C. gratefully acknowledges the Department of Chemistry of the Norwegian University of Science and Technology (NTNU) for funding of her Ph.D. research. The Norwegian Research Council (NFR) is acknowledged for financial support from the eVITA (Grant No. 205273), GASSMAKS (Grant No. 182536), and KOSKII (Grant No. 177322) programmes, as well as for CPU resources granted through the NOTUR supercomputing programme. The University of Bergen is acknowledged for financial support through the Nanoscience programme and ChemAxon (http://www.chemaxon.com) for free academic use of the Marvin package.

■ REFERENCES

(1) Manly, C. J.; Louise-May, S.; Hammer, J. D. *Drug Discovery Today* **2001**, *6*, 1101−1110.
(2) Xu, H.; Agrafiotis, D. K. *Curr. Top. Med. Chem.* **2002**, *2*, 1305−1320.
(3) Schneider, G.; Böhm, H.-J. *Drug Discov. Today* **2002**, *7*, 64−70.
(4) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.
(5) Schneider, G. *Nat. Rev. Drug Discov.* **2010**, *9*, 273−276.
(6) Klebe, G. *J. Mol. Med.* **2000**, *78*, 269−281.
(7) Hartenfeller, M.; Schneider, G. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 742−759.
(8) Comba, P.; Kerscher, M. *Coord. Chem. Rev.* **2009**, *253*, 564−574.
(9) Wang, R. X.; Gao, Y.; Lai, L. H. *J. Mol. Model.* **2000**, *6*, 498−516.
(10) Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.
(11) Schneider, G.; Lee, M.-L.; Stahl, M.; Schneider, P. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 487−494.
(12) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15*, 2785−2794.
(13) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387−1391.
(14) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190−195.
(15) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; Clair, J. L. S.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. *Science* **2010**, *329*, 309−313.
(16) Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, 1975.
(17) Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Cambridge, MA, 1992.
(18) Schwefel, H.-P. *Evolution and Optimum Seeking*; Wiley-Interscience, New York, 1995.
(19) Fogel, L. J.; Owens, A. J.; Walsh, M. J. *Artificial Intelligence through Simulated Evolution*; John Wiley & Sons: New York, 1966.
(20) Lameijer, E.-W.; Bäck, T.; Kok, J. N.; Ijzerman, A. P. *Nat. Comput.* **2005**, *4*, 177−243.
(21) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. J. *J. Med. Chem.* **2003**, *46*, 2765−2773.
(22) Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1079−1087.
(23) Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. *J. Med. Chem.* **2005**, *48*, 2457−2468.
(24) Fechner, U.; Schneider, G. *J. Chem. Inf. Model.* **2006**, *46*, 699−707.
(25) Dey, F.; Caflisch, A. *J. Chem. Inf. Model.* **2008**, *48*, 679−690.
(26) Durrant, J. D.; Amaro, R. E.; McCammon, J. A. *Chem. Biol. Drug Des.* **2009**, *73*, 168−178.
(27) Holena, M.; Linke, D.; Rodemerck, U. *Catal. Today* **2011**, *159*, 84−95.
(28) Baumes, L. A.; Serna, P.; Corma, A. *Appl. Catal. A: Gen.* **2010**, *381*, 197−208.
(29) Kang, S.; Clerc, F.; Farrusseng, D.; Mirodatos, C.; Woo, S. I.; Park, S. *Top. Catal.* **2010**, *53*, 2−12.
(30) Kreutz, J. E.; Shukhaev, A.; Du, W. B.; Druskin, S.; Daugulis, O.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2010**, *132*, 3128−3132.
(31) Maldonado, A. G.; Hageman, J. A.; Mastroianni, S.; Rothenberg, G. *Adv. Synth. Catal.* **2009**, *351*, 387−396.

(32) Burello, E.; Rothenberg, G. *Adv. Synth. Catal.* **2005**, *347*, 1969–1977.
(33) Cundari, T. R.; Deng, J.; Zhao, Y. *Ind. Eng. Chem. Res.* **2001**, *40*, 5475–5480.
(34) Buda, C.; Flores, A.; Cundari, T. R. *J. Coord. Chem.* **2005**, *58*, 575–585.
(35) Hay, B. P.; Firman, T. K. *Inorg. Chem.* **2002**, *41*, 5502–5512.
(36) Hageman, J. A.; Westerhuis, J. A.; Frühauf, H.-W.; Rothenberg, G. *Adv. Synth. Catal.* **2006**, *348*, 361–369.
(37) Occhipinti, G.; Meermann, C.; Dietrich, H. M.; Litlabø, R.; Auras, F.; Törnroos, K. W.; Maichle-Mössmer, C.; Jensen, V. R.; Anwander, R. *J. Am. Chem. Soc.* **2011**, *133*, 6323–6337.
(38) Werner, J.-P.; Burger, P. *J. Phys. Chem. A* **2011**, *115*, 13885–13895.
(39) Weill, N.; Corbeil, C. R.; De Schutter, J. W.; Moitessier, N. *J. Comput. Chem.* **2011**, *32*, 2878–2889.
(40) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. *Dalton Trans.* **2009**, 8183–8196.
(41) Sparta, M.; Børve, K. J.; Jensen, V. R. *J. Am. Chem. Soc.* **2007**, *129*, 8487–8499.
(42) Burello, E.; Rothenberg, G. *Int. J. Mol. Sci.* **2006**, *7*, 375–404.
(43) Occhipinti, G.; Bjørsvik, H.-R.; Jensen, V. R. *J. Am. Chem. Soc.* **2006**, *128*, 6952–6964.
(44) Burello, E.; Farrusseng, D.; Rothenberg, G. *Adv. Synth. Catal.* **2004**, *346*, 1844–1853.
(45) Deeth, R. J.; Foulis, D. L. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4292–4297.
(46) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard III, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
(47) Hoveyda, A. H.; Zhugralin, A. R. *Nature* **2007**, *450*, 243–251.
(48) Kozuch, S.; Shaik, S. *Acc. Chem. Res.* **2011**, *44*, 101–110.
(49) Loving, K.; Alberts, I.; Sherman, W. *Curr. Top. Med. Chem.* **2010**, *10*, 14–32.
(50) Han, Y. Q.; Steinbeck, C. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 489–498.
(51) Steinbeck, C.; Han, Y. Q.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493–500.
(52) Trnka, T. M.; Grubbs, R. H. *Acc. Chem. Res.* **2001**, *34* (1), 18–29.
(53) Dias, E. L.; Nguyen, S. T.; Grubbs, R. H. *J. Am. Chem. Soc.* **1997**, *119*, 3887–3897.
(54) Fürstner, A.; Ackermann, L.; Gabor, B.; Goddard, R.; Lehmann, C. W.; Mynott, R.; Stelzer, F.; Thiel, O. R. *Chem.—Eur. J.* **2001**, *7*, 3236–3253.
(55) Samojłowicz, C.; Bieniek, M.; Grela, K. *Chem. Rev.* **2009**, *109*, 3708–3742.
(56) Trnka, T. M.; Morgan, J. P.; Sanford, M. S.; Wilhelm, T. E.; Scholl, M.; Choi, T.-L.; Ding, S.; Day, M. W.; Grubbs, R. H. *J. Am. Chem. Soc.* **2003**, *125*, 2546–2558.
(57) Vougioukalakis, G. C.; Grubbs, R. H. *Chem. Rev.* **2010**, *110*, 1746–1787.
(58) Schwab, P.; France, M. B.; Ziller, J. W.; Grubbs, R. H. *Angew. Chem., Int. Ed.* **1995**, *34*, 2039–2041.
(59) Nguyen, S. T.; Johnson, L. K.; Grubbs, R. H.; Ziller, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 3974–3975.
(60) Scholl, M.; Ding, S.; Lee, C. W.; Grubbs, R. H. *Org. Lett.* **1999**, *1*, 953–956.
(61) Huang, J. K.; Stevens, E. D.; Nolan, S. P.; Petersen, J. L. *J. Am. Chem. Soc.* **1999**, *121*, 2674–2678.
(62) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743.
(63) Höskuldsson, A. *J. Chemom.* **1988**, *2*, 211–228.
(64) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173–1213.
(65) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(66) Fontaine, F.; Pastor, M.; Sanz, F. *J. Med. Chem.* **2004**, *47*, 2805–2815.
(67) Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F. *J. Med. Chem.* **2005**, *48*, 2687–2694.
(68) Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci* **2000**, *11* (Suppl 2), S29–S39.
(69) King, R. D.; Muggleton, S. H.; Srinivasan, A.; Sternberg, M. J. E. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 438–442.
(70) Buttingsrud, B.; Ryeng, E.; King, R. D.; Alsberg, B. K. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 361–373.
(71) Buttingsrud, B.; King, R. D.; Alsberg, B. K. *J. Chemom.* **2007**, *21*, 509–519.
(72) Kadyrov, R.; Wolf, D.; Azap, C.; Ostgard, D. J. *Top. Catal.* **2010**, *53*, 1066–1072.
(73) Chung, C. K.; Grubbs, R. H. *Org. Lett.* **2008**, *10*, 2693–2696.
(74) Fournier, P.-A.; Collins, S. K. *Organometallics* **2007**, *26*, 2945–2949.
(75) Fiesselmann, H.; Sasse, K. *Chem. Ber.-Recl.* **1956**, *89*, 1775–1791.
(76) Hirano, K.; Urban, S.; Wang, C.; Glorius, F. *Org. Lett.* **2009**, *11*, 1019–1022.
(77) Deeth, R. J.; Anastasi, A. E.; Wilcockson, M. J. *J. Am. Chem. Soc.* **2010**, *132*, 6876–6877.
(78) Baber, J. C.; Feher, M. *Mini-Rev. Med. Chem.* **2004**, *4*, 681–692.
(79) Chu, C.; Alsberg, B. K. *J. Chemom.* **2010**, *24*, 399–407.
(80) Stewart, J. J. P. *MOPAC2009*; Stewart Computational Chemistry, Colorado Springs, CO, USA, http://openmopac.net (2008).
(81) von Grotthus, M.; Koczyk, G.; Pas, J.; Wyrwicz, L. S.; Rychlewski, L. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 757–761.
(82) Cragg, G. M.; Newman, D. J.; Snader, K. M. *J. Nat. Prod.* **1997**, *60*, 52–60.
(83) Breinbauer, R.; Vetter, I. R.; Waldmann, H. *Angew. Chem., Int. Ed.* **2002**, *41*, 2878–2890.
(84) Adlhart, C.; Chen, P. *J. Am. Chem. Soc.* **2004**, *126*, 3496–3510.
(85) Marvin 5.3.4, ChemAxon Ltd., http://www.chemaxon.com.