DOI: 10.1002/chem.200802192

Design of a Genetic Algorithm for the Simulated Evolution of a Library of Asymmetric Transfer Hydrogenation Catalysts

Nicolas Vriamont,[a] Bernadette Govaerts,[b] Pierre Grenouillet,[c] Claude de Bellefon,*[c] and Olivier Riant*[a]

Abstract: A library of catalysts was designed for asymmetric-hydrogen transfer to acetophenone. At first, the whole library was submitted to evaluation using high-throughput experiments (HTE). The catalysts were listed in ascending order, with respect to their performance, and best catalysts were identified. In the second step, various simulated evolution experiments, based on a genetic algorithm, were applied to this library. A small part of the library, called the mother generation (G0), thus evolved from generation to generation. The goal was to use our collection of HTE data to adjust the parameters of the genetic algorithm, in order to obtain a maximum of the best catalysts within a minimal number of generations. It was namely found that simulated evolution's results depended on the selection of G0 and that a random G0 should be preferred. We also demonstrated that it was possible to get 5 to 6 of the ten best catalysts while investigating only 10% of the library. Moreover, we developed a double algorithm making this result still achievable if the evolution started with one of the worst G0.

Keywords: asymmetric catalysis · genetic algorithm · simulated evolution · transfer hydrogenation

Introduction

Over the last decades, the importance of chirality has been widely recognized in the medicinal, pharmaceutical and agricultural chemistry. It makes it a centre of interest for synthetic organic chemist trying to develop methodologies that provide high levels of enantiodiscrimination, and asymmetric catalysis is by far the most appealing way to get highly enantioselective chemical transformations of molecules of high interest.[1]

In this paper, we focus on the particular case of carbonyl reductions, with regards to the importance of chiral alcohols as essential building blocks.[2] The development of a new asymmetric catalyst, is a challenging and time-consuming task: A good catalyst must combine a robust structure and proper kinetics, which requires the adjustment of a large number of parameters. A common way to develop new catalysts is to imagine a general structure, which combines a metallic centre with a chiral ligand that leads to the formation of a chiral catalyst that could efficiently promote a reaction, in terms of both *ee* and yield. Analogs of this general structure are then synthesized and tested, which may then provide general trends for improving the catalyst design.[3] Better structures can then be synthesized to obtain enhanced or at least satisfactory performances. Certainly, generalities can be drawn regarding the connection between the structure of catalysts and their potential application.

The "rate-limiting step" of such a procedure is the time a chemist can work within a defined period. Therefore, the part of the chemical space (chiral ligands, metallic precursors, solvents, concentrations, temperatures, pressures, etc.) that can be explored remains rather small. High-throughput screenings and high-throughput experiments techniques (HTS and HTE) have been set up to help chemists perform-

[a] N. Vriamont, Prof. O. Riant
Departement de Chimie Organique et Médicinale
Université catholique de Louvain, Bâtiment Lavoisier
Place Pasteur, 1, 1348 Louvain-la-Neuve (Belgium)
Fax: (+32) 104-741-68
E-mail: olivier.riant@uclouvain.be

[b] Prof. B. Govaerts
Institut de Statistique
Université catholique de Louvain
20 voie du roman pays, 1348 Louvain-la-Neuve (Belgium)

[c] P. Grenouillet, Dr. C. de Bellefon
Laboratoire de Génie des Procédés Catalytiques
CNRS-ESCPE Lyon, BP2077, 69616 Villeurbanne (France)
Fax: (+33) 4-7243-1754
E-mail: claude.debellefon@lgpc.cpe.fr

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200802192.

ing more synthesis and, consequently, more tests within a given time. Combinatorial chemistry has also been fruitfully combined with HTE/HTS to allow researchers to explore a wider chemical space.[4] It is now possible to get access to some structures faster, as well as speeding up the evaluation phase of catalysts (Figure 1).

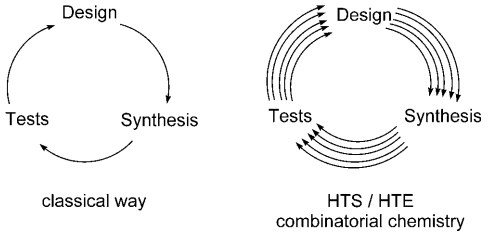
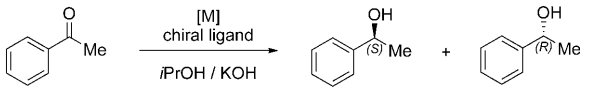


Figure 1. Comparison of the usual steps in the classical way and the HTE way on engineering structures.

Despite the distinct advantage in the principles of combinatorial chemistry and the alarming rise in HTE/HTS techniques, they still suffer from the same limitations as classical experiments: All arrays have to undergo synthesis and evaluation before any statement concerning a high-quality catalyst(s) can be released. Even though the chemical space is now wider, it still has to be fully explored.

Faster access to relevant catalysts might arise from the ability to predict catalyst behavior directly from its structure. The efforts made in this field occasionally gave very attractive results, but such an approach still requires high-quality data and time-consuming calculations.[5] Alternatively, scientists were inspired by nature for improving the design of catalysts. Every lifeform on Earth is the result of an evolution over billions of years, and results from the amazing capacity of life to accommodate to the environment by self-modification processes. The genetic code of live organisms is organized in building blocks, the position of which can endure some variations. Sporadically, these variations produce a slightly modified organism, which is better adapted to its environment. The adaptation of such concepts to the codification of possible variations of the backbone of a synthetic target has led some groups of combinatorial chemists to see them as pieces of a large puzzle that can infinitely re-arrange.[6] The possibility to mimic Nature's machinery with evolutionary chemistry is an attractive approach, which allows transmitting only the most interesting features during the optimization of a structure for a chosen chemical or biological property.[7] This approach has been investigated by several groups and designated as "directed evolution"[8] or "simulated evolution"[9] depending on the chosen pathway for the evolution process.

In this paper, we decided to apply the principle of simulated evolution to a library of chiral catalysts that have been used in asymmetric hydride transfer reduction of ketones (Scheme 1), a reaction extensively explored by Noyori et al.[10]

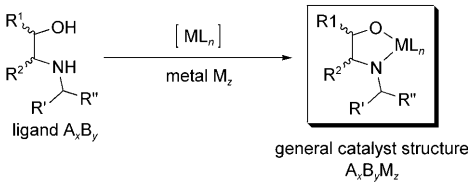


Scheme 1. Metal-catalyzed hydride transfer to acetophenone, as a benchwork reaction.

To validate the methodology, we first considered the synthesis of whole building blocks of the catalysts and HTE evaluation of all the possible combinations. We then constructed a genetic algorithm that was applied to a small part of the library, with systematic variation of the adjustable parameters of the genetic algorithm. We performed simulated evolutions to evaluate the ability of our algorithm to find the best catalysts without testing the entire library. Our main result is that it was sufficient to evaluate about 10% of the whole library to get the best catalysts.

Results and Discussion

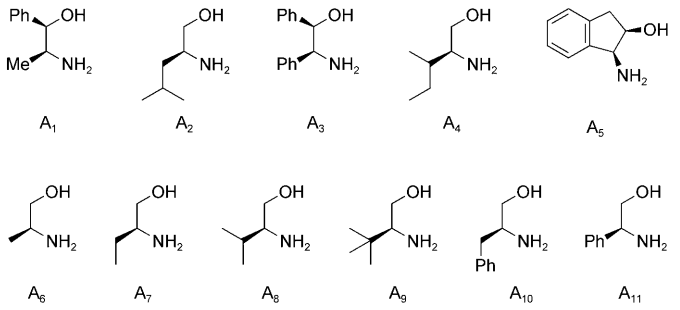
Framework of simulated evolution: We first built a library of chiral catalysts by considering a family of ligands with a chiral β-amino alcohol structure bearing a substituent on the nitrogen atom. Such ligands have been reported to provide high activity and enantioselectivity in the ruthenium asymmetric catalyzed transfer hydride reaction to prochiral ketones and can be easily prepared from commercially available starting materials.[11] We introduced chemical diversity on those structures through alkylation of the primary amine group by reductive amination with various aldehydes and ketones, yielding a sub-library of chiral ligands A_xB_y. The third variation was introduced through the metal complex precursor M_z, giving the final library of chiral catalysts $A_xB_yM_z$ (Scheme 2).



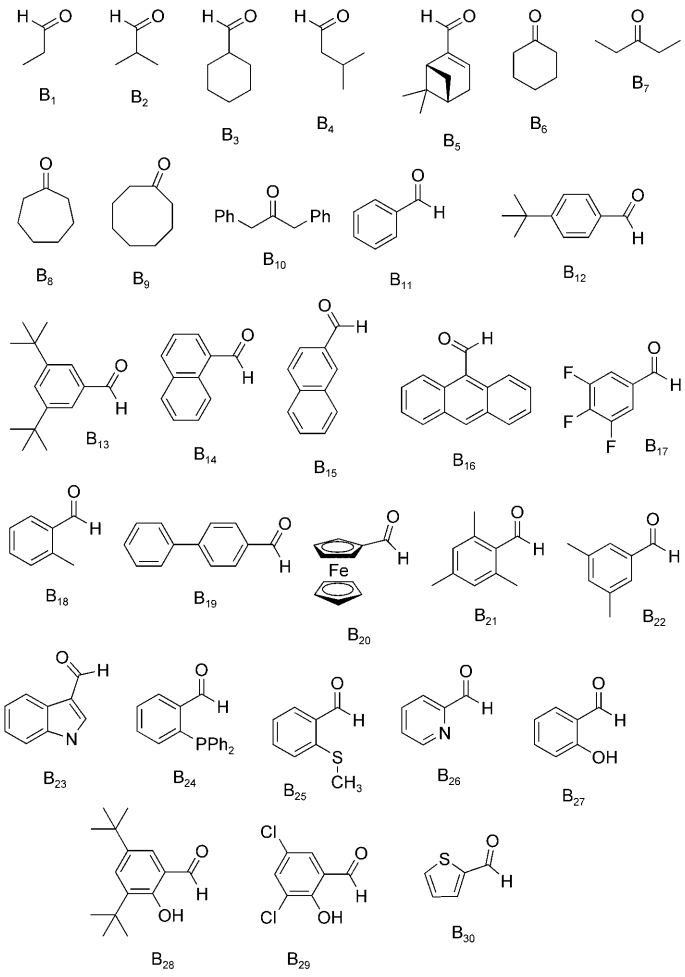
Scheme 2. Generation of a library of chiral catalysts from a metal precursor and a chiral β-amino-alcohol ligand.

A set of 11 enantiopure β-amino alcohols (A_1-A_{11}) was selected according to their commercial availability and/or simplicity of synthesis (from the corresponding amino-acids), and diversity in the substituent of the β-amino alcohol backbone (Scheme 3).[12]

We introduced the substituents on the nitrogen atom through reductive alkylation involving various aldehydes and ketones. The set of 30 carbonyl groups (B_1–B_{30}) selected for this transformation is given in Scheme 4.

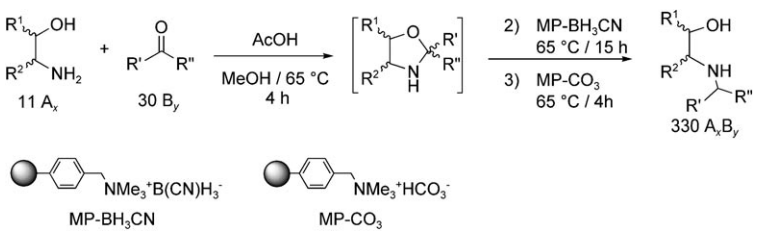


Scheme 3. The set of β-amino alcohols from A_1–A_{11}.



Scheme 4. The set of carbonyl groups from B_1–B_{30}.

The reductive amination was carried out by parallel liquid phase synthesis using polymer supported reagents in order to speed up the synthesis of the sub-library and to avoid purification of each chiral ligand. About 0.2 mmol of each ligand was prepared by reaction of the amino-alcohol with 1 equivalent of the carbonyl compound and 1 equivalent of acetic acid in methanol, which generated a mixture of oxazolidine and imine. The cyanoborohydride supported reagent was then added to reduce the intermediate oxazolidine. The alkylated ligand was released by addition of an excess of carbonate supported reagent, followed by filtration and evaporation of the solvent (Scheme 5).

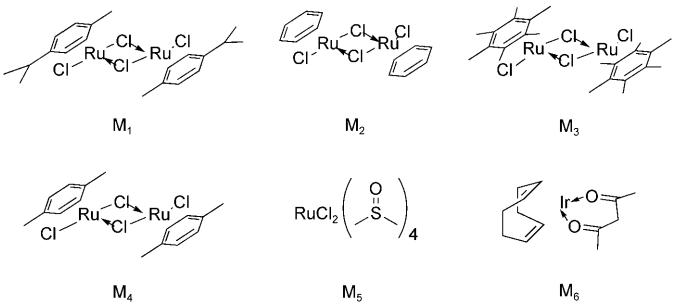


Scheme 5. Preparation of a sub-library of chiral ligands A_xB_y.

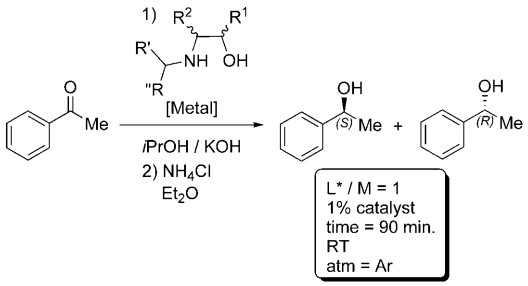
The purity of the 330 ligands was evaluated by ^{1}H NMR analysis of a statistical choice of molecules of the library (see the Supporting Information, average purity: 76%). A procedure for the formation of "in situ" catalysts was also chosen by mixing the amino-alcohol ligands with six metal precursors M_1–M_6 (Scheme 6). M_1 and M_6 were commercially available; M_2, M_3 and M_4 were synthesized according to literature procedures,[13] as well as for M_5.[14]

Evaluation by high throughput experiments (HTE): The combination of 330 chiral ligands and 6 metallic precursors provided a library of 1980 catalysts. Each catalyst was tested regarding the reduction of acetophenone by isopropanol in the presence of potassium hydroxide by using high throughput experiments (HTE). Reactions (Scheme 7) were carried out in standard 2 mL glass vials under an inert atmosphere, and all the operations were monitored by a robot (Gilson 215). The stock solutions of metal, ligands, acetophenone, and KOH were freshly prepared before use and stored under argon. Catalyst solutions were prepared by mixing equimolecular amount of ligand and metal precursor solution (1:1 ratio, 1 mol% of metal loading), and the solutions were aged for 30 minutes at room temperature. After injection of acetophenone, each reaction was initiated by injection of the KOH solution, with a total volume of 400 μL for each vial (substrate concentration, 0.1 M). Reactions were performed for 90 minutes at room temperature under argon. Each reaction was stopped by addition of aqueous ammonium chloride solution and Et_2O, which caused a phase separation. We used an optimized pressure and temperature GC program that allowed us to perform 20 analyses per hour.

For a given catalyst tested within specified conditions, gas chromatography gave us direct access to the conversion rate and the enantiomeric excess. Although optimization of these two factors is required to develop a performing catalytic system, genetic algorithms are usually designed for the optimization of a single parameter. The weight attributed to the conversion with respect to the *ee* could possibly be worked out thanks to a desirability index,[15] but would not necessa-



Scheme 6. The set of metallic precursors from M_1–M_6.



Scheme 7. Enantioselective hydrogen transfer to acetophenone.

rily match with our desire, which is to give a priority to enantiomeric excess over the conversion rate of the reaction. It was necessary to create a single parameter that accounted for both conversion and enantioselectivity. Thus, for each $A_xB_yM_z$ catalyst, we calculated the value of a normalized performance factor (NPF). The NPF is defined as the sum of the enantioselectivity and twice the conversion rate, divided by the sum of the enantioselectivity and twice the conversion rate of the best catalyst of the library (Figure 2). A catalyst NPF was thus comprised between 0 and 1.

Results: We first focus on the results obtained with the A_1 series, which provided some of the best catalysts. The NPF obtained with the 6 metallic precursors and some of the nitrogen substituents are depicted in Figure 3. The NPF obtained with M_1, M_2 and M_4 are from good to excellent, in particular for the best catalyst of the series, $A_1B_{19}M_2$, whose structure is depicted in the inset of the Figure 3. The NPF for $A_1B_{19}M_2$ was then 1, and resulted from a conversion rate of 36% and an enantioselectivity of 73%. The catalysts involving

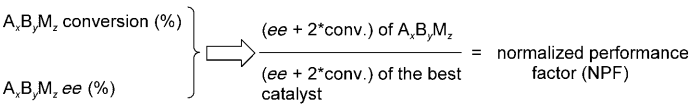


Figure 2. Definition of the normalized performance factor (NPF).

M_3 and M_6 displayed a weak activity, whereas those involving M_5 were almost inactive.

The A_1 and A_7 series provided the best catalysts. Some other series (A_2, A_3, A_6, and A_{11}) gave medium results. Their conversion rates as well as their enantioselectivities led to an NPF value typically around 0.55 (to be compared to an NPF of 1 for the best catalyst). The average enantioselectivity of the best catalysts of these series ranged from 50 to 65%. Here again, M_3, M_5, and M_6 provided poorly active catalysts. Most of the catalysts derived from the A_4, A_5, A_8, A_9, and A_{10} series were almost inactive. The A_5 series was by far the worst one, the best catalyst being only associated to an NPF of 0.06. Considering all the results, we identified the best catalysts, which are listed in Table 1 with their related NPF, conversion rate and enantioselectivity (for details of the data of other catalysts, see the Supporting Information). The three best structures, associated to the first line of the table, are also shown at the bottom of the Table 1. The 9 first catalysts (NPF column) are linked to the highest conversion rates, larger than 30%. Catalysts $A_7B_4M_1$ and $A_7B_{19}M_1$ had the highest conversion rates of the library, up to 61%. If we continued to explore this table until the bottom of the library, we would find the first catalysts with conversion rates lower than 20% and 10% at the 26th and 108th positions respectively. The *ee* of the best catalysts are all higher than 83%, the best ones being very close to 90%. Here again, would we descend the table, first catalysts with enantioselectivity lower than 80% and 50% would be at the 26th and 93rd positions respectively.

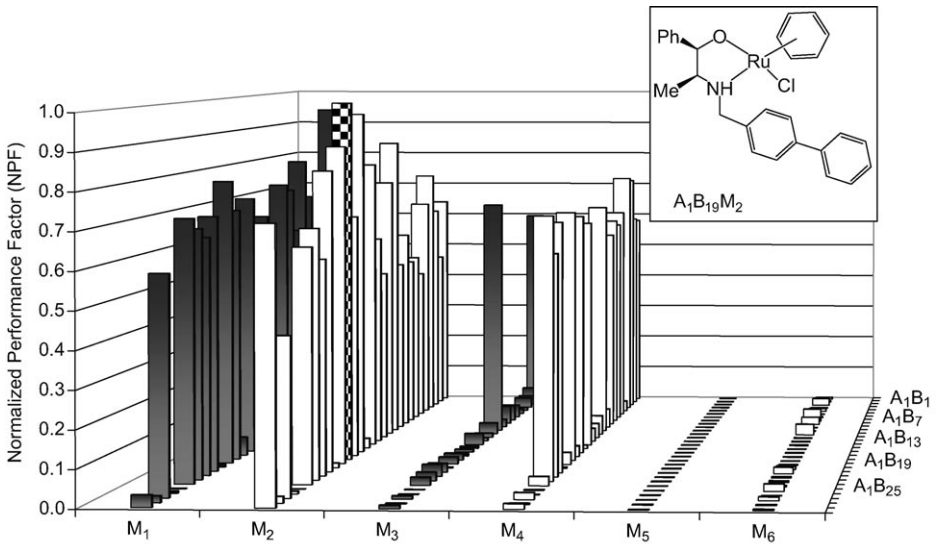
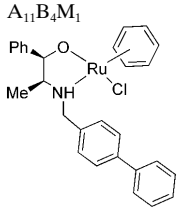
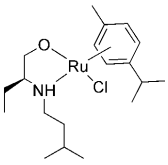
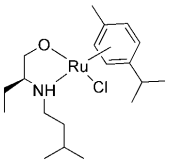


Figure 3. NPF of the 180 catalysts from the A_1 series.

Table 1. List of the 20 best catalysts.

Entry	Performance					
	NPF		Conversion [%]		*ee* [%]	
1	$A_1B_{19}M_2$	**1.00**	$A_7B_4M_1$	**61.5**	$A_1B_3M_4$	**88.9**
2	$A_1B_{17}M_2$	0.96	$A_7B_{19}M_1$	56.0	$A_1B_{11}M_3$	88.5
3	$A_1B_4M_1$	0.95	$A_7B_8M_1$	43.9	$A_1B_{20}M_1$	88.3
4	$A_7B_4M_1$	0.91	$A_7B_1M_1$	38.2	$A_1B_{11}M_4$	88.2
5	$A_7B_{19}M_1$	0.89	$A_7B_6M_1$	37.9	$A_1B_{22}M_4$	88.1
6	$A_1B_{20}M_2$	0.88	$A_1B_{19}M_2$	36.4	$A_1B_4M_4$	88.1
7	$A_1B_{11}M_2$	0.86	$A_1B_{17}M_2$	36.3	$A_1B_{17}M_1$	88.0
8	$A_1B_{22}M_2$	0.82	$A_7B_{13}M_1$	35.7	$A_1B_{24}M_4$	87.2
9	$A_1B_{15}M_2$	0.81	$A_7B_3M_2$	35.5	$A_1B_{22}M_1$	86.7
10	$A_1B_9M_1$	0.80	$A_3B_2M_1$	27.5	$A_1B_{11}M_1$	86.6
11	$A_1B_{20}M_1$	0.78	$A_1B_{20}M_2$	26.2	$A_1B_{25}M_1$	86.5
12	$A_1B_{12}M_2$	0.75	$A_1B_{11}M_2$	26.2	$A_1B_4M_1$	86.5
13	$A_1B_{12}M_1$	0.74	$A_1B_4M_1$	25.8	$A_1B_{12}M_1$	86.2
14	$A_1B_4M_2$	0.74	$A_3B_1M_1$	24.9	$A_1B_5M_1$	86.1
15	$A_1B_4M_4$	0.73	$A_1B_{22}M_2$	24.8	$A_1B_3M_1$	86.1
16	$A_1B_{17}M_1$	0.72	$A_7B_{23}M_1$	24.7	$A_1B_{17}M_4$	85.7
17	$A_1B_{11}M_1$	0.72	$A_1B_{15}M_2$	24.1	$A_1B_{19}M_4$	85.6
18	$A_1B_{30}M_2$	0.72	$A_7B_{10}M_1$	22.2	$A_1B_{19}M_1$	85.5
19	$A_1B_3M_4$	0.72	$A_3B_3M_4$	22.0	$A_1B_{24}M_1$	84.9
20	$A_{11}B_4M_1$	0.71	$A_3B_1M_2$	21.9	$A_1B_{15}M_1$	83.3
						
	$A_1B_{19}M_2$		$A_7B_4M_1$		$A_1B_3M_4$	

Attention can be paid to the constitutive building blocks of the catalysts of the Table 1. An analysis of the structures of the catalysts with the best conversion rates and enantioselectivities reveals that A_1 and M_1 have a 67.5% and 57.5% rate of occurrence respectively, which indicates that A_1 and M_1 bear interesting features regarding the catalytic process. Unfortunately, a similar trend could not be found for "B" (carbonyl groups).

The evaluation of the whole library is depicted in Figure 4 in which the NPF values are listed in ascending order. The figure shape clearly depicts a zone of inactivity of the catalysts, or at best, rather poor activity, as well as a discontinued line around 1600 evaluations. This discontinuity originates from an arbitrary threshold of 5% conversion required for an accurate measurement of the enantioselectivity. Thus, when a conversion of less than 5% was measured, the enantioselectivity was assessed to 0%. This Figure shows that only about 40% of the whole library displays a catalytic activity. Although this could appear at first sight as a poorly effective library, it seems especially representative of what could be expected from a large library of chiral catalysts when evaluated by optimizing a target reaction.

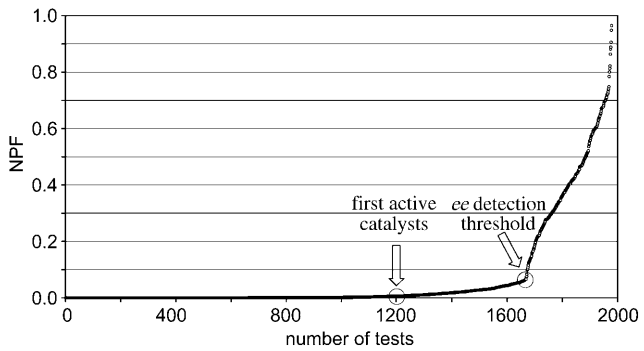


Figure 4. Classification of the NPF in ascending order.

Simulated evolution: The main goal of our study was to develop a genetic algorithm (GA) that reproduces an evolution system and optimizes a small collection of chiral catalysts randomly chosen in a larger library. We used our collection of HTE data to adjust the GA parameters, to obtain a maximum of the best catalysts in a minimal number of generations. Starting from around 1% of the whole library, our method should avoid the use of expensive HTS/HTE technologies and could be carried out using simple laboratory techniques, as a limited number of catalysts synthesis and following evaluations are required at each generation. Moreover, no molecular descriptors are pre-required and a simple "genetic code" can be drawn from the building blocks of our catalysts.

The principle of the genetic algorithm is depicted in Figure 5. A set of catalysts (usually 24 catalysts) was consid-

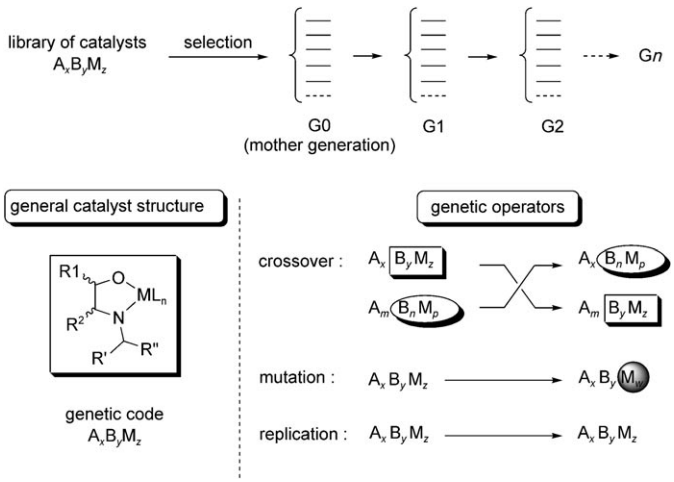


Figure 5. Principle of a genetic algorithm applied to a library of $A_xB_yM_z$ catalysts.

ered as the mother generation (G0), defined as the first generation from which the GA starts. The evolution process usually stopped when about 200 HTE had been performed, which corresponds to an evaluation of about 10% of the library. The evolution process between each generation consisted in the application of three genetic operators (Figure 5). The crossover operator (first operator) generated two catalysts by rearranging two catalysts of the previous generation. The mutation operator (second operator) produced new catalysts by a random variation of one of the three components of the catalysts. The evolution ended with the replication operator, which kept the best catalysts throughout the evolution process.

The Figure 6 depicts the building of the genetic algorithm (GA) and the iterative loop associated to the evolution process. The GA started by creating a mother generation,

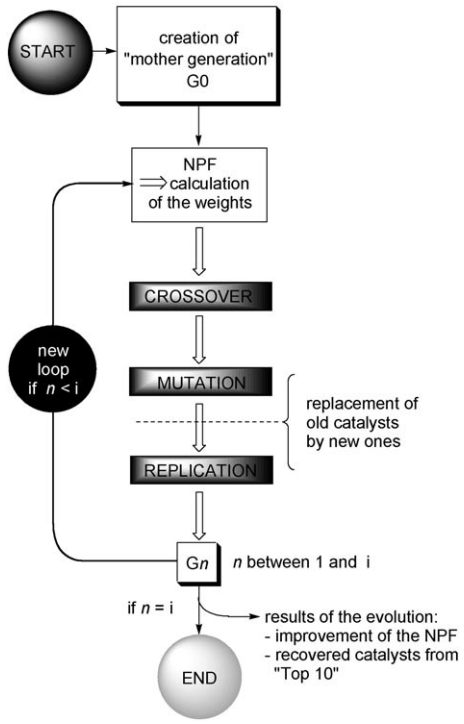


Figure 6. Building the genetic algorithm.

which was extracted from the whole library either randomly or by the chemist. Each catalyst of the whole library was associated to an NPF. Each catalyst of the mother generation was weighted by its NPF, allowing the GA to order the catalysts from the worst (smallest NPF) to the best one (highest NPF). The high-ranked catalysts were selected with a higher probability for crossover operation. Then the GA applied the mutation step, randomly within the generation. Both crossover and mutation rates could be specified, as explained below. After the crossover and mutation steps, some of the mother generation catalysts were replaced by new catalysts, the lowest NPF catalysts having a highest probability of being replaced. The replication step was then applied, for transmitting the unmodified catalysts to the next generation, and the evolution process was resumed. Two additional features of the GA should be noted. First, the algorithm could detect a catalyst that already had been proposed; in this case, the GA searched for another catalyst. Second, the GA generated a minimum of six new combinations per evolution loop, which corresponds to a feasible number of catalysts synthesis and subsequent evaluation by the chemist.

The following parameters of the GA were adjusted for optimization of the evolution process:

G0: Mother generation. This set of catalysts was chosen either randomly or by the chemist.

R: Size of each generation of catalysts. The value of R was 24, 30, 36 or 48 (see below).

RE: Rate of exchange, or crossover rate. Since a crossover step involved 2 catalysts, 2 RE new catalysts were generated at this stage.

RM: Rate of mutation. RM new catalysts were produced at each generation.

RR: Number of catalysts which could be replaced during an evolution loop. RR could be the number of catalysts generated by crossover and mutation; in this case, the new catalysts replaced the poorest catalysts of the previous generation. RR could also be larger than the sum of 2 RE and RM. In this case, the replacements occurred within the RR poorest catalysts, with a higher probability for the worst catalysts.

NI: Number of evolution loops in a completed evolution process.

The genetic algorithm was programmed using the R 2.6.1 software available for free from the R Foundation for Statistical Computing.[16] The first step was to create the table of catalysts, described by their respective genetic code. This table was linked to a second table containing the NPF's of each catalyst. For the second step, the parameters of the GA were written in R 2.6.1 code.[17] The implementation of the GA resulted in an evolution process throughout our library of asymmetric catalysts. Thus, the goal was to use our collection of data to simulate the evolution of sub-libraries of catalysts, trying to improve the efficacy of our catalysts from generation to generation.

Selection of the mother generation (G0): We first considered simulated evolutions in which G0 was randomly extracted from the whole library. To obtain a statistically reliable result, 20 simulations were performed for each set of catalysts. The GA was first implemented with the following parameters: R = 24; RE = 3; RM = 2; RR = 16; NI = 20. The results of a typical simulation are depicted in Figure 7 wherein every dot stands for a generation. The Figure 7A

(lower curve) gives the mean catalyst NPF value as a function of the number of evaluated catalysts (HTE), and the evolution of the mean NPF value of the 10 best catalysts of a given generation (Figure 7A, upper curve). 176 HTE were carried out within 20 evolution loops, that is, 24 HTE for the mother generation and 8 HTE per evolution loop. The performance of an evolution process can also be estimated by considering the number of created catalysts that are in the top 10 list of the entire library (Figure 7B).

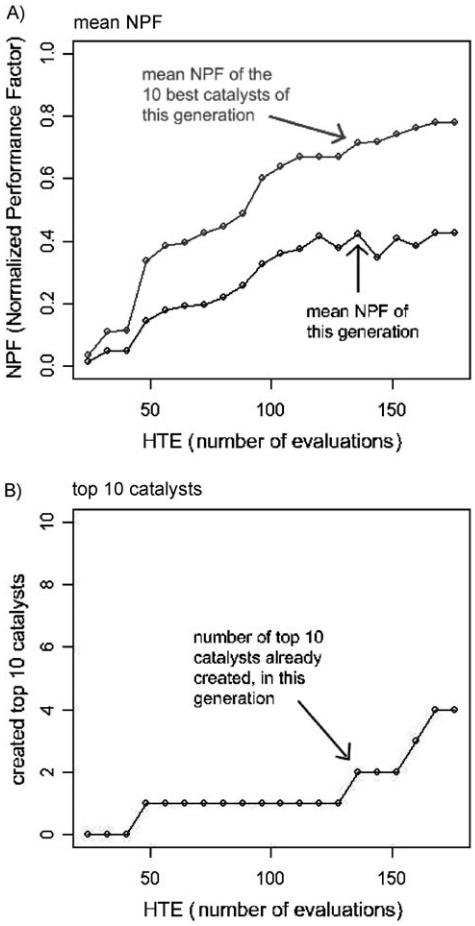


Figure 7. Simulated evolution with random G0. GA parameters: R = 24; RE = 3; RM = 2; RR = 16; NI = 20.

Alternatively, we directly selected the mother generation. We choose to introduce chemical diversity in G0, assuming that it would be a prerequisite for successful evolutions. We divided the constitutive catalysts building blocks (A, B, and M) into families (Figure 8). We selected the β-amino alcohols A_3, A_5 and A_7 as representatives of three chemical families. Each amino-alcohol (A) was combined with carbonyl groups (B) randomly selected among aliphatic aldehydes, ketones, aromatic aldehydes and aromatic aldehydes with a potentially coordinating heteroatom. Finally, the associated metal precursors (M) were split into two groups: the four $[RuCl_2(arene)]_2$ and the two remaining complexes. The resulting combinations were supposed to maximize the chemical diversity in a mother generation of 24 catalysts.

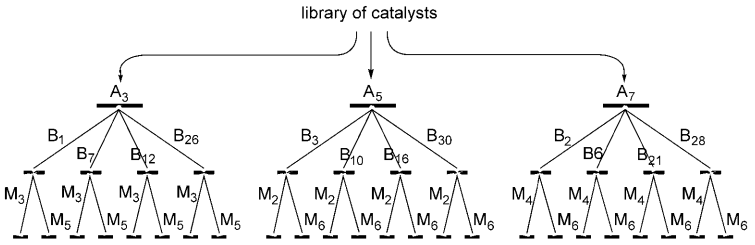


Figure 8. Selection of the G0 members.

This mother generation was submitted to the GA, which, however, was unable to perform a single evolution loop because most of the catalysts were associated to a null NPF. Therefore the GA was unable to decide which catalysts of G0 had to undergo crossover and mutation steps. Nevertheless, by randomly choosing the B and M building blocks associated to A_3, A_5 and A_7, the GA was capable to complete the evolution process.

We compared the results of two sets of 20 evolutions starting from either random or partially-selected mother generation. For this purpose, five parameters were evaluated:

1) The mean number of top 10 catalysts found at the end of the evolutions (mean top 10).
2) The mean number of tests (MNT), which were required to find one of the ten best catalysts. The MNT can be considered as a measure of the speed of the evolution process to create a catalyst from the top 10, even if the lower the MNT, the faster the evolution creates catalysts from top 10.
3) The standard deviation of the MNT (σ_{MNT}).
4) The occurrence of the best catalyst structure ($A_1B_{19}M_2$), given in percent.
5) The percentage of evolutions that were unable to find any of the top 10 catalysts (no top 10).

Results are given in Table 2.

The highest number of best catalysts was obtained from simulations starting from random mother generations, which also triggered the lowest MNT, meaning the highest evolu-

Table 2. Comparison of two sets of evolutions with random or selected G0.

G0	Mean top 10	MNT[b]	σ_{MNT}	$A_1B_{19}M_2$ [%]	no top 10 [%]
random[a]	3.8	62	56	35	35
selected[a]	2.7	79	54	30	30

[a] Results given for 20 simulated evolutions in every set. [b] The lower the MNT, the faster the evolution is.

tion speed. The standard deviations of the MNT's were high and rather similar, as well as the occurrence of $A_1B_{19}M_2$ and the percentage of evolutions without top 10 catalysts. Thus, a random choice of the mother generation is statistically more efficient than trying to guess a "right" G0.

Influence of the parameters of the genetic algorithm: We investigated the effect of the cross-over (RE) and mutation (RM) rates on the GA performance, the mother generation being randomly chosen and the size R of the generation being fixed. Since a minimum rate of mutation was necessary for introducing novelty, we first considered evolutions that involved only one mutation (RM = 1) and crossover rates RE ranging from 1 to 4. We performed 10 evolutions for each RE value, and noticed only a slight increase of the GA performance with increasing RE. We then considered evolution processes with RE = 1 and RM = 2, 4, 6 and 8, and for each value of RE we performed 10 evolutions. Increasing RM affected the GA performance only slightly. For instance, the mean number of top 10 catalysts for RM = 2 and RM = 8 was 1.4 and 2.4 respectively, whereas the MNT was 53 and 101 respectively, in both cases with a large standard deviation (σ_{MNT} of 29 and 80 respectively). These results can be understood by considering that, at low cross-over rate, a novelty introduced by a mutation has a low probability of being transmitted, because the probability of being replaced by next mutations before rearrangement is high.

We also performed simulated evolutions in which the preference was given to either crossover, with RE = 4 and RM = 2, or mutation, with RE = 2 and RM = 4. Since such rates could enable the GA to find more catalysts from the top 10, we increased the size of the mother generation G0 from 24 to 30, in order to be sure that the top 10 catalyst found by the GA were preserved during the evolution process. The number of required tests was thus increased to 220. Both evolutions provided similar results (Table 3). In particular, the MNT does not depend on the RE/RM ratio, only σ_{MNT} is slightly reduced when the priority is given to mutations.

Table 3. Comparison of two sets of evolutions with preference to either crossover or mutation rates.

Preference	Mean top 10	MNT[b]	σ_{MNT}	$A_1B_{19}M_2$ [%]	no top 10 [%]
crossover (RE)[a]	2.4	100	73	15	10
mutation (RM)[a]	2.6	79	43	15	0

[a] Results given for 20 simulated evolutions in every set. [b] The lower the MNT, the faster the evolution is.

We also addressed the impact of both high crossover and mutation rates. Ten simulated evolutions were performed with RE = 4 and RM = 8 in order to estimate the effect of these large rates on the 24 members of the mother generation. As shown in Figure 9 and Table 4, the GA performance was increased. The mean number of top 10 catalysts was 5.7, the MNT was 59, and $A_1B_{19}M_2$ was found in 70% of the evolutions. In most cases, however, the mean NPF remained rather low (Figure 9 A, lower curve) as expected with so high perturbation rates, when the mean NPF value of the 10 best catalysts of a given generation (Figure 9A, upper curve) kept rising. Thus, high perturbation rates promoted the finding of top 10 catalysts rather than an increase of the mean NPF. The number of required test was also largely increased, and 344 evaluations were necessary to drive each simulated evolution.

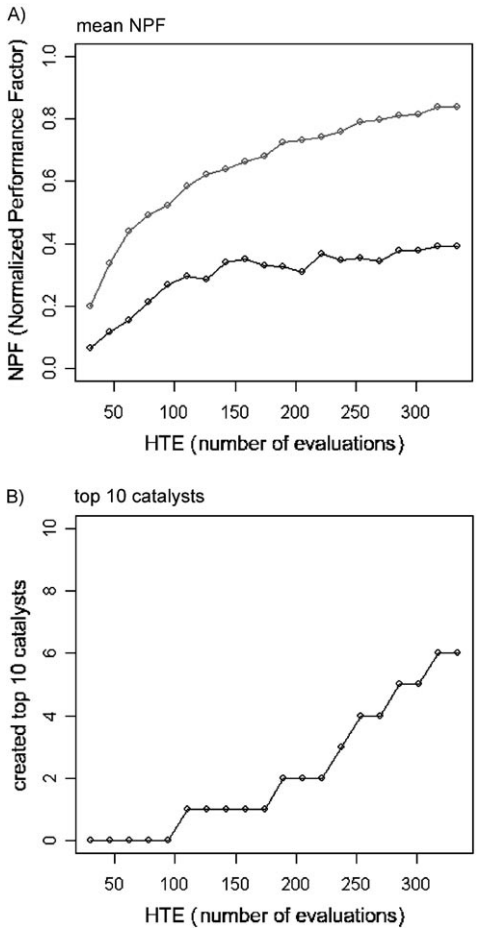


Figure 9. Simulated evolution with high crossover and mutation rates. GA parameters: R = 30; RE = 4; RM = 8; RR = 20; NI = 19.

Table 4. Set of evolution with high crossover and mutation rates.

Preference	Mean top 10	MNT[b]	σ_{MNT}	$A_1B_{19}M_2$ [%]	no top 10 [%]
high RE and high RM[a]	5.7	59	22	70	0

[a] Results given for 10 simulated evolutions in every set. [b] The lower the MNT, the faster the evolution is.

We also modified the number of catalysts, which were replaced at each generation (RR). As explained before, the smaller RR, the more the replacements happened within the less competitive catalysts. Ten evolution processes were performed for RR = 6, 12 and 18. The cross-over and mutation rates were RE = 2 and RM = 2, and the mother genera-

tion G0 was randomly selected. Simulated evolution results were not affected by RR variations (data not shown). The explanation lies on the size of replacement space compared to the proportion (about 20%) of the library that contains moderate to efficient catalysts (Figure 4). One can assume that a randomly selected G0 contains the same proportion of efficient catalysts. The maximum size of the replacement space we tested corresponds to RR = 18, that is, a replacement space of 75% of each generation. Since an influence of RR on the first evolution loops is expected only if RR is larger than 80% of the G0 size, simulated evolution results were therefore not affected by RR variations. It also implies that the algorithm should be robust with regards to RR variations during the next evolution loops.

At last, we studied the effect of the size of the mother generation G0. We expected that increasing the size of G0 would incorporate more chemical diversity and thus increase the GA performance, in the cost of more HTE required at each generation. The G0 sizes we implemented were 24, 36 and 48, the catalysts being randomly selected. Taking into account the lack of influence of the RR variations on the GA performance, we selected a maximum RR (14, 26 and 38 respectively, so the 10 best catalysts could eventually be recovered and kept throughout the simulated evolution experiments). Finally, NI was set to 40 to bring to light any generation size effect. The mutation and crossover rates were RE = 2 and RM = 3. To our surprise, the size of G0 did not seem to have any influence on the GA performance (Table 5). The MNT was unaffected by G0 size variations, as well as the number of top 10 recovered catalysts, which nevertheless was more than doubled due to a larger NI. Thus, a mother generation of 24 catalysts was sufficient to find the best catalysts. The simulated evolution processes performed with these RE and RM intermediate values provided good results. Finally, the size of G0 might have more influence on simulated evolution performed within a larger library, as a G0 of 24 members, that is, more than 1% of this library, is still a large mother generation. Note that each simulation process led to at least one catalyst from the top 10, which gives the reason why the column dedicated to processes with zero top 10 catalysts is no longer presented.

Table 5. Variation of the size of G0.

Size of G0	Mean top 10	MNT[b]	σ_{MNT}	$A_1B_{19}M_2$ [%]
24[a]	6.5	40	14	90
36[a]	6.5	41	9	80
48[a]	6.1	48	15	70

[a] Results given for 10 simulated evolutions in every set. [b] The lower the MNT, the faster the evolution is.

Simulated evolution with a double algorithm (DA): We now address a real case of optimization as must be managed by a simulated evolution process: The whole library can surely be pictured by the chemist, but only a few catalysts would really be synthesized and evaluated. On one hand, we learned from former simulations that it is advantageous to start with a randomly selected mother generation. On the other hand, this first generation could contain only inefficient catalysts, which can seriously impede the evolution process, as demonstrated previously when we selected the mother generation.

The efficiency of an optimization process is thus directly linked to the quality of the mother generation. But we also know, since we evaluated the whole library, that it contains about 75% of inefficient catalysts. The probability P that n members of G0 originate from this zone of inactivity can be approximated by Equation (1).

$$P(\chi = n) = \left(\frac{m}{N}\right)^n \quad (1)$$

In which N is the total number of catalysts and m the number of catalysts in the zone of inactivity. For the library ($m/N \approx 0.75$), the probability that the mother generation contains only inefficient catalysts is thus $P(n{=}24) \approx 0.1\%$. Even if this probability is low, we have already shown that a simulated evolution starting from such a mother generation is unsuccessful. We forced the GA to select G0 randomly, but solely from the zone of inactivity, that is, catalysts built with A_1, A_7 and A_{11} were excluded from G0. The GA parameters were RE = 2, RM = 3 and RR = 14. In addition, the GA could perform a maximum of 199 evaluations, which is close to 10% of the library. As shown in Figure 10, the performance of the GA was low: first generations were associated to small NPF and only two of top ten catalysts were found.

To incorporate more chemical diversity and thus to improve the performance of the evolution process, one could increase the number of catalysts in the mother generation, from 24 to 48 for instance. However, it would prevent the GA from meeting our requirement of 10% catalysts to be tested. One could also increase RE and RM, but we have shown that only 3 catalysts from top 10 were found within the 200 first experiments ($3 \approx 5.7$*(200 HTE/340 HTE), Table 4) which is a rather low improvement.

To deal with simulated evolutions that could start from a mother generation containing only inefficient catalysts, we developed a double algorithm (DA) which was evaluated by simulated evolution. The first step of this algorithm consisted in turning G0 into G1 using 14 mutations. Thus, the 10 best catalysts from G0 were transmitted to G1, and the other catalysts underwent a mutation, which introduced more chemical diversity in the mother generation. After this high-mutation-rate step, the probability that all catalysts from G1 came from the inefficient part of the library dramatically decreased to about 0.001%, that is, a very small probability. We then applied to G1 the "common" algorithm with RE = 2 and RM = 3 for the next 23 generations and for a total of 199 evaluations (including the 14 tests necessary for the G0–G1 high-mutation step). As depicted in Figure 11, the results of the GA were significantly improved: the mean NPF was increased, as well as the number of top

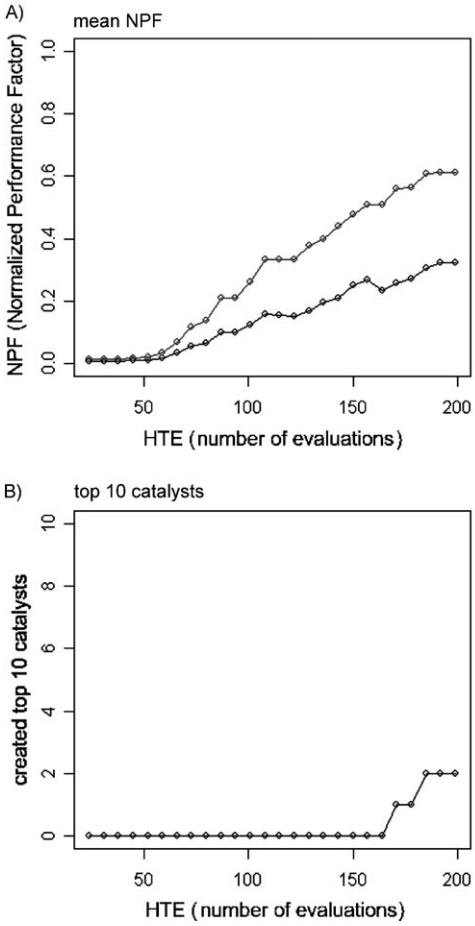


Figure 10. Simulated evolution starting from random G0 taken in the inactivity zone. GA parameters: R = 24; RE = 2; RM = 3; RR = 14; NI = 25.

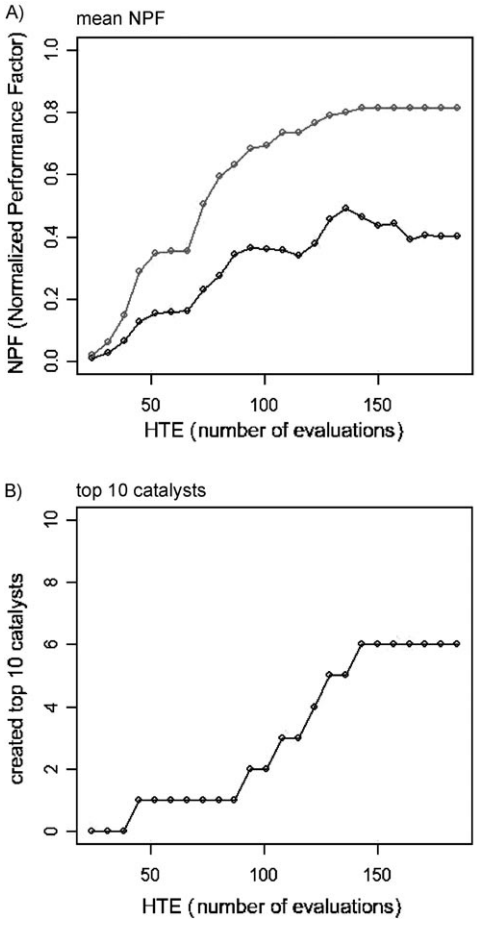


Figure 11. Simulated evolution starting from random G0 taken in the inactivity zone and using a double algorithm (DA). 1st GA parameters: R = 24; RE = 0; RM = 14; RR = 14; NI = 1. 2nd GA parameters: R = 24; RE = 2; RM = 3; RR = 14; NI = 23.

10 catalysts. The introduction of new catalysts by mutation was thus sufficient to ensure a successful evolution process.

The table 6 summarizes the results obtained with/without the double algorithm. When the high-mutation step was not applied to the mother generation randomly selected among the zone of inefficiency (entry 1, Table 6), the mean number of top 10 catalysts was low, and the MNT depicted a slow evolution process, whereas the associated σ_{MNT} reveals dissimilar behaviors throughout evaluations. When the double algorithm was applied to the same mother generation (entry 2, Table 6), the performance of the GA was quantitatively enhanced: on average more than five catalysts from top 10 were found, the process was almost three times faster, with a MNT of 36, roughly constant throughout evaluations, as demonstrated by the value of σ_{MNT}. Moreover, $A_1B_{19}M_2$ was found for 70% of the evolutions. Finally, we compared the results obtained from a double algorithm starting either from the zone of inefficient catalysts or from a G0 randomly selected amongst the whole library (entry 3, Table 6). Comparison of the data shows similar performances, which confirms the efficiency of the double algorithm technique.

Table 6. Comparison of sets of evolutions made without and with DA.

Entry	G0	DA	Mean top 10	MNT[b]	σ_{MNT}	$A_1B_{19}M_2$ [%]
1[a]	G0 from z.i.[c]	no	2.1	99	55	10
2[a]	G0 from z.i.[c]	yes	5.2	36	9	70
3[a]	random G0[d]	yes	6.5	40	14	90

[a] Results given for 10 simulated evolutions in every set. [b] The lower the MNT, the faster the evolution is. [c] G0 randomly selected among catalysts from the zone of inefficiency. [d] G0 randomly selected among the library.

Conclusion

In this study, we developed a genetic algorithm that mimics an evolution process. This algorithm was devoted to the creation of the best catalysts within a library designed for the asymmetric-hydrogen transfer to acetophenone. To both validate and optimize our algorithm, we considered a library of 1980 catalysts that were evaluated by HTE, and from which we found out the 10 best catalysts. Our genetic algorithm was able to find most of these best catalysts evaluating only 10% of the whole catalyst library.

The algorithm gave the best results when the mother generation was randomly chosen. The results of the simulated evolutions were found to be only slightly affected by variations of the algorithm parameters (size of the mother generation, mutation and cross-over rates, number of evolution loops…). Those results also show that the algorithm introduced enough chemical diversity during the evolution process. For a mother generation of 24 catalysts, however, the best results were obtained with crossover and mutation rates of 2 and 3 respectively. Using these values, the algorithm was typically able to find 3–5 catalysts of the top 10, the number of catalysts to be synthesized and evaluated ranging from 176 to 350.

In addition, we developed a double algorithm (DA) designed for dealing with libraries that contain a large number of inefficient catalysts. The use of this methodology enabled us to drive profitable simulation experiments, even with the worst mother generations. It ensured finding more than five catalysts from top ten, while keeping the threshold of a maximum of 10% evaluation of the library. On the other hand, the double algorithm has a predictive interest: if after one or two runs of high mutation step no candidates with a minimum activity have been found, the suitability of the library should be considered, as the probability of being in a local minimum has been sharply decreased. Therefore, the double algorithm can also be a decision tool.

Our findings show that simulated evolution is a fast and inexpensive optimization process. In this particular, a time reduction up to 70% could have been provided, compared to HTS/HTE-based catalyst design. Moreover, for evaluating small libraries, simulated evolution can be carried out using low throughput devices (e.g., Radley®) rather than much more expensive high throughput materials. Finally, the GA was built and run using a freeware software. Work is now currently in progress to use the new genetic algorithm to find efficient catalytic systems for new challenging catalytic enantioselective reactions.

Experimental Section

General procedure of the A_xB_y ligand synthesis: To a solution of aminoalcohol A_x (0.2 mmol, 1 equiv) and aldehyde (2 mmol, 1 equiv) in 2 mL of MeOH was added AcOH (0.012 mL, 0.2 mmol, 1 equiv). The reaction mixture was stirred and heated to reflux for 4 h. Resin supported BH_3CN (1.5 equiv) was then added and reflux continued for 15 h with occasional orbital stirring. Resin supported CO_3 (1.1 equiv) was finally added, and reflux continued for an additional 4 h. The reaction mixture was then filtrated and supported reagents washed with 3 mL of a hot solution of $MeOH/CH_2Cl_2$ (1:1). The solvents were evaporated under vacuum and the mixture analyzed by ^1H NMR.

General procedure of the $A_xB_yM_z$ catalyst evaluation: Into a vial flushed with Ar and capped with a Teflon system was injected 4×10^{-7} mol (0.01 equiv) of metal precursor M_z (diluted in isopropanol/cosolvent: 3:1) directly followed by injection of 4×10^{-7} mol (0.01 equiv) of the chiral ligand A_xB_y diluted in isopropanol. 30 minutes later, acetophenone (4×10^{-5} mol, 1 equiv) diluted in isopropanol was added to the preformed catalyst solution, directly followed by injection of 4×10^{-6} mol (0.1 equiv) of KOH in isopropanol. The reaction was allowed to progress for 90 minutes at room temperature. The addition of saturated NH_4Cl quenched the reaction, and Et_2O was added to the mixture. Chemical analysis of the reaction mixture was performed by gas chromatography.

Acknowledgements

The "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture" (FRIA) from Belgian government is gratefully acknowledged for their financial support of this research. The authors also want to thank the "R Foundation for Statistical Computing" for the free use of the R 2.6.1 software. Dr Pierre Desbiolles is also gratefully acknowledged for help with the corrections of the manuscript.

[1] For a review, see: a) H.-U. Blaser, B. Pugin, F. Spindler, *J. Mol. Catal. A* **2005**, *231*, 1–20; b) H.-U. Blaser, F. Spindler, M. Studer, *Appl. Catal. A* **2001**, *221*, 119–143. For selected exemples, see: c) O. Pàmies, J.-E. Bäckvall, *Chem. Rev.* **2003**, *103*, 3247–3261; d) R. Schmid, R. Verger, *Angew. Chem.* **1998**, *110*, 1694–1720; *Angew. Chem. Int. Ed.* **1998**, *37*, 1608–1633; e) K. Wünsche, U. Schwanenberg, U. Bornscheuer, H. Meyer, *Tetrahedron: Asymmetry* **1996**, *7*, 2017–2022; f) H.-U. Blaser, *Chem. Commun.* **2003**, 293–297; g) W. S. Knowles, *J. Chem. Educ.* **1986**, *63*, 222–225; h) I. Ojima, N. Clos, C. Bastos, *Tetrahedron* **1989**, *45*, 6901–6939; i) H.-U. Blaser, H. Buser, K. Coers, H. Hanreich, E. Jalett, B. Jelsch, B. Pugin, F. Schneider, F. Spindler, A. Wegmann, *Chimia* **1999**, *53*, 275–280; j) H. Cotton, T. Elebring, M. Larsson, L. Li, H. Sörensen, S. von Uge, *Tetrahedron: Asymmetry* **2000**, *11*, 3819–3825.

[2] a) T. Ohkuma, D. Ishii, H. Takeno, R. Noyori, *J. Am. Chem. Soc.* **2000**, *122*, 6510–6511; b) H. Ngo, W. Lin, *J. Org. Chem.* **2005**, *70*, 1177–1187; c) T. Hamada, T. Torii, K. Izawa, R. Noyori, T. Ikariya, *Org. Lett.* **2002**, *4*, 4373–4376; d) T. Ohkuma, M. Koizumi, K. Muniz, G. Hilt, C. Kabuto, R. Noyori, *J. Am. Chem. Soc.* **2002**, *124*, 6508–6509; e) T. Ohkuma, T. Hattori, H. Ooka, T. Inoue, R. Noyori, *Org. Lett.* **2004**, *6*, 2681–2683; f) T. Ohkuma, N. Utsumi, K. Tsutsumi, K. Murata, C. Sandoval, R. Noyori, *J. Am. Chem. Soc.* **2006**, *128*, 8724–8725; g) T. Ohkuma, K. Tsutsumi, N. Utsumi, N. Arai, R. Noyori, K. Murata, *Org. Lett.* **2007**, *9*, 255–257; h) A. Zanotti-Gerosa, W. Hems, M. Groarke, F. Hancock, *Platinum Met. Rev.* **2005**, *49*, 158–165.

[3] a) Y. Xu, G. Clarkson, G. Docherty, C. North, G. Woodward, M. Wills, *J. Org. Chem.* **2005**, *70*, 8079–8087; b) I. Sarvary, F. Almqvist, T. Fredj, *Chem. Eur. J.* **2001**, *7*, 2158–2166; c) J. Le, B. Pagenkopf, *J. Org. Chem.* **2004**, *69*, 4177–4180; d) H. Zhang, C. Yang, Z. Donga, J. Gao, H. Nakamura, K. Murata, T. Ikariya, *Chem. Commun.* **2003**, 142–143; e) H. Mimoun, J.-Y. de Saint Laumer, L. Gianinni, R. Scopelliti, C. Floriani, *J. Am. Chem. Soc.* **1999**, *121*, 6158–6166; f) D. Matharu, D. Morris, A. Kawamoto, G. Clarkson, M. Wills, *Org. Lett.* **2005**, *7*, 5489–5491; g) Y. Kanazawa, Y. Tsuchiya, K. Kobayashi, T. Shiomi, J. Itoh, M. Kikuchi, Y. Yamamoto, H. Nishiyama, *Chem. Eur. J.* **2006**, *12*, 63–71; h) A. Patti, S. Pedotti, *Tetrahedron: Asymmetry* **2003**, *14*, 597–602; i) T. Fang, J. Xu, D. Du, *Synlett* **2006**, 1559–1563; j) T. Ohkuma, H. Ooka, S. Hashiguchi, T. Ikariya, R. Noyori, *J. Am. Chem. Soc.* **1995**, *117*, 2675–2676.

[4] For principles of combinatorial chemistry and some applications in catalysis, see: a) J. de Vries, L. Lefort, *Chem. Eur. J.* **2006**, *12*, 4722–4734; b) C. Gennari, U. Piarulli, *Chem. Rev.* **2003**, *103*, 3071–3100; c) K. Ding, H. Du, J. Long, *Chem. Eur. J.* **2004**, *10*, 2872–2884; d) O. Lavastre, F. Bonnette, L. Gallard, *Curr. Opin. Chem. Biol.* **2004**, *8*, 311–318; e) K. Burgess, H. Lim, A. Porte, G. Sulikowski, *Angew. Chem.* **1996**, *108*, 192–194; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 220–222; f) F. Taran, C. Gauchet, B. Mohar, S. Meunier, A. Valleix, P. Renard, C. Créminon, J. Grassi, A. Wagner, C. Miokowski, *Angew. Chem.* **2002**, *114*, 132–135; *Angew. Chem. Int. Ed.* **2002**, *41*, 124–127. For examples of lead compounds discovered from MCR's, see: g) A. Golebiowski, S. Klopfenstein, D. Portlock, *Curr. Opin. Chem. Biol.* **2001**, *5*, 273; h) A. Dömling, *Chem. Rev.* **2006**, *106*, 17–89.

[5] For examples in activity or property prediction, see: a) G. Schneider, M. Nettekoven, *J. Comb. Chem.* **2003**, *5*, 233–237; b) A. Ragusa, J. Hayes, M. Light, J. Kilburn, *Eur. J. Org. Chem.* **2006**, 3545–3549; c) M. Reetz, G. Haderlein, K. Angermund, *J. Am. Chem. Soc.* **2000**, *122*, 996–997; d) J. Hageman, J. Westerhuis, H.-W. Frühauf, G. Rothenberg, *Adv. Synth. Catal.* **2006**, *348*, 361–369; e) J. Torrecilla, F. Rodriguez, J. Bravo, G. Rothenberg, K. Seddon, I. Lopez-Martin, *Phys. Chem. Chem. Phys.* **2008**, *10*, 5826–5831. For examples in enantiomeric excess prediction, see: f) M. Hoogenraad, G. Klaus, N. Elders, S. Hooijschuur, B. McKay, A. Smith, E. Damen, *Tetrahedron: Asymmetry* **2004**, *15*, 519–523; g) J. Aires-de-Sousa, J. Gasteiger, *J. Comb. Chem.* **2005**, *7*, 298–301; h) J. Ianni, V. Annamalai, P. Phuan, M. Panda, M. Kozlowski, *Angew. Chem.* **2006**, *118*, 5628–5631; *Angew. Chem. Int. Ed.* **2006**, *45*, 5502–5505; i) M. Kozlowski, S. Dixon, M. Panda, G. Lauri, *J. Am. Chem. Soc.* **2003**, *125*, 6614–6615; j) N. Wilmot, M. Marsella, *Org. Lett.* **2006**, *8*, 3109–3112; k) A. Taggi, A. Hafez, T. Dudding, T. Lectka, *Tetrahedron* **2002**, *58*, 8351–8356; l) D. Gleich, R. Schmid, W. Herrmann, *Organometallics* **1998**, *17*, 2141–2143; m) G. Alagona, C. Ghio, *J. Organomet. Chem.* **2005**, *690*, 2339–2350; n) S. Bahmanyar, K. Houk, H. Martin, B. List, *J. Am. Chem. Soc.* **2003**, *125*, 2475–2479; o) P. Ha-Yeon, K. Houk, *Synthesis* **2005**, 1533–1537; p) M. Garcia-Garibay, K. Houk, A. Keating, C. Cheer, M. Leibovitch, J. Scheffer, L. Wu, *Org. Lett.* **1999**, *1*, 1279–1281; q) S. Tomic, B. Bertosa, B. Kojic-Prodic, I. Kolosvary, *Tetrahedron: Asymmetry* **2004**, *15*, 1163–1172; r) I. Gridnev, T. Inamoto, *Acc. Chem. Res.* **2004**, *37*, 633–644.

[6] Z. Gartner, *Pure Appl. Chem.* **2006**, *78*, 1–14.

[7] a) L. Weber, *Drug Discovery Today* **2002**, *7*, 143–147; b) C. Fernando, J. Rowe, *J. Theor. Biol.* **2007**, *247*, 152–167.

[8] a) M. Reetz, *Tetrahedron* **2002**, *58*, 6595–6602; b) M. Reetz, L. Wang, M. Bocola, *Angew. Chem.* **2006**, *118*, 1258–1263; *Angew. Chem. Int. Ed.* **2006**, *45*, 1236–1241; c) F. El Oualid, H. van der Elst, I. Leroy, E. Pieterman, L. Cohen, B. Burm, H. Overkleeft, G. van der Marel, M. Overhand, *J. Comb. Chem.* **2005**, *7*, 703–713; d) C. Gerlach, M. Münzel, B. Baum, H.-D. Gerber, T. Craan, W. Diederich, G. Klebe, *Angew. Chem.* **2007**, *119*, 9265–9269; *Angew. Chem. Int. Ed.* **2007**, *46*, 9105–9109.

[9] a) K. Huang, C. Chen, D. Lü, *App. Cat. A* **2001**, *219*, 61–68; b) K. Huang, X. Zhan, F. Chen, D. Lü, *Chem. Eng. Sci.* **2003**, *58*, 81–87; c) W. Bannwarth, E. Felder, *Combinatorial Chemistry* **2000**, Wiley-VCH, Weinheim; d) A. Leach, R. Bryce, A. Robinson, *J. Mol. Graph. Modell.* **2000**, *18*, 358–367; e) S. Shi, Z. Peng, J. Kostrowicki, G. Paderes, A. Kuki, *J. Mol. Graphics Modell.* **2000**, *18*, 478–496; f) P. Labute, *J. Mol. Graphics Modell.* **2000**, *18*, 464–477; g) C. Klanner, D. Farusseng, L. Baumes, M. Lengliz, C. Mirodatos, F. Schüth, *Angew. Chem.* **2004**, *116*, 5461–5463; *Angew. Chem. Int. Ed.* **2004**, *43*, 5347–5349; h) M. Karplus, S. So, *J. Med. Chem.* **1996**, *39*, 1521–1530; i) A. Tropsha, W. Zheng, *Comb. Chem. High Throughput Screening* **2002**, *5*, 111–123; j) A. Ruskinko III, S. Young, D. Drewry, S. Gerritz, *Comb. Chem. High Throughput Screening* **2002**, *5*, 125–133; k) D. Drewry, S. Gerritz, J. Linn, *Tetrahedron Lett.* **1997**, *38*, 3377–3380; l) R. Fox, A. Roy, S. Govindarajan, J. Minshull, C. Gustafsson, J. Jones, R. Emig, *Protein Eng.* **2003**, *16*, 589–597; m) G. Grubert, S. Kolf, M. Baerns, I. Vauthey, D. Farrusseng, A. van Veen, C. Mirodatos, E. Stobbe, P. Cobden, *Appl. Catal. A* **2006**, *306*, 17–21; n) Y. Watanabe, T. Umegaki, M. Hashimoto, K. Omata, M. Yamada, *Catal. Today* **2004**, *89*, 455–464; o) Y. Yokobayashi, K. Ikebukuro, S. McNiven, I. Karube, *J. Chem. Soc. Perkin Trans. 1* **1996**, 2435–2437; p) L. Weber, S. Wallbaum, C. Broger, K. Gubernator, *Angew. Chem.* **1995**, *107*, 2452–2454; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2280–2282; q) K. Illgen, T. Enderle, C. Broger, L. Weber, *Chem. Biol.* **2000**, *7*, 433–441; r) L. Weber, *Curr. Opin. Chem. Biol.* **1998**, *2*, 381–385; s) J. Singh, M. Ator, E. Jaeger, M. Allen, D. Whipple, J. Soloweij, S. Chowdhary, A. Treasurywala, *J. Am. Chem. Soc.* **1996**, *118*, 1669–1676; t) N. Budin, S. Ahmed, N. Majeux, A. Caflish, *Comb. Chem. High Throughput Screening* **2001**, *4*, 661–673; u) R. Sheridan, S. SanFeliciano, S. Kearsley, *J. Mol. Graphics Modell.* **2000**, *18*, 320–334; v) W. Zhang, K. Yano, I. Karube, *BioSystems* **2007**, *88*, 35–55; w) J. Beckers, F. Clerc, J. Blank, G. Rothenberg, *Adv. Synth. Catal.* **2008**, *350*, 2237–2249.

[10] For the details of the mechanism, see: a) M. Palmer, M. Wills, *Tetrahedron: Asymmetry* **1999**, *10*, 2045–2061; b) J.-E. Bäckvall, R. Chowdhury, *J. Chem. Soc. Chem. Commun.* **1991**, 1063–1064; c) A. Pfaltz, *Chimia* **2004**, *58*, 49–50; d) D. Sterk, M. Stephan, B. Mohar, *Tetrahedron Lett.* **2004**, *45*, 535–537; e) J. Cossy, F. Eustache, P. Dalko, *Tetrahedron Lett.* **2001**, *42*, 5005–5007; f) P. Peach, D. Cross, J. Kenny, I. Mann, I. Houson, L. Campbell, T. Walsgrove, M. Wills, *Tetrahedron* **2006**, *62*, 1864–1876; g) T. Ohkuma, H. Ooka, T. Ikariya, R. Noyori, *J. Am. Chem. Soc.* **1995**, *117*, 10417–10418; h) T. Ohkuma, H. Doucet, T. Pham, K. Mikami, T. Korenaga, M. Terada, R. Noyori, *J. Am. Chem. Soc.* **1998**, *120*, 1086–1087; i) M. Yamakawa, H. Ito, R. Noyori, *J. Am. Chem. Soc.* **2000**, *122*, 1466–1478; j) J.-E. Bäckvall, *J. Organomet. Chem.* **2002**, *652*, 105–111; k) J. Samec, J.-E. Bäckvall, P. Andersson, P. Brandt, *Chem. Soc. Rev.* **2006**, *35*, 237–248; l) R. Wisman, J. de Vries, B. Deelman, H. Heeres, *Org. Process Res. Dev.* **2006**, *10*, 423–429; m) J. Johnson, J.-E. Bäckvall, *J. Org. Chem.* **2003**, *68*, 7681–7684; n) C. de Bellefon, N. Tanchoux, *Tetrahedron: Asymmetry* **1998**, *9*, 3677–3686; o) O. Pàmies, J.-E. Bäckvall, *Chem. Eur. J.* **2001**, *7*, 5052–5058; p) K. Haack, S. Hashiguchi, A. Fujii, T. Ikariya, R. Noyori, *Angew. Chem.* **1997**, *109*, 297–300; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 285–288; q) T. Ikariya, K. Murata, R. Noyori, *Org. Biomol. Chem.* **2006**, *4*, 393–406; r) R. Chowdhury, J.-E. Bäckvall, *J. Chem. Soc. Chem. Commun.* **1991**, 1063–1064; s) A. Yim, M. Wills, *Tetrahedron* **2005**, *61*, 7994–8004; t) R. Noyori, M. Yamakawa, S. Hashiguchi, *J. Org. Chem.* **2001**, *66*, 7931–7944; u) M. Yamakawa, I. Yamada, R. Noyori, *Angew. Chem.* **2001**, *113*, 2900–2903; *Angew. Chem. Int. Ed.* **2001**, *40*, 2818–2821; v) C. Sandoval, T. Ohkuma, N. Utsumi, K. Tsutsumi, K. Murata, R. Noyori, *Chem. Asian J.* **2006**, *1*, 102–110. For estimation of transition states′ energies, see R. Noyori, M. Yamakawa, S. Hashiguchi, *J. Org. Chem.* **2001**, *66*, 7931–7944.

[11] a) K. Everaere, J.-F. Carpentier, A. Mortreux, M. Bulliard, *Tetrahedron: Asymmetry* **1999**, *10*, 4083–4086; b) K. Everaere, A. Mortreux, M. Bulliard, J. Brussee, A. van der Gen, G. Nowogrovki, J.-F. Carpentier, *Eur. J. Org. Chem.* **2001**, 275–291; c) M. Palmer, T. Walsgrove, M. Wills, *J. Org. Chem.* **1997**, *62*, 5226–5228; d) D. Petra, J. Reek, J. Handgraaf, E. Meijer, P. Dierkes, P. Kamer, J. Brussee, H. Schoemaker, P. van Leeuwen, *Chem. Eur. J.* **2000**, *6*, 2818–2829; e) I. Schiffers, T. Rantanen, F. Schmidt, W. Bergmans, L. Zani, C. Bolm, *J. Org. Chem.* **2006**, *71*, 2320–2331.

[12] a) J. Kanth, M. Periasamy, *J. Org. Chem.* **1991**, *56*, 5964–5965; b) M. McKennon, A. Meyers, *J. Org. Chem.* **1993**, *58*, 3568–3571.

[13] M. Bennett, A. Smith, *J. Chem. Soc. Dalton* **1974**, 233.

[14] I. P. Evans, A. Spencer, G. Wilkinson, *J. Chem. Soc. Dalton Trans.* **1973**, 204–209.

[15] B. Govaerts, C. Le Bailly de Tilleghem, *Discussion Paper* **2005**, *0532*, 1–14.

[16] http://www.r-project.org

[17] For details of the canonic GA used for our simulated evolution experiments, see supporting information.

Received: October 23, 2008
Published online: May 12, 2009

