

Backbone Diversity Analysis in Catalyst Design

Ana G. Maldonado,^a Jos A. Hageman,^b Sergio Mastroianni,^c
and Gadi Rothenberg^{a,*}

^a Van't Hoff Institute of Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
Fax: (+31)-(0)20-525-5604; phone: (+31)-(0)20-525-6963
g.rothenberg@uva.nl

^b Biometris, Wageningen University & Research Centre, Bornsesteeg 47, 6708 PD Wageningen, The Netherlands

^c Rhodia CRL, 85 Av Frères Perret, 69190 St Fons, France

Received: September 17, 2008; Revised: December 18, 2008; Published online: February 4, 2009

Abstract: We present a computer-based heuristic framework for designing libraries of homogeneous catalysts. In this approach, a set of given bidentate ligand-metal complexes is disassembled into key substructures (“building blocks”). These include metal atoms, ligating groups, backbone groups, and residue groups. The computer then rearranges these building blocks into a new library of virtual catalysts. We then tackle the practical problem of choosing a diverse subset of catalysts from this library for actual synthesis and testing. This is not trivial, since ‘catalyst diversity’ itself is a vague concept. Thus, we first define and quantify this diversity as the difference between key structural parameters (descriptors) of the catalysts, for the specific reaction at hand. Subsequently, we propose a method for choosing diverse sets of

catalysts based on catalyst backbone selection, using weighted D-optimal design. The computer selects catalysts with different backbones, where the difference is measured as a distance in the descriptors space. We show that choosing such a D-optimal subset of backbones gives more diversity than a simple random sampling. The results are demonstrated experimentally in the nickel-catalysed hydrocyanation of 3-pentenitrile to adiponitrile. Finally, the connection between backbone diversity and catalyst diversity, and the implications towards *in silico* catalysis design are discussed.

Keywords: catalysis descriptors; combinatorial chemistry; homogeneous catalysis; ligand design; structure-activity relationship

Introduction

Sustainability and sustainable development are high on our agenda in the 21st century. The tools of green chemistry and catalysis play a key role here, as they can provide mankind with bona fide solutions to environmental problems.^[1,2] Importantly, chemists now also have suitable experimental tools for approaching these problems.^[3,4] In homogeneous catalysis, the combination of laboratory automation,^[5-7] advanced modelling and data mining algorithms puts us on the brink of *in silico* catalyst design.^[8] Much progress has been made since the 1990s in quantitative structure/activity relationships (QSAR) and quantitative structure/property relationship (QSPR) models.^[9,10] To realise the goal of true *in silico* catalyst design, however, we must be able to assemble and screen virtual libraries of ligands and ligand-metal complexes (Figure 1). This new mode of operation requires also a new mode of thinking, especially where large libraries are concerned. Here, catalyst selection is the major prob-

lem. Generating millions of structures *via* computer is easy, but how should one choose the candidates for synthesis and testing? Moreover, how should one select the right components for assembling these virtual libraries?

The pharmaceutical industry has solved an analogous problem in the field of drug design. They used topological modelling, as well as selection and classification methods for screening large virtual libraries of drug candidates.^[11] These methods aim to identify groups of features that embody drug efficiency. Such two-dimensional descriptor models are less common in catalysis, where mechanistic studies rely primarily on quantum mechanical (QM) calculations.^[12] The reason for this is simple. You get what you pay for, and topological modelling is computationally cheap. Thus, the results of QSAR/QSPR models based on topological descriptors are less accurate than those based on QM ones. Nevertheless, unlike QM calculations, topological descriptors give a realistic approach for modelling large catalyst libraries. Recently, we

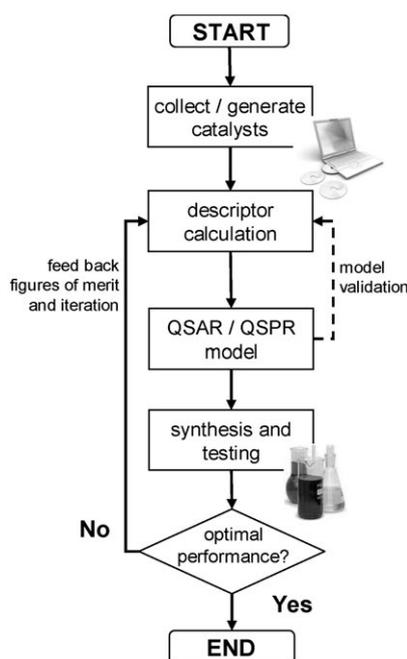


Figure 1. Flow chart showing the screening of virtual libraries and subsequent iterative modelling and laboratory validation. In each iteration, promising candidates are synthesised and tested, and results are fed back to the model. Note that the sets of catalysts that are actually synthesised are much smaller than the virtual library. The primary challenge, therefore, is choosing the right candidates for building the library and validating the model.

showed that by combining chemical principles, topological descriptors, and statistical analysis, one can predict important parameters in such libraries.^[13] Here we apply the same principles to the complex problem of ligand library diversity analysis.

Practically speaking, a library of catalysts should be as diverse as possible, while still maintaining a realistic size.^[14] Unfortunately, although the term “diversity” is widely used in homogeneous catalysis, it is ill-defined. Everyone agrees that a diverse library of ligands is ‘one that gives a good spread of results’. In practice, however, diversity is problem-related, since different chemical reactions are often sensitive to different parameters (see below for details). When dealing with large libraries (thousands of candidates and more), one cannot estimate the diversity by simply looking at the structures. One way to solve this problem is by defining diversity indices, that can be calculated automatically.^[15] This requires descriptors that quantify the ligands’ structure-activity relationships. These descriptors should be chemically relevant, and simple and easy to calculate.^[13,16]

Organometallic complexes with bidentate ligands are well established as catalysts in academia and industry.^[17,18] Thanks to the chelate effect, they are more stable and robust than monodentate complexes.

Moreover, the bidentate structure offers ample opportunities for fine-tuning and optimising a variety of stereoelectronic effects.^[19,20] Some notable examples include Noyori’s BINAP ligands,^[21] the Josiphos ligands used in the Ciba-Geigy metolachlor process,^[22] Jacobsen’s catalyst,^[23,24] and the Keim complexes used in the Shell higher olefins process (SHOP).^[25] Although there are several important studies on bidentate ligand descriptors,^[26,27] there is only scant information regarding the problem of ligand diversity for library assembly and catalyst selection.^[15] Here, we examine measures of diversity for bidentate ligands, with a specific emphasis on the ligand backbone. As an example, we study the nickel-catalysed hydrocyanation of 3-pentenitrile to adiponitrile. We also discuss how these measures can help chemists take better decisions when applying *in silico* catalyst design.

Results and Discussion

Relating Backbone Diversity and the Catalyst Space

One can envisage diversity in homogeneous catalysis as a spread of ligand-metal complexes over a given catalyst space, or as a lack of similarity within a set of catalysts. Saying that “library **I** is more diverse than library **II**” can mean that library **I** covers a larger volume of the catalyst space. Alternatively, it can mean that the catalysts in **I** are better spread in the same space. To illustrate this, we divide the system in three spaces (Figure 2). The first space, **A**, is the catalyst space. It is a grid containing a large number of metal-ligand complexes. Each point in space **A** denotes one catalyst that has been tested experimentally. Space **B**, or descriptor space, contains the values of the catalysts’ descriptors (internal parameters such as backbone flexibility, partial charge on the metal atom, polarity or lipophilicity) as well as the reaction conditions (external parameters such as temperature, pres-

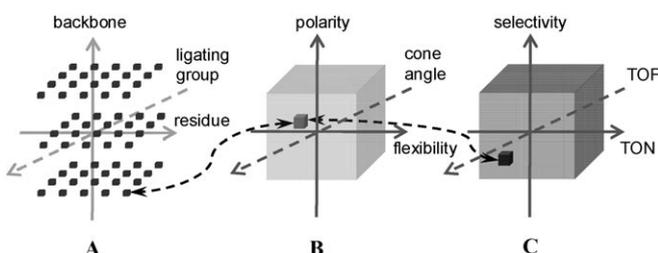


Figure 2. Simplified representation of the multi-dimensional spaces **A**, **B**, and **C**, containing the catalysts, the molecular descriptor values, and the figures of merit, respectively. For clarity, we show only three axes per space. In reality, the spaces are multi-dimensional, with each dimension representing a structural feature, a descriptor or a figure of merit, respectively.

sure, and solvent type). All of these parameters may influence the reaction outcome, and a model may include both internal and external descriptors. Finally, **C** is the space of the figures of merit (FOM, e.g., the TON, TOF, product selectivity, price, and so forth). Note that while space **A** is a discrete grid, **B** and **C** are continuous, and are arranged such that each dimension represents one property. Essentially, spaces **B** and **C** are connected *via* structure/activity and structure/property relationships.

Traditionally, chemists define diversity in space **A**, by comparing molecular structures and especially functional groups. However, relevant differences between catalysts are much better represented by distances in space **B**, the descriptor space. Thus, instead of evaluating molecules based on their 'Chemdraw® structures', we should compare them based on their descriptor values. Choosing the right descriptors is a crucial step, because diversity is problem-related. One reaction can be controlled by the bite angle, while for another, the reaction temperature and pressure may be the key parameters (broader discussions on descriptor types^[28] and descriptor selection^[27,29] are published elsewhere). Here, we focus on the backbone diversity of bidentate ligands. The reason for this is that the ligand backbone often dictates the final size, flexibility, and electronic properties of the complex. These are directly related to the reaction pocket environment and the catalytic performance.^[27] Moreover, the concept of derivative synthesis is based on a common backbone, and indeed different ligand types are often referred to by their backbone designation (e.g., Salen or BINAP catalysts). Importantly, ligand libraries are often prepared in practice *via* derivative synthesis, i.e., by assembling various ligating and residue groups onto a small number of scaffolds.^[30] This means that backbone diversity holds the key to catalyst diversity.

Case Study: Ni-Catalysed Hydrocyanation

With the above set-up in mind, we turned to a real-life example of a homogeneously catalysed reaction: the nickel-catalysed hydrocyanation of 1,3-butadiene to adiponitrile (Figure 3).^[31] This important industrial process forms the basis of Nylon 6.6 and Nylon 6 production. Despite the fact that this is a well known process, catalyst optimisation remains a challenge.^[32,33] We focus here on the second step, the hydrocyanation of 3-pentenenitrile to adiponitrile (Figure 3, *bottom*).

For our initial library (space **A**), we selected 115 biphosphite and biphosphine ligand-nickel complexes from over 200 articles and patents.^[34] Each catalyst was divided into two ligating groups, a backbone, and up to four residue groups. This division gives a standard framework for backbone comparison. Figure 4 shows an example of such catalyst structures, denoting

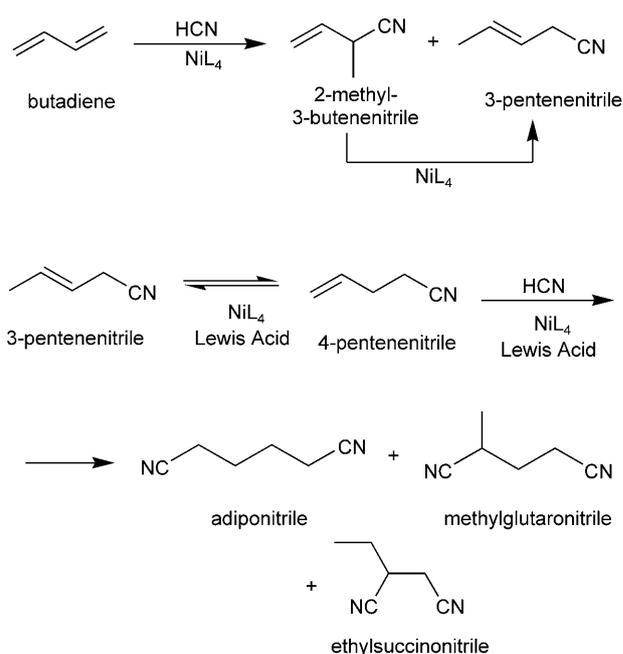


Figure 3. The nickel-catalyzed hydrocyanation of butadiene to 3-pentenenitrile (*top*) and the subsequent hydrocyanation of 3-pentenenitrile to adiponitrile (*bottom*).

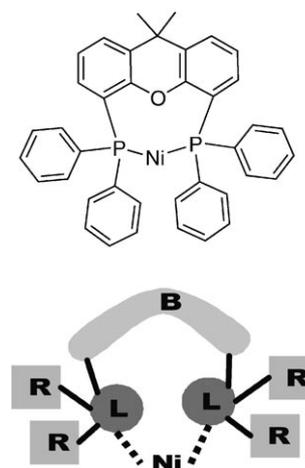


Figure 4. Example of a biphosphine-nickel complex, one of the initial set of 115 catalysts (*top*); cartoon showing the formal division into ligating (**L**), backbone (**B**), and residue groups (**R**, *bottom*).

the different building blocks.^[8] In total, our dataset of 115 ligands reduces to 25 ligating groups, 60 residue groups and 42 backbones. The backbone structures are shown in Figure 5. Even with such small numbers, assessing the backbone diversity by simply looking at the structures is difficult. The task becomes impossible when one moves from backbones to actual catalysts, since the 115 catalysts by no means cover the entire combination space. The entire catalyst space contains over 10^{60} compounds.^[35] Even the number of

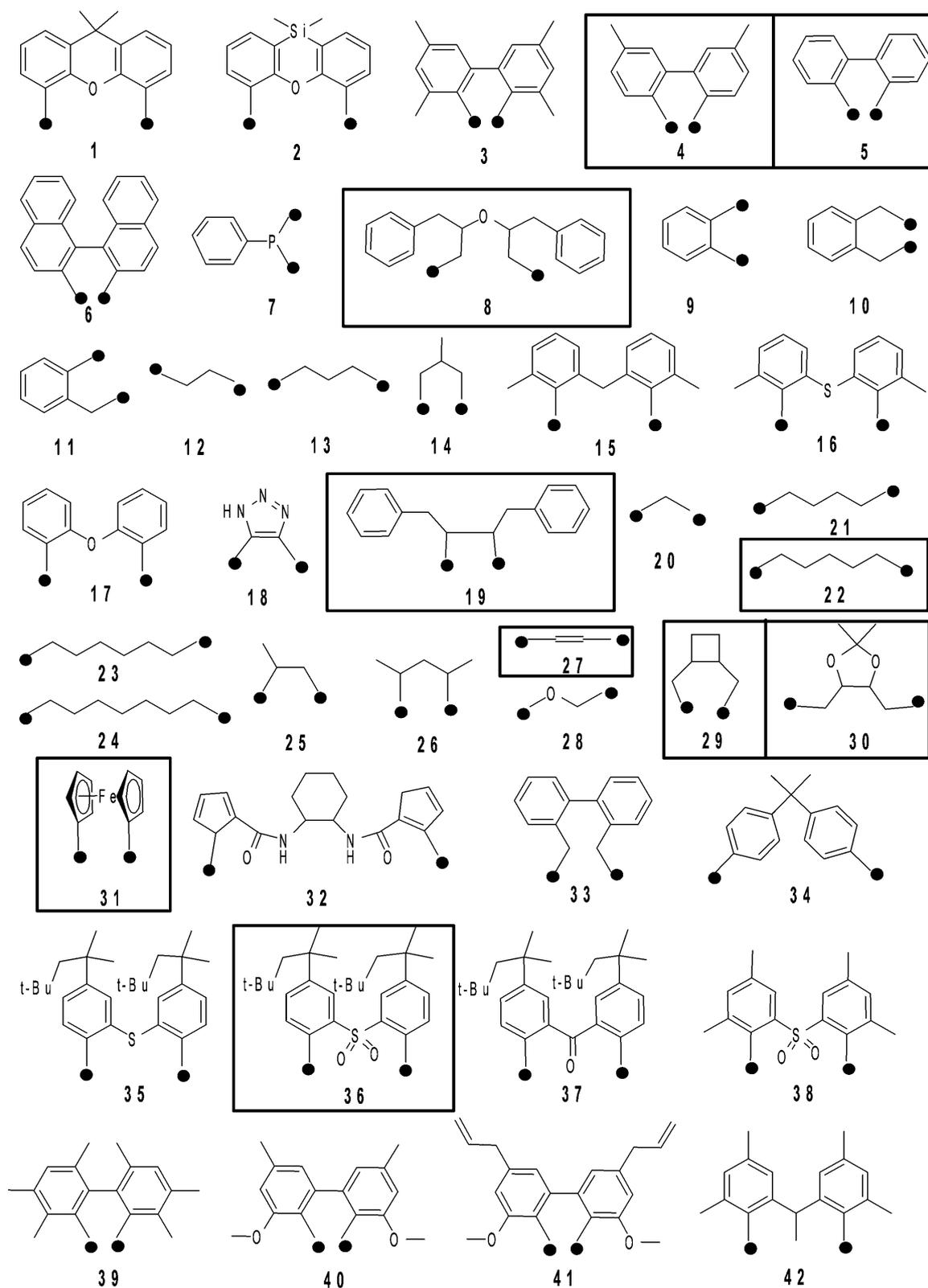
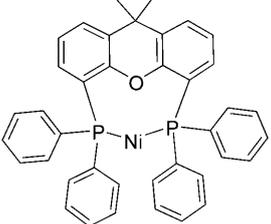
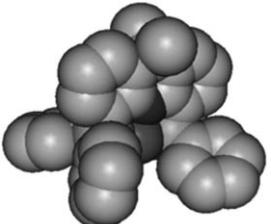


Figure 5. Structures of the 42 backbones used in this study. The backbones belonging to the D-10 subset (*vide infra*) are framed.

Table 1. Examples of typical molecular descriptors.

Descriptor type	Typical representation	Examples
1D	$C_{39}H_{32}NiOP_2$	Molecular weight; atom counts
2D		Fragment counts; topological indices; connectivity
3D		Molecular surface; molecular volume; interaction energies

building block permutations is an overwhelming 3.4×10^9 . Such numbers of catalysts are beyond computational treatment, let alone synthesis. Thus, we limit our diversity analysis to the backbones only. Our objective is to see whether this analysis can help us select a representative subset of backbones, which would then be used for synthesising a representative test set of catalysts.

For all the backbones, we calculate a set of selected descriptors. Choosing relevant descriptors is a challenge.^[16,36–38] Descriptors can be determined either from the molecular structure, or the physical properties.^[14] Table 1 shows some typical examples of molecular descriptors for the complex shown in Figure 4. Depending on cost limits and the problem at hand, one can use 2D ('quick and dirty') descriptors and/or 3D (more accurate, but also more costly) descriptors.

Drug design companies use different strategies for guiding descriptor selection. One is selecting a particular set of descriptors that show good performance on a given problem.^[11,39] Alternatively, you can first calculate a large number of descriptors and then remove any correlated ones.^[40,41] Another different approach is letting the computer choose the optimal combination of descriptors by testing all the possible combinations. In our previous studies of bidentate ligand-nickel complexes, we showed that the most important topological descriptors were related to the bite angle and flexibility,^[13] while the most relevant electronic descriptors were related to charge densities.^[27] Here, we calculated a total of 168 topological, constitutional, geometric and electrostatic descriptors for all the 42 backbones. This set was then ranked down to three

main attributes: size, flexibility, and polarity. This gives a matrix, where each row corresponds to a backbone and each column to a descriptor. Plotting these attributes gives a 3D backbone "map". Such maps allow us to compare the diversity of backbone sets, and ultimately of ligands and catalysts libraries.

Figure 6 shows the three-dimensional "map" for the 42 backbones of the bidentate ligand-Ni complexes. The two-dimensional projections on the different descriptor axes are shown in Figure 7. These maps tell us how "different" the backbones are from each other. If two dots are far away from each other (e.g., **19** and **20**), the backbones are different. Conversely, if the dots are close to each other (e.g., **35** and **36**), the backbones are similar. Looking at Figure 6, we see four main clusters. Some of these are easily explained. Backbones **35**, **36** and **37**, all contain a very bulky substituted *para*-benzene moiety. Ferrocene, (backbone **31**) due to its particular electronic and steric properties, is an outlier. The two large clusters are clearly influenced by the size parameter, as shown in Figure 7 (highlighted in gray in the middle and bottom graphs).

To demonstrate the experimental validity of our hypothesis, we chose three catalysts, derived from the backbones shown in Figure 5. Two figures of merit were measured experimentally for each catalyst: 3-pentenitrile conversion and product selectivity (see the Experimental Section for details). The three ligand-Ni complexes (structures **43–45** in Figure 8) are based on backbones **15**, **6**, and **31**, respectively.

Our hypothesis states that if the catalyst backbones are close to each other in the descriptor space, **B**, then these catalysts will show similar figures of merit. Comparing the FOM values in Figure 8, with the three-dimensional descriptor plot in Figure 6 we see that **43** and **44**, belong to the same backbone cluster. The structures may differ, but their properties are similar. Happily, both ligands indeed give similar selectivity for adiponitrile (73.0% and 73.5%, respectively). The conversion rates are less similar, but still within 10%. Conversely, complex **45** is based on a very different backbone, when comparing descriptor values, and gives very different figures of merit.

Choosing the "Right" Backbones for Diversity

All the possible combinations of a set of backbones, ligands and residues represent on one hand a small part of the whole chemical space, and on the other hand an impossible large set for synthesis and testing. One must choose a small catalysts subset, and this is precisely where the usefulness of our analysis comes in. Analysing the graphs shown in Figure 6 and Figure 7, we can select a representative subset of backbones, and from these generate a diverse set of

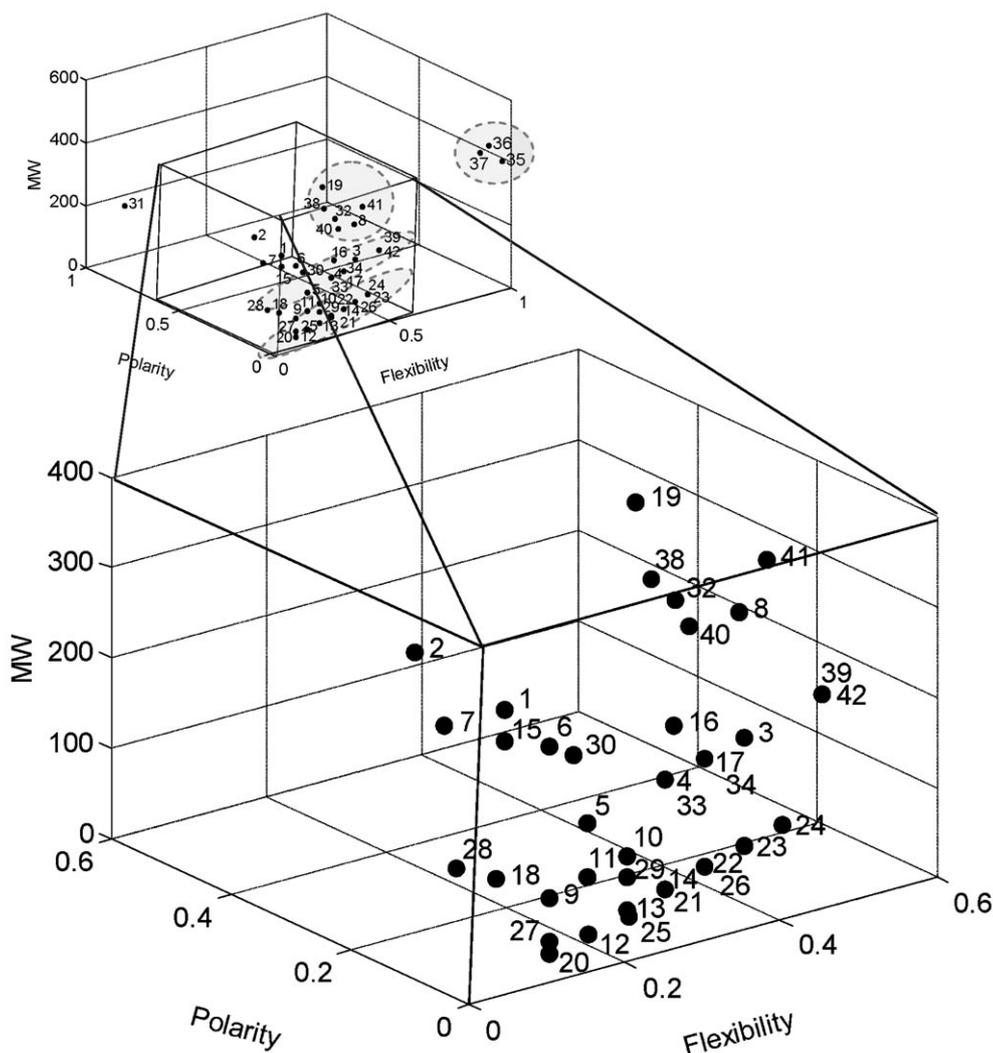


Figure 6. Three-dimensional descriptor space plot of the full set of 42 backbone structures (*top*), highlighting the four main clusters. The zoom-in view shows the distribution of the structures in the main cluster (*bottom*).

new ligands for synthesis purposes. This can be done, for example, using D-optimal design^[42,43]. We partition the space into n equal partitions, choosing from each partition a number of backbones that depends on the backbone density in that partition. Such a weighted selection gives each partition the correct influence on the results (a good analogy is the US presidential elections, where each state has a number of electoral votes that is proportional to the number of its inhabitants).

Such a selection gives an optimal coverage of the catalyst space for a given number of experiments. There is less chance of missing a good candidate, compared with a random sampling of the catalyst space. Importantly, the distances in the descriptor space reflect the backbones' similarity. This means that if synthesising the combination $\{L_1(R_1, R_2)-B-L_2-(R_3, R_4)\}$ is difficult, you can use an alternative backbone, B' , providing that the distance BB' in the de-

scriptor space is short. For example, if synthesising a complex with backbone **36** is difficult, we could probably use backbone **35** instead (see Figure 6).

Exploring Diversity

To analyse the backbone distribution in the descriptor space, we used the average distance of each backbone to all other backbones.^[15] This is the mean intermolecular distance, $\delta_{average}$, given by Eq. (1) (where d_{ij} is the

$$\delta_{average} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{N-1} \sum_{j=1}^N d_{ij} \right) \quad (1)$$

distance between points i and j , and N is the number of backbones). If $\delta_{average}$ is large, the backbones in the

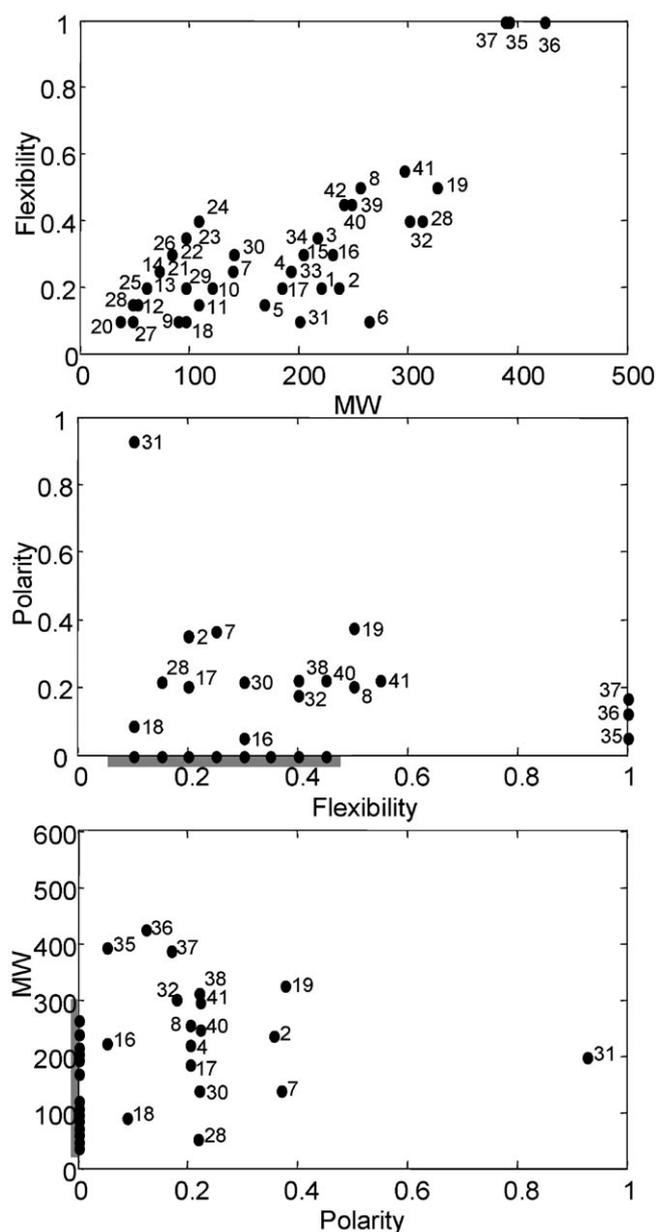


Figure 7. Two-dimensional projections of the 3D plot shown in Figure 6, on the MW/flexibility (*top*), flexibility/polarity (*medium*) and polarity/MW (*bottom*) axes. The shaded areas represent the main backbone cluster {1, 3, 5, 6, 9, 10, 11, 12, 13, 14, 15, 20, 21, 22, 23, 24, 25, 26, 27, 29, 33, 34, 39, 42}.

library are further apart. This means that the library covers a larger volume of space **B**.

However, $\delta_{average}$ does not tell us *how* the backbones are distributed. For this, we must also calculate the average distance from each backbone to its nearest neighbour. This parameter, δ_{min} , is defined by Eq. (2)

$$\delta_{min} = \frac{1}{N} \sum_{i=1}^N d_{min,i} \quad (2)$$

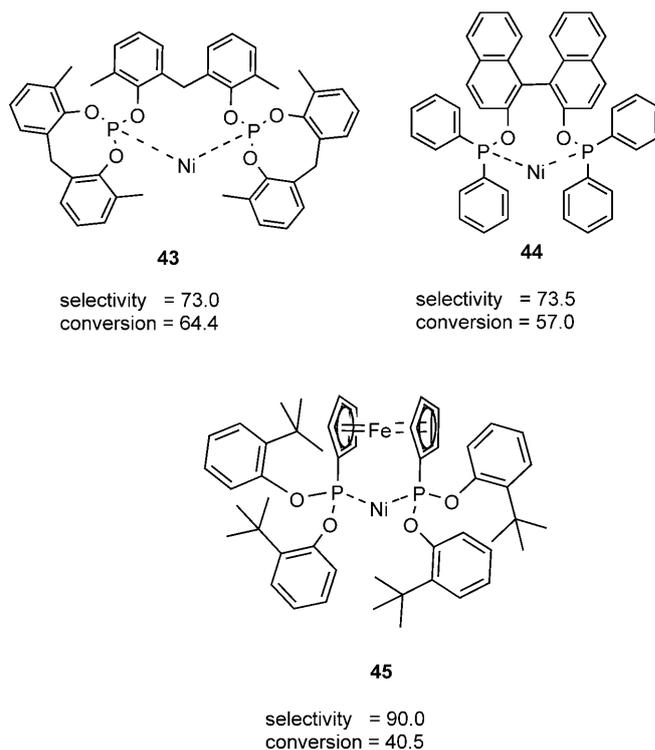


Figure 8. Catalyst structures used in the experimental demonstration, with their corresponding figures of merit.

(where $d_{min,i}$ is the distance between point i and its nearest neighbour). δ_{min} gives an idea of the clustering of the backbones in space **B**. If δ_{min} is small, each backbone has at least one other backbone close by. Conversely, if δ_{min} is large, the backbones are on average far apart.

A diverse library should be large, with the backbones distributed well over the space. Thus, the product $\delta_{average} \times \delta_{min}$ is a good measure of library diversity. Detailed discussions of $\delta_{average}$ and δ_{min} , complete with worked-out examples, are published elsewhere.^[1,15]

Table 2 compares the $\delta_{average} \times \delta_{min}$ values for three backbone sets: The full set of 42 backbones, a randomly chosen set of 10 backbones (R-10), and a set of 10 backbones chosen using weighted D-optimal design in the descriptor space (D-10). Note that selecting the latter two subsets requires multiple sampling (see the Experimental Section for details). The

Table 2. Diversity indices for various backbone sets.

Dataset	$\delta_{average}$	δ_{min}	$\delta_{average} \times \delta_{min}$
Full set	118.0	7.2	858
R-10 ^[a]	140.6	20.1	2829
D-10	139.4	30.9	4308

^[a] Representative average set from a population of 500 random subsets, see Experimental Section for details.

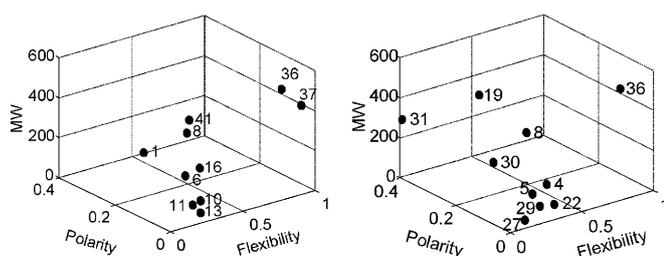


Figure 9. Backbone three-dimensional descriptor space for the selected subsets: R-10 (left) and D-10 (right).

$\delta_{average} \times \delta_{min}$ value for the full set of 42 backbones is low, mainly due to the lower value of the δ_{min} parameter in comparison with those of R-10 and D-10. If δ_{min} is small, there is a high clustering of the data. This is shown in Figure 6 and Figure 7 for the set of 42 backbones. The main clusters are in the lower polarity (0–0.3) and lower flexibility (0–0.2) regimes. Indeed, the high average $\delta_{average} \times \delta_{min}$ for the R-10 backbone subset is an evidence of the poor diversity of the 42 backbones subset, since any subset of 10 random backbones will represent better the structural and properties features of the data. The R-10 and D-10 $\delta_{average} \times \delta_{min}$ values are very different. The D-10 subset is more diverse. Interestingly, the main difference is in δ_{min} , while $\delta_{average}$ remains almost constant for both subsets.

When studying the random subsets we observed a pattern for some subsets that showed high diversity. All the subsets with $\delta_{average} \times \delta_{min}$ values >4000 contained a density weighted selection of backbones from the four clusters shown in Figure 9. This confirms that density-weighted D-optimal design in the descriptor space is indeed an optimal backbone selection method.

Conclusions

We have shown that *in silico* catalysis design can assist chemists in the difficult task of ligand selection. The huge diversity of the multi-dimensional ligand

space can be tackled by isolating the main backbone motifs from ligand libraries, creating a representative set of manageable proportions. This set can then be modeled using 2D and 3D descriptors. However, for synthesis purposes, a subset of backbones must be chosen. This choice is not trivial, but it can be done by selecting structures that are far away from each other in the descriptor space. As shown here, this is easily done using weighted D-optimal design. This selection technique gives more diverse subsets than random sampling. The final backbone set can then be combined with selected ligating and residue groups, generating a diverse library of homogeneous catalysts.

Experimental Section

Computational Methods

Ligand Construction and Optimisation

The 42 backbone structures used in the analysis were chosen manually from a dataset of 115 biphosphite and biphosphine ligand-nickel complexes, taken from papers and patents^[34] on the Ni-catalysed hydrocyanation of butadiene. Each structure was decomposed following the building block scheme described in Figure 4. Ligand geometry optimisation (for calculating the 3D descriptors) was performed using Hyperchem.^[44] We used the MM+ force field in combination with a conjugate gradient optimisation method (Polak–Ribiere). The backbone descriptors size, flexibility and polarity (see Table 3) were computed with the Codessa software package^[45] and analysed using Matlab scripts.^[46]

Table 4. Diversity indexes for five runs for 100 different combinations of 10 random backbones.

R-10 (100 times)	1	2	3	4	5	Average
$\delta_{average} \times \delta_{min}$	2877	2634	2840	2976	2677	2800

Table 3. The three influential backbone descriptors.

Descriptor type	Name (units)	Description
Constitutional	Molecular weight (g mol^{-1})	Sum of the individual isotopic masses of all the atoms of a molecule.
Topological	Kier flexibility index ^[47] (no units, normalised descriptor)	$\Phi = \frac{{}^1k \cdot {}^2k}{N_{SA}}$ where 1k and 2k are the Kier shape indices (a function of non-hydrogen atom counts and the number of paths of length n in the molecular graph) and N_{SA} the number of non-hydrogen atoms in the molecule.
Electrostatic	Polarity parameter ^[48] (electron charge density e , normalised)	$P = Q_{max} - Q_{min}$ where Q_{max} is the most positive and Q_{min} the most negative atomic partial charge in the molecule, respectively. ^[48]

Procedure for Choosing a Representative Random Subset of Backbones

To obtain a statistically consistent measure for the R-10 dataset, a single combination of backbones is not sufficient. Instead, we first compute the average $\delta_{average} \times \delta_{min}$ for 100 randomly selected different combinations of 10 backbones. We repeat the calculation five times, and then compute the average value as shown on Table 4. After studying the compositions of some random sets of 10 backbones out of the 100 sets, we chose a group of 10 random backbones with $2700 < \delta_{average} \times \delta_{min} < 2900$.

Procedure for Selecting a Weighted D-Optimal Subset of Backbones

For the D-optimal subset selection, we partitioned the backbone descriptor space in nine rectangles. However, due to the uneven distribution of the backbones in this space, picking one backbone from each rectangle yields relatively low diversity indices ($\delta_{average} \times \delta_{min} \approx 3000$). To avoid this, we selected from every rectangle a number of backbones that is proportional to the number of total backbones in that rectangle. Subsequently, we chose *which* backbones should be selected by searching for backbone combinations that increase the diversity of the subset. This is done by studying the compositions of top-performing random subsets of 10 backbones out of 100 sets. We observe that all the random high-scoring subsets have always some backbones from selected clusters. In this case, the 14 highest-performing random sets ($4200 < \delta_{average} \times \delta_{min} < 5800$) have always one or two backbones of the cluster {35, 36, 37}. Other backbones that appear often are: 24 (7 times in the top 14 sets), 21 (7 times), 29 (6 times), and 8 (5 times).

Experimental Methods

CAUTION! Hydrogen cyanide and acetone cyanohydrin are highly toxic! They should be used only in a well-ventilated and monitored fume hood or a dry-box, by teams of at least two technically qualified persons.

Materials and Instrumentation

All reactions were performed under argon in a glove-box. Several HCN Dräger monitors controlled the level of HCN inside and outside the hoods. GC analyses were performed using a Hewlett Packard HP6890 gas chromatograph with a $250 \mu\text{m} \times 30 \text{m}$ capillary column. All chemicals were purchased from commercial sources and used as received. All products are known compounds and were identified by comparing their spectral properties and GC retention times to those of commercial samples.

Procedure for Catalytic Hydrocyanation of 3-Pentenenitrile

The catalyst precursor, $\text{Ni}(\text{COD})_2$ (0.5 mmol, 138 mg), the ligand (2.5 mmol), 3-pentenenitrile (15 mmol, 1.21 g) and ZnCl_2 (0.5 mmol, 68 mg) were added to a 60-mL Schlenk glass tube equipped with a septum stopper. The mixture was stirred and heated to 70°C . Acetone cyanohydrin (previously dried over molecular sieves) was fed to the mixture by a pressure syringe (0.45 mL h^{-1}). The addition was stopped after 3 h. The mixture was cooled to 25°C , diluted with acetone and analysed by GC. HCN levels were permanently monitored in the laboratory and reaction hood and kept under 2 ppm. Excess acetone cyanohydrin was disposed of by slowly adding to aqueous sodium hypochloride.

Acknowledgements

We thank Rhodia for financial support.

References

- [1] G. Rothenberg, *Catalysis: Concepts and Green Applications*, Wiley-VCH, Weinheim, 2008.
- [2] P. T. Anastas, J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, 2000.
- [3] G. Li, *J. Org. Chem.* 2002, 67, 3634–3650.
- [4] D. J. Cole-Hamilton, *Catalyst Separation, Recovery And Recycling: Chemistry And Process Design*, Springer Verlag, Berlin, Heidelberg, 2006.
- [5] M. T. Reetz, *Angew. Chem.* 2001, 113, 292–320; *Angew. Chem. Int. Ed.* 2001, 40, 284–310.
- [6] M. Baumann, I. R. Baxendale, S. V. Ley, C. D. Smith, G. K. Tranmer, *Org. Lett.* 2006, 8, 5231–5234.
- [7] J. G. de Vries, A. H. M. de Vries, *Eur. J. Org. Chem.* 2003, 799–811.
- [8] J. A. Hageman, J. A. Westerhuis, H.-W. Frühauf, G. Rothenberg, *Adv. Synth. Catal.* 2006, 348, 361–369.
- [9] E. Burello, G. Rothenberg, *Int. J. Mol. Sci.* 2006, 7, 375–404.
- [10] P. Adriaensens, A. Pollaris, D. Vanderzande, J. Gelan, J. L. White, M. Kelchtermans, *Macromolecules* 2000, 33, 7116–7121.
- [11] A. Böcker, G. Schneider, A. Teckentrup, *QSAR Comb. Sci.* 2004, 23, 207–213.
- [12] P. W. N. M. van Leeuwen, K. Morokuma, J. H. van Lenthe, *Theoretical Aspects of Homogeneous Catalysis*, Vol. 18, Springer-Verlag, New York, 1995.
- [13] E. Burello, G. Rothenberg, *Adv. Synth. Catal.* 2005, 347, 1969–1977.
- [14] A. G. Maldonado, J. P. Doucet, M. Petitjean, B.-T. Fan, *Mol. Diversity* 2006, 10, 39–79.
- [15] J. A. Westerhuis, J. A. Hageman, H.-W. Frühauf, G. Rothenberg, *Chim. Oggi* 2007, 25, 28–32.
- [16] M. Urbano-Cuadrado, J. J. Carbo, A. G. Maldonado, C. Bo, *J. Chem. Inf. Model.* 2007, 47, 2228–2234.
- [17] B. Cornils, W. A. Herrmann, *Applied Homogeneous Catalysis with Organometallic Compounds*, Wiley-VCH, Weinheim, 1996.

- [18] H.-U. Blaser, B. Pugin, F. Spindler, M. Thommen, *Acc. Chem. Res.* **2007**, *40*, 1240–1250.
- [19] P. W. N. M. van Leeuwen, *Homogeneous Catalysis: Understanding the Art*, Kluwer, Dordrecht, **2004**.
- [20] M. W. Van Laren, C. J. Elsevier, *Angew. Chem.* **1999**, *111*, 3926–3939; *Angew. Chem. Int. Ed.* **1999**, *38*, 3715–3717.
- [21] R. Noyori, *Adv. Synth. Catal.* **2003**, *345*, 15–32.
- [22] H.-U. Blaser, *Adv. Synth. Catal.* **2002**, *344*, 17–31.
- [23] M. Corsi, *Synlett* **2002**, 3127–3128.
- [24] T. Flessner, S. Doye, *J. Prakt. Chem.* **1999**, *341*, 436–444.
- [25] W. Keim, *New J. Chem.* **1994**, *18*, 93–96.
- [26] P. W. N. M. van Leeuwen, P. C. J. Kamer, J. N. H. Reek, *Pure Appl. Chem.* **1999**, *71*, 1443–1452.
- [27] E. Burello, P. Marion, J.-C. Galland, A. Chamard, G. Rothenberg, *Adv. Synth. Catal.* **2005**, *347*, 803.
- [28] N. Nikolova, J. Jaworska, *QSAR Comb. Sci.* **2003**, *22*, 1006–1026.
- [29] E. Burello, D. Farrusseng, G. Rothenberg, *Adv. Synth. Catal.* **2004**, *346*, 1844–1853.
- [30] O. Lavastre, F. Bonnette, L. Gallard, *Curr. Opin. Chem. Biol.* **2004**, *8*, 311–318.
- [31] C. A. Tolman, R. J. McKinney, W. C. Seidel, J. D. Druliner, W. R. Stevens, *Adv. Catal.* **1985**, *33*, 1–46.
- [32] L. Bini, C. Muller, J. Wilting, L. von Chrzanowski, A. L. Spek, D. Vogt, *J. Am. Chem. Soc.* **2007**, *129*, 12622–12623.
- [33] A. P. V. Gothlich, M. Tensfeldt, H. Rothfuss, M. E. Tau-chert, D. Haap, F. Rominger, P. Hofmann, *Organometallics* **2008**, *27*, 2189–2200.
- [34] The data for the 115 biphosphine-and biphosphite-nickel complexes were taken from the following patents: *US Patent* 2005/0090677; *US Patent* 6,355,833; *US Patent* 7,022,866; *US Patent* 7,067,685; *US Patent* 2006/0258874; *US Patent* 2007/0060766; *European Patent* EP1001928, *European Patent* EP1000022; *US Patent* 6,846,945; *European Patent* EP0815073.
- [35] T. Langer, G. Wolber, *Pure Appl. Chem.* **2004**, *76*, 991–996.
- [36] A. Fernandez, C. Reyes, T. Y. Lee, A. Prock, W. P. Giering, C. M. Haar, S. P. Nolan, *J. Chem. Soc. Perkin Trans. 2* **2000**, 1349–1357.
- [37] D. Woska, A. Prock, W. P. Giering, *Organometallics* **2000**, *19*, 4629–4638.
- [38] M. R. Wilson, A. Prock, W. P. Giering, A. L. Fernandez, C. M. Haar, S. P. Nolan, B. M. Foxman, *Organometallics* **2002**, *21*, 2758–2763.
- [39] L. Xue, J. Bajorath, *Comb. Chem. High Throughput Scr.* **2000**, *3*.
- [40] E. Burello, C. Bologna, V. Freccer, S. Miertus, *Mol. Phys.* **2002**, *100*, 3187–3198.
- [41] D. C. Whitley, M. G. Ford, D. J. Livingstone, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160–1168.
- [42] R. L. Tranter, *Design and Analysis in Chemical Research*, CRC Press, Sheffield, **2000**.
- [43] For more information on D-optimal design see these two tutorial papers, aimed at a general scientific audience: a) D-Optimal designs. Engineering Statistics Handbook Online. Section 5.5.2.1, available on <http://www.itl.nist.gov/div898/handbook/pri/section5/pri521.htm>; b) Statistical Molecular Design – A core concept in multivariate QSAR and combinatorial technologies, **2003**, available on <http://www.chemometrics.se/images/stories/pdf/april2003.pdf>.
- [44] *Hyperchem for Windows* (Molecular Modeling System), Version 7.5, Hypercube, Inc., **2002**, <http://www.hyper.com/>.
- [45] *Codessa*, Version 2.642, **1994**, <http://www.codessa-pro.com>; see also: A. R. Katritzky, V. S. Lobanov, M. Karelson, R. Murugan, M. P. Grendze, J. E. Toomey Jr., *Rev. Roum. Chim.* **1996**, *41*, 851.
- [46] *Matlab*, Version R2007a, The Mathworks, Inc., **1984–2007**, <http://www.mathworks.com/>.
- [47] L. B. Kier, in: *Computational Chemical Graph Theory*, (Ed.: D. H. Rouvray), Nova Science Publishers, New York, **1990**.
- [48] The polarity parameter gives a better representation of the backbones in this case than the dipole moment; see: K. Osmialowski, J. Halkiewicz, A. Radecki, R. Kalisz, *J. Chromatog.* **1985**, *346*, 53.