DOI: 10.1002/adsc.201200453

Quantitative Comparison of Chiral Catalysts Selectivity and Performance: A Generic Concept Illustrated with Cyclododecanone Monooxygenase as Baeyer–Villiger Biocatalyst

Michael J. Fink,[a,e] Daniela V. Rial,[b,e] Petra Kapitanova,[a] Alenka Lengar,[a] Jessica Rehdorf,[c] Qiong Cheng,[d] Florian Rudroff,[a,]* and Marko D. Mihovilovic[a,]*

[a] Institute for Applied Synthetic Chemistry, Vienna University of Technology, Getreidemarkt 9/163-OC, 1060 Vienna, Austria
Fax: (+43)-1-588-01-163-99; e-mail: frudroff@ioc.tuwien.ac.at or mmihovil@pop.tuwien.ac.at

[b] Área Biología Molecular, Departamento de Ciencias Biológicas, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002 LRK Rosario, Argentina

[c] Department of Biotechnology & Enzyme Catalysis, University of Greifswald, Institute of Biochemistry, Felix-Hausdorff-Str. 4, 17487 Greifswald, Germany

[d] Biological and Chemical Sciences and Engineering, Central Research and Development, E. I. DuPont de Nemours, Inc., Wilmington, Delaware 19880-0328, USA

[e] These authors contributed equally to this work.

Received: May 22, 2012; Published online: ■■ ■■, 0000

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/adsc.201200453.

Abstract: Within this work a generic tool for chiral catalyst evaluation is established based on the application-oriented properties activity and selectivity; the concept aims at quantitatively comparing catalyst performance in general on a multitude of substrates. It is designed and intended to serve as decision guidance for challenges in catalysis and comprehensible information extraction from already recorded but unrefined data sets. The underlying algorithm assigns function points to catalytic entities *via* a statistically solid model possessing high flexibility and generates a relative ranking. This is coupled to an automated iterative refinement process towards maximum information content of results employing Shannon entropy optimization. Consequently, the developed workflow facilitates high distinguishability between catalysts even in low-scattering data sets. The numerical ranking is complemented by a clearly arranged graphic representation permitting facile and reliable visual interpretation of generality or niche capabilities of catalysts. Usefulness of the title concept is demonstrated by the performance evaluation of cyclododecanone monooxygenase, a highly versatile Baeyer–Villiger enzyme. To retain broad applicability, an open-source MATLAB® script is provided in electronic form.

Keywords: asymmetric synthesis; biocatalysis; catalyst evaluation; function point analysis; Shannon entropy

Introduction

The quest to identify a well suited catalyst for a given chemical transformation remains a key issue in applied synthetic chemistry.[1] While fundamental research aims at understanding the underlying chemical nature of catalysis, properties such as activity, selectivity and stability are usually awarded higher importance within applied research and process development when seen from a result-oriented point of view.[2] With this goal in mind all disciplines of catalysis – biological,[3] metal-assisted,[4] inorganic,[5] and organic[6] – provide ever more chemical entities capable of augmenting reaction rates, thereby producing an enormous wealth of data on catalyzed transformations.

A challenge directly connected to the discovery of catalysts is represented by the extraction of information contained in the reaction data. However, in many cases, such extracts often gravitate to merely cataloguing instead of properly analyzing obtained results on all quality levels from lab report to perennial review article. One clear objective of such analyses should be the identification of potent candidates from a group of catalysts for a certain reaction class enabling informed decisions on both further develop-

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ment of the compound itself and the proper selection for a given application. In a broader context, data analysis should therefore not be limited to crude qualitative assessment (e.g., activity as a digital function or stereopreference); it should offer quantitative results allowing the comparison of more or less active and more or less selective catalysts on detailed levels. Moreover, the calculation model should be based on sound statistics to avoid unwanted bias and results should be presented in an easily comprehensible fashion.

Evaluation systems with mathematical rationale and graphic result output have been devised during recent years for the purpose of catalyst discovery; for example, high-throughput screening assays coupled with color-coded matrices for efficient characterization of biocatalysts were proposed by Reymond and co-workers.[7] However, a similar tool for facile *comparative* analyses of catalysts on a *diverse* set of substrates has not been published, so far, according to our best knowledge.

Building on the prerequisites outlined above, we present a quantitative evaluation model for chiral catalysts in combination with a generally intelligible graphical results representation. The algorithm condenses information on activity and stereoselectivity from a multitude of substrate-catalyst pairings into concise rankings of the catalytic entities.

Development of this concept was inspired by a detailed characterization of the substrate profile of cyclododecanone monooxygenase (CDMO), a recent addition to the Baeyer–Villiger monooxygenase (BVMO) family.[8] Since their first description in the scientific literature in 1976[9] and their first application in biocatalysis in 1988,[10] the collection of BVMOs has grown to approx. 40 native and mostly very promiscuous enzymes of various bacterial and fungal origins. Detailed studies on such biocatalytic entities led to a large collection of transformation data.[11] Despite a plethora of research efforts to explore the limits of promiscuity and peaks of performance, most reported activity and selectivity figures are traditionally referenced to early discovered BVMOs, namely cyclohexanone monooxygenase (CHMO)[9,12] and cyclopentanone monoxygenase (CPMO),[13] rather than compared to the currently best performing enzyme.

This comparison mode may well be suited whenever novel substrate transformations or significant improvements of known reactions are published. However, this approach compromised our interpretation of results obtained with CDMO: the enzyme seemed to be a generally excellent biocatalyst for Baeyer–Villiger oxidations with interesting but rather few new activities.[14] While interpretation of generated screening data was easily possible for expected trends and single transformations, the multitude of results and lack of organized reference archives precluded us from drawing any generally viable conclusion beyond the obvious. Our hypothesis of broad applicability needed statistical support and shall serve as a case study for the title concept.

Results and Discussion

We will first describe the algorithm leading to a ranking of nine BVMOs[15] in eight substrate classes in the following sections. The conclusions from this part of the study will then lead to a description of CDMO's relative position within this group of biocatalysts. Highlights and strong features of the title BVMO will be discussed as a second part. Detailed results on biocatalyst performance by CDMO investigating 83 substrates are deposited in the Supporting Information.

Development of the Algorithm

Our proposed ranking system essentially derives from Function Point Analysis, a theoretical tool in information technology (IT) developed to evaluate the functionality of software or any other IT system by methodically assigning points to certain properties.[16] Multiple entities can then be graded by summing up their function points. We adapted this concept to our needs and priorities (activity and stereoselectivity; defined bias towards selectivity, see below).

A dataset of 65 substrates[8a,14,17,18] was selected and grouped into structural and reaction-type classes (Table 1). Within a sequence of logical steps, the work-flow for catalyst assessment is initiated by the definition of a point scoring function, followed by data processing and statistics. The model is then iteratively refined in each substrate class and optimized to maximal information content before the final result for each enzyme can be plotted in a clearly arranged graphical form.

From the perspective of a synthetic organic chemist the optimum enzyme is characterized by both high activity and impeccable enantioselectivity. Failing this combined criterion, we considered selectivity as more relevant owing plainly to practical reasons: mixtures of substrate and product are generally much easier to separate than mixtures of product isomers. Furthermore, the best biocatalysts should be awarded disproportionately high when compared with mediocre ones, allowing identification of the most proficient enzyme for each substrate class. Operation stability was not accounted for due to lack of comparable data in this catalyst class. These are the only preconceptions we intended to incorporate into our model. Otherwise unbiased data evaluation was implemented by introduction of distribution-independent statistics and

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Substrate classes, number of substrates and highest scoring enzymes.

Reaction Type	Substrate Class		Number of Substrates	Best Scoring BVMO
Desymmetrizations	1	Prochiral cyclobutanones	6	$CHMO_{Rhodo1}$
	2	Prochiral cyclohexanones	13	$CDMO_{Rhodo}$
	3	Prochiral polycyclic ketones	6	$CPMO_{Coma}$
Kinetic Resolutions	4	Racemic Cycloketones	5	$CHMO_{Arthro}$
	5	Racemic linear β-amino ketones	9	$CDMO_{Rhodo}$
	6	Racemic linear β-hydroxy ketones	10	$CPMO_{Coma}$
Regiodivergent Transformations	7	Optically pure terpenones	6	$CHMO_{Acineto}$/$CHMO_{Brachy}$
	8	Racemic fused cyclobutanones	10	$CDMO_{Rhodo}$
Σ			**65**	

an optimization protocol derived from information theory.

Total points of each enzyme per substrate class Pts_{TTL} are calculated according to Eq. (1).

$$Pts_{TTL} = Pts_{Activity} \cdot Pts_{Selectivity} \quad (1)$$

Multiplicative combination of the two factors was chosen to comply with our constraint that only the most active and most selective enzyme should attain the maximal score.

Organic chemists commonly report the measurement unit *conversion*. This can be treated as equivalent to activity, as long as the crucial reaction parameters time, substrate, and enzyme concentration are kept constant (see Experimental Section).

Unfortunately, conversion values of several BVMO-catalyzed reactions are not reported numerically, but already pre-filtered in four categories:

- No conversion — 0 Points
- 1–50% conversion — 1 Point
- >50–90% conversion — 2 Points
- >90% conversion — 3 Points

As raw numerical data for all substrates were hardly accessible to us, we decided to use this function directly for the assignment of activity points, since it was largely compliant with our requirements.

For the selectivity points function a different strategy was applied. Enantiomeric excess values in our data are not normally distributed with largely deviating ranges (Figure 1), requiring a statistically adjusted scoring function to ensure impartial evaluation of selectivity.

To implement this, statistical measures per substrate were used instead of fixed numerical thresholds: the lowest value (ee_{min}), median *ee* ($q_{0.5}$), two variable interval markers at the x_1- and x_2-quantile levels (q_{x1} and q_{x2}), and best *ee* (ee_{max}):

- $ee = 0$ — 0 Points
- $0 < ee < q_{0.5}$ — 0.5 Points
- $q_{0.5} \leq ee < q_{x1}$ — 1 Point
- $q_{x1} \leq ee < q_{x2}$ — variable Points
- $q_{x2} \leq ee < max$ — variable Points
- $ee = max$ — 5 Points

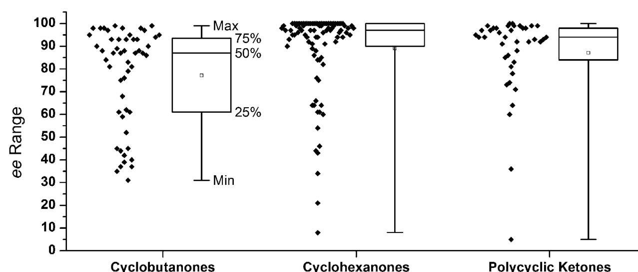


Figure 1. Box plots for enantiomeric excess in three desymmetrization substrate classes, indicating a non-uniform and non-Gaussian distribution of values.

The additional variable intervals were introduced to allow finer classification within the top 50% of enzymes; they also serve as basis for model refinement and optimization. Additionally, an equal-rank rule was introduced, stipulating that only the second best category of points is scored if ee_{max} is reached by multiple enzymes.

We became aware of the fact that willful setting of the variable quantiles would not lead to higher distinguishability in all substrate classes, since the statistical ranges differed too much. It was concluded to incorporate this decision into the mathematical model together with an automated iterative refinement procedure.

For this optimization task an exponential fit function (Figure 2, black dashed line) was generated based on the three fixed thresholds and their respective defined number of points (ee_{min}, $q_{0.5}$ and ee_{max} with 0, 0.5 and 5.0 points; black squares in Figure 2). Quantiles q_{x1} and q_{x2} could now easily be varied between $q_{0.5}$ and ee_{max} with their corresponding points

© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

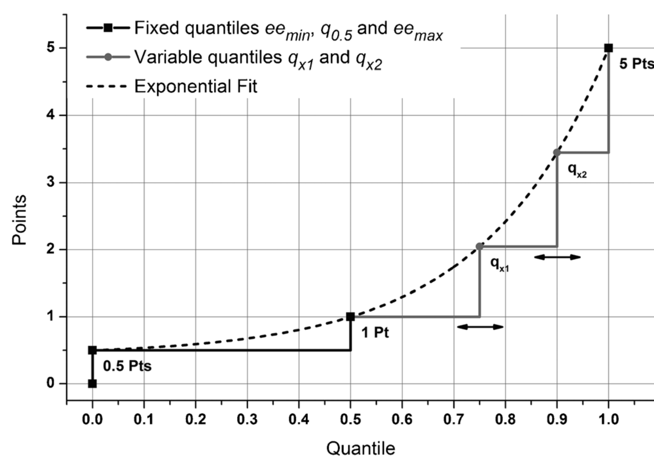


Figure 2. Horizontal spline step function for points scored from enantioselectivity including the exponential fit function and variable quantiles q_{x1} and q_{x2}.

being determined by read-out of the ordinate value on the fit function (grey dots in Figure 2).

Subsequently, the scoring algorithm was run separately with each pair from an exhaustive pool of combinations of q_{x1} and q_{x2}, followed by calculation of the Shannon entropy H of total points in each iteration – the crucial step in the optimization.[19] This way, quantiles leading to H_{max} and therefore maximum information content within the given substrate class were identified and the evaluation was eventually finalized with these values.

In most substrate classes this had quite dramatic effects on the outcome in terms of information density. Point distributions from performing the optimization both to maximal and to minimal H in substrate class 1 are shown in Figure 3. Setting q_{x1} and q_{x2} for minimal Shannon entropy results in three enzymes reaching top rank ($CHMO_{Brevi1}$, $CHMO_{Rhodo1}$ and $CHMO_{Xantho}$), consequently making it difficult to decide which BVMO would be best suited for such a synthetic challenge. Three others ($CHMO_{Acineto}$, $CHMO_{Arthro}$ and $CHMO_{Brachy}$) form an undifferentiated second tier.

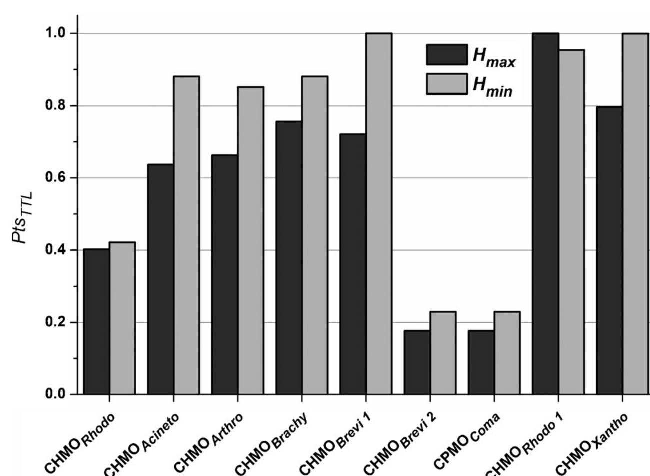


Figure 3. Comparison of points scored in substrate class 1 with maximal ($q_{0.78}$ and $q_{0.83}$) and minimal ($q_{0.60}$ and $q_{0.87}$) Shannon entropy (points are normalized to the highest value).

The picture changes substantially when the algorithm is iterated towards highest information content: $CHMO_{Rhodo1}$ clearly performs best, with all but the bottom two biocatalysts in distinct ranks. It is evident that fixation of variable thresholds based on experience or just user arbitration could hardly lead to such unbiased solutions (H_{max} thresholds: 0.78 and 0.83; H_{min} thresholds: 0.60 and 0.87). Even more, the computational approach recognized some elusive trends and generated reasonable results in all substrate classes.

With this largely unbiased and refined model at hand, we were further concerned with straightforward visualization of the scoring results. Thus, points were normalized to the highest value of each class and then plotted in radar charts, where every axis represents one category of substrates. This illustration mode allows both comparison of catalysts in single disciplines (axis to axis) and evaluation of generality (area to area) in a condensed but comprehensible fashion (Figure 4).

Scoring Results of BVMOs

Interpretation of the graphic results immediately suggests the following conclusions:

(i) There is no pareto-optimal biocatalyst among the tested BVMOs (scoring highest in all categories).

(ii) Phylogenetic relationships and proposed clustering in BVMO subfamilies[18a] are reflected in the compounded activity-selectivity scores: members of one cluster show similar area patterns in the radar charts.

(iii) Low all-round applicability (small area) does not necessarily downgrade the usefulness of an enzyme, as can be seen with $CPMO_{Coma}$. Despite relatively sluggish performance in six out of eight substrate classes, this BVMO is ranked best in the remaining two, making it an indispensable niche catalyst.

(iv) The use of $CHMO_{Acineto}$ as reference standard (as frequently observed in previous comparative studies) is clearly misleading, as it is outperformed in terms of generality by essentially all other CHMO-type enzymes ($CHMO_{Arthro}$, $CHMO_{Brachy}$, $CHMO_{Rhodo1}$, $CHMO_{Xantho}$).

The first result was to be expected: although cycloketone-converting BVMOs (CHMO-type, CPMO-type and CDMO enzymes) do perform well with linear ketones too, affinity to their natural substrate

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

These are not the final page numbers!

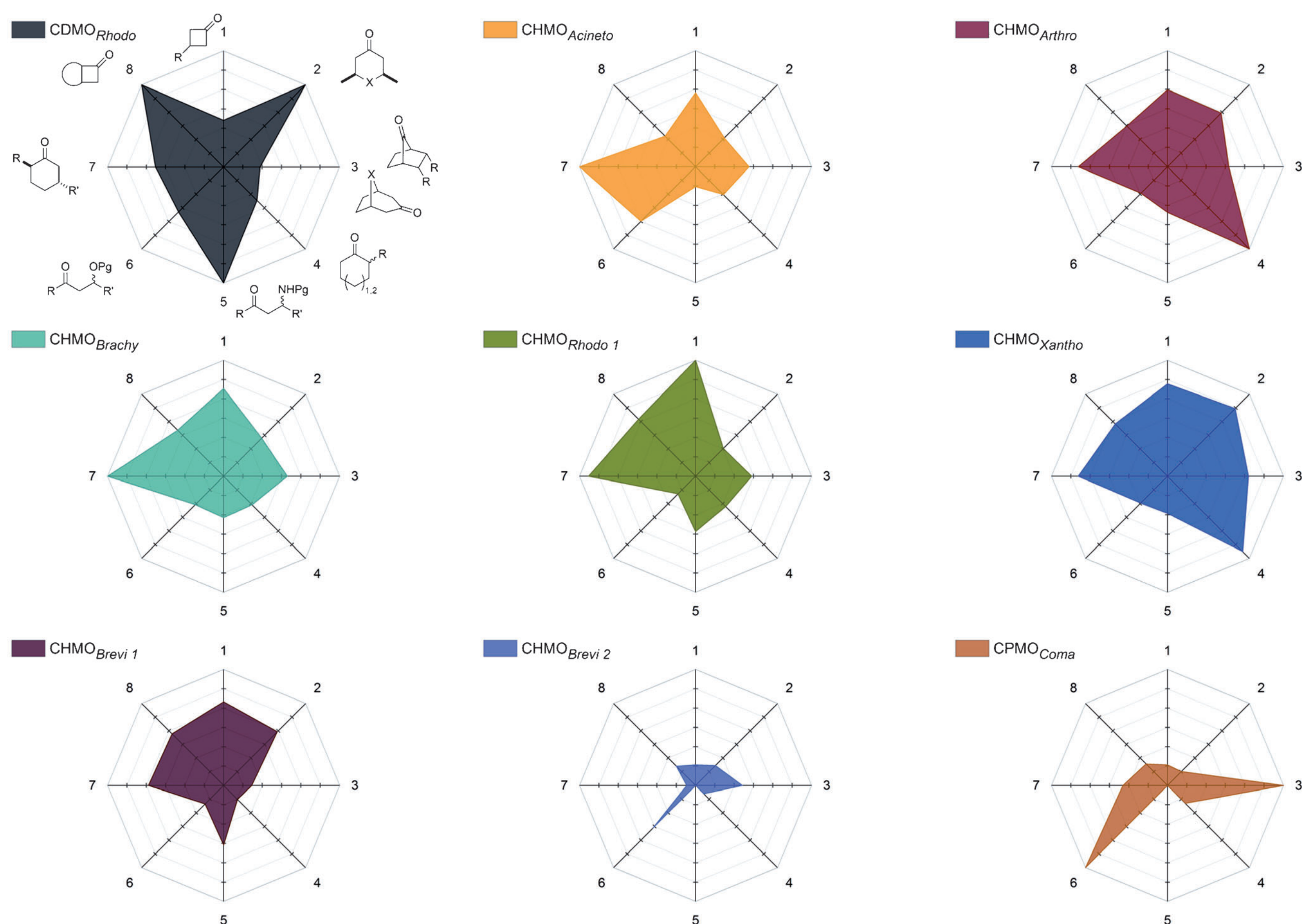


Figure 4. Radar charts for nine BVMOs, sorted according to their phylogenetic relationship ($CDMO_{Rhodo}$; $CHMO_{Acineto}$, $CHMO_{Arthro}$, $CHMO_{Brachy}$, $CHMO_{Rhodo1}$, $CHMO_{Xantho}$; $CHMO_{Brevi1}$; $CHMO_{Brevi2}$, $CPMO_{Coma}$). Each axis represents one substrate class, starting clockwise from the top. Points are normalized to unity by division through the highest value of each category.

motif is apparent from the graphs. Unfortunately, we were not able to include prominent linear ketone-converting BVMOs (*p*-hydroxyacetophenone monooxygenase HAPMO,[20] phenylacetone monooxygenase PAMO[21]) in this study to show their complimentary behavior, since there are far fewer comprehensive data records available in the literature for those biocatalysts.

The second conclusion stands in direct relation to the former. Our hypothesis of linking sequence information with substrate specificity and enantioselectivity is supported by this novel data analysis.[18a] According to their phylogenetic relationships the nine evaluated BVMOs are grouped into four clusters: large ring converting BVMOs (outstanding performance in substrate classes 2, 5 and 8), CHMO-type BVMOs (excellent performance in substrate classes 4 and 7), CPMO-type BVMOs (poor general but top ranking performance in classes 3 and 6) and $CHMO_{Brevi1}$ (above average performance in substrate classes 1, 2, 7 and 8) as stand-alone branch. Within these clusters, general and specific activity is largely comparable, whereas the clusters themselves differ distinctly.

Third, it is evident from the results of $CPMO_{Coma}$ that a lack in generality can indeed be compensated with excellent activity and selectivity in one substrate class in order to qualify as a useful catalyst. This is even multiplied when the proficiency of the narrow-ranged enzyme is orthogonal to all others. At this point one important caveat of our model needs to be mentioned: singular highlight transformations that protrude from the overall performance of a catalyst are mostly suppressed by inherent statistical smoothing of our algorithm. Thus, higher ranking of one biocatalyst only indicates higher probability of efficient and selective transformations with *any substrate* of a *given class*.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

This leads to our fourth and last direct interpretation of Figure 4: $CHMO_{Acineto}$ is not a very versatile BVMO, so referencing of biotransformation efficiency to values obtained with this enzyme should be discontinued as it distorts the comparison with any new value. Instead, the best known BVMO for the transformation should be cited.

Directing our focus from general interpretation now to individual scrutiny, it turned out that $CHMO_{Xantho}$ is the most versatile biocatalyst of this set, scoring good to excellent ranks in all but the linear substrate classes. It however never reaches a top position. This trend continues with the other four CHMO-type enzymes: their strength apparently lies with bio-oxidations of cyclic ketones. Performance graphs of $CHMO_{Brevi1}$ and the CPMO-type enzymes ($CHMO_{Brevi2}$ and $CPMO_{Coma}$) show a different pattern: while the first BVMO still proves to be generally applicable in four categories, the latter seem to be performing on an inferior level at first glance. This must be put into perspective, as those two enzymes are usually enantiocomplementary to CHMO-type BVMOs, however often with lower selectivity. Table 1 summarizes the best-performing biocatalysts of each category.

Eventually, we were delighted to find our initial hypothesis confirmed to a large extent: CDMO is indeed a versatile BVMO when compared to eight other cycloketone-converting enzymes. In this series it is the only BVMO ranked best in three substrate categories (desymmetrizations of cyclohexanones, kinetic resolutions of linear β-amino ketones and regiodivergent biotransformations of fused cyclobutanones). It scored average points in three other categories (desymmetrizations of cyclobutanones, kinetic resolutions of β-hydroxy ketones and regiodivergent terpenone oxidations), leaving deficiencies in kinetic resolutions of cyclic carbonyl substrates and desymmetrizations of polycyclic ketones.

Although the active site of CDMO was described to be capable of accommodating large compounds,[8b] for example, the native substrate cyclododecanone, the last result did not come as a surprise. Similar behavior was already found with the close homologue cyclopentadecanone monooxygenase, which showed comparably unsatisfactory performance with sterically demanding ketones.[22] These findings and the fact that BVMOs with smaller native substrates ($CPMO_{Coma}$,[23] but most prominently $CHMO_{Xantho}$[18c,l]) readily perform such transformations lend to speculate that there are still more intricate details influencing BVMO promiscuity in the reaction mechanism[24] of the enzymatic Baeyer–Villiger oxidation that have not been understood so far.

Indicative Biotransformations for the Catalytic Proficiency of CDMO

This section highlights selected examples from the substrate pool that demonstrate the general performance of CDMO and support its top rank among the evaluated biocatalyts. The first core competence of the title enzyme lies with desymmetrizations of prochiral cyclohexanones and pyranones (substrates **7**–**19**; for compound assignment refer to the Supporting Information). Overall enantioselectivity is excellent in this substrate class, as can be seen in the scatter plots in Figure 1. Median *ee* of all 117 biotransformations used in the scoring algorithm lies at 97%, making it difficult to point out examples with significant improvement over known results.

Exactly this situation can serve as showcase scenario for our analysis, visualized as heat map in Figure 5. CDMO is superior or at least equal to the best value obtained from one of the other BVMOs in most cases and therefore the most versatile in this structural class.

Kinetic resolutions of linear β-amino ketones are the second strength of CDMO (Figure 6). Within this study it is the only enzyme that accepts substrates **37** and **38** (HAPMO from *Pseudomonas fluorescens* ACB also converts these compounds, although with negligible enantiodiscrimination).[14b] Application of this feature and related biotransformations in the stereoselective synthesis of β-amino acids has been demonstrated in a previous study.[14]

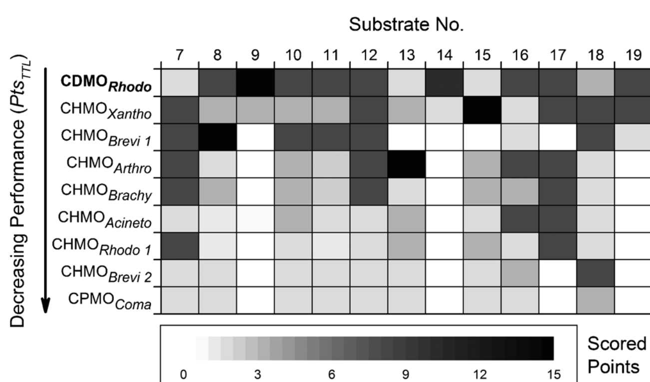


Figure 5. Heat map of scored points (Pts_{TTL}) for 13 prochiral cyclohexanones (substrates **7**–**19**). Enzymes are ordered with decreasing total performance from top down. Darker shades of grey indicate higher score.

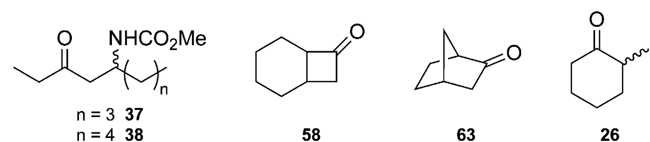


Figure 6. Substrates with major improvements in enantioselectivity using CDMO as biocatalyst.

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Additionally, CDMO is presented as the biocatalyst of choice for regiodivergent oxidations of racemic fused bicycloketones. These reactions proceed *via* parallel kinetic resolutions: one enantiomer is converted to the expected oxygen-insertion product (*normal* lactone, governed by higher nucleophilicity of the migrating carbon atom) while the other gives rise to the *abnormal* lactone due to stereoelectronic effects. Both reaction rates are usually within the same order of magnitude and therefore these biotransformations are not regarded as classic kinetic resolutions; they are allowed to proceed to full conversion, yielding two constitutionally isomeric (regioisomeric) but optically pure products in equal amounts in an ideal case.

Enantioselectivity for the normal lactone product of **58** could be improved to 98% with CDMO *vs.* 84% with $CHMO_{Xantho}$, the previously best known result. Furthermore, a synthetically useful regiodivergent enzymatic Baeyer–Villiger oxidation of norcamphor **63** was observed for the first time. Whereas all other tested BVMOs preferentially produce normal lactones with low enantioselectivity (best value 42% *ee* with $CHMO_{Brevi2}$), 98% optical purity was reached with CDMO, however at a very low reaction rate (<50% conversion after 24 h).

The last example, 2-methylcyclohexanone **26**, comes from a weaker part in the enzyme's profile. Nevertheless CDMO was again the only representative out of nine biocatalysts to show reasonable enantiodiscriminiation of the substrate enantiomers ($E=125$) (this substrate was already part of an earlier study on CDMO.[8a] We were however not able to reproduce the published value of $E \geq 200$, but attributed this inconsistency to the known high sensitivity of calculated enantiomeric ratios E to analytical error at high values.[25]) and significantly outperforms the best published BVMO (CPDMO from *Pseudomonas* sp. HI-70, $E=41$).[22] In stark contrast to this highlight, all other tested 2-substituted cycloketones were either not at all or just unselectively converted by CDMO, resulting in a low score for this compound category.

Conclusions

In this work we have presented a novel generic tool for quantitative and comparative evaluation of chiral catalysts based on available numerical reaction data. The focus is set on performance in terms of activity and enantioselectivity with intended bias towards higher selectivity. The multiplicative approach of point distribution allows facile incorporation of other parameters (e.g., operational stability) when comparable data sets are available. An algorithm was developed following the principles of Function Point Analysis to rank the catalysts, coupled to simple descriptive statistics of raw data values in form of quantile thresholds, thus enabling the model to process data with non-uniform or non-Gaussian distribution with minimized influence of user preferences. Furthermore, iterative optimization of the thresholds *via* calculation of the Shannon entropy provides results with maximum information content. Consequently, the most promiscuous catalysts with overall high activity and selectivity are ranked best and optimum distinguishability is achieved within any given test set. Finally, the outcome is visualized in easily intelligible radar charts eliminating the need to interpret numerical values for broad conclusions (Figure 7).

The analysis serves the purpose of decision guidance for users in both catalyst selection and development from an application-oriented point of view, discerning the most versatile catalytic entity for either one to many or even all substrate classes, thereby forming a pareto-optimal solution in the latter case (marked as red dotted line in Figure 7). Identification of a small number of potent enzymes may provide cost-efficient entry points to scientists into a new catalytic reaction class or indicate possibilities for co-operative catalyst design.[26] For this purpose a general open-source MATLAB® script file, suitable for adaptation to other transformations or catalyst types, is available for download free of charge on our website.[27]

We have illustrated the usefulness of our concept in a case study ranking cyclododecanone monooxygenase as Baeyer–Villiger biocatalyst within a group of nine BVMOs. CDMO's catalytic proficiency has previously only been sparsely characterized, albeit with promising results. Concluding from an elaborate investigation comprising 83 ketones of various structural classes we postulated that CDMO might be a highly versatile biocatalyst. This hypothesis could

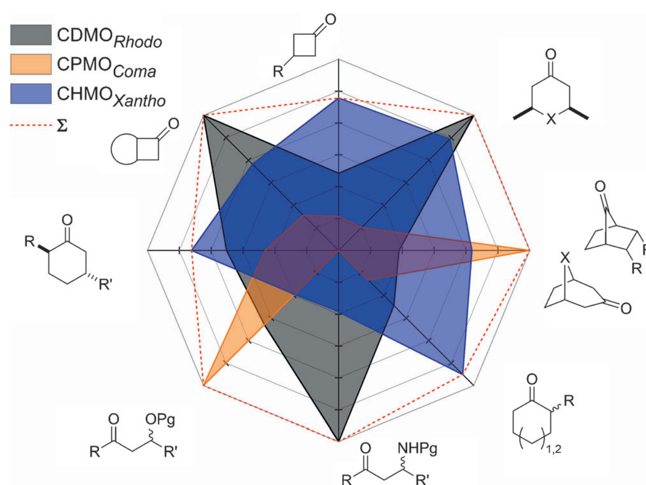


Figure 7. Comparative illustration of the functionality of three BVMOs. Their joint performance (marked as red dotted line) almost fully covers the functional space set in this study.

© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

then be strongly supported by the developed statistical evaluation method. Moreover, the analysis revealed that with a combination of only three out of the nine enzymes virtually the entire BVMO function space can be covered with high probability (Figure 7). We believe that the presented analysis is adaptable to any comparable data record and are convinced that similar results could be found for other reaction types, independent of the catalysts' mode of action.

Experimental Section

Numerical reaction data were obtained from literature sources or unpublished data from preparative or analytical scale biotransformations using BVMOs as whole-cell biocatalysts in recombinant host organisms. Conditions used for these reactions were consistent (reaction time, temperature, substrate concentration, cell growth conditions). Biotransformation and synthetic procedures and physical data of novel compounds can be found in the Supporting Information.

Development of the algorithm and all statistical calculations were carried out in MATLAB® R2011b using standard functions. All graphs were plotted in OriginPro 8.6. For each substrate class calculation is performed according to the following logic steps:

(i) Load matrices for conversion and enantiomeric excess.

(ii) Determine fixed statistical values from *ee* matrix for each substrate (quantiles $q_{0.0}=ee_{min}$, $q_{0.5}$, $q_{1.0}=ee_{max}$, number of ee_{max}).

(iii) Determine variable quantiles q_{x1} and q_{x2} for each substrate, starting with arbitrary values for x_1 and x_2 between 0.5 and 1.0 excluding limits.

(iv) Assign selectivity points (Pts_{Sel}) for each transformation according to the intervals defined in the scoring step function. This function was generated by horizontal spline connection of points on the exponential fit function (Figure 2). Fixed thresholds Pts_{Sel} (0.0)=0.5, Pts_{Sel} (0.5)=1 and Pts_{Sel} (1.0)=5 were fitted using OriginPro 8.6 with residuals below 10^{-15} [Eq. (2)]:

$$Pts_{Sel}(x) = 0.4285 + 0.714e^{4.1588x} \quad (2)$$

(v) Calculate total points (Pts_{TTL}) by multiplication of Pts_{Sel} with activity points (Pts_{Act}), derived from binned conversion values (see Section "Development of the Algorithm").

(vi) Sum up points for each enzyme, normalize to unity by division through the highest value.

(vii) Calculate Shannon entropy H according to Eq. (3) with i as catalyst index number.

$$H = -\sum_i Pts_{TTL_i} \cdot \log_2 Pts_{TTL_i} \quad (3)$$

(viii) Iterate steps 3–7 with all combinations of q_{x1} and q_{x2} between 0.50 and 1.00 excluding limits ($q_{x1} < q_{x2}$). Stepping size of 0.01 was used in this study, resulting in 1176 combinations. Smaller intervals did not lead to an improvement of the maximal Shannon entropy H_{max}.

(ix) Determine quantile set resulting in H_{max} and re-iterate steps 3–7 to generate the final result matrices (Pts_{TTL} per substrate and enzyme; sum of Pts_{TTL} per enzyme for graphical output).

For kinetic resolution and regiodivergent oxidation substrate classes this protocol was adapted for compliance with the sequence described above:

Kinetic Resolutions

In lieu of enantiomeric excess the enantiomeric ratio E was used for assignment of Pts_{Sel}, as it condenses selectivity information from this reaction mode into a single descriptor. E values were calculated from *ee* values of substrate and product, or *ee* of substrate/product and conversion according to Sih's equations[25a] using a software tool developed by Faber et al.[25b] Since conversion data for kinetic resolutions are generally available in numeric form in contrast to the other categories, we extended our statistical evaluation from Pts_{Sel} to Pts_{Act} as well. Conversion values c were binned in three intervals in compliance with desymmetrization values: All other operations remain unchanged.

- $c = 0$ 0 Pts_{Act}
- $0 < c < q_{0.5}$ 1 Pts_{Act}
- $q_{0.5} \leq c < q_{0.9}$ 2 Pts_{Act}
- $c \geq q_{0.9}$ 3 Pts_{Act}

Regiodivergent Biotransformations

Reactions of this type require three selectivity descriptors: enantiomeric excess of normal lactone products ee_N, abnormal lactone products ee_{ABN} and the regioisomeric ratio $r=[N]/[ABN]$. Lacking an established method to compound these values, Pts_{Sel} were calculated for both normal and abnormal products (including separate substrate statistics) and then weighted by multiplication with r [Eq. (4)]:

$$Pts_{Sel} = Pts_{Sel_N} \cdot r + Pts_{Sel_{ABN}} \cdot (1-r) \quad (4)$$

This way, enzymes with perfect regiodivergent behavior (producing both regioisomers in equal amounts with excellent optical purity) are ranked highest.

Regiodivergent oxidations of terpenones were carried out with enantiomerically pure starting materials. Since epimerization of ketones or lactones is usually not observed, optical purity is preserved throughout the reaction and the regioisomeric ratio r remains as only measure of selectivity. Pts_{Sel} were calculated according to the original desymmetrization script, using regioisomeric excess *re* instead of *ee*.

Author Contributions

Q.C. kindly provided the expression system of CDMO. D.V.R. performed the major share of substrate acceptance and performance screens. P.K., A.L. and J.R. performed various parts of substrate synthesis and screening work with

 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

These are not the final page numbers!

CDMO and other BVMOs to complete record sets used. F.R. sparked the idea of quantitative catalyst ranking. M.J.F. developed and implemented all necessary methods – statistical and graphical. F.R. programed the open-source MATLAB script provided in electronic form. M.D.M. conceptually designed the substrate profiling program for BVMOs as well as approaches to link biocatalyst performance with biochemical properties of enzymes. M.D.M., M.J.F. and F.R. co-wrote the manuscript. All authors commented on the manuscript.

Acknowledgements

We thank Prof. Dr. Peter Filzmoser with the Institute of Statistics and Probability Theory, Vienna University of Technology, for valuable advice and critical assessment of our work. M.J.F. is generously funded by EU-FP7 within the project OxyGreen (grant no. 212281). D.V.R. was funded by the Austrian Science Fund FWF (project number P-18945). At present D.V.R. is a staff member of Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

References

[1] I. Chorkendorff, J. W. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics*, Wiley-VCH, Weinheim, **2003**.

[2] J. Hagen, *Industrial Catalysis: A Practical Approach*, 2nd edn., Wiley-VCH, Weinheim, **2006**.

[3] T. Hudlicky, *Chem. Rev.* **2011**, *111*, 3995–3997.

[4] J. A. Gladysz, *Chem. Rev.* **2011**, *111*, 1167–1169.

[5] G. Bertrand, *Chem. Rev.* **2010**, *110*, 3851–3851.

[6] B. List, *Chem. Rev.* **2007**, *107*, 5413–5415.

[7] a) J.-L. Reymond, D. Wahler, *ChemBioChem* **2002**, *3*, 701–708; b) J.-P. Goddard, J.-L. Reymond, *J. Am. Chem. Soc.* **2004**, *126*, 11116–11117; c) J. Grognux, J.-L. Reymond, *ChemBioChem* **2004**, *5*, 826–831.

[8] a) B. G. Kyte, P. Rouviere, Q. Cheng, J. D. Stewart, *J. Org. Chem.* **2004**, *69*, 12–17; b) K. Kostichka, S. M. Thomas, K. J. Gibson, V. Nagarajan, Q. Cheng, *J. Bacteriol.* **2001**, *183*, 6478–6486.

[9] N. A. Donoghue, D. B. Norris, P. W. Trudgill, *Eur. J. Biochem.* **1976**, *63*, 175–192.

[10] M. J. Taschner, D. J. Black, *J. Am. Chem. Soc.* **1988**, *110*, 6892–6893.

[11] a) G. de Gonzalo, M. D. Mihovilovic, M. W. Fraaije, *ChemBioChem* **2010**, *11*, 2208–2231; b) H. Leisch, K. Morley, P. C. K. Lau, *Chem. Rev.* **2011**, *111*, 4165–4222; c) V. Alphand, R. Wohlgemuth, *Curr. Org. Chem.* **2010**, *14*, 1928–1965; d) M. D. Mihovilovic, *Curr. Org. Chem.* **2006**, *10*, 1265–1287.

[12] Y. C. J. Chen, O. P. Peoples, C. T. Walsh, *J. Bacteriol.* **1988**, *170*, 781–789.

[13] a) M. Griffin, P. W. Trudgill, *Eur. J. Biochem.* **1976**, *63*, 199–209; b) P. W. Trudgill, *Methods Enzymol.* **1990**, *188*, 77–81; c) H. Iwaki, Y. Hasegawa, S. Wang, M. M. Kayser, P. C. K. Lau, *Appl. Environ. Microbiol.* **2002**, *68*, 5671–5684.

[14] a) J. Rehdorf, M. D. Mihovilovic, U. T. Bornscheuer, *Angew. Chem.* **2010**, *122*, 4609–4611; *Angew. Chem. Int. Ed.* **2010**, *49*, 4506–4508; b) J. Rehdorf, M. D. Mihovilovic, M. W. Fraaije, U. T. Bornscheuer, *Chem. Eur. J.* **2010**, *16*, 9525–9535.

[15] CDMO from *Rhodococcus ruber* SC1, CHMO from *Acinetobacter calcoaceticus* NCIMB 9871, CHMO from *Arthrobacter* sp. BP2, CHMO from *Brachymonas petroleovorans*, two CHMOs from *Brevibacterium epidermis* HCU, CHMOs from *Rhodococcus pyridinovorans* Phi1 and *Rhodococcus ruber* Phi2, CHMO from *Xanthobacter* sp. ZL5 and CPMO from *Comamonas* sp. NCIMB 9872. References for isolation and cloning in recombinant hosts can be found in the Supporting Information.

[16] A. J. Albrecht, in: *Proceedings of the Joint SHARE, GUIDE and IBM Application Development Symposium*, Monterey, CA, **1979**, pp 83–92

[17] Detailed biotransformation data was partly obtained from published records and new data and can be found in the Supporting Information.

[18] a) M. D. Mihovilovic, F. Rudroff, B. Grötzl, P. Kapitan, R. Snajdrova, J. Rydz, R. Mach, *Angew. Chem.* **2005**, *117*, 3675–3679; *Angew. Chem. Int. Ed.* **2005**, *44*, 3609–3613; b) F. Rudroff, J. Rydz, F. H. Ogink, M. Fink, M. D. Mihovilovic, *Adv. Synth. Catal.* **2007**, *349*, 1436–1444; c) D. V. Rial, D. A. Bianchi, P. Kapitanova, A. Lengar, J. B. van Beilen, M. D. Mihovilovic, *Eur. J. Org. Chem.* **2008**, *2008*, 1203–1213; d) S. Wang, M. M. Kayser, H. Iwaki, P. C. K. Lau, *J. Mol. Catal. B: Enzym.* **2003**, *22*, 211–218; e) M. Vogel, U. Schwarz-Linek, in: *Bioorganic Chemistry. Highlights and New Aspects*, (Eds.: U. Diederichsen, T. K. Lindhorst, B. Westermann, L. A. Wessjohann), Wiley-VCH, Weinheim, **1999**, pp 102–110; f) M. D. Mihovilovic, G. Chen, S. Wang, B. Kyte, F. Rochon, M. M. Kayser, J. D. Stewart, *J. Org. Chem.* **2001**, *66*, 733–738; g) M. D. Mihovilovic, R. Snajdrova, B. Grötzl, *J. Mol. Catal. B: Enzym.* **2006**, *39*, 135–140; h) M. D. Mihovilovic, R. Snajdrova, A. Winninger, F. Rudroff, *Synlett* **2005**, 2751–2754; i) R. Snajdrova, I. Braun, T. Bach, K. Mereiter, M. D. Mihovilovic, *J. Org. Chem.* **2007**, *72*, 9597–9603; j) M. D. Mihovilovic, B. Muller, M. M. Kayser, P. Stanetty, *Synlett* **2002**, 700–702; k) J. D. Stewart, K. W. Reed, J. Zhu, G. Chen, M. M. Kayser, *J. Org. Chem.* **1996**, *61*, 7652–7653; l) D. V. Rial, P. Cernuchova, J. B. van Beilen, M. D. Mihovilovic, *J. Mol. Catal. B: Enzym.* **2008**, *50*, 61–68; m) J. Rehdorf, A. Lengar, U. T. Bornscheuer, M. D. Mihovilovic, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3739–3743; n) P. Cernuchova, M. D. Mihovilovic, *Org. Biomol. Chem.* **2007**, *5*, 1715–1719; o) M. D. Mihovilovic, P. Kapitan, P. Kapitanova, *ChemSusChem* **2008**, *1*, 143–148; p) M. D. Mihovilovic, F. Rudroff, B. Grotzl, P. Stanetty, *Eur. J. Org. Chem.* **2005**, 809–816; q) M. D. Mihovilovic, B. Muller, M. M. Kayser, J. D. Stewart, J. Frohlich, P. Stanetty, H. Spreitzer, *J. Mol. Catal. B: Enzym.* **2001**, *11*, 349–353; r) M. D. Mihovilovic, B. Groetzl, W. Kandioller, A. Muskotal, R. Snajdrova, F. Rudroff, H. Spreitzer, *Chem. Biodiversity* **2008**, *5*, 490–498.

[19] a) C. E. Shannon, in: *Key Papers in the Development of Information Theory*, (Ed.: D. Slepian), IEEE Press,

© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

New York, NY, **1974**; b) C. E. Shannon, *Bell Syst. Tech. J.* **1948**, *27*, 379–423; c) C. E. Shannon, *Bell Syst. Tech. J.* **1948**, *27*, 623–656.

[20] a) N. M. Kamerbeek, M. J. H. Moonen, J. G. M. van der Ven, W. J. H. van Berkel, M. W. Fraaije, D. B. Janssen, *Eur. J. Biochem.* **2001**, *268*, 2547–2557; b) M. D. Mihovilovic, P. Kapitan, J. Rydz, F. Rudroff, F. H. Ogink, M. W. Fraaije, *J. Mol. Catal. B: Enzym.* **2005**, *32*, 135–140; c) J. Rehdorf, C. L. Zimmer, U. T. Bornscheuer, *Appl. Environ. Microbiol.* **2009**, *75*, 3106–3114.

[21] M. W. Fraaije, J. Wu, D. P. H. M. Heuts, E. W. Hellemond, J. H. L. Spelberg, D. B. Janssen, *Appl. Microbiol. Biotechnol.* **2005**, *66*, 393–400.

[22] M. J. Fink, T. C. Fischer, F. Rudroff, H. Dudek, M. W. Fraaije, M. D. Mihovilovic, *J. Mol. Catal. B: Enzym.* **2011**, *73*, 9–16.

[23] M. D. Mihovilovic, D. A. Bianchi, F. Rudroff, *Chem. Commun.* **2006**, 3214–3216.

[24] a) R. Orru, H. M. Dudek, C. Martinoli, P. D. E. Torres, A. Royant, M. Weik, M. W. Fraaije, A. Mattevi, *J. Biol. Chem.* **2011**, *286*, 29284–29291; b) I. Polyak, M. T. Reetz, W. Thiel, *J. Am. Chem. Soc.* **2012**, *134*, 2732–2741.

[25] a) C. S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299; b) K. Faber, H. Hönig, A. Kleewein, *Selectivity 1.0*, a program for the calculation of the enantiomeric ratio *E*, available free of charge at the author's website: http://borgc185.kfunigraz.ac.at.

[26] H. L. van Beek, G. d. Gonzalo, M. W. Fraaije, *Chem. Commun.* **2012**, *48*, 3288–3290.

[27] http://www.ias.tuwien.ac.at/mdm/.

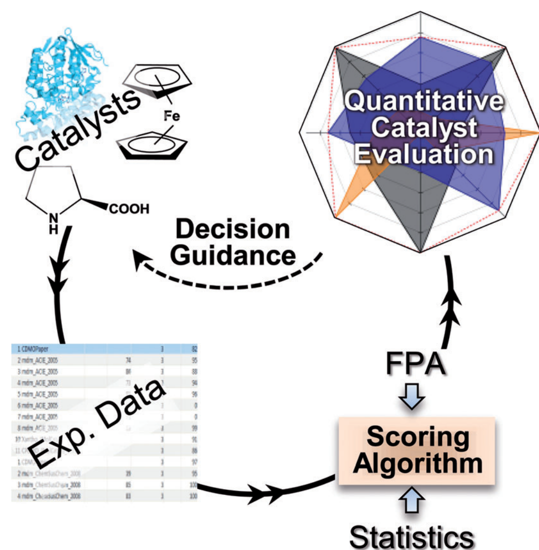
 © 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

These are not the final page numbers!

Quantitative Comparison of Chiral Catalysts Selectivity and Performance: A Generic Concept Illustrated with Cyclododecanone Monooxygenase as Baeyer–Villiger Biocatalyst

Adv. Synth. Catal. **2012**, *354*, 1 – 11

Michael J. Fink, Daniela V. Rial, Petra Kapitanova, Alenka Lengar, Jessica Rehdorf, Qiong Cheng, Florian Rudroff,* Marko D. Mihovilovic*



© 2012 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim