Development of 96 multiple injection-GC-MS technique and its application in protein engineering of natural and non-natural enzymatic reactions

Anja Knorrscheidt,[a] Pascal Püllmann,[a] Eugen Schell,[a] Dominik Homann,[a] Erik Freier[b] and Martin J.Weissenborn*[a, c]

Abstract: *Directed evolution requires the screening of enzyme libraries in biological matrices. Available assays are mostly substrate or enzyme specific. Chromatographic techniques like LC and GC overcome this limitation, but require long analysis times. The herein developed multiple injections in a single experimental run (MISER) using GC coupled to MS allows the injection of samples every 33 s resulting in 96-well microtiter plate analysis within 50 min. This technique is implementable in any GC-MS system with autosampling. Since the GC-MS is far less prone to ion suppression than LCMS, no chromatographic separation is required. This allows the utilisation of an internal standards and the detection of main and side-product. To prove the feasibility of the system in enzyme screening, two libraries were assessed: i) YfeX library in an E. coli whole cell system for the carbene-transfer reaction on indole revealing the novel axial ligand tryptophan, ii) a library of 616 chimeras of fungal unspecific peroxygenase (UPO) in S. cerevisiae supernatant for hydroxylation of tetralin resulting in novel constructs. The data quality and representation are automatically assessed by a new R-script.*

Introduction

Directed evolution mimics the natural selection process of Darwinian evolution in the laboratory. Random gene libraries are created and the corresponding enzyme variants are assessed for their altered and desired properties. The improved variants are thereafter submitted to further rounds of mutagenesis and screening.[1] The library size in a classical directed evolution approach consist of 500 – 2000 variants per round requiring a rapid analytical tool to quantify the activity of the different enzyme variants. Highly sensitive and robust assays are needed since starting activities are often below 10 turnovers (TON)[2] and the reactions need to be screened in complex biological matrices *i.e.* cell lysates, whole cell systems or cultivation supernatants.
The overwhelming majority of successfully employed enzyme assays relies on changes in absorption or fluorescence wavelengths during the course of the reaction of either participating model substrates or of additionally coupled enzyme cascade reactions.[3] To further expand the portfolio of assaying techniques, a flexible, substrate unspecific analytical tool would be of high interest since UV/vis or fluorescence-based assays suffer from the requirement of spectroscopically active compounds. These fit-for-purpose compounds, however, can differ substantially from the actual substrate of interest yielding enhanced variants for a differing substrate conversion.[4] The ideal assay would allow the screening with the exact substrate of interest, be highly sensitive as well as exceedingly reproducible.
A flexible and sensitive analytical technique is provided by chromatographic technologies such as liquid (LC) or gas chromatography (GC). While these techniques often provide the necessary sensitivity and are applicable to a wide range of substrates, they suffer from long analysis times preventing a high-throughput screening with several hundreds of samples. Developments in the last 15 years delivered faster chromatographic technologies, like the UHPLC (ultra high performance liquid chromatography) reducing analytical time by using sub-2-micron particle in their packed columns and high pressure of up to 1000 bar.[5]
A system capable of high-throughput GC analysis was introduced by Boeker *et al.*: the flow field thermal gradient gas chromatography (FF-TG-GC). The first prototype of an FF-TG-GC instrument was constructed in 2015 by combining the classical temperature gradient with an additional spatial temperature gradient.[6] This approach is based on theoretical calculations of heat transport processes and led to a fast analysis in less than 60 seconds for one sample finding its application for example at lowered temperatures for explosive substances.[7]
The analysis of multiple, overlapping samples in one chromatogram, coined multiplexing, has been introduced in 1967 by Izawa for GC[8] and was extended by Trapp and co-workers for a high-throughput approach. This technique bases on a high number of short injections in fast succession, in which the pattern for all sample injections is coded by a pseudo-random binary sequence derived from Hadamard matrices.[9] This pattern allows decoding of the resulting highly overlapping peaks in a so-called multiplexed chromatogram into single chromatograms and represents a very powerful technology. Trapp *et al.* were able to show a throughput enhancement factor of 38. The method relies on short but highly reproducible injections with regard to injection length, time and volume. They therefore employed a specifically

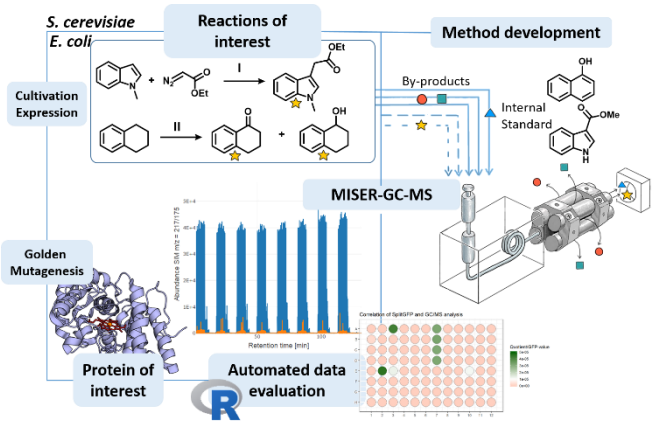

Figure 1. MISER-GC-MS and its implementation within directed evolution enabling investigation of two case studies: I) carbene-transfer reaction, and II) hydroxylation of 1,2,3,4-tetrahydronaphthalene.

[a] Anja Knorrscheidt, Pascal Püllmann, Eugen Schell, Dominik Homann, Jun.-Prof. Dr. Martin J. Weissenborn
Leibniz Institute of Plant Biochemistry
Bioorganic Chemistry
Weinberg 3, 06120 Halle (Saale), Germany
E-mail: martin.weissenborn@ipb-halle.de
[b] Dr. Erik Freier
Leibniz-Institut für Analytische Wissenschaften -ISAS- e.V.
Otto-Hahn-Str. 6b, 44227 Dortmund, Germany
[c] Jun.-Prof. Dr. Martin J. Weissenborn
Institute of Chemistry, Martin-Luther-University Halle-Wittenberg
Kurt-Mothes-Str. 2, 06120 Halle (Saale), Germany
Supporting information for this article is given via a link at the end of the document

built, highly-sophisticated injector system to achieve this throughput enhancement.[9a]

Welch *et al.* introduced MISER—multiple injections in a single experimental run—in 2010 as a method to enhance throughput in LC measurements without the need of special equipment or expert knowledge.[10] The MISER technique relies on injecting several different samples under isocratic conditions into one chromatography run requiring baseline separation of the consecutive peaks.[11] This enabled the performance of long chromatographic separations in a high-throughput manner as usually peak-free ("silent") areas of the chromatogram are used for multiple injections. While MISER was originally published as LC technique,[11] it has already been extended to LCMS[12] and was applied in kinetic measurements,[11] analysis of enantiopurity,[13] and determining ingredients in different matrices.[14]

As a screening device for directed evolution, we required a medium to high throughput (>500 samples/day) technique, which is sensitive—down to low µM product concentrations—and allows the analysis of a broad range of products. The method of MISER-GC-MS has not yet been employed for these applications or large sample numbers,[15] but its properties seem highly suitable for the application in directed evolution for several reasons: i) the quantification of target substances by GC-MS extracted from crude lysates, cell supernatants or whole cells is far less prone to ion suppression by matrix effects than by LC-MS systems,[16] ii) MISER allows a distinct throughput enhancement for GC without the requirement of special equipment, iii) it can be combined with internal standards to compensate for deviations in injection time, volume and length, or even in concentration due to sample evaporation, iv) MISER-GC-MS also enables the quantification of multiple molecules with overlapping peak areas, eliminating the need for chromatographical separation and therefore raising the throughput enhancement factor further and v) the MS-based detection is highly sensitive and allows low range detections.

Herein, we report the development of a new MISER-GC-MS strategy paving the way for an assay-independent platform for enzymatic reactions applicable for any GC-MS equipped with an autosampler (Figure 1). We have performed MISER-GC-MS in the screening of a non-natural carbene-transfer reaction in *Escherichia coli* (*E. coli*) leading to the discovery of tryptophan as an axial heme ligand. To demonstrate the versatility of the MISER-GC-MS system a second reaction in a different biological system has been investigated: The fungal unspecific peroxygenase (UPO) catalysing a hydroxylation reaction performed in yeast cell (*S. cerevisiae*) supernatant. In the latter example 672 transformants (including controls) have been screened within a day using a sample injection frequency of 33 s while quantifying three molecules in parallel and thereafter normalising the enzyme concentration with a split-GFP assay. To further facilitate the data evaluation, a freely available R-script has been written allowing the analysis, quality control and correlation to split-GFP signal in "one click".

Results and Discussion

Development of autosampler and GC-MS methods for highly reproducible 96-well analysis in biological matrices.

A standard autosampler (AOC-5000, Shimadzu, Kyoto, JP) was modified for fast injection application by decoupling the autosampler from the GC instrument. The read-out-signal of the GC was suppressed enabling independently controlling of the autosampler. The method development commenced by altering various conditions at the autosampler. The various settings like post-cleaning with isopropanol after each injection, filling speed (3 µL/s) and filling strokes for each microtiter plate led to significantly improved standard deviations from initially 51.6 %—no post-cleaning, 10 µL/s filling speed, no filling strokes—to 6.5 % (Table S2). The data were achieved using GC conditions of 230 °C and a split ratio of 60. The GC-MS system was then further developed for the best MISER conditions using the product of an enzymatic reaction of interest, ethyl 3-indoleacetate, as well as methyl indole-3-carboxylate as internal standard. MISER-GC-MS method development was performed employing three different ethyl 3-indoleacetate concentrations (20, 50 and 90 µM) in triplicates. Due to the continuous injection of up to 96 samples into one experimental run, the methods have to be performed under an isocratic temperature profile. Further parameters, which were investigated were the split ratio, the MS mode (SIM or Scan) and the injection interval (Figure 2, Table S3, Figure S6). The lowest oven temperature of 150 °C led to extensive peak tailing and poor baseline separation therefore leading to problems in automated area integration. Utilising an temperature of 190 °C excellent baseline separation and enhanced peak shape could be obtained as well as an improved standard deviation at 20 µM ethyl 3-indoleacetate from 8.6 % to 4.0 %. In order to position the injection peak, between different analytes peaks, the injection interval was altered from 67 s to 82 s. This further lowered the standard deviation and allowed the increase of the split ratio up

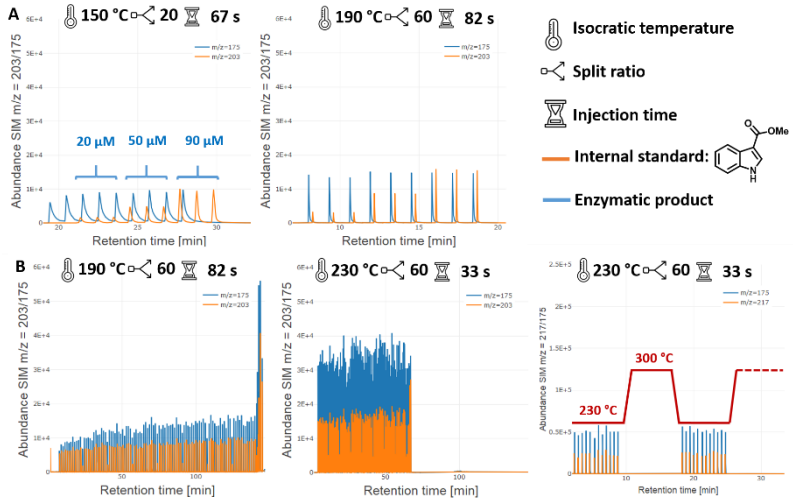

Figure 2. A) The method development with the analytes ethyl 3-indoleacetate (m/z 203) and methyl indole-3-carboxylate (m/z 175) with three different concentrations of ethyl 3-indoleacetate (20, 50 and 90 µM). **B)** Transferring the best conditions into 96-well format using a biological matrix (*E. coli* lysate). An alternative "stacked" method was developed with 3-(*N*-methyl-indole)acetate (m/z 217).

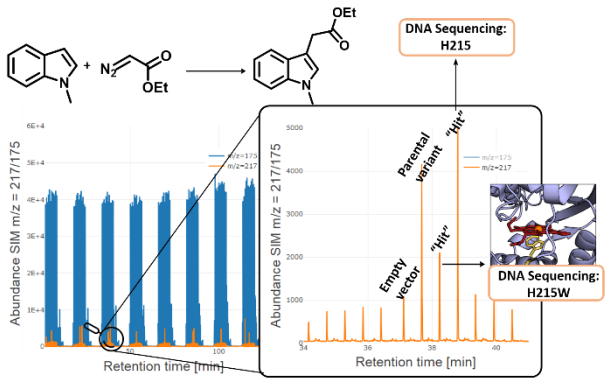

Figure 3. Case Study I: 96-well analysis by the previously developed stacked method for MISER-GC-MS. As controls the empty vector was included in column 6 and the parental YfeX variant in column 7, which harbours the axial ligand H215.

to 60 resulting in excellent standard deviations of 1.0 % for 20 µM ethyl 3-indoleacetate.

With these optimised analytical conditions in hand, we further challenged the system by using *E. coli* cell lysates spiked with ethyl 3-indoleacetate and analysed the resulting samples in 96-well experiments using methyl indole-3-carboxylate as the internal standard. As further quality control and for calibration purposes ethyl 3-indoleacetate (20 µM) extracted from a buffer system was injected after each microtiter plate row (12 samples) and after 96 wells calibration standards (2 x 20 µM, 2 x 50 µM, 2 x 90 µM) were injected as well. The system was assessed with the best conditions of the GC system (190 °C, split ratio 60, and 82 s injection) leading to a standard deviation of 4.0 % (Figure 2B, left). This could be even further improved to a standard deviation of 2.5 % by increasing the temperature to 230 °C that moreover allowed a shorter injection interval of 33 s (Figure 2B, middle, Table S6).

Since the developed methods shall be applied in the screening of non-natural enzyme activities, which in general suffer from low turnovers and require high substrate loading, the contamination of the GC-column could pose a substantial problem within 96 injections. Therefore an additional method, which includes a heating cycle after 12 injections (one microtiter plate row) has been developed as previously shown by Welch *et al*.[15b]

Case Study I: Screening of a YfeX enzyme library in an *E. coli* whole cell system for improved or retained activities of carbene transfer reactions revealed tryptophan as axial heme ligand

With these highly reproducible data in a biological matrix in hand, the new MISER-GC-MS technique was applied to YfeX in a whole cell screening of a focused mutant library. The dye-decolorizing peroxidase YfeX from *E. coli* was previously shown to perform non-natural carbene-transfer reactions such as carbonyl olefination[17] and C—H functionalisation.[18] The starting activity for the latter reaction was previously improved by an alanine scan within the active site.[18] We were subsequently interested in the influence of the axial ligand in the activity of YfeX on the C—H functionalisation reaction. The axial ligand complexes the heme iron and substantially influences its redox potential and electrophilicity and hence the overall activity of the occurring heme-carbenoid complex.[19] Amino acids having side chains, which are similar in size to wildtype histidine and harbour a lone electron pair—i.e. serine, threonine, cysteine—are preferred and others will most likely result in complete loss of activity. To study the axial ligand in YfeX, we saturated the axial ligand residue histidine 215—using Golden Mutagenesis[20]—and screened the resulting library for the occurrence of other functional axial ligands in whole cell reactions by stacked MISER-GC-MS using the carbene-transfer reaction on 1-methyl-indole. As a control for the background reaction cells harbouring the empty plasmid (pAGM22082) were included in this plate in column 6. In column 7 the parental YfeX variant was placed as positive control exhibiting H215 as natural axial heme ligand. To our delight, the results demonstrated the feasibility of MISER-GC-MS for the screening of this reaction. The parental YfeX variant could be clearly distinguished from the background reaction included as empty plasmid control (Figure 3). A new variant was identified, which carried a highly unusual tryptophan as axial heme ligand. To proof that the MISERgram revealed a “true positive” result, the corresponding plasmid was freshly transformed, expressed and purified by His-Tag chromatography. These results confirmed YfeX-H215W showing only slightly reduced activities compared to the parental variant (Figure S11).

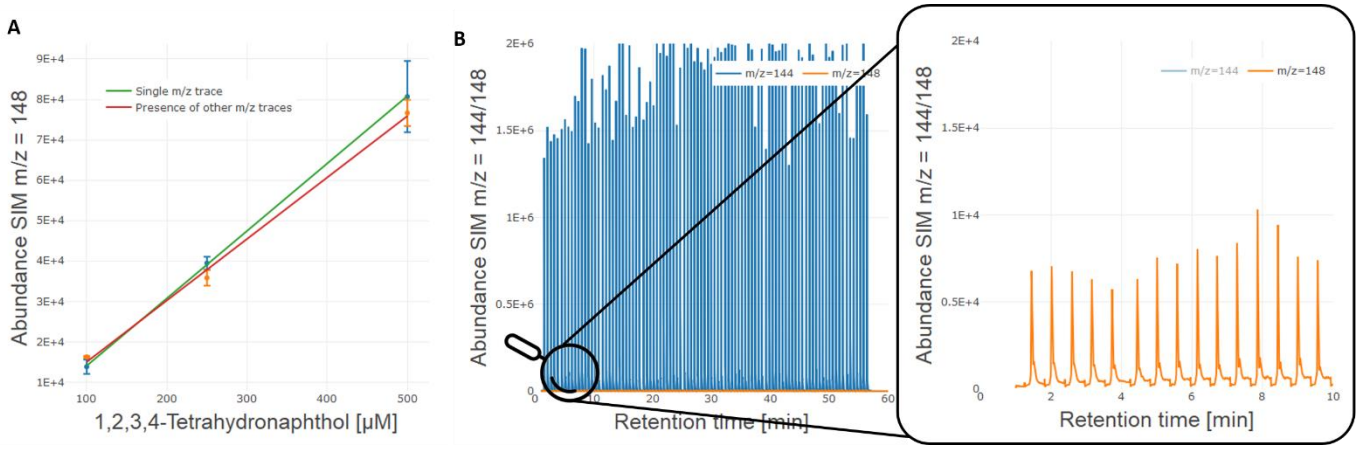

Figure 4. A) Extraction and mass spectrometric analysis of 1,2,3,4-tetrahydronaphthol in presence and absence of the other two compounds within a MISER experiment, **B)** MISER-Gram of the hydroxylation of 1,2,3,4-tetrahydronaphthalene with 96 biological replicates in microtiter plate format with a standard deviation of 9.7 %.

Case study II: Screening of 672 transformants of a fungal unspecific peroxygenase (UPO) library in S. cerevisiae supernatant analysing three analytes simultaneously

To demonstrate that the MISER-GC-MS method can be readily applied to other biological, enzyme as well as reaction systems, the screening was applied to a shuffled library (unpublished results) of fungal unspecific peroxygenases (UPO),[21] which can be heterologously secreted from *S. cerevisiae* and catalyse amongst other reactions the hydroxylation and further oxidation of 1,2,3,4-tetrahydro-naphthalene (tetralin). Whereas the previous case study proved to allow the simultaneous quantification of the internal standard and the product, we here aimed toward the analysis of three molecules with one sample injection every 33 s. To determine the feasibility of the quantification of three molecules by MISER-GC-MS the samples 1,2,3,4-tetrahydronaphthol, α-tetralone and 1-naphthol were injected in three different concentrations and the MS response compared to the obtained values when injecting only one analyte at a time (Figure 4, A). The data for each analyte revealed excellent correlation to the analysis with the simultaneous analysis of all molecules, hence allowing the desired simultaneous analysis of the three target molecules. One previously identified UPO variant was 96 times expressed (biological replicates) in microtiter plate format and the corresponding UPO containing supernatant tested for the conversion of tetralin and analysed by MISER-GC-MS. The aim was to validate that the MISER-GC-MS technology enabled the analysis of the entire plate with the individually expressed variant with an overall standard deviation of less than 10 %. We were delighted to see that the analysis of the entire 96 well plate within 48 minutes showed a standard deviation of only 9.7 % for the formation of 1,2,3,4-tetrahydro-naphthol (Figure 4, B). We now created a chimera library consisting of 243 possible gene combinations and screened 672 transformants produced and secreted in *S. cerevisiae* supernatant. In the seven different 96 well plates, four previously discovered novel, active UPO variants have been placed. All four variants could be identified within the screening as well as >30 newly identified hits (Figure 5).
In order to be able to deconvolute the actual protein amount from the protein activity within a single microtiter plate, each variant was equipped with a C-terminal fused split-GFP tag (GFP-11) resulting in a quantitative fluorescence signal upon recombination with GFP-1-10.[22] Correlation of GC-MS with split-GFP data revealed several hits showing a decreased secretion compared to PaDa-I as WT and decreased activity but yielding a comparable or even higher activity-to-protein-amount factor.
Due to the resulting huge amount of data and to ensure high quality and correct peak integration in each microtiter plate, an automatised R-script has been written. The script assesses the data from the MISERgram for reproducibility (based on internal standard) and based on the injection interval whether every peak has been correctly integrated or injected. All data are displayed as chromatogram were each *m/z* signal can be individually selected and zoomed in via hover-visualization. The quotient of internal standard and product peak is built and illustrated as bar chart and colour-coded microtiter plate for fast data evaluation. Lastly, the data are correlated to the split-GFP signal giving an expression independent activity signal. All data are shown in one final output html file.

Conclusion

The development of MISERGC-MS was shown for its application in directed evolution.

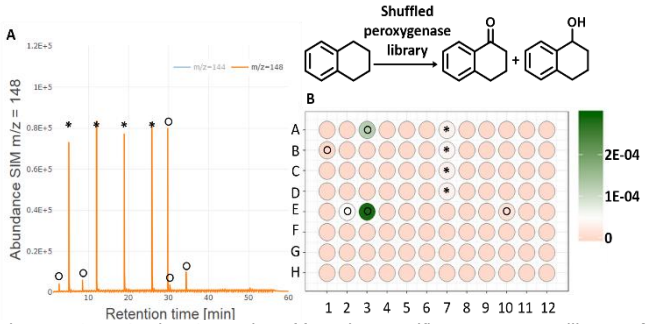

Figure 5. Case StudyII: Screening of fungal unspecific peroxygenase library. **A)** MISERgram of microtiter plate 1, **B)** automated R-script evaluation of plate 1(○=identified hit, ∗=PaDa-I).

Two methods were developed: the stacked method with particular relevance to low activities and large amounts of side-products including a cleaning step after every 12th injection to improve the quality of the acquired data. The second method for the screening of natural enzymatic reactions, which generally has only few side-products, was run with 33 s injection interval allowing microtiter plate analysis within 50 min. Both systems were employed by screening two different enzymes and reactions in two biological systems demonstrating the applicability of the MISER-GC-MS system. For the YfeX catalysed carbene-transfer reaction a highly unusual variant with tryptophan as an axial ligand could be identified. For the screening of the chimera library of fungal unspecific peroxygenase (UPO) the system allowed the analysis of seven plates within a day resulting in new peroxygenase chimeras as catalysts. The additional R-script provides the user with quick a quality control and data evaluation, also taking into account the protein quantification based on a split-GFP assay.
The demonstrated MISER-GC-MS technology can be implemented into any laboratory with a GC-MS equipped with an autosampler.

Acknowledgements

E.F. gratefully acknowledges the financial support by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and M.J.W, A.K., E.S. and E.F thank the Bundesministerium für Bildung und Forschung („Biotechnologie 2020+ Strukturvorhaben: Leibniz Research Cluster“, 031A360B and 031A360E) for generous funding. P.P. thanks the Landesgraduiertenförderung Sachsen-Anhalt for a PhD scholarship.
The authors thank Miguel Alcalde (CSIC Madrid) and Dirk Holtmann (DFI Frankfurt) for providing the peroxygenase genes and Benjamin Jones for substantial support on Figure 1.

Keywords: Directed Evolution; High throughput; protein engineering, low matrix effect, carbene transfer

[1] a) S. B. Kan, R. D. Lewis, K. Chen, F. H. Arnold, *Science* **2016**, *354*, 1048-1051; b) F. H. Arnold, *Angew. Chem.* **2018**, *130*, 4212-4218, *Angew. Chem. Int. Ed.* **2018**, *57*, 4143-4148; c) H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nat. Biotechol.* **1998**, *16*, 258-261; d) M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891-903; e) Y. Cen, W. Singh, M. Arkin, T. S. Moody, M. Huang, J. Zhou, Q. Wu, M. T. Reetz, *Nat. Commun.* **2019**, *10*, 3198.

[2] P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **2013**, *339*, 307-310.

[3] a) J. L. Reymond, V. S. Fluxa, N. Maillard, *Chem. Commun.* **2009**, 34-46; b) A. Holt, M. M. Palcic, *Nat. Protoc.* **2006**, *1*, 2498-2505.

[4] M. T. Reetz in *Directed Evolution of Selective Enzymes: Catalysts for Organic Chemistry and Biotechnology*, Wiley-VCH, **2016**.

[5] L. Nováková, H. Vlčková, *Anal. Chim. Acta* **2009**, *656*, 8-35.

[6] P. Boeker, J. Leppert, *Anal. Chem.* **2015**, *87*, 9033-9041.

[7] J. Leppert, M. Hartel, T. M. Klapotke, P. Boeker, *Anal. Chem.* **2018**, *90*, 8404-8411.

[8] K. Izawa, K. Furuta, T. Fujiwara, N. Suyama, *Ind. Chim. Belge* **1967**, *32*, 223-226.

[9] a) O. Trapp, Angew. Chem. **2007**, *119*, 5706, *Angew. Chem. Int. Ed.* **2007**, *46*, 5609-5613; b) O. Trapp, *J. Chrom A* **2010**, *1217*, 6640-6645; c) A. F. Siegle, O. Trapp, *Anal. Chem.* **2014**, *86*, 10828-10833; d) M. R. Wunsch, R. Lehnig, C. Janke, O. Trapp, *Anal. Chem.* **2018**, *90*, 9256-9263.

[10] C. J. Welch, X. Gong, W. Schafer, E. C. Pratt, T. Brkovic, Z. Pirzada, J. F. Cuff, B. Kosjek, *Tetrahedron: Asymmetry* **2010**, *21*, 1674-1681.

[11] C. J. Welch, X. Y. Gong, W. Schafer, E. C. Pratt, T. Brkovic, Z. Pirzada, J. F. Cuff, B. Kosjek, *Tetrahedron: Asymmetry* **2010**, *21*, 1674-1681.

[12] R. Papp, U. Andersson, L. D. Cantin, *J. Pharm. Biomed. Anal.* **2013**, *77*, 94-99.

[13] K. Zawatzky, M. Biba, E. L. Regalado, C. J. Welch, *J. Chrom. A* **2016**, *1429*, 374-379.

[14] a) C. J. Welch, E. L. Regalado, E. C. Welch, I. M. K. Eckert, C. Kraml, *Anal. Methods* **2014**, *6*, 857-862; b) C. Welch, E. Regalado, C. Kraml, E. Welch, M. Welch, H. Semmelhack, D. Almstead, A. Kress, N. Hidalgo, M. Kress, *LC GC N. Am.* **2015**, *33*, 262-269.

[15] a) J. P. Vistuba, M. Piovezan, M. G. Pizzolatti, A. M. Rebelo, M. S. Azevedo, L. Vitali, A. C. Costa, G. Amadeu Micke, *J. Chrom. A* **2013**, *1274*, 159-164; b) W. Schafer, H. Wang, C. J. Welch, *J Sep. Sci.* **2016**, *39*, 2978-2985.

[16] L. Silvestro, I. Tarcomnicu, S. R. Savu, in *Tandem Mass Spectrometry-Molecular Characterization*, IntechOpen, **2013**.

[17] M. J. Weissenborn, S. A. Low, N. Borlinghaus, M. Kuhn, S. Kummer, F. Rami, B. Plietker, B. Hauer, *ChemCatChem* **2016**, *8*, 1636-1640.

[18] K. J. Hock, A. Knorrscheidt, R. Hommelsheim, J. M. Ho, M. J. Weissenborn, R. M. Koenigs, *Angew. Chem.* **2019**, *131*, 3669-3673, *Angew. Chem. Int. Ed.* **2019**, *58*, 3630-3634.

[19] a) K. Chen, S. Q. Zhang, O. F. Brandenberg, X. Hong, F. H. Arnold, *J. Am. Chem. Soc.* **2018**, *140*, 16402-16407; b) T. Hayashi, M. Tinzl, T. Mori, U. Krengel, J. Proppe, J. Soetbeer, D. Klose, G. Jeschke, M. Reiher, D. Hilvert, *Nat. Catal.* **2018**, *1*, 578-584; cT. K. Hyster, F. H. Arnold, Isr. J. Chem. **2015**, *55*, 14-20.

[20] P. Püllmann, C. Ulpinnis, S. Marillonnet, R. Gruetzner, S. Neumann, M. J. Weissenborn, *Sci. Rep.* **2019**, *9*, 10932.

[21] a) R. Ullrich, J. Nuske, K. Scheibner, J. Spantzel, M. Hofrichter, *Appl. Environ. Microbiol.* **2004**, *70*, 4575-4581; b) P. Molina-Espeja, E. Garcia-Ruiz, D. Gonzalez-Perez, R. Ullrich, M. Hofrichter, M. Alcalde, *Appl. Environ. Microbiol.* **2014**, *80*, 3496-3507; c) M. Faiza, S. F. Huang, D. M. Lan, Y. H. Wang, *BMC Evol. Biol.* **2019**, *19*, 19; d) S. J. Willot, E. Fernandez-Fueyo, F. Tieves, M. Pesic, M. Alcalde, I. Arends, C. B. Park, F. Hollmann, *ACS Catal.* **2019**, *9*, 890-894.

[22] J. Santos-Aberturas, M. Dorr, G. S. Waldo, U. T. Bornscheuer, *Chem. Biol.* **2015**, *22*, 1406-1414.