University of Groningen

Redesign of Baeyer–Villiger Monooxygenases for Synthetic Applications

Dudek, Hanna M.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2013

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Dudek, H. M. (2013). *Redesign of Baeyer–Villiger Monooxygenases for Synthetic Applications*. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 04-06-2022

Chapter 1

General introduction and scope of the thesis

Protein engineering for biocatalysis

Biocatalysis employs enzymes for catalysing chemical reactions. Replacement of traditional chemical methods by the use of enzymes can be advantageous for many reasons. First, enzymes often show great selectivity. In the active site of the enzyme, protein environment dictates interactions of a substrate with the enzyme and thus the structure of a product. For this reason, enzymes can perform functionalisation of complex molecules at very specific sites. They also promote formation of a specific product over alternative products or stereoisomers. Therefore, biocatalytic processes may result in manufacture of highly pure products, which simplifies further processing. This is crucial in many cases, for example, in the pharmaceutical industry, where high standards for product purity have to be maintained. Next, enzymes operate mostly in water solutions, at moderate temperatures and pH, and under atmospheric pressure. Also, application of enzymes often eliminates the need for use of heavy-metal-containing catalysts, organic solvents, and toxic or dangerous compounds. Thus, it can help to increase process safety and reduce waste production. Notably, if needed, enzymes resistant to high temperatures, extreme pH values, or organic solvents, can be found in nature or engineered. Therefore, enzymes allow great versatility of operating conditions. Overall, application of biocatalysts may lead to more efficient and environmentally-friendly chemical processes.

In the quest for enzymes suitable for industrial processes, two approaches are pursued. First, researchers screen available enzyme collections and look for new enzymes in various environments, often characterised by extreme conditions (temperature, pH, salt concentration, or presence of pollutants). Second, it is possible to engineer enzymes according to the given needs. This approach is becoming more and more common due to developments in many fields including molecular biology, screening technologies, mechanistic studies, and structure–function prediction. Moreover, it is plausible that naturally occurring enzymes do not cover all required reactions and process conditions. Therefore, protein engineering is often the only way to obtain the desired biocatalyst. Traditionally, two paths have been followed within protein engineering: rational design and directed evolution (Figure 1). Rational design requires knowledge of the 3D-structure of the enzyme and understanding of the catalytic mechanism in order to select targets for mutagenesis and propose suitable changes in the enzyme (Figure 1A). Nowadays, computational methods, and in particular molecular dynamics simulations, are used to predict effects of mutations. On the extreme of the rational approach, there is *de novo* design of enzymes. This is a challenging methodology, and biocatalysts obtained so far present very low activities compared to existing enzymes.

On the other hand, directed evolution is a powerful method for modifying traits of proteins without detailed knowledge of the protein structure and mechanism.

Thanks to advances in molecular biology and activity-screening techniques, the idea of mimicking natural selection in a test tube has been brought into laboratory practice. Directed evolution has been applied, among others, to improve stability of enzymes, boost the activity, introduce new reaction types, and increase the selectivity. A clear advantage of directed evolution over rational engineering is no requirement for structural or mechanistic information. Directed evolution is based on creating diversity in a purely random manner and screening this diversity in order to find an improved variant, which becomes a template for the next engineering round (Figure 1B). The cycle of introducing random mutations and screening is repeated until a satisfactory change in the properties of an enzyme is achieved. While directed evolution requires little knowledge about the target protein, challenges lie in creating unbiased libraries and developing efficient screening methods.

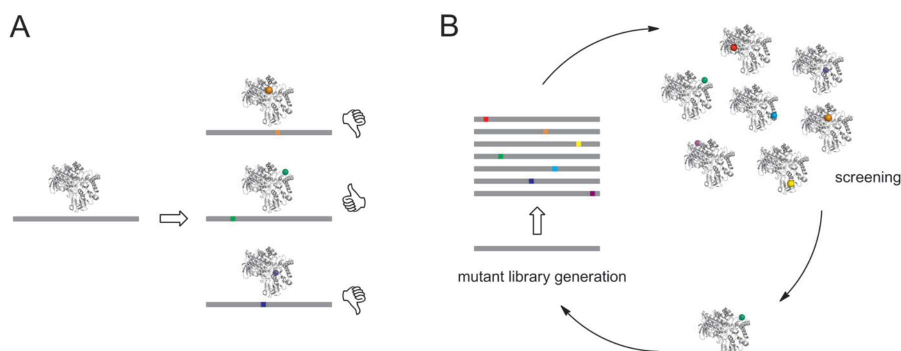

Figure 1. Rational design vs. directed evolution. A. In rational design, mutations are introduced at specific sites, and mutants are tested one by one for the desired improvements. B. In directed evolution, libraries of variants containing random mutations are generated and screened for mutants with improved properties, which then become templates for the next evolution round.

Generation of perfectly random mutant libraries, that is, libraries in which every amino acid can be changed to any other amino acid with equal probability, represents a difficulty. Error-prone PCR, a method commonly used to prepare random libraries, bears serious limitations due to the inherent DNA polymerases bias and the necessity of changing consecutive nucleotides to cover the full mutational spectrum. However, methods allowing generation of less and less biased libraries are being developed, for example, SeSaM (Mundhada et al., 2011). When it is desired to introduce mutations at specific positions of a protein, miscellaneous site-specific mutagenesis methods are available. Most of these techniques are based on the use of mutagenic primers (Ruff et al., 2013).

Testing a mutant library in which every position is fully randomised is an impossible task both in terms of library preparation and screening. Thus, the mutational space to be probed has to be limited. The key question is how to

narrow it down in order to match the screening method capacity while providing enough diversity to achieve the desired improvement. How many mutations should be on average introduced in a gene? Should they be randomly distributed, or should they cluster in certain regions of a protein? Smaller libraries are easier to handle, as with the number of variants in the library, the number of clones required to cover a certain portion of the library is increasing rapidly (Reetz et al., 2008). Also, most mutations have deleterious effect on the stability/activity, and introduction of many mutations can destabilise or inactivate an enzyme. On the other hand, when altering only a few positions in a protein, beneficial synergistic effects of mutations may be missed.

Table 1. Comparison of advantages and limitations of different screening assay formats.

Method	Throughput	Advantages	Limitations
GC, HPLC, MS, NMR	10^2–10^4	Analysis of enantioselectivity	Require specialised equipment Low throughput
Micro-titer plate	10^3–10^5	Accurate	Time- and labour-consuming
Solid phase	10^4–10^6	Easy handling Cost-efficient	Low accuracy
FACS	10^7–10^8	Ultra-high throughput	Low accuracy Detection based on fluorescence
Growth selection	$>10^8$	Ultra-high throughput	Difficult to set up

The availability of a good screening assay is an essential requirement for directed evolution. The assay should be reliable: it should allow identification of active variants ("hits") while no false positive or false negative hits should appear. Preferably, it should be time- and cost-efficient in order to screen large libraries. Various formats have been employed for assay development (Table 1). Assays in multi-well plates are commonly used in combination with colorimetric or fluorometric detection of product formation/substrate depletion. Multi-well-plate assays typically allow screening 10^3–10^5 variants (Tee and Schwaneberg, 2007). Chromatographic (GC, HPLC) and spectroscopic (MS, NMR) methods are being adapted for screening libraries. While they provide detailed information on the reaction yield and the (enantio)purity of the product, sample preparation and analysis in a high-throughput manner may be troublesome. Screening colonies grown on agar plates, paper filters, or membranes offers a throughput of at least one order of magnitude higher than multi-well plates (10^4–10^6, Tee and Schwaneberg, 2007). In this setup, colonies expressing active variants acquire coloration, which allows discrimination between active and inactive clones. Easy handling and typically low cost of this technique are compromised by a relatively low accuracy. However, a solid phase screen can be used as a pre-screen before using more precise, for example, multi-well-plate-based methods. Fluorescence-activated cell sorting

allows an impressive throughput of 10^7–10^8 clones screened within hours (Agresti et al., 2010). In this method, detection has to be based on fluorescence, which is an obvious limitation. As the detection is applied on single cells, variations in protein expression and cell physiology may affect the accuracy of the screening. Applying selection assays rather than screening is another powerful approach. In selection assays, only cells harbouring improved enzyme variant are able to grow and divide. This is achieved by, for instance, producing an essential nutrient or removing a toxic compound by the active enzyme variant. However, selection assays are difficult to establish and require careful optimisation. Also, they are typically specific for a certain compound and, therefore, have limited applicability.

From another point of view, screening assays can be performed using whole cells, cell extracts, (partially) purified protein, secreted/surface-exposed proteins, and cell-free systems which use mRNA display or cell-free protein synthesis. Again, different advantages and limitations are associated with each format, for example, some substrates may not enter intact cells while preparation of cell extracts is a laborious step. Using proteins secreted to the periplasmic space/culture medium or exposed on the cell surface relieves the substrate accessibility issue and separates the target reaction from the cell metabolism, but some proteins may not be exported in active form.

In summary, a wide range of screening assays is available, offering different throughput, accuracy, and complexity in terms of required equipment and sample preparation. Automation and downscaling thanks to, among others, application of microfluidic devices, are constantly helping to increase the throughput of screens. Advantages and limitations in screening capacity, amount of acquired information, and reliability are to be considered before choosing a screening method for a given task. A direct assay of limited throughput may be advantageous when multiple parameters have to be taken into account, including activity/stereoselectivity/side product formation, for example, in preparation of pharmaceutical products. Lastly, it is often highlighted that reaction conditions under which the screening is performed should reflect as much as possible the target process.

Recently, one can observe a shift of focus within the directed evolution field. Researchers tend to switch from completely random approaches (such as error-prone PCR) towards more rational concepts (Lutz, 2010). Such semi-random approaches allow design of small, focused libraries with increased frequency of positive hits (“smaller but smarter libraries”). Targeting a subset of active-site residues and using reduced codon sets instead of full randomisation helps to minimise screening effort and maximise the success rate. Knowledge-driven redesign helps to overcome the pitfalls of both rational design and directed evolution.

Many considerations have to be taken into account when designing a protein engineering experiment, and there are no general solutions. The choice of

a mutagenesis method and a screening method depends on the property to be changed, desired improvements, and available resources. In some of the most successful engineering campaigns which aimed at developing industrial processes for manufacturing pharmaceuticals (atorvastatin, sitaglitptin), various mutagenesis methods have been combined: error-prone PCR, site-directed mutagenesis, and gene shuffling (Fox et al., 2007; Savile et al., 2010). Finally, it is assumed that different paths can lead to the desired improvements in the protein characteristics.

Aims and outline of the thesis

Baeyer–Villiger monooxygenases (BVMOs) constitute an interesting group of monooxygenases, which are capable of inserting an oxygen atom into a C–C bond and performing stereoselective heteroatom oxidations. Despite many exciting advances in the research on BVMOs, several challenges remain. In particular, applications of BVMOs in industrial-scale synthesis require improvements of activity and selectivity of these enzymes. Phenylacetone monooxygenase (PAMO) is a prototype BVMO, and it is a convenient model for engineering studies since the crystal structure and the catalytic mechanism of this enzyme were solved. PAMO is also thermostable, resistant to organic solvents, and well expressed in *Escherichia coli*, while its limited substrate scope represents a serious drawback. The goal of the research described in this thesis was threefold. First, by various site-directed and semi-random methods, we aimed at improving activity of PAMO with different substrates. Second, we strove to develop new tools for effective engineering of BVMOs. Third, we sought to increase the knowledge on BVMOs, which would aid future redesign studies of these enzymes.

Chapter 2 provides an overview of the developments in the BVMO-related research in the last 10 years. In particular, newly cloned enzymes, structural studies, approaches for coenzyme regeneration, protein engineering efforts, and applications of BVMOs in organic synthesis are discussed.

Most Type I BVMOs, including PAMO, present strict preference towards NADPH over NADH. In **Chapter 3**, the effect of several mutations on the coenzyme specificity of PAMO was probed. Attempts to increase the activity of PAMO with NADH and, therefore, to change its coenzyme specificity are presented.

Chapter 4 describes a mutational analysis of the predicted substrate binding site of PAMO. The structure of PAMO was compared with a model of cyclopentanone monooxygenase. Based on the comparative analysis, 15 positions potentially important for the substrate recognition were selected. Activity, regio-, and enantioselectivity of thirty single and multiple mutants with several substrates were investigated. These results indicated new hot-spots determining the substrate specificity of PAMO.

An efficient activity-screening method forms the basis of every directed evolution project. In **Chapter 5**, a novel method for screening BVMOs is introduced. This method allows screening libraries in whole-cell format, and it can be applied in combination with any substrate. Periplasmic expression of PAMO and application of phosphate as an indirect reporter of BVMO activity resulted in a generic and reliable screening method. Evaluation of this method by using model reactions as well as screening small libraries is presented.

The conclusions from **Chapter 4** as well as the analysis of the structure of PAMO in complex with $NADP^+$ allowed us to design a library of PAMO mutants. The library was screened using the new, phosphate-based method (**Chapter 5**), which led to the identification of a mutant with expanded substrate acceptance profile. Design, construction, and screening of the library as well as characterisation of the isolated mutant are described in **Chapter 6**.

References

Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.C., Marquez, M., Klibanov, A.M., Griffiths, A.D., and Weitz, D.A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proc. Natl. Acad. Sci. USA *107*, 4004–4009.

Fox, R.J., Davis, S.C., Mundorff, E.C., Newman, L.M., Gavrilovic, V., Ma, S.K., Chung, L.M., Ching, C., Tam, S., Muley, S., et al. (2007). Improving catalytic function by ProSAR-driven enzyme evolution. Nat. Biotechnol. *25*, 338–344.

Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. Curr. Opin. Biotechnol. *21*, 734–743.

Mundhada, H., Marienhagen, J., Scacioc, A., Schenk, A., Roccatano, D., and Schwaneberg, U. (2011). SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR. ChemBioChem *12*, 1595–1601.

Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. ChemBioChem *9*, 1797–1804.

Ruff, A.J., Dennig, A., and Schwaneberg, U. (2013). To get what we aim for—progress in diversity generation methods. FEBS J. *280*, 2961–2978.

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., et al. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science *329*, 305–309.

Tee, K.L., and Schwaneberg, U. (2007). Directed evolution of oxygenases: Screening systems, success stories and challenges. Comb. Chem. High Throughput Screen. *10*, 197–217.