Chapter 1.
Introduction

Part I: Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability

Robert J. Floor, Hein J. Wijma and Dick B. Janssen

Protein engineering strategies for increasing stability can be improved by replacing random mutagenesis and high-throughput screening by approaches that include bioinformatics and computational design. Mutations can be focused on regions in the structure that are most flexible and involved in the early steps of thermal unfolding. Sequence analysis can often predict the position and nature of stabilizing mutations, and may allow the reconstruction of thermostable ancestral sequences. Various computational tools make it possible to design stabilizing features, such as hydrophobic clusters and surface charges. Different methods for designing chimeric enzymes can also support the engineering of more stable proteins without the need of high-throughput screening.

Part of this chapter has been published:

*Current Opinion in Structural Biology* 2013, (23), 588–594
Introduction

Stability under practical conditions is one of the most critical properties when biotechnological applications of proteins are explored\[1\]. High thermostability is often accompanied by good performance under unfavorable conditions, such as the presence of cosolvents. Improved thermal stability may also correlate with higher expression yields in heterologous hosts, improved long-term survival under mild conditions, increased ability to remain active in non-aqueous solvents, and a longer half-life under harsh industrial process conditions\[2\]. For medicinal proteins, thermostability may correlate with high serum survival times\[3\]. In protein engineering, many function-gaining mutations, such as modified enzyme selectivity, diminish stability. As a result, thermostable proteins can tolerate a larger number of mutations than mesostable proteins\[4\], and give better results when used as a starting point in protein engineering\[5\]. For these reasons, and because thermostable variants often are not available from the natural biodiversity, increasing thermostability is an important goal of protein engineering.

This review focuses on recent developments in protein engineering aimed at enhancing thermostability. A number of well-established techniques in protein thermostabilization by mutagenesis have been developed, providing important insights into the structural features that govern thermostability. These principles and examples have been reviewed elsewhere\[4, 6\]. On the other hand, directed evolution has emerged as a powerful approach, especially in the absence of structural information\[6b, 6e\]. However, when high-throughput screening is not possible, random approaches that are frequently used in directed evolution may become inefficient. This triggers a quest for more focused methods, for example, by incorporating information provided by bioinformatics and structural analysis. The first section of this review highlights recent progress in identifying target positions for mutagenesis and substitutions that enhance stability. The next sections describe recent advances in protein stabilization by employing sequence comparisons, computational design, loop grafting and chimera building.

Locating critical sites for stability improvement

Mutations that strongly enhance protein thermostability often appear to cluster in a particular region of the protein, whereas similar substitutions introduced elsewhere in the protein may have a negligible effect on thermostability\[6a\]. The explanation for this phenomenon is that the strongly stabilizing mutations are located at a spot where the protein starts to unfold\[6a\]. Mutations that improve local structural stability also improve global stability of a protein, especially
Introduction

when they are positioned in a region that unfolds at a critical step in a kinetically controlled unfolding pathway. However, the pathways of unfolding are usually unknown, making a knowledge-driven approach difficult. Yet, several methods have been explored to predict which positions are most likely to contribute to thermostability and should be targeted by mutagenesis\[7\].

An approach that employs crystallographic information to identify target positions is the B-fit method\[7a\], in which residues with the highest crystallographic B-factors are selected for mutagenesis. The rationale is that highly flexible residues are more likely to unfold in an early stage, and that substitutions at these positions have a higher chance to stabilize the folded state. Several successes have been reported (Table 1-1) \[7b-e\]. A recent investigation of an esterase that was improved by the B-fit method showed that the observed increase in temperature at which 50% of the enzyme was inactivated by heating was in fact due to improved refolding after heat treatment\[7f\]. Whereas the mutant esterase can refold after heating (95°C), it unfolds at an even lower temperature than the wild-type esterase, suggesting that the B-fit approach can also yield substitutions that increase reversibility of unfolding rather than enhancing thermostability.

Besides crystallographic B-factors, which may be influenced by crystal contacts and solvent conditions, various theoretical methods can be used to predict flexible regions that may serve as targets for stabilization. A recent example is provided by the stabilization of xylanase, for which flexible residues were located either by MD simulations at different temperatures or by molecular mechanics based flexibility predictions\[8\] (http://www.flexweb.asu). Using the molecular mechanics based flexibility analysis, computationally designed point mutations were introduced at 8 positions in the sequence of the 185 aa protein, which resulted in an improvement of the $T_{M,app}$ by 2°C and a 15-fold longer half-life at 50°C\[9\]. Employing the results of MD simulations and computational design, five different mutations were incorporated in the same xylanase, which gave a 4°C higher apparent melting temperature and a 30-fold longer half-life\[10\].

Constraint Network Analysis (CNA) can provide insight in the unfolding mechanisms of proteins, including the identification of residues involved in critical unfolding steps\[11-1\]. In CNA, hydrogen bonds act as constraints of the structure and residues that are connected by a continuous network of hydrogen bonds are considered to be folded. The strength of each H-bond is calculated after which the H-bonds are gradually removed to simulate an increasing temperature. The CNA simulations indicated that the disappearance of particular H-bonds signals the abrupt disintegration of the original interaction cluster into
much smaller clusters, which corresponds to melting of the protein. The H-bonding residues involved in the collapse of the cluster of citrate synthase, isopropylmalate dehydrogenase, and thermolysin-like protease agreed with the residues that are known to be critical for thermostability improvement.

<table>
<thead>
<tr>
<th>Targeted protein</th>
<th>T_{M,app} increase (°C)</th>
<th>Part of protein targeted by mutagenesis (%)</th>
<th>Origin of stabilizing mutations</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>esterase</td>
<td>+9^A</td>
<td>1</td>
<td>consensus library</td>
<td>[7b]</td>
</tr>
<tr>
<td>epoxide hydrolase</td>
<td>+21^A</td>
<td>3</td>
<td>NDT codon library</td>
<td>[7c]</td>
</tr>
<tr>
<td>α-amino ester hydrolase</td>
<td>+7^A</td>
<td>3</td>
<td>consensus design</td>
<td>[7d]</td>
</tr>
<tr>
<td>heme peroxidase</td>
<td>+4</td>
<td>3</td>
<td>computational design</td>
<td>[7e]</td>
</tr>
<tr>
<td>xylanase</td>
<td>+2</td>
<td>4</td>
<td>computational design</td>
<td>[9]</td>
</tr>
<tr>
<td>xylanase</td>
<td>+5</td>
<td>3</td>
<td>computational design</td>
<td>[10]</td>
</tr>
<tr>
<td>haloalkane dehalogenase</td>
<td>+18</td>
<td>4</td>
<td>random and site-directed</td>
<td>[12]</td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>+7</td>
<td>3</td>
<td>site-directed mutagenesis</td>
<td>[13]</td>
</tr>
<tr>
<td>thioredoxin</td>
<td>+32</td>
<td>NA</td>
<td>ancestral method</td>
<td>[14]</td>
</tr>
<tr>
<td>triosephosphate isomerase</td>
<td>+8</td>
<td>8</td>
<td>ancestral method</td>
<td>[15]</td>
</tr>
<tr>
<td>β-amy lase</td>
<td>+9</td>
<td>4</td>
<td>ancestral method</td>
<td>[16]</td>
</tr>
<tr>
<td>acylphosphatase GTPase</td>
<td>+9</td>
<td>5</td>
<td>computational design</td>
<td>[17]</td>
</tr>
<tr>
<td>methionine amino peptidase</td>
<td>+18</td>
<td>2^D</td>
<td>computational design</td>
<td>[18]</td>
</tr>
<tr>
<td>xylanase</td>
<td>+8</td>
<td>3</td>
<td>computational design and consensus</td>
<td>[19]</td>
</tr>
<tr>
<td>cellophylhydrase</td>
<td>+9^A</td>
<td>10</td>
<td></td>
<td>[20]</td>
</tr>
<tr>
<td>methyl parathion hydrolase</td>
<td>+12</td>
<td>3</td>
<td>computational design</td>
<td>[21]</td>
</tr>
<tr>
<td>sesquiterpenoid synthase</td>
<td>+45</td>
<td>2</td>
<td>computational design</td>
<td>[22]</td>
</tr>
<tr>
<td>Baeyer-Villiger monooxygenase</td>
<td>+16^B</td>
<td>NA</td>
<td>chimeric enzyme</td>
<td>[23]</td>
</tr>
<tr>
<td>aldo-keto reductase</td>
<td>-</td>
<td>NA</td>
<td>chimeric enzyme</td>
<td>[24]</td>
</tr>
<tr>
<td>polyol dehydrogenases</td>
<td>+12^B</td>
<td>NA</td>
<td>chimeric enzyme</td>
<td>[25]</td>
</tr>
<tr>
<td>cellulase</td>
<td>+8.5^AC</td>
<td>NA</td>
<td>SCHEMA</td>
<td>[26]</td>
</tr>
<tr>
<td>cytochrome P450</td>
<td>+14.5^AC</td>
<td>NA</td>
<td>chimeric enzyme</td>
<td>[27]</td>
</tr>
<tr>
<td>green fluorescent protein</td>
<td>+5</td>
<td>NA</td>
<td>circularization</td>
<td>[28]</td>
</tr>
</tbody>
</table>

^A Based on 50 % residual activity after heat treatment, values of which can differ extremely from the T_{M,app} measured by other techniques. ^B Compared to the melting temperature of the thermolabile parent of the chimera. ^C Compared to the melting temperature of the thermostable parent of the chimera. ^D Only residues near void in the hydrophobic core were targeted.

Recently, two groups reported that the mutation of residues at a tunnel to the enzyme active site provided highly stabilized variants. For a haloalkane dehalogenase it was possible to increase the T_{M,app} by 19°C using this method.
yielding an enzyme variant that was also less inhibited by DMSO, which was attributed to more difficult entrance to the active site. For a β-galactosidase, mutation of tunnel residues yielded an enzyme with a 10°C increase of the optimal temperature for catalytic activity. The authors concluded that selective targeting of the active site entrance tunnel should be a broadly applicable strategy to stabilize enzymes[12]. A tunnel to an active site may need to be flexible in order to allow substrate to enter the active site, which can explain the success of this method.

**The ancestral and the consensus methods**
The ancestral stabilization method allows for the stabilization of proteins without the use of structural information. The method is based on the hypothesis that far ancestors of current organisms were thermophiles that possessed proteins, which were more thermostable than extant homologs. Thus, mutations for improving stability are derived from predicted sequences of common ancestors of a protein family. This method has been developed from 1990 onwards[29] and was successfully used to stabilize several proteins. Recent progress includes a study on the stabilization of the B-subunit of the DNA gyrase, in which the ancestral method was compared to the consensus method[30]. The consensus method replaces amino acids by residues that are most frequently present in multiple sequence alignments. The constructed ancestral protein was 4–5 °C more stable ($T_{M,app}$) compared to the protein created on basis of the consensus method. The ancestral method was also used for recreating a thioredoxin that should have occurred in organisms that existed about a billion years ago, instead of millions of years ago as explored used in other studies[14]. The proposed recreated precambrian thioredoxin variants were 32°C more stable in $T_{M,app}$ than modern thioredoxins and were more active under acidic conditions which supposedly existed in the precambrian era.

The half-life of activity at 85°C of a thermostable cellulase was increased 14-fold using the consensus method. Interestingly, most protein sequences that were used to create the consensus alignment originated from mesostable organisms, but they could still contribute to the stability of a thermostable protein[31]. Consensus mutations were also shown to enhance the evolvability of a Kemp eliminase. Such mutations were included alongside functional mutations to counteracting destabilizing mutations that were selected because they contributed to the function, in this case the Kemp elimination[32]. Without such stabilizing mutations, it was found to be impossible to evolve a computationally designed Kemp eliminase.
It is possible to increase the success rate of the ancestral and consensus methods by including structural information. As an example, consensus mutations combined with structural information on B-factors and inter-subunit disulphide bonds, led to a 7°C increase in the half-life of activity of an α-amino ester hydrolase\textsuperscript{[7d]}. For the ancestral method, a study on the stabilization of a triosephosphate isomerase introduced two new parameters (residues that are correlated to other residues and residues that were nearly invariant and therefore might have ‘hidden correlations’ to other residues) to exclude residues from being targeted for mutagenesis\textsuperscript{[15]}. Using this approach 90% of the predicted mutations were found to be stabilizing, leading to a final variant that was 8°C more stable in $T_{M,app}$ whereas a variant which was designed without use of these extra parameters was 2°C less stable. A study on the stabilization of a β-amylase combined the ancestral method with structural information and introduced a parameter to account for the extent to which the residues surrounding the target residue are conserved\textsuperscript{[16]}. This parameter correlated well with the stability effects of substitutions.

**Computational design to stabilize enzymes**

With computational protein design, the effect of mutations is calculated based on the 3D-structure of the protein. An appropriate energy function implemented in the algorithm is used to discover mutations that stabilize the folded protein structure. The use of computational design can reduce the screening effort required to find a desired stability enhancement by orders of magnitude as compared to random directed evolution methods. Two different approaches are currently employed in the computational design of thermostable enzymes. Mutations are either selected on the basis of their predicted effect on the overall folding energy (ΔΔ$G^{\text{fold}}$), or computational tools are applied to improve particular features that stabilize proteins.

Several methods can improve stabilizing features in a protein, such as loops, hydrophobic clusters, and surface charges. In all these approaches, the energy function within the software is used to verify that the designed protein structure is lower in energy than the native structure. One approach is the computational design of a favorable network of positive and negative charges on the protein surface. This way, it was possible to increase the $T_{M,app}$ of an acylphosphatase and a GTPase by +9°C\textsuperscript{[17]}. By developing a design algorithm that used the Rosetta energy function\textsuperscript{[33]} to find the most favorable positions to introduce mutations that increase the net charge of a protein (supercharging\textsuperscript{[34]}), it was possible to increase the heat-survival of an antibody\textsuperscript{[35]}. The stabilized
antibodies were also monomeric in solution, unlike the native antibody, which formed aggregates. A commonly used computational approach involves improving the hydrophobic packing interactions between buried protein residues\(^{9-10,18-19}\). Software to detect hydrophobic networks was used to analyze hydrophobic clusters in different xylanases. Introduction of promising mutations into the target xylanase gave three mutations which increased the \(T_{M,\text{app}}\) by 8°C\(^{19}\). An algorithm to introduce hydrophobic mutations to fill protein voids was experimentally demonstrated to give a \(T_{M,\text{app}}\) increase of 18°C when applied to a methionine amino peptidase\(^{18}\).

The second approach for using computational design to enhance stability does not aim to improve one particular type of interaction but generates all types of mutations that should be favorable for stability according to the implemented energy function. This work is mostly done with energy functions that are specifically parameterized to predict differences in \(\Delta G_{\text{fold}}\) due to point mutations. Examples of suitable algorithms include FoldX\(^{36}\), PoPMuSiC\(^{37}\), and PreTherMut\(^{21}\). Application of FoldX in combination with consensus design resulted in an improvement of the \(T_{M,\text{app}}\) of a cellobiohydrolase by 9°C\(^{20}\). For a feruloyl esterase, the \(T_{M,\text{app}}\) did not increase by mutations that were designed by PoPMuSiC, but the half-life of the enzyme at 50°C was improved 10-fold\(^{38}\). With PreTherMut, the \(T_{M,\text{app}}\) of a methyl parathion hydrolase was improved by 12 °C \(^{39}\). Recently, using a similar algorithm, the stability \((T_{M,\text{app}})\) of a sesquiterpenoid synthase was increased by a phenomenal 45°C\(^{22}\). However, the resulting enzyme gave more side-products and the catalytic activity was reduced.

**Chimeric enzymes, truncation and circularization**

Thermostability can also be enhanced by making chimeric proteins consisting of subdomains of different proteins. This way, favorable features from different parents can be combined. For example, chimeric variants of Baeyer–Villiger monoxygenases (BVMOs) were obtained by combining subdomains from BVMOs with different stabilities. Some chimeric enzymes were 10–15°C more stable than the thermolabile parent enzymes, while retaining or even improving the substrate selectivity of a thermolabile BVMO\(^{23}\). Several stabilizing chimeras have been created of different xylanases\(^{40}\) and of xylanases and glucanases\(^{41}\). A chimera created by grafting loops from an archaeal enzyme onto a human aldo-keto reductase showed that it is possible to combine favorable features, in this case stability and activity, of archaeal and human proteins\(^{24}\). Such a stabilized chimera can even be obtained without using a thermostable variant, as demonstrated by combining domains of two mesostable polyol
dehydrogenases[25]. The resulting chimera was 12°C more stable than its parents. Finally, a chimeric enzyme can also be constructed by proteins with a low sequence similarity (<30%), as was shown by replacing parts of a (βα)₈-barrel protein by parts of flavodoxin-like proteins[42]. The replacement was possible since the structure of flavodoxin-like proteins resembles a part of the (βα)₈-barrel. The resulting chimera between the two proteins adopted a fold strongly resembling the (βα)₈-barrel scaffold and was thermostable.

A sophisticated method for creating such chimeras is the SCHEMA approach[43], which uses an algorithm to calculate the most favorable crossover positions between subdomains. The calculation selects positions that minimize the number of interactions that are broken when combining subdomains. The method was applied to the stabilization of a cellulase by recombination of an enzyme from three different parents[26]. By analyzing and modeling the thermostabilities of 54 characterized members of a family of fungal cellobiohydrolases, one of the subdomains appeared critical for stability and its replacement led to an 8.5°C increase in apparent melting temperature ($T_{M,app}$). In another study, the SCHEMA method was combined with an active learning algorithm, which analyzed data from a training set of chimeras to predict a large collection of stabilizing chimeras[44]. This method revealed a negative correlation between the isoelectric point of a chimeric enzyme and its long-term stability, which was used to construct enzymes with enhanced long-term stability.

Recently, a chimeric P450 enzyme was created which was 14.5°C more stable compared to its parent[27]. This thermostable enzyme was created by modeling the protein fitness landscape and constructing the most stabilizing variants. Bayesian statistics were used to predict, to a reasonable extent, the mean and variance fitness of any sequence.

Other possible stabilization methods are truncation and cyclization. It is well known that the removal of thermolabile parts or even whole domains from a protein can significantly improve its stability[45]. Recent achievements include the stabilization of an endo-β-1,4-glucanase by removing a cellulose binding domain, which in this case is not needed for biotechnological use of the enzyme. This truncation resulted in a threefold increase in half-life of the protein at 65 °C, while the catalytic activity increased. Circularization of the proteins sequence can also enhance stability. The N-terminus and C-terminus are often the most flexible part of a protein and easily become the targets of proteolytic enzymes. Therefore, connecting the C-terminus and N-terminus can stabilize a protein. Such a circularized protein can be created by introducing the protein of interest between parts of a naturally split intein from the dnaE gene[46]. This method was
used to circularize a GFP mutant, resulting in a 5°C more stable protein \cite{28}. This circularized protein was also highly resistant to proteolysis.

**Conclusions and outlook**

Even though directed evolution is a highly successful method for improving protein stability, there is a demand for additional powerful techniques, especially in case of proteins that cannot be expressed and assayed in high-throughput format \cite{47}. Sequence-based bioinformatics approaches and computational methods exploiting protein structures are especially successful. When integrated into directed evolution protocols, these newer methods can boost the efficiency and reliability of directed evolution protocols. For example, the B-fit approach and the consensus method assist the design of small sets of variants (or smart libraries) from which improved enzymes may be obtained by screening only small numbers. Furthermore, when three-dimensional structures are available, energy calculations may identify substitutions that improve hydrophobic cores, or introduce surface charges that prevent aggregation. This again leads to smaller numbers of mutants to be tested experimentally, and a high abundance of desired variants among those sets.

![Figure 1-1](image.png)

**Figure 1-1.** $T_{M,\text{app}}$ improvements obtained by protein engineering of enzymes from 2007 to 2012. Primary publications were searched on Web of Science using different key word combinations to find protein engineering studies aimed at increasing enzyme thermostability. If a publication was a follow-up study in which an earlier stabilized enzyme was further improved, the additional stabilization is plotted.
An important remaining challenge is to obtain large improvements of thermostability. Currently, the increase in thermostability by mutagenesis typically amounts to 2–15°C in $T_{M,\text{app}}$ (Figure 1-1). For many applications this is acceptable because even a modest increase often corresponds to a >10-fold longer life-time (e.g. [9-10,48]). However, the differences in $T_{M,\text{app}}$ between natural enzymes that display hyperthermostability and their mesostable homologs are much larger, indicating that there is still a lot of room for improving engineering techniques.
Part II:  

Rational design of enzyme enantioselectivity  

Enantioselective enzymes can be used in the biocatalytic production of enantiopure chiral compounds, either by kinetic resolution of racemic mixtures or in asymmetric synthesis from prochiral precursors. The degree of enantiodiscrimination of enzymes can be improved by protein engineering, for which directed evolution employing global or localized random or semi-random mutagenesis has emerged as a powerful strategy. However, directed evolution projects usually require screening of a large number of mutants and suffer from poor predictability of the final outcome. This review discusses rational approaches that aim to reduce the required screening effort, either by integrating sequence information in library design or by using protein structure information to preselect promising mutations. We discuss various methods for computational prediction of mutations that improve enantioselectivity or that can rank enzymes variants according to their probability of carrying out stereoselective conversions. The biotechnological implications of these approaches are discussed.
Methods to improve enzyme enantioselectivity

The capability of enzymes to catalyze stereoselective reactions makes it attractive to apply them in the preparation of chiral pharmaceutical products[49]. Since a high enantiomeric excess is usually required, the degree of stereodiscrimination is a key property that may allow or prohibit the application of an enzyme in a biocatalytic production process. Fortunately, protein engineering methods have been developed that make it possible to enhance or even invert the enantioselectivity of enzymes[50]. The most successful protein engineering technique to increase enzyme enantioselectivity is directed evolution, which uses several cycles of (semi)-random mutagenesis and screening to discover mutants with the desired catalytic properties. Examples are summarized in several reviews[50a, 51].

The current manuscript focuses on the incorporation of rational approaches in the improvement of enzyme enantioselectivity. When libraries are constructed by random mutagenesis, only a fraction of the mutants will have beneficial properties since most mutations are neutral or detrimental for activity and selectivity[52]. As a result, typically several hundred to several thousand mutants need to be screened to discover the desired variants (Table 1-2). When high-throughput assays are available, e.g. using a fluorogenic substrate, this may be feasible, but screening for improved enantioselectivity often requires time-consuming chiral chromatography[50a]. To solve this problem and make it possible to discover improved variants with low-throughput screening, efforts are being made to construct libraries that are enriched in variants that carry mostly useful mutations. This can be achieved by focusing on a restricted number of positions, e.g. residues located around the active site. Furthermore, restricted sets of amino acids can be incorporated at the selected target positions, instead of testing all 20 proteinogenic amino acids at each position. Restricting mutagenesis to positions and substitutions that have the potential to improve enantioselectivity may result in libraries that contain less inactive or otherwise useless variants, and therefore are more suitable for the discovery of desired mutants with less screening. The design of such rational libraries can be based on inspection of structures, on comparison of sequences of homologous proteins, or it may incorporate results of computational analysis of enzyme-substrate interactions[53]. These strategies and examples illustrating the merits and challenges of different approaches to rational library design are discussed below.

Two main ways exist to enzymatically produce enantiopure compounds. First, enantioselective conversion of a racemic mixture can yield a single product enantiomer and a single substrate enantiomer if only one of the two enantiomers
is converted. This process is called a kinetic resolution. The origin of this type of enantioselectivity is preferred binding or faster conversion of one enantiomer over the other\textsuperscript{[54]}. Well known examples are lipases\textsuperscript{[55]}, esterases\textsuperscript{[56]}, epoxide hydrolases\textsuperscript{[57]}, halohydrin dehalogenase\textsuperscript{[58]}, alcohol dehydrogenases\textsuperscript{[59]}, hydroxynitrile lyase\textsuperscript{[60]} and haloalkane dehalogenase\textsuperscript{[61]}.

<table>
<thead>
<tr>
<th>Category</th>
<th>Method</th>
<th>Required knowledge</th>
<th>Number of variants screened</th>
<th>Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure or sequence inspired directed evolution</td>
<td>Iterative site mutagenesis</td>
<td>Enzyme structure or homology model</td>
<td>466-3000</td>
<td>Lipases\textsuperscript{[62]}, epoxide hydrolase\textsuperscript{[67]}, limonene epoxide hydrolase\textsuperscript{[63]}, monoamine oxidase\textsuperscript{[64]}, transaminase\textsuperscript{[65]}, P450 enzymes\textsuperscript{[66]}</td>
</tr>
<tr>
<td>Phylogeny based methods</td>
<td>Enzyme structure and multiple sequence alignment</td>
<td>1600-5500</td>
<td>Haloalkane dehalogenase\textsuperscript{[67]}, esterases\textsuperscript{[56, 68]}</td>
<td></td>
</tr>
<tr>
<td>Rational design</td>
<td>Manual design</td>
<td>Enzyme structure, substrate binding position and catalytic mechanism</td>
<td>A</td>
<td>Lipases\textsuperscript{[69]}</td>
</tr>
<tr>
<td>MD simulation or docking aided</td>
<td>Enzyme structure or homology model and catalytic mechanism</td>
<td>A</td>
<td>Lipases\textsuperscript{[70]}, hydroxyl nitrile lyase\textsuperscript{[71]}, ω-transaminase, arylmalonate decarboxylase\textsuperscript{[72]} non-heme Fe\textsuperscript{3+} dependent hydroxylase\textsuperscript{[73]}</td>
<td></td>
</tr>
<tr>
<td>Motif-guided sequence searches</td>
<td>Rational design of improved mutations, sequence motifs and sequence databases</td>
<td>21</td>
<td>Transaminase\textsuperscript{[74]}</td>
<td></td>
</tr>
<tr>
<td>In silico approaches</td>
<td>Docking and simulation aided</td>
<td>Enzyme structure, substrate docking catalytic mechanism and simulation parameters</td>
<td>A</td>
<td>Lipases\textsuperscript{[75]}, diisopropyl fluorophosphatase\textsuperscript{[76]}</td>
</tr>
</tbody>
</table>

\textsuperscript{A} Only successful mutations reported.

A more attractive process is asymmetric synthesis, where a chiral center is formed when an enzyme converts in a stereoselective manner a prochiral compound. Examples are the asymmetric conversion of one of two equivalent groups of a substrate, resulting in desymmetrization of the molecule\textsuperscript{[77]}, and the asymmetric addition of substituents to a double bond, as in ketoreductase or aminotransferase reactions. The latter is caused by a difference in reaction rates at the two sides of a pro-chiral center, which is related to a specific binding mode of the substrate in the active site, i.e. with the re or si face of the reactive carbon
exposed to the group that is added\textsuperscript{[65]}. Also differences in regioselectivity during the conversion of meso substrates can cause stereoselective product formation, e.g. during the selective hydrolysis of alicyclic epoxides by limonene epoxide hydrolase\textsuperscript{[78]}.  

Both for asymmetric synthesis starting with a prochiral or meso compound, and kinetic resolution of a mixture of enantiomers, the enzyme enantioselectivity can be expressed by the $E$-value\textsuperscript{[79]}, which is defined as, respectively:

\begin{equation}
E = \frac{V_R}{V_S}, \quad \text{and} \quad E = \frac{\left( \frac{k_{cat}}{K_M} \right)_R}{\left( \frac{k_{cat}}{K_M} \right)_S}
\end{equation}

These $E$-values are catalyst properties, which may be influenced by mutations and reaction conditions. In case of a asymmetric conversion, product $ee$ is independent of the degree of substrate conversion and can be used to describe enzyme properties. In case of kinetic resolution of a racemic mixture, substrate and product $ee$ are not constant over time, and $ee$ should not be treated as a biocatalyst property.

**Structure-inspired directed evolution**

Sequence- or structure-inspired site-saturation mutagenesis continues to be a successful tool in directed evolution. By careful selection of the number of positions that is mutated, library size is reduced without reducing the number of improved variants. This approach uses structural inspection or sequence alignments to predict positions where mutations are expected to have an effect on enantioselectivity without destruction of catalytic activity\textsuperscript{[56, 68, 80]}. Such influential positions may also be discovered by employing a random method such as error-prone PCR, after which saturation mutagenesis may be used to test all proteinogenic amino acids. Instead of random mutagenesis, structural information may be used, either in the form of a crystal structure if available, or by constructing a homology model. Suggestions for specific mutations can be obtained by comparing structures of homologous enzymes with different enantioselectivities or by analyzing the docked or experimental substrate binding modes\textsuperscript{[68, 80c]}. Multiple sequence alignment of homologous enzymes can also be used to restrict the number of substitutions at each target position\textsuperscript{[56, 80d, 80e]}.  

Iterative saturation mutagenesis (ISM)\textsuperscript{[81]} is a well-established structure-inspired directed evolution method\textsuperscript{[57, 62a, 63, 66, 80b]}. Based on structural information, the active site is divided into different subsites, each consisting of one to three amino acids. The different positions within an individual subsite are
Introduction

mutated combinatorially and the best combination mutant obtained from one subsite is used as the initial variant when exploring mutants at the next site\textsuperscript{[81]}. Such iterations of mutagenesis and screening were effective in discovering enantioselective mutants in libraries that were much smaller as compared to the random libraries commonly explored in directed evolution experiments\textsuperscript{[57, 62a, 63, 66c, 80b, 82]}. When positions are individually screened in iterative rounds, the total diversity explored is the sum of the number of mutants screened at each round, while in combinatorial site-saturation mutagenesis, a much larger diversity can be examined\textsuperscript{[80b]}

By using the ISM method, it is possible to change the enantioselectivity without having a detailed understanding of the interactions that determine enantiodiscrimination. ISM was used to improve the enantioselectivity of a lipase from \textit{Pseudomonas aeruginosa} for the kinetic resolution of \textit{rac}-2-phenylpropanoic acid \textit{p}-nitrophenyl ester (Scheme 1-1). Four mutations in two different subsites resulted in an improvement of the selectivity from $E=2$ to 234\textsuperscript{[62a]}. In hindsight, the mutations contributing to higher selectivity were proposed to stabilize the oxyanion interactions of the tetrahedral intermediate of the faster reacting enantiomer and to increase the space available for binding the large substituent of the (S)-enantiomer. The same approach was used to improve the enantioselectivity of an \textit{Aspergillus niger} epoxide hydrolase for the kinetic resolution of racemic glycidyl phenyl ether. By the ISM method mutant was obtained, which possessed an $E$ of 160 in this conversion, while the wild type was only moderately selective with an $E$ of 4.6\textsuperscript{[57]}. This improvement was partially caused by the unintended insertion of 13 residues by primer insertion.

\[ \text{Scheme 1-1. Kinetic resolution of rac-2-phenylpropanoic acid p-nitrophenyl ester by an (S)-selective \textit{P. aeruginosa} lipase.} \]

This ISM method can also be used to improve the enantioselectivity during asymmetric synthesis. The enzyme limonene epoxide hydrolase from \textit{Rhodococcus erythropolis} DCL 14 was improved for the asymmetric synthesis of alicyclic epoxides (Figure 1-2A)\textsuperscript{[63]}. Starting from a wild type enzyme, which was only slightly enantioselective for the hydrolysis of cyclopentene oxide into (\textit{R},\textit{R})-cyclopentanediol diol with an \textit{ee} of 14\%, two different highly selective enzyme variants were discovered, one which produces (\textit{S},\textit{S})-cyclopentanediol with an \textit{ee} of 93\% and one that produces (\textit{R},\textit{R})-diol with an \textit{ee} of 80\%. Each variant carried
four different mutations, which were suggested to reshape the active site and thereby reposition the substrate (Figure 1-2B)\textsuperscript{[63]}. For this enzyme, the regioselectivity of nucleophilic water attack on the meso-epoxide determines the stereoconfiguration of the product. The mutations appeared to significantly shorten the distance between the catalytic water molecule and one epoxide ring carbon atom compared to the other, enhancing the regioselectivity for nucleophilic water attack at the closest carbon atom.

**Figure 1-2.** A) Asymmetric synthesis of cyclopentene epoxide determined by the regioselectivity of the nucleophilic attack by water in the active site of limonene epoxide hydrolase on either carbon atom 1 or 2. B) The positions of the water molecule and catalytic residues 53 and 132 are highlighted (pdb: 1NWW). The other apolar amino acids that line the active site cavity and influence the position of the substrate during catalysis are indicated. Residues with similar color (except green) were mutated as one subsite during the ISM approach to influence regioselectivity of this enzyme.

In many systems, an important source of enantioselectivity is the steric hindrance between active site residues and the slower reacting enantiomer. Structures containing a docked substrate can be used to identify positions where mutations can increase this hindrance. This approach was used to improve the enantioselectivity of a monoamine oxidase from *A. niger* to become enantioselective for the deracemization of 4-chlorobenzhydrylamine (Scheme 1-2) and 1-phenyltetrahydroisoquinoline, both compounds are important building blocks in the synthesis of pharmaceuticals containing chiral amine groups\textsuperscript{[64]}. The employed strategy aims to stoichiometrically convert a racemic amine into one enantiomer of the same compound. Therefore it uses cycles of enzyme-catalyzed selective oxidation of the amine to an imine and non-selective reduction back to the amine using sodium borohydride, resulting in the
accumulation of a single non-converted enantiomer of the starting compound, comparable to complete conversion to one product enantiomer in a dynamic kinetic resolution [83]. Two large residues flanking the active site were proposed to have steric interactions that influence the binding of 4-chlorobenzhydrylamine enantiomers. Mutations of these residues could create more space in the active-site pocket and thereby allow for better binding of the chloro-phenyl substituent of the desired (S)-enantiomer. Subsequently, site-saturation mutagenesis was performed at these two positions. Screening resulted in the discovery of five variants that were able to convert 4-chlorobenzhydrylamine with high enantioselectivity. This study demonstrated a large improvement in enzyme enantioselectivity for the production of highly useful chiral amines.

Scheme 1-2. Deracemization of 4-chlorobenzhydrylamine by a monoamine oxidase and a non-selective reducing agent (NaBH₄)[64]. The bold arrow indicates the net formation of one enantiomer of the starting material by the used procedure.

The applicability of structure-inspired mutagenesis can be enhanced by using homology models in addition to crystal structures. For example, Savile and co-authors modified a transaminase for the asymmetric production of the drug sitagliptin, which has an amine group at a chiral carbon atom (Scheme 1-3)[65]. Residues that interact with the substrate were predicted by docking the substrate in a homology model. Subsequently, site saturation mutagenesis was performed at these residues. This resulted in the discovery of a mutant that could synthesize the (R)-enantiomer of sitagliptin with an ee of more than 99.5%, while the wild type enzyme did not produce the desired enantiomer[65].

Scheme 1-3. Kinetic resolution of prositagliptin ketone to produce sitagliptin, catalyzed by a transaminase.
Another factor that can determine the enantioselectivity of an enzyme is the position of a cofactor relative to the substrate. To reverse the enantioselectivity of P450 enzymes the ISM approach was used to change the positions of both the substrate and the oxygen-activating heme cofactor. Tang and co-authors used this approach to improve the P450pyr enzyme from *Sphingomonas* sp. HXN-200 for the asymmetric hydroxylation of 1-benzylpyrrolidine\[66a\]. The authors screened mutants at 17 positions around the active site and discovered a variant with an ee of 85\% for the production of the (R)-enantiomer and a variant with an ee of 42\% for the (S)-enantiomer. This was a remarkable achievement, since reversing the enantioselectivity is often quite difficult to achieve. The same enzyme was also improved for the selective hydroxylation of N-benzyl pyrrolidine to (S)-N-benzyl 3-hydroxypyrrolidine (Scheme 1-4)\[66b\]. The ISM method was successfully used to introduce mutations that change the orientation of the benzyl ring and the distance between the heme O-atom and the pro-(S) or pro-(R) H-atom of the substrate. These examples show that the ISM method is applicable for the engineering of P450s, which are very promising enzymes for biocatalysis in view of their ability to oxidize many substrates in a far more specific manner than chemical oxidation catalysts.

![Scheme 1-4. Asymmetric synthesis of (S)-N-benzyl 3-hydroxypyrrolidine by an engineered cytochrome P450.](image)

**Directed evolution steered by multiple sequence alignments**

Sequence data can hold significant information on the type of mutations that can be introduced during directed evolution to achieve a certain effect. To significantly reduce the number of unstable or inactive variants, several methods use phylogeny-derived information to make a preselection of the amino acids to target in directed evolution. The use of restricted codon sets is a way to introduce a subset of amino acids at selected positions, which can be an efficient method to screen the sequence diversity while reducing the number of unstable or inactive variants\[56, 67-68\].

Restricted codon sets was used to engineer a haloalkane dehalogenase from *Rhodococcus rhodochrous* to become enantioselective for the hydrolysis of the pro-chiral compound 1,2,3-trichloropropane\[67\]. In this study, small libraries
Introduction

based on structural information and sequence alignments were used for directed evolution. In total, thirteen to seventeen residues located close to or more distant from the active site were mutated. Most of the mutations were proposed to have steric effects that influence the binding of the substrate. This resulted in the discovery of two variants that either produced (R)-2,3-dichloropropan-1-ol with an ee of 90% or (S)-2,3-dichloropropan-1-ol with an ee of 97%, respectively [67].

In another example of the use of phylogeny based information, the composition of a restricted codon set was based on the observed frequencies of residues among homologous proteins in the 3DM family database (Table 1-3) [84]. Using a preselection of residues based on this database, an esterase from *Pseudomonas fluorescens* was improved for the kinetic resolution of 3-phenylbutyric acid ethyl ester [56]. A variant containing three mutations was obtained that had an *E* of 80 for the production of the (R)-enantiomer [56].

**Table 1-3. Use of a restricted codon set during the protein engineering of an esterase from *P. fluorescens* for the asymmetric hydrolysis of 3-phenylbutyric acid ethyl ester**

<table>
<thead>
<tr>
<th>Position</th>
<th>Desired amino acids</th>
<th>Disallowed amino acids</th>
<th>Codon&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Encoded amino acids&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
</table>

<sup>a</sup>KBS, RBC, KKK and DYA are ambiguous DNA triplets (K: T or G base; B: C, T or G; S: G or C; R: A or G; D: A, G or T; Y: C or T).<br> <sup>b</sup>Disallowed amino acids in the encoded amino acids are shown in bold.<br> <sup>c</sup>Number of desired amino acids encoded as compared to total number of amino acids encoded. Data from Jochens et al. [56].

**Rational design based on structural inspection**

Rational design aims to predict specific substitutions to improve enzyme enantioselectivity based on inspection of enzyme structures and enzyme substrate-complexes [69-70, 71-72, 86]. This requires a detailed understanding of the structural determinants of enantioselectivity. An important possibility is to introduce mutations that increase steric hindrance for the undesired enantiomer. This strategy was used to engineer a lipase from *Burkholderia cepacia* for the kinetic resolution of racemic 1-phenylethanol [69a]. Changing an Ile near the active site into a larger Phe residue increased the *E*-value for the (S)-enantiomer from 88 to 156 [69a]. The same authors also reported the construction of an (R)-selective variant of the enzyme for the resolution of 1-phenyl-1-hexanol [69b]. For
this, two mutations were predicted, one of which was supposed to repulse the (S)-enantiomer, while the other would create space for the binding of the (R)-enantiomer. The designed variant was found to have an $E$-value of over 200, while the $E$ of the wild type for this conversion was 5[^69b].

The rational design of steric hindrance can also be inspired by sequence comparisons, which was demonstrated in another study on a lipase from *Candida rugosa*. The authors observed that the sequences of isoenzymes of this lipase, which display different enantioselectivities, vary at two important positions[^70a]. Subsequently, using MD simulations they confirmed that these residues governed enantioselectivity. Residues in the oxyanion hole had different interactions with the tetrahedral intermediate formed after reaction with the (R)-substrate compared to the (S)-substrate. Furthermore, a phenylalanine has $\pi-\pi$ stacking interactions with the aromatic substituent of the ester and thereby influences the position where the substrate binds in the active site. Based on this analysis, mutants were constructed to improve the enantioselectivity during the kinetic resolution of octyl 2-bromo-2-phenylacetate (Scheme 1-5). This resulted in the discovery of several enantioselective variants; the best two variants were either (R)-selective with an $E$ of 45 or (S)-selective with an $E$-value of over 200[^70a].

A possible approach to improve the enantioselectivity during kinetic resolution is to increase the space available for the binding of the desired enantiomer in the active site. This was used to increase the (S)-selectivity of a *Candida antarctica* lipase for the kinetic resolution of secondary alcohols containing a bulky side-groups, such as 1-phenylethanol[^70b]. Docking studies revealed that the phenyl group of the (S)-substrate would sterically clash with the enzyme. Accordingly, mutations were designed to create space for the binding of this substituent. The wild type possessed an $E$ of more than 200 for the (R)-enantiomer, while the obtained mutant was slightly (S)-selective with and $E$-value of 6.6 for the conversion of 1-phenylethanol[^70b]. Although the change in enantioselectivity was not spectacular, inverting the enantioselectivity by rational engineering is an important achievement.

![Scheme 1-5](image)

**Scheme 1-5.** Kinetic resolution octyl 2-bromo-2-phenylacetate by a lipase from *C. rugosa.*
Docking studies can also be used to identify interactions that influence the orientation of a substrate in an active site during asymmetric synthesis. Subsequently, mutations can be predicted that change the orientation of the substrate in the active site and result in the preferred formation of one enantiomer of the product. This approach was used to engineer a hydroxynitrile lyase from *Prunus amygdalus* for asymmetric synthesis of (R)-2-hydroxy-4-phenylbutene nitrile\[^{71}\]. Formation of the (R)-product would be impaired by repulsive interactions between the active site residues and the alkyl side-chain of (R)-product (Figure 1-3). Therefore, substitutions were predicted which increased the space available for improved binding of the substrate in the required orientation and the resulting mutants indeed showed enhanced enantioselectivity for the production of (R)-2-hydroxy-4-phenylbutyronitrile.

![Figure 1-3](image_url)

**Figure 1-3.** The use of docking to investigate repulsive interactions between the potential product and the active site. For a hydroxynitrile lyase, this analysis revealed clashes between the (R)-product and active site residues, which indicates that the substrate cannot bind in the pro-(R) configuration. The authors thus proposed mutagenesis at the highlighted positions to change substrate orientation. Figure from Weis *et al.*\[^{71}\].

A different study used docking to analyze binding of a substrate in either the pro-(R) or pro-(S) orientation during asymmetric synthesis of 2-aminotetralin by the \(\omega\)-transaminase from *Chromobacterium violaceum*. By analyzing the outcomes of these docking experiments, it was discovered that substrate was able to bind in a pro-(S) configuration, but a pro-(R) configuration was not observed. Subsequently, two mutations were proposed that enhanced the binding of the substrate in the pro-(R)-orientation while it reduced binding.
in the pro-\( (S) \) orientation. Experimentally, this changed the \( E \)-value for the synthesis of 2-aminotetralin from 3.9 \( (S) \) to 63 \( (R\)-selective)\[^{[86]}\].

Another approach to invert the enantioselectivity during asymmetric synthesis is to change the position of the most important catalytic residue to a mirrored location within the active site. Such an approach was used to reverse the enantioselectivity of a decarboxylase from \textit{Alcaligenes bronchisepticus} for the asymmetric decarboxylation of arylmethylmalonate to arylpropionate (Scheme 1-6). No structure of the enzyme was available, so all rational design was based on a homology model. By changing the position of the cysteine nucleophile from the \( si \)-face to the \( re \)-face of the enolate intermediate, the authors could almost completely inverse the enantioselectivity, from an \( E \)-value of 65 for the \( (R) \)-product to an \( E \)-value of 49 for the \( (S) \)-product\[^{[72]}\]. However, the obtained \( (S) \) selective mutant was 200-fold less active as the wild-type protein.

![Scheme 1-6](image)

**Scheme 1-6.** Decarboxylation of rac-2-methyl-2-(naphthalen-2-yl)malonic acid to \( (R) \)-2-(naphthalen-2-yl)propanoic acid, catalyzed by a decarboxylase from \textit{A. bronchisepticus}.

**Molecular dynamics simulations combined with visual inspection**

Detailed computational modeling of the interaction of the substrate with the surrounding amino acids can be used to identify interactions, including those related to substrate-induced conformational changes that steer the binding of the substrate during kinetic resolution, or influence the orientation of the substrate during asymmetric synthesis. Subsequently mutations can be deduced that result in more enantioselective catalysis. The structure of the resulting mutants can be predicted and analyzed using the same modeling approaches, to yield cycles of design and modeling. This approach was used to improve the enantioselectivity of a lipase from \textit{Candida antarctica} for esterification of 1-halo-2-octanols compounds during a kinetic resolution\[^{[75b]}\]. A combination of molecular docking and interaction energy calculations showed that there are no selective interactions governing the binding of the halogen moiety in the wild type enzyme; therefore there was no preference for binding either enantiomer. Subsequently, positions where mutations could introduce such interactions were visually identified. At these positions, mutations were introduced \textit{in silico}, interaction energies of resulting mutants were calculated, and the resulting
structures were analyzed visually\cite{75b}. Four mutants were predicted based on this inspection and were tested experimentally. This resulted in the discovery of two enantioselective mutants. The best variant possessed a two-fold higher $E$-value of 28 for the resolution of 1-chloro-2-octanol.

A different strategy uses the structure of a transition state inhibitor to predict the orientation of the substrate and product during catalysis. MD simulations of the enzyme with this transition state inhibitor bound gave insight into the dynamics of these interactions. Based on these simulations, mutations were predicted that should hinder catalysis with the undesired ($R$)-enantiomer. This approach was employed to change the selectivity of an esterase from *Bacillus subtilis* for the kinetic resolution of the tertiary alcohol 3-phenylbut-1-yn-3-yl acetate (Scheme 1-7)\cite{75c}. Simulations revealed important hydrogen bonding interactions that were expected to influence the reaction rate with different enantiomers in the active site. Based on this analysis, a Glu residue was predicted to be a key element of a hydrogen-bonding network that stabilized the transition state for catalysis with the ($R$)-enantiomer. Experimentally, modifying this hydrogen-bonding network by replacing this Glu by a Phe indeed inverted the enantioselectivity for the kinetic resolution of 3-phenylbut-1-yn-3-yl acetate. The wild type is ($R$)-selective with an $E$ of 5, while this mutant is ($S$)-selective with an $E$-value of 2. Thus, only a small change in the enantioselectivity of the enzyme was obtained by this method. However, the discovered position was important for ($R$)-selectivity, since replacing this Glu by a smaller Ala or Asp residue enhanced the ($R$)-selectivity of the enzyme tenfold up to $E$-values of 56 and 46\cite{75c}.

A similar approach was used to invert the enantioselectivity of a diisopropyl fluorophosphatase to enhance its applicability for the degradation of several nerve gases\cite{76}. The wild type enzyme is able to hydrolyze the organophosphorus nerve agents sarin, cyclosarin and soman. However, it preferably degrades the less toxic ($R$)-enantiomers of these compounds, while it has a very low activity for the more toxic ($S$)-enantiomers. Therefore, the enzyme was engineered to be more active for the degradation of the ($S$)-enantiomer of cyclosarin, while maintaining some activity for the ($R$)-enantiomer. To achieve this goal, both enantiomers of the substrate cyclosarin were docked into the active site and the interactions between the enzyme and the substrate were investigated using docking and MD simulations. These experiments revealed a binding cleft in the active site for the alkoxy-substituent of the substrate, which allowed for the binding of the ($R$)-enantiomer but not of the ($S$)-enantiomer (Figure 1-4). Based on this MD model, several mutations were proposed which
allowed for the binding of alkoxy-substituents of the \((S)\)-substrate in a mirrored position in the active site. This resulted in the prediction of a \((S)\)-selective quadruple mutant. Experimentally, the mutant indeed was \((S)\)-selective (\(E\)-value of 3.77), while the wild type was \((R)\)-selective (\(E\)-value of 50). The obtained mutant enzyme was also \((S)\)-selective for the conversion of the analogous organophosphorus compound sarin, which broadened the applicability of this enzyme in bioremediation\([76]\).

\[
\begin{align*}
\text{Ph} & \quad \text{Ac} \\
\text{H}_{2}\text{O} & \quad \text{Ac} \\
\text{Ph} & \quad \text{Ac}
\end{align*}
\]

**Scheme 1-7.** Kinetic resolution 3-phenylbut-1-yn-3-yl acetate by a lipase from *B. subtilis*.

---

A different approach employed a combination of MD simulations, density functional theory calculations and docking simulations to find interactions that influence the orientation of the substrate in the active site during asymmetric synthesis. Pratter and co-authors used these methods to change the enantioselectivity of a mononuclear non-heme Fe\(\text{II}\)-dependent hydroxylase from *Streptomyces coelicolor* for the enantioselective hydroxylation of phenylpyruvic acid to \((R)\)-mandelate\([73]\). First, they discovered in their simulations that a hydrophobic pocket selectively bound the aromatic substituent of the \((S)\)-

---
product, but did not accommodate the (R)-product. Second, the aromatic substituent of the (R)-product would collide with a tyrosine residue of the enzyme. To change the enantioselectivity of this enzyme, ten variants were constructed which would better accommodate the aromatic ring of the (R)-product and would not be able to bind the (S)-product. The combined introduction of the three best mutations resulted in a variant that produced (R)-mandelate with an ee of 95%, a 9,300-fold change in specificity compared to the wild type enzyme\[73\]. This study demonstrated that is possible to obtain a very large specificity change in enantioselectivity by combining computational methods.

An important step in directed evolution is to efficiently combine mutations discovered in parallel rounds. An interesting solution used computational modeling to analyze possible combinations and select the best ones. This approach combined mutations of a lipase from *Burkholderia cepacia* with enhanced enantioselectivity for the kinetic resolution of 2-chloroethyl-2-bromophenylacetate\[75a\]. All possible double mutants were modeled by MD simulations and path-planning algorithms. These algorithms quantify the effects of mutations on the ability of a substrate to enter the active-site. During this computational procedure both enantiomers of the substrate were docked into the predicted structures and the substrate orientations during MD simulations were analyzed. Furthermore, the movement of enantiomers into the active site was modeled using path-planning algorithms, in order to discover if mutations improved the entrance of substrates into the active site. The thirteen best behaving double mutants were constructed experimentally and three indeed possessed very high enantioselectivity. The best variant possessed an $E$ of 178 for the kinetic resolution of 2-chloroethyl-2-bromophenylacetate, a 15-fold improvement compared to the parent enzyme\[75a\].

**In silico prediction, analysis and selection of mutations**

The computational methods discussed so far rely on visual interpretation of the computationally modeled structures or simulations to predict mutations that improve enantioselectivity\[62b, 75-76\]. They do not predict the enantioselectivity qualitatively, but provide qualitative interpretations of structural changes leading to changes in selectivity. This limits their throughput and makes it difficult to quantify and compare effects of predicted mutations between different experiments. It remains a challenge to select mutations that improve enzyme enantioselectivity computationally instead of by rational selection or laboratory screening. An *in silico* ranking method would select mutations based
on their predicted effects on the binding of different enantiomers, on the predominant orientation of the substrate in the active site, or on differences in reaction energy barriers with substrate enantiomers or substrate orientations. The latter approach requires quantification of the chemical reaction step by methods such as quantum mechanics/molecular mechanics (QM/MM)\cite{87} or the quantification of near attack conformations (NAC)\cite{88}.

**QM/MM transition state complementarity**
QM/MM can predict transition state energy barriers and protein structures with a high accuracy based on quantum mechanical calculations\cite{78,87,89} and therefore could be a promising method for \textit{in silico} analysis and prediction of enantioselective mutants. However, it requires large computational resources and therefore its applicability is limited to small systems with the size of a few hundred atoms.

QM/MM has not yet been used to compare a large number of mutants \textit{a priori} but has been used to rationalize the outcome of directed evolution experiments aiming for enantioselective enzymes\cite{78,89}. Mutants of limonene epoxide hydrolase that were enantioselective in the asymmetric hydrolysis of cyclopentane epoxide were analyzed by this technique\cite{63,78}. For this, the authors calculated a transition state model for the hydrolysis of the oxirane ring by water in the active site. Subsequently, transition state energy barriers were calculated for the nucleophilic attack by a water molecule at either the C1 or C2 carbon atom by the wild type enzyme and the evolved mutants. These calculations confirmed the experimentally obtained results; \((R,R)\)-selective mutants had a lower energy barrier for attack on the C2 atom, while \((S,S)\)-selective variants had a lower energy barrier for attack on the C1 carbon atom. The calculations also showed that steric repulsions between the mutated residues and the substrate resulted in different substrate orientations, which caused these differences in energy barrier. The applied QM/MM method was able to predict these energy differences qualitatively but not quantitatively (Figure 1-5A)\cite{78}, thus it will require more detailed models of the active site to predict enantioselective mutations \textit{a priori}.

**NAC calculations**
The analysis of near attack conformations (NACs) is another method to quantify differences in reactive conformations that could be used for \textit{in silico} selection of enantioselective mutants. This analysis is based on the NAC theory formulated by Bruice and co-workers\cite{88}. In a NAC, the enzyme-substrate interactions are
geometrically similar to those of the transition state\textsuperscript{[90]} and the substrate-enzyme complex has to pass through the NAC to reach the transition state of the reaction. A NAC is considered to be formed when the angles, distances and dihedrals are within a set of defined criteria, for example the reacting atoms are within van der Waals contact distance and the angles of the reacting atoms within $15^\circ$ from those in the transition state. During an MD simulation, the fraction of the simulation time that the enzyme and substrate form NACs can be quantified\textsuperscript{[91]}. The NACs that occurred during an MD simulation of the ribozyme self-cleavage reaction were shown to correlate well with experimental results\textsuperscript{[91]}.

By comparing the occurrence of NACs for complexes of an enzyme with the enantiomers of a substrate during separate simulations, a ratio of NACs for the enzyme-enantiomer combinations can be obtained, which is representative for the enantioselectivity of an enzyme during kinetic resolution\textsuperscript{[61]}. A prediction for the \textit{ee} value of the product during asymmetric synthesis can be obtained by comparing NACs for the attack on different atoms or for different orientations of a prochiral compound. The NAC method has not yet been applied for the \textit{a priori} prediction of enantioselectivity, but it has been used to rationalize the differences in enantioselectivity observed during the kinetic resolution of $\alpha$-bromoamides by a haloalkane dehalogenase\textsuperscript{[61a]}, $\alpha$-bromoesters and $\beta$-bromoalkanes by different haloalkane dehalogenases\textsuperscript{[92]} and for the hydrolysis of glycidyl phenyl ether enantiomers by an epoxide hydrolase\textsuperscript{[61b]}.

In the first example, the enantioselectivity of a haloalkane dehalogenase from \textit{Sphingomonas paucimobilis} UT26 for the kinetic resolution of different $\alpha$-bromoamides was predicted by comparing NAC frequencies for the substrate enantiomers\textsuperscript{[61a]}. During this study, NACs were defined as substrate orientations that would allow nucleophilic attack of the aspartate oxygen on the halogen-substituted carbon atom and facilitate the development of negative charge on the halogen atom. The criteria were expressed as distances and angles between the substrate, the catalytic oxygen atom of the nucleophilic Asp, and the polar hydrogen atoms of the Asn and Trp that stabilize the transition state by polarizing the halogen atom (Figure 1-5B). Using these criteria, it was found that the ratio between the NAC frequencies observed in MD trajectories with $(R)$ or $(S)$-N-benzyl-2-bromobutanamide correlated well with the observed enantioselectivity (Figure 1-5C)\textsuperscript{[61a]}.

In a second example, NACs were used to rationalize the outcome of a directed evolution experiment that improved the enantioselectivity of an epoxide hydrolase for the kinetic resolution of glycidyl phenyl ether\textsuperscript{[61b]}. For the enzyme-substrate complex of the mutant and wild type enzyme, the distance
between the nucleophilic oxygen atom of the Asp and the C2 carbon atom of the epoxide was used to define the NAC. This distance was recorded during MD simulation of the enzyme with either the \((R)\)- or \((S)\)-epoxide bound. The difference between these values correlated well with the experimentally observed differences in enantioselectivity for the kinetic resolution of glycidyl phenyl ether between the two variants\(^{[61b]}\). The use of NACs for the computational analysis and selection of mutations that improve the enantioselectivity is explored further in Chapter 6 of this thesis.

**Figure 1-5.** Two different methods to quantify enantioselectivity of an enzyme.  
**A)** The correlation between the \(\Delta\Delta E\) (differences in energy barriers for the attack on either the C1 or C2 carbon atom, Figure 1-2A) calculated by QM/MM for this reaction, and the experimentally observed compared \(\Delta\Delta E\) (based on observed \(ee\) values) between the wild type and several mutant enzymes, figures obtained from\(^{[78]}\).  
**B)** NAC criteria for the nucleophilic attack of the aspartate oxygen on an electrophilic carbon atom of a haloalkane in haloalkane dehalogenase.  
**C)** Example of a screening of NACs using MD simulations to investigate the dehalogenation of \(t\)-butyl 2-(2-bromopropanamido)acetate. The observed nucleophilic attack angle \(\theta_1\) versus simulation time is plotted and compared between enantiomers. The grey area indicates angles that satisfy the NAC criteria. Panels B and C are from Westerbeek *et al.*\(^{[61a]}\)
Conclusion

The work discussed in this review has shown that structure- or phylogeny-inspired directed evolution is a mature approach, which has been successfully used to improve the enantioselectivity of variety of enzymes\textsuperscript{[56-57, 62a, 63-68, 80b]}. Rational design was successful only for a limited number of enzymes, mostly lipases and transaminases\textsuperscript{[69-70, 71-72, 86]}, mainly because prediction by visual inspection of mutations that influence enantioselectivity remains a challenge of which success partially depends on subjective and poorly defined criteria. To replace it by computational methods, several groups have used MD simulations for inspection of mutants \textsuperscript{[62b, 73, 75-76]}, which resulted in the correct prediction of enantioselective variants. An important future development would be the complete \textit{in silico} design of enantioselective mutants, in which multiple mutations are introduced in a single round. Research focusing on the application of the NAC analysis\textsuperscript{[61]} for the prediction of enantioselective mutants is described in Chapter 6 of this thesis.

Structural investigation of proteins with engineered enantioselectivity

Analysis of structures of mutant enzymes by protein crystallography can reveal changes in spatial organization of the active site and differences in substrate binding mode, and thereby provide an explanation for the functional interactions underlying changes in enzyme enantioselectivity\textsuperscript{[61b, 64]}. Crystal structures are also used to explain the enantioselectivity of natural enzymes\textsuperscript{[93]}. Proteins can be co-crystallized or soaked with ligands to reveal enzyme-substrate interactions in detail\textsuperscript{[93a]}. As an alternative, the substrate can be computationally docked into the active site to reveal discriminating enzyme-substrate interactions\textsuperscript{[61b, 64, 76]}.

In case of rationally engineered enzymes, protein crystallography can clarify if the observed kinetic effects are indeed related to the intended structural changes. This is not always the case. Docking experiments with (R)- and (S)-cyclosarin in the structures of an engineered diisopropyl fluorophosphatase did not reveal a preference for binding of one or the other enantiomer\textsuperscript{[76]}, which was in contrast to the original goal to improve the binding of the (S)-cyclosarin (Figure 1-4). It was subsequently proposed that the enantioselectivity was improved due to a different mechanism; the mutant enzyme was suggested to stabilize the phosphoenzyme intermediate of the reaction with the (S)-enantiomer more than the intermediate with the (R)-enantiomer\textsuperscript{[76]}.

Furthermore, elucidated crystal structures containing bound substrates or analogs can be used to investigate how structure-inspired directed evolution
improves enantioselectivity. MD simulations on these structures could reveal the increase or decrease of NACs or changes in important distances or angles between catalytic residues during simulations. This was done for an engineered epoxide hydrolase, which revealed that a change in the distance between the carboxyl group of the catalytic Asp and the C2 atom of the (R)- or (S)-epoxide caused the improvement in enantioselectivity\cite{61b}. Furthermore, several rearrangements in the active site occurred when mutations were introduced, which caused steric clashes between the undesired (R)-enantiomer and several active-site residues. The flexibility of a part of the active site that allows for binding of the (S)-enantiomer was also increased. These structural observations were supported by kinetic studies, which both concluded that the main reason for the improved (S)-selectivity was a reduction in the reaction rate of the (R)-enantiomer\cite{61b}.

During rational design an often-used technique is to change the residues flanking the active site cavity and thereby influence the binding of enantiomers in the active site. The structure of a resulting variant was investigated in a study on a monoamine oxidase constructed for kinetic resolution of 4-chlorobenzhydrylamine\cite{64}. The crystal structure of this variant revealed that the enantioselectivity was improved by the enlargement of a cavity next to the active site, which was responsible for the binding of the phenyl group of (R)-4-chlorobenzhydrylamine. The volume of this cavity was significantly increased by a flanking Trp to Gly mutation. This enhanced the binding of the (R)-enantiomer of the substrate (Scheme 1-2, Figure 1-6A). Furthermore, the reaction rate was improved by four mutations that increased the diameter of the substrate access channel. Smaller residues replaced several larger residues lining this channel. An introduced hydrogen-bonding interaction repositioned a Tyr residue away from the access tunnel. This allows the substrate to enter the active site more readily (Figure 1-6A)\cite{64}.

Another study highlighted the importance of mutating residues that flank the active site and orient the substrate during asymmetric synthesis. The crystal structure of a mononuclear non-heme FeII-dependent hydroxylase\cite{73} rationally engineered for the enantioselective hydroxylation of phenylpyruvic acid to (R)-mandelate showed that introduced Phe and Met residues occupy a pocket responsible for the binding of the aromatic substituent of the pro-(S)-mandelate in the wild type enzyme. The binding of the substituent of the substrate placed in the pro-(R)-conformation at a mirrored position was made possible by changing a Tyr residue into Ala. Thus, the elucidated structure confirmed that the intended structural effects occurred and changed the enantioselectivity as expected.
A crystal structure has also been used to validate the result of a rational protein engineering study that used a homology model. A ligand-bound structure of the engineered arylmalonate decarboxylase described above revealed that two mutations resulted in the removal of the cysteine nucleophile on one side of the active site and the introduction of a cysteine nucleophile at a mirrored position in the active site (Scheme 1-6, Figure 1-6B)\(^{[93a]}\). During the reaction this caused the protonation of the re-face of the enolate intermediate by the SH-group of the cysteine, in contrast to the protonation of the si-face in the catalytic cycle of the wild type enzyme. These differences in protonation result in a difference in enantioselectivity\(^{[93a]}\). Thus, the structure confirmed that the improvements of the enantioselectivity were obtained by the structural effects predicted from a homology model\(^{[72]}\).

**Figure 1-6.** Examples of crystal structures that reveal structural effects improving the enantioselectivity of engineered enzymes. **A)** A crystal structure of an engineered monoamine oxidase with the docked substrate 4-chlorobenzhydrylamine. Figure obtained from Ghislieri et al.\(^{[64]}\). Several mutations increase the size of the substrate entrance tunnel, while other mutations allow binding of the phenyl group of the \((R)\)-enantiomer. **B)** Comparison of wild type and mutant structures of ligand-bound arylmalonate decarboxylase. In the second figure the mutant structure (in white) is overlaid with the wild type structure. The position of the cysteine nucleophile in the mutant (C188) is mirrored as compared the wild type enzyme (C74), resulting in a mirrored attack on the ligand. Figure obtained from Obata et al.\(^{[93a]}\).
References


Introduction


Introduction
