Chapter 7

7. Summary and Outlook

This thesis describes the improvement of catalytic performance of an epoxide hydrolase in terms of its thermostability and enantioselectivity. The epoxide hydrolase used in this work was originally isolated from \textit{A. niger} strain LCP521 and successfully cloned into \textit{E.coli} by Arand and coworkers [42]. This enzyme is highly important as it shows exceptionally fast substrate turnover compared to other epoxide hydrolases in addition to being highly enantioselective toward a few industrially important epoxides [43]. However, many other important substrates are yet not turned over with such high selectivity. Furthermore, its crystal structure has been resolved at 1.8 Å resolution, revealing the two 44kDA subunits in the asymmetric unit. All these factors make ANEH a suitable candidate for directed evolution studies. More importantly, this experimental platform serves as a means to perform methodology development in directed evolution.

ANEH has been reported to be used as enzymatic alternative for preparation of epoxides of high optical purity, mainly in the case of styrene oxide derivatives [37; 45]. However, as most of epoxides have limited solubility and stability in aqueous solutions, the reactions was normally carried out with addition of organic solvents. Nevertheless, the restricted ANEH stability in such harsh reaction condition makes it not widely used in industrial applications. Therefore, the initial task was to increase its stability (thermal stability) via the directed evolution approach.

Our group has developed recently a methodology, known as B-FIT (B-factor iterative test), to increase enzyme stability. The principle of B-FIT is based on the idea of increasing the enzyme stability by making it more rigid or in other word less flexible, which can be achieved by targeting the residues with high degrees of flexibility. In the present work regarding ANEH, six out of sixteen libraries generated based on B-FIT resulted in moderate improvement in \( T_{50} \) values. Several additional libraries based on other strategies (e.g. such as salt bridge introduction, helix capping, N- or C-terminal stabilization, or loop stabilization) were also generated. However, no positive results were found in such libraries.

As in other ISM studies, in order to further enhance ANEH thermalstability, successive rounds of saturation mutagenesis were transversed. Unlike the frequently applied strategy, "pick the best and continue", three different types of mutants as parental sequences were employed. Five pathways, using the most-improved-mutant (2 pathways), moderate-improved-mutant (2 pathways), and the least improved mutant (1 pathway) obtained after initial round of saturation mutagenesis, were generated. Results from ISM indicated that the mutants became progressively more resistant to thermal denaturation in each round; however the accomplished improvements were varied among these five pathways. Surprisingly, the final best thermomutant originated from a disfavored pathway (with the least-improved-mutant as parental sequence), which under the traditional directed evolution procedure usually would not be chosen. The other pathways, which were generated using the most- or moderate-improved parent, however, end up giving mutants with similar moderate thermostability improvements. The final best
mutant acquired 21°C increase in $T_{50}^{60}$ values and 80 fold improvements in half-life at 60°C. Meanwhile, the other thermostable mutants demonstrated 10 - 14°C increase in $T_{50}^{60}$ values and 20-30 fold improvements in half-life at 60°C. These thermomutants not only exhibit a shifted optimum temperature, but also higher activity at higher temperature in comparison to the wild type. After thorough sequence and structural comparison of thermostable mutants, it appears that the final best mutant (that originated from the least-improved-parent) acquired different mutations in comparison to the others. This might demonstrate the possibility that all those thermostable mutants might have evolved in different ways against the thermal pressure.

In addition to the saturation mutagenesis, the site directed approach in an attempt to rapidly obtain ANEH thermostable mutants by simply combining the beneficial mutations that were acquired in the first ISM round was also carried out. The $T_{50}^{60}$ values obtained by some site directed mutants were comparable to those obtained by the ISM approach, but the combination of some mutations resulted in inactive variants. This result underscores the additional benefits of the ISM approach by giving the enzyme the opportunity to choose other amino acids in order to avoid dead end pathways (or being trapped in evolutionary local minima). In summary, all these results show that ISM based on B-FIT Is a useful approach for increasing protein thermostability.

In addition to thermostabilization, ANEH was also subjected to other directed evolution studies, e.g. for improving its enantioselectivity or broadening its substrate scope towards non-natural substrates. The approach used in these experiments is called CASTing (combinatorial active site saturation test), developed recently in the Reetz group, which basically targets iteratively residues near by the enzyme binding pocket. ANEH CASTing was previously successful in increasing its enantioselectivity, generating a highly enantioselective mutant (LW202) having E-value of 115 [59]. However, in that study neither the full protein sequence space was covered nor was the exploration of other possible pathway. Therefore, in an effort to mainly evaluate the efficiency of ISM approach, a thorough exploration of all possible pathways in four rounds of ISM was performed. For this purpose, a total of 64 libraries (480 clones were screened for each library, 30720 clones in total) were generated. Each round consists of saturation mutagenesis followed by HPLC based screening and selection of the most enantioselective variant, which was then used as the starting point for the next step.

In contrast to other studies, which are mainly based on step by step site-specific-mutagenesis, the present approach conducts step by step site-saturation-mutagenesis. The other goal of this study was to reveal the real dynamics of the evolutionary process *per se* by analyzing the evolutionary intermediates or the influence of epistatic interactions in the protein fitness landscape, in addition to finding other mutants which show similar or even better enantioselectivity profiles compared to that of LW202. The number of residues allowed to be changed (four library sites, each sites consist of two amino acids) and the type of amino acids introduced was deliberately chosen to be relatively small (NDT codon degeneracy, 12 amino acids) so as to reduce the screening effort to a reasonable amount. Despite these limitations, 12 out of 24
pathways led to highly enantioselective mutants, having E-values above 80. Many of these 12 favored pathways originated from library D, which actually attained the lowest improvement in initial ISM round. On the contrary only one pathway stemmed from library sites which gain the highest improvement in first round of ISM, library B. The highest improvement was achieved via pathway originating from library F, pathway FBED, leading to mutant GUY-228 with an E-value of 158. This pathway also appeared to be the shortest mutational trajectory in the sense that only three ISM rounds were necessary to elevate the selectivity factor. The productivity of the approach was also proven by the high number of improved mutants encountered in exploring all of those libraries, a total of 69 hits (31 mutants which have E-values above 80 and 37 mutants with E-values 50-80) were obtained.

From the fitness pathway landscape generated in this study, two different types of trajectories connecting the wild type and the best mutant obtained for each pathway are encountered. The first type demonstrates a smooth continuous decrease in the free energy, whereas the second type of pathway is the one characterized by the existence of so called “local minima”. 16 out of 24 pathways show the characteristic of first type of trajectory whereas the remaining eight pathways belong to the second one. This experimental result exhibits the high probability in ISM approach to find enzyme with improved properties (67% probability is encountered). It also supports the previously reported hypothesis that this problem can be overcome simply by choosing another set of mutations at a previous stage and place the evolutionary process back on an energetically favored trajectory (backtracking) [62].

Further sequence and structural comparisons among the best mutants gained in 24 pathways revealed the similarities and also the differences among them. A “phylogenetic” tree, constructed based on multiple protein sequence alignment, showed that the best mutants can be divided into three major groups. Furthermore, it was also shown that the mutant obtained in pathway FBDE appears to be the one which is the closest distance in terms of sequence with the previously obtained variant LW202. Both mutants apparently display also similar E-values. The comparison of the model structure of several mutants having different E-values revealed the pattern of substitutions they acquired. The highly enantioselective mutants tend to acquire bulky amino acid residues, introducing steric hindrance, which prevents the binding of the disfavored R-enantiomer, whereas smaller size substitutions were acquired by low enantioselective mutants.

To summarize, our results in exploring the 24 evolutionary pathways not only revealed the efficiency of the ISM strategy. They can be used in the future to develop a predictive method which circumvents the need for trial and error as to which pathway should be chosen. Further characterization (e.g. enzyme purification, substrate scope of the best mutants, and kinetic parameters) of mutants that show similar selectivity profile as the LW202 still has to be performed.

Like in other directed evolution experiments, the ultimate goal of the library generation-screening process was to obtain catalysts which show better performance in industrial processes. It was therefore the aim to
generate ANEH mutants which not only can resist high temperatures (high stability), but also show high selectivity during the hydrolytic kinetic resolution. Two different approaches were tested. In the first, the beneficial mutations that had proven to affect ANEH enantioselectivity, namely the nine sets of mutations that lead to LW202, were introduced by means of site directed mutagenesis to the thermostable mutants. In the second approach, mutations known to increase ANEH stability were introduced into LW202. Both of these strategies can be viewed as combining B-FIT and CAST by a rapid site directed approach. From the two strategies, the first, i.e. introducing the enantioselective mutations that lead to LW202 to the thermostable mutants gave better results than the second. Two mutants which show high stability and excellent enantioselectivity profiles were generated, GUY-095 and GUY-104 having E-values of 87 and 78; $T_{50}$ values of 58 and 57°C, respectively.

An additional attempt to apply CASTing for inverting ANEH enantioselectivity was also part of this thesis. Initial round of saturation mutagenesis resulted in inverting the selectivity towards the disfavored $R$-enantiomer. Despite the inversion, the E-value obtained is still relatively low (E-value ($R$) = 7). Further rounds of ISM are hence required. Finally, ISM in the form of CASTing was recently applied successfully in other studies as well ([58; 287; 292; 293; 294; 295; 296; 297; 298]). Thus, ISM is rapidly advancing to the best, fastest, and most efficient method in directed evolution.