Chapter 1

General introduction

Antibiotics

Modern healthcare strongly depends on antibiotics. In the past decades the research for new antibiotics has yielded not only many novel compounds, but also insight in their mechanisms of action. Different antibiotics affect several essential processes in the prokaryotic cell. They can interfere with DNA, RNA or protein synthesis, or in the construction of the cell wall or membrane. One of the first identified and still most important classes of antibiotics are the β-lactam antibiotics, which interfere with cell wall synthesis. They account for about 65% of the worldwide antibiotic prescriptions and their total sales are now estimated to be US$ 15 billion [1].

History of β-lactam antibiotics

Although Alexander Fleming already discovered penicillin in 1928 [2], it was not until the second world war that penicillin G (penG) was isolated, characterised and produced at a large scale. This resulted in improved antibacterial therapy around the world. Unfortunately, since this wonderdrug killed a lot of different bacteria, it was used almost unlimited although effects on the natural equilibrium of micro organisms had not been properly investigated. Further research revealed that the instability of penG hampers easy administration and some pathogens are not as sensitive to penG. Moreover, the abundant use has resulted in the occurrence of novel variants of pathogenic bacteria resistant towards this compound. This has created the need for new β-lactam antibiotics that are more effective and more stable [3]. The substitution of the side chain of penG, phenyl acetic acid, with other side chains like D-phenylglycine and D-p-hydroxyphenylglycine, resulted in the semi-synthetic β-lactam antibiotics ampicillin and amoxicillin (Figure 1), respectively, which are more stable and can be orally administered. Substitutions at C2 of the penicillin core resulted in new β-lactam antibiotics to which resistant bacteria are sensitive.

Another step forward in the quest for new antibiotics was the use of derivatives of cephalosporin C (CPC, Figure 1). The isolation of CPC from *Acremonium chrysogenum* (formerly known as *Cephalosporium acremonium*) was first described in 1955 [4]. Although CPC can not be used clinically, since its antibacterial activity is too low in the body, it was an interesting compound because of its *in vitro* activity towards gram negative bacteria. Unfortunately, the production of CPC turned out to be much more difficult than penG since the formation of this compound in the fungus is not uniform. *A. chrysogenum* produces the closely related intermediates deacetylcephalosporin C and penicillin N in considerable amounts as well [5]. These intermediates hardly display any antibiotic action, but are difficult to separate from CPC, resulting in low yields of the desired product.

Since the antibiotic properties of cephalosporins and penicillins were similar *in vitro*, the relationship between them was readily studied and the chemical relationship

between their skeletons was discovered. In 1963 Morin *et al*. described the expansion of the five membered thiazolidine ring of penicillin G to a six membered dihydrotiazine ring, resulting in the cephalosporin core 7-aminodesacetoxycephalosporanic acid (7-ADCA) [6]. This finding was the first step towards the clinically used semi-synthetic variants of cephalosporins, cephalexin and cefadroxil (Figure 1), which have the same side chains as ampicillin and amoxicillin respectively.

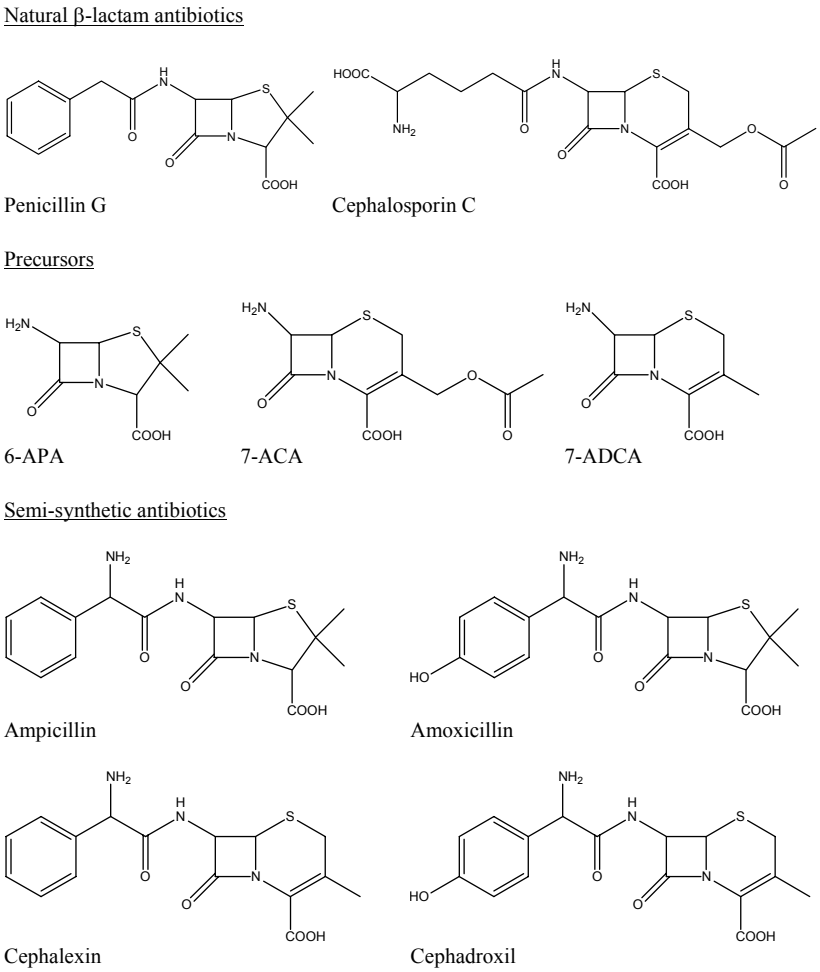

Figure 1. Some (semi-synthetic) β-lactam antibiotics and their precursors

Production of β-lactam antibiotics

The processes for large scale production of β-lactam antibiotics have gone through several drastic changes over the past 50 years. It is therefore a very nice example of the insights and achievements in biotechnology over the past decades [1,5]. Up to the 90's, the only 'bio' component in the production of these β-lactam antibiotics was the fermentation of *P. chrysogenum* and *A. chrysogenum*. All modification were accomplished by chemical synthesis. The phenyl acetyl side chain was cleaved off and the remaining 6-aminopenicillanic acid (6-APA) was protected and coupled to the appropriate side chain. In one variation to obtain cephalosporins, the five membered ring of penG was first expanded chemically to get phenylacetyl-7-ADCA using hazardous chemicals at extremely low temperatures, thereby using a lot of energy. In the early 1960's penicillin G acylase was found to hydrolyse the side chain from the β-lactam core. This process was, however, very inefficient and expensive, since the enzyme was produced at low yields and discarded after each cycle. Consequently, the chemical routes for production of β-lactam antibiotics were further optimised and used until the 1990's. At the same time, research towards biocatalytic routes remained an active field, not in the least due to the fact that the chemical approach produces a lot of waste. The production of cephalexin, for example, which has an annual consumption of about 3000 tons nowadays, generates a waste of 30-40 kg per kg of end product [5]. Since environmental issues have become more important and the cost for waste disposal has increased enormously in the last decade, it is worthwhile to look into more sustainable processes to produce semi-synthetic antibiotics.

The first enzyme to be used in the production of semi-synthetic β-lactam antibiotics was the penG acylase mentioned above. Yields of this enzyme have been highly improved thanks to genetic engineering. It was used for both the hydrolysis of the phenyl acetic side chain from penG and the synthesis of a new aromatic side chain onto 6-APA. Since the deacylation of penG is an equilibrium process, choosing the right conditions will direct the enzymatic reaction towards either hydrolysis or synthesis. A lot of research effort was put into optimising reaction conditions for the enzymatic synthesis of semi-synthetic antibiotics, but this lies outside the scope of this thesis [7-9]. The hydrolytic process became more efficient by screening different bacterial and fungal strains for the acylase and using recombinant DNA technologies in order to get a more efficient, stable and highly produced enzyme. Furthermore, fermentation conditions for the production of penicillin acylase and reaction conditions for hydrolysis of the substrate were optimised [10]. The immobilisation of the enzyme resulted in the possibility of recycling the biocatalyst and thereby reducing enzyme costs [3]. The production of the right stereo-isomer of the side chain was also optimised. By the use of an enzyme which produces only the wanted isomer of the side chain and racemisation of the non-used substrate, conversions of > 95% can be obtained [5].

Optimising production of cephalosporins has been pursued along two different routes, since both 7-aminocephalosporanic acid (7-ACA) and 7-ADCA are precursors for semi-synthetic cephalosporins, like cephalexin, cephaclor, cephuroxim, cephotaxim and cephixim. First, the production of 7-ACA from CPC produced in *A. chrysogenum* was optimised. In order to get 7-ACA, the D-α-amino adipyl side chain has to be hydrolysed from CPC. Since the chemical hydrolysis of this side chain is very expensive, extensive screening of organisms around the world was performed in order to find an enzyme capable of hydrolysing the side chain of CPC. However, this has not resulted in the identification of an efficient enzyme yet. Although different groups claimed to have isolated a cephalosporin acylase, the activities of these enzymes towards CPC are at least a factor 100 slower than towards glutaryl-7-ACA [11-13]. Therefore, a two step approach is used nowadays. First, a D-amino acid oxidase removes the amino group from the side chain. A spontaneous conversion of the resulting ketoadipyl-7-ACA into glutaryl-7-ACA supplies the substrate for the glutaryl acylase. These enzymes are also immobilised and re-used in subsequent cycles of deacylating CPC [14].
The other route towards semi-synthetic cephalosporins starts from 7-ADCA. The production of 7-ADCA out of 6-APA by ring expansion was approached. Since *A. chrysogenum* and other fungi, like *Streptomyces clavuligerus*, produce β-lactam antibiotics with 6-membered rings, the genes for this expansion process were readily discovered and cloned. Unfortunately, the produced enzymes were not capable to expand the 5-membered ring of 6-APA, penG or penicillin V *in vitro* [15]. Mutagenesis and recombination of the expansion genes towards these substrates as well as optimising process conditions were pursued and slowly led to some successes, which lie outside the theme of this thesis [16,17]. In another approach adipic acid instead of phenyl acetic acid was fed to the *P. chrysogenum* strain containing the native gene, resulting in a high production of adipyl-7-ADCA [18]. Cleaving off the adipyl chain results in a cost effective production of 7-ADCA. Although a true adipyl acylase is not available, the cephalosporin acylases mentioned above are capable of hydrolysing the adipic side chain from adipyl-7-ADCA to some extent. In order to make this process economically feasible, enzymatic reaction conditions were optimised [19].
A similar approach was used in *A. chrysogenum*. It was found that the organism produces several different β-lactam compounds because of the double activity of an expandase/hydrolase enzyme encoded by the cefEF gene. To discard the unwanted compounds, the normal route for production of CPC was blocked by inactivating the step from penicillin N to deacetylcephalosporin C, and introducing the cefE gene from *S. clavuligerus*. This gene encodes an expandase enzyme without hydrolytic activity, resulting in a more pure end product, desacetoxycephalosporin C (= amino adipyl-7-ADCA) [20].

β-lactam acylases

All β-lactam acylases mentioned above are part of the family of penicillin amidases (EC 3.5.1.11) and are found in bacteria, yeast and fungi. Several members of this large class of enzymes are widely used in industry, mainly in the production of 6-APA and the synthesis of semi-synthetic β-lactam antibiotics, but also in peptide synthesis and the resolution of racemic mixtures of chiral compounds [10].

All acylases show structural similarity and they are members of the N-terminal nucleophile (Ntn) hydrolase family [21,22]. This family consists of enzymes containing one to four heterodimers. The gene is transcribed into one propeptide which folds and cleaves itself into the heterodimer. Upon cleavage, the N-terminal amino acid of the β-subunit is liberated. This amino acid, a Ser, Thr or Cys, is the active site residue that performs the nucleophilic attack. In cephalosporin and penG acylases the N-terminal residue is a Ser. The OH-group on the side chain of this Ser attacks the peptide bond between the β-lactam core and its side chain. The free terminal NH of the same Ser and a water molecule assist in this nucleophilic attack. The side chain is released and the β-lactam core is covalently attached to the Ser. An attack of a second water molecule releases the core and brings the enzyme back to its original state (Figure 2).

β-lactam acylases are divided into several subfamilies. This classification used to be based on their substrate specificities towards β-lactam antibiotics [23]. In this classification the enzymes are roughly divided into two classes: the penicillin acylases and the cephalosporin acylases. Since the natural function of these enzymes is still not

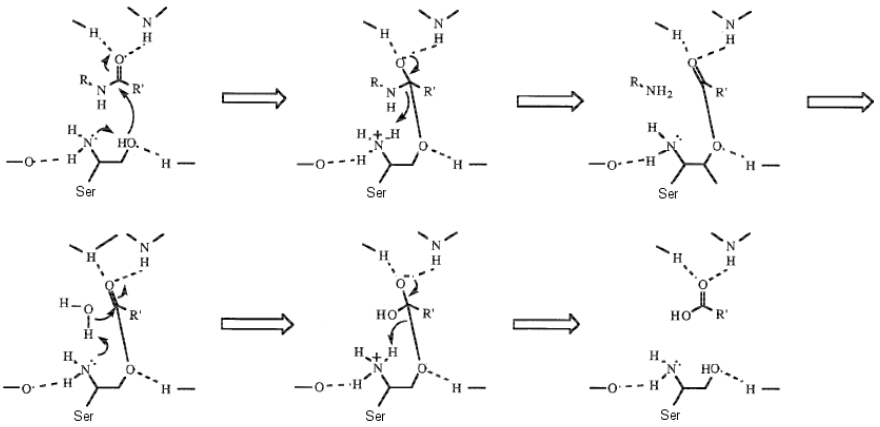

Figure 2. Catalytic mechanism of glutaryl acylase.
The nucleophilic oxygen of Ser199 donates its proton to its own α-amino group and attacks the carbon from the peptide bond of the substrate. The tetrahedral intermediate is stabilised in the oxyanion hole. The core is released upon the donation of the proton of the α-amino group to the nitrogen of the scissile bond. Then, the hydroxyl group of an incoming water attacks the carbonyl carbon of the side chain and donates its proton to the α-amino group again. The reaction is complete upon the donation of this proton to the side chain, to release it from the enzyme.
R = core (β-lactam core or leucine), R'= side chain (glutaryl, adipyl or amino adipyl).

fully understood, all cloned cephalosporin acylases are nowadays classified into five classes according to their gene structure, molecular masses and enzyme properties [11]. The enzymes in the different classes do exhibit different substrate specificities towards glutaryl-7-ACA and CPC. The activity towards CPC is rather low, only 0% to 4% relative to glutaryl-7-ACA hydrolysis. The cephalosporin acylases of class I and III show the highest activity towards CPC [24]. The enzymes of class I are transcribed into a precursor protein of 70 kDa, which comprises a signal peptide, an α-subunit of 16 kDa, a spacer peptide and a β-subunit of 54 kDa (Figure 3). The exact number of amino acids in the signal and spacer peptides differs per enzyme. The acylases that belong to this group are isolated from *Pseudomonas* SY-77 [13], *Pseudomonas* GK-16 [25], *Pseudomonas* C427 [26], *Pseudomonas* sp. 130 [27] and *Pseudomonas diminuta* KAC-1 [12] and show high similarity to each other (> 90% identity). Despite these similarities every enzyme was numbered in a different way. Since this thesis deals with the glutaryl acylase of *Pseudomonas* SY-77, the numbering of all base pairs and amino acids in it are according to the numbering of this enzyme [28]. Crystal structures of both the mature and precursor structure from *Pseudomonas* GK-16 [29] and *P. diminuta* KAC-1 [30,31] have been determined. The structures are almost identical except for the spacer peptide in the precursor structure. These crystal structures revealed that the spacer peptide is most likely autocatalytically cleaved, which was already proposed in biochemical studies. The cleavage of the side chain from glutaryl-7-ACA is a typical Ntn-hydrolysis (Figure 2). During this reaction the amino acids Asn442 and Val268 most likely form the anion hole to stabilise this reaction [30]. The substrate is positioned towards the active site serine by hydrophobic residues around the aliphatic side chain, while charged and hydrophilic residues deeper inside the enzyme interact with the negatively charged head of the substrate. The β-lactam core is positioned outside of the enzyme and seems to have very little interaction with the enzyme [32]. This explains the observation that the enzyme can hydrolyse all kinds of substrates with a glutaryl side chain, like glutaryl-7-ACA and glutaryl-6-APA, but also glutaryl-leucine and glutaryl-serine and esters with glutaryl side chains [28,33]. This feature makes the glutaryl acylase an ideal candidate to be optimised for the sustainable production of the world largest market of intermediates for antibiotics resulting in a totally "green" route by protein engineering.

Figure 3. A glutaryl acylase gene
ss, signal sequence; α, α-subunit; sp, spacer peptide; β, β-subunit

Protein engineering of biocatalysts

Enzymes are able to catalyse all kinds of chemical reactions. They can perform conversions in minutes or even seconds which would take hundreds of years without their interference, which means that they are able to enhance the rate of chemical

reactions with a factor up to 10^{17} [34]. Furthermore, they catalyse reactions which are difficult to perform by chemical methods, like the enantio- or regioselective hydrolysis or addition of chiral groups. All of these features are generally displayed at room temperature under mostly aqueous conditions. Consequently, there is a strong industrial interest in the replacement of traditional chemical processes with bioconversions. The research towards the use of biocatalysts is mainly driven by the exploration of sustainable technologies for the production of chemicals (green routes) and production of more selective and complex active ingredients in a pharmaceutical and agrobiological context [35]. This remains a big challenge, since new biocatalytic processes also have to compete economically with the well-established chemical processes which are optimised for years. Although many complicated chemical reactions can be efficiently performed by biocatalysts, nature demands different properties from enzymes than industry does. In nature most reactions occur at moderate temperatures in aqueous media, while an enzyme in an industrial process usually needs to be as stable as possible in an environment of higher temperatures, high substrate concentrations, sheering forces and organic solvents. Therefore, most enzymes found in soil and water may display the desired activity, but are generally not suited for industrial use [36]. Furthermore, for numerous industrial chemical processes an adequate enzyme can not be readily found in nature, implying that there is a need for novel biocatalysts. Enzymes with the desired activity under industrial conditions can be obtained by optimising process conditions and by protein engineering.

Rational design of new enzyme functions

The most obvious way of generating novel enzyme functions is starting from enzymes with related properties or a small activity towards the desired substrate [37]. The largest effects in protein engineering are obtained by mutating the key residues for enzyme activity, which are often present in the active site where they play a role in the actual enzymatic reaction or may be involved in stabilisation or binding of the substrate. Changing of these residues may improve the enzyme towards the wanted activity. The identification of important active site residues has been speeded up by the crystallisation of numerous enzymes in the past years. The crystal structure of an enzyme usually gives a clue about the residues that are involved in catalysis, especially when the substrate or a similar component is co-crystallised. Interactions between substrate and enzyme or between amino acids in the enzyme itself can be deduced from distances between the linked atoms. Unfortunately, crystal structures show only one possible state the enzyme can accept, which may imply that certain dynamic shapes of the enzyme will never be seen. This makes it difficult to fully comprehend the catalytic process from a single crystal structure and may lead to ignoring important residues. Nevertheless, the crystal structure of an enzyme or a close relative thereof is a good start to search for residues that should be changed in order to modify the activity of an enzyme. Modelling of the altered protein can help a researcher in choosing the right alternatives. The changes in the enzyme can be made at the DNA level by site-directed mutagenesis and the resulting

mutant enzyme can be produced and tested on the desired substrate. A lot of progress has been made in this field as can be seen in the many research articles and reviews about this subject. Unfortunately, rational design predictions do not always result in the desired activities, since the current knowledge about structure-function relation is still not good enough [38]. Crystallisation of mutant enzymes gives the opportunity to compare the new structure with the original enzyme and in an iterative approach the model and prediction of important residues can be improved. Progress is also being made in the field of bio-informatics and the algorithms of computer programs predicting the structure of a mutant enzyme are still becoming better [39,40]. It is, however, obvious that large modifications in enzyme properties can not be achieved by the change of only one or a few residues. The prediction of implications of drastic changes including several amino acids for the enzyme structure and activity, however, go far beyond our present knowledge and computational power.

Directed evolution of proteins

Many enzymes and enzyme classes are still not crystallised yet. This leaves a large number of enzymes unexplorable by rational design. Fortunately, this gap is filled by a newer branch of protein engineering: random mutagenesis. Mother nature has given us a good example how new enzymes with new activities can be evolved from existing ones. In the last decades, research has been focused on mimicking and accelerating this process in the laboratory. Since this evolution process is performed at the DNA level it was first called molecular evolution. But the technique itself is only a means to an end; the evolution of an enzyme is always directed towards a certain wanted activity by the researcher. Therefore, it nowadays is called directed evolution. All directed evolution experiments do have the same general scheme to discover the best mutant enzyme.

Methods of engineering, mutagenesis

In nature, evolution and creation of new functionalities is achieved by mutagenesis, recombination and survival of the fittest. The constantly continuing process of mutagenesis is the easiest one to copy in the lab. Mutations can be introduced at specific places using site-directed or at random mutagenesis throughout the gene. During site-directed mutagenesis an amino acid can be changed into one or more other amino acids (site-saturation mutagenesis). The mutagenesis is generally performed using oligonucleotides with altered bases. Introduction of these oligos is normally PCR based and is a common tool in DNA manipulation [41]. Different variations to this theme have also been described [42-44]. Although most site-directed mutagenesis protocols provide for the creation of only one mutation, it is not limited to this one amino acid change. Different methods to mutagenise more than one amino acid in one experiment have been developed, like QuikChange Multi Site-Directed Mutagenesis [45] and TAMS technology [46]. Although both methods claim to be highly efficient for multiple-site

mutagenesis, the efficiency of mutagenesis decreases drastically with every additional mutational site [45]. Site directed mutagenesis *in situ* can be achieved by transplacement of the mutation from template plasmids to phage DNA using cross over events [47].
Random mutagenesis can be performed *in vivo* or *in vitro*. Mutagenesis *in vivo* is performed by transforming a plasmid containing the gene to be mutagenised to a mutator *E. coli* strain. These strains lack DNA repair mechanisms [48] or contain a modified polymerase with lower fidelity [49] resulting in the introduction of mutations in total DNA during growth. The optimal growth time has to be established, since replicating plasmids for more cycles in the mutator strain will introduce more mutations in the gene of interest, but also in the chromosome, reducing viability of the bacteria. *In vitro* mutagenesis can be divided into two categories. The first strategy is based on modifications of the PCR reaction and called error-prone PCR (epPCR). Mutations are introduced by providing non-optimal conditions for a DNA polymerase, which will lead to more mismatches during extension and thus mutations in the gene of interest. The frequency of mutations introduced by this epPCR can be controlled by the amount of Mg^{2+}, by addition of Mn^{2+}, by using unbalanced or high concentrations of dNTPs, by the amount of template, by the kind of polymerase and finally by varying the extension time of the PCR. The effect of these parameters are different on every template and have to be established for each gene to be mutagenised [50]. To circumvent mutational bias, the original protocol was changed by modifying PCR conditions [51] or using a mutated polymerase which is supposed to give unbiased mutations [52]. In another *in vitro* random mutagenesis approach triphosphate derivatives of nucleoside analogues or universal bases are used in the PCR. These analogues can pair with different bases, thereby introducing all mutations possible. Mutation frequencies can be controlled by the amount and nature of nucleoside analogues and the number of PCR cycles [53,54].
Besides mutational bias there is also bias present in the coding of amino acids, because of the degeneracy in the genetic code [55]. In order to circumvent this bias, methods that are based on the mutagenesis of codons rather than single amino acids were established, like MAX [56] and RID [57]. In the latter approach up to five codons can be randomly deleted and inserted, which is also a well known feature in evolution.
Mutation frequency is an important parameter in directed evolution. More mutations increase the possibility of finding a totally new activity, but also disrupt the overall structure of the enzyme. A mutation frequency of 1-3 mutations per gene, or in case of a large gene per 1000 bps, is commonly regarded as a good rate for finding improvements of the desired activity of an enzyme. Hypermutation is thought to result in too many non-functional mutants and is therefore hardly used, although some nice results have been reported [58,59].

Methods of engineering, recombination

In nature, more radical changes in evolution originate from recombination. The crossovers between chromosomes of ones father and mother in the production of eggs and sperm, and the highly variable parts of antibodies are good examples of this

process. Almost 10 years ago, Pim Stemmer was the first one to recognise this and use it for mimicking evolution [60,61]. The technique is called shuffling and is based on the mixing and subsequent joining of different small DNA fragments in order to form a complete new gene (Figure 4). In the process of shuffling, both recombination frequency and homology are important. A high degree of recombination is important to get all possible combinations of mutations. The homology is needed for the reconstruction of a full-length gene, which will be translated into a protein that should be able to fold into a soluble and active enzyme. Since the small DNA fragments prime each other in order to become a full-length gene, they need to have an overlap of at least 14 bps [61]. This implicates that mutations which are closer together have less chance of being separated.

In order to find the best protein it is important to start from an comprehensive library. This library can be created by random mutagenesis or by using the diversity already present in nature. This last method is called family-shuffling [62] and uses similar genes from the same organism (orthologues) or the same genes from different organisms (paralogues). Since the best results are obtained with a lot of different parents which have an average homology > 80%, family shuffling is more appropriate for eukaryotic gene families, which usually have more members. In prokaryotic genomes appropriate related genes suited for gene shuffling are often absent. Recently, the exploitation of the metagenome has largely expanded the available sequence space of prokaryotic enzymes [63]. This involves the use of randomly picked sequences from soil or water and shuffling the gene of interest with homologous genes with unknown functions.

In the last decade at least 20 different methods were developed to be able to shuffle genes with lower homology, to improve mutant libraries by negative selection for wild type sequences or to get a less biased library [64] (Table 1). Most new methods have been invented to solve drawbacks and failures using the original protocol. Although these procedures aim to solve one problem, they usually appear to create another one. For example, it is difficult to start with highly different parent genes and still reach a high recombination frequency. Methods aiming at higher recombination frequencies start with more homologues parents, thereby introducing fewer mutations. So they have a smaller chance to result in enzymes with totally new activities. On the other hand, methods have been developed to recombine genes without any homology. These processes result mainly in only one crossover, although some methods do generate multiple crossovers at fixed places. Another way of obtaining a high recombination frequency is starting from synthetic oligonucleotides. The advantage of this method is that any mutant can be constructed and thereby the largest possible sequence space can be explored. An additional advantage is the possibility to use other codons than the original ones in order to obtain more homology. Furthermore, the preferred codon usage for the expression host can be applied in the synthetic oligonucleotides. The largest disadvantages of synthetic methods are the high costs and the large size of the library, which quickly exceeds the most elaborate screening and selection methods [65]. Other conflicting parameters seem to be speed and bias. Recombination methods that aim for

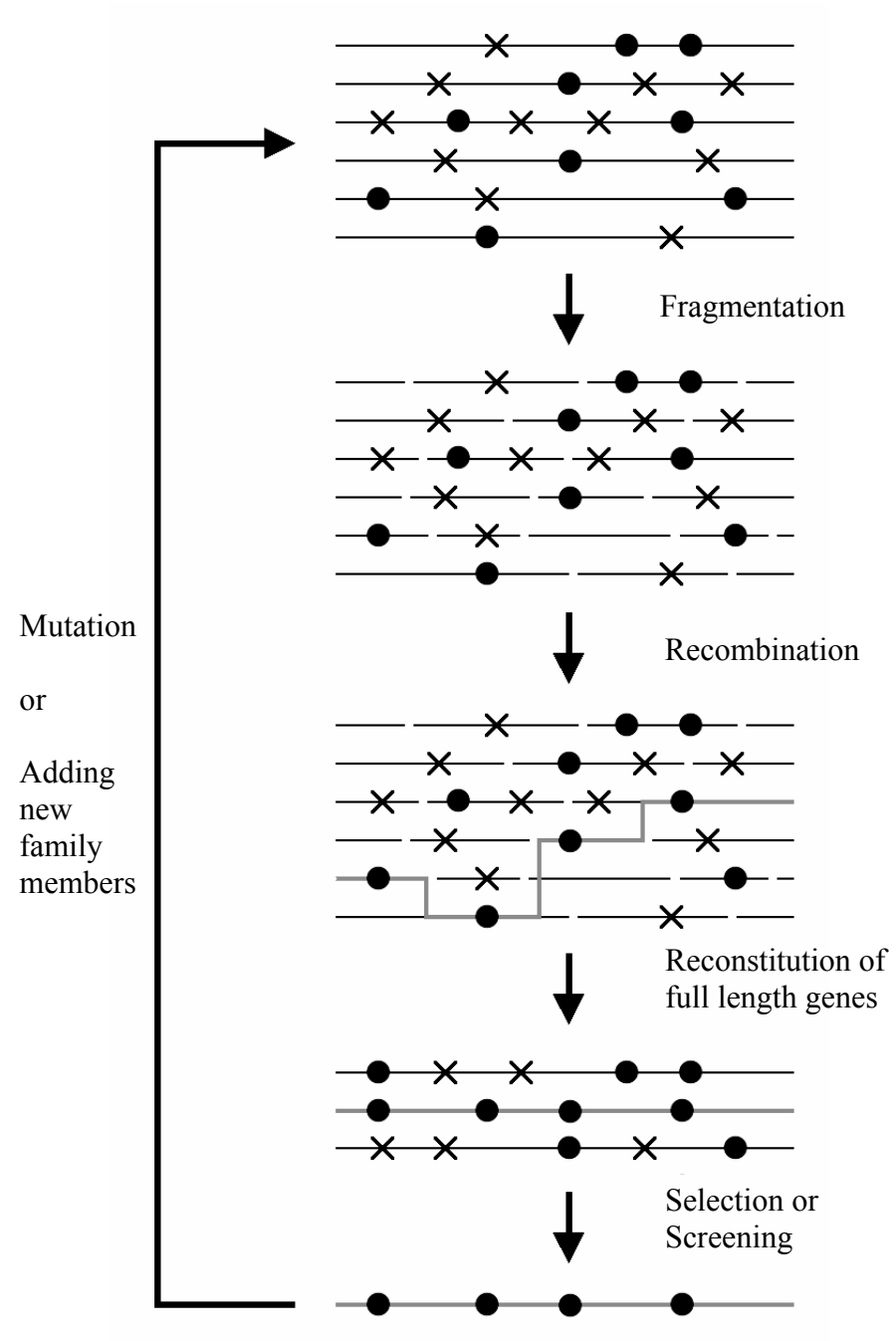

Figure 4. The basic scheme of shuffling.
The starting pool of mutant DNA can be either a mutated gene or a family of related genes.

the generation of unbiased libraries all consist of numerous steps in order to achieve this, resulting in more elaborate procedures, while quick protocols normally result in more wild type background and a biased library. Most of the mentioned methods resulted in good libraries of mutants. Therefore, the choice for one or another strategy is usually led by the size of the protein, the goal of the research, the existence of homologous proteins, the selection and/or screening capacity and practical issues like the equipment and expertise in the research group. Although many different protocols have been described in the last decades, they can be divided into six categories of underlying ideas which each have their own advantages and disadvantages (Table 1).

Table 1. Different recombination techniques with their main advantages and disadvantages

Group	Technique	Members [ref]
Shuffling	Recombination of small fragments based on homology in the sequence between mutations that stem from all kinds of mutagenesis strategies or different family members. Aims for high recombination, but difficult to separate close mutations.	Shuffling [60,61] Family shuffling [62] RE cut shuffling [75] ssDNA shuffling [76] Mn^{2+} DNase cut [77] Endonuclease V cut [78] RPR [79] RETT [80] SCRATCHY [81]
Full length parent shuffling	Recombination of small fragments from different origin using one or more full length parent strands. Higher recombination frequency, but more elaborate.	StEP [82] RACHITT [83]
Single cross over	Recombination of non homologous genes by ligating front and back of two different genes, selection of new genes on size. Recombination possible between low or non-homologous genes, but only one recombination point.	(THIO)ITCHY [84,85] SHIPREC [86] SCRATCHY [81]
Domain swapping	Recombination of structural, functional or less homologous parts of different family members. More active enzymes in the resulting library, but only a few recombination points, which are hard to find.	Exon shuffling [87] DOGS [88] SISDC [89] YLBS [90] SCOPE [91]
In vivo recombination	Recombination using the gap repair system of yeast or de recE/recT system of E. coli. High yield, since no ligation necessary, but specialised vectors and multiple steps necessary.	CLERY [92] ET-recombination [93]
Synthetic shuffling	Recombination of (un)known mutations in synthetic oligonucleotides. Recombination of close mutations possible, but expensive and good selection necessary.	Single step shuffling [94] DHR [95] Synthetic shuffling [96] ADO [97]

Selection of new enzymatic functions

The success of a directed evolution experiment highly depends on the method that is used to select the best mutant enzyme. Since most directed evolution experiments generate a huge mutant library, it is very important to develop an efficient method to search this library for the desired property. Both selection and screening strategies have been developed for all kind of enzyme functions. The big challenge in these strategies is making the improved function quantifiable. Enzymatic assays have to be sufficiently sensitive and specific to identify positive mutants [66].

Selection mimics the natural survival of the fittest strategy and is the most efficient method to find the best mutant, since only mutants of interest will appear. Unfortunately, this approach is not possible for all enzymatic activities. Selection is based on the fact that mutants with the desired enzyme function have an advantage over wild type enzymes. For *in vivo* selection this means that only enzyme activities with a growth or survival advantage can be used. Only a few industrially interesting enzymes are essential for the bacterial cell themselves, so most selection methods are based on enzymatic activities which produce a product that is essential for growth of the expression host. The coupling of the desired enzymatic reaction to survival in the selection step often requires the development of complex, nontrivial and intelligent assays [67]. Sometimes, this means that the substrates in these selection systems are not the desired substrates, but analogues thereof. This may result in the selection of undesired mutants with activity towards the analogue and not towards the wanted substrate. For example, directed evolution of penG acylase towards a glutaryl acylase was performed with a glutaryl-leucine substrate as sole leucine source for the leucine deficient bacteria. Mutants that were selected on this substrate could, however, not hydrolyse the desired substrate glutaryl-7-ACA [68]. It is, therefore, very important to carefully choose the selection substrate, since the first law of directed evolution is: “you get what you select/screen for” [69].

In vitro selection is usually based on the binding of the enzyme to the desired substrate or a transition state analogue, although strategies in which catalytic properties are used for selection are also described [70]. These methods are mostly based on a physical linkage between phenotype and genotype. The first established and most used technique is phage display, which has been successfully used for finding improved enzymes. In this system the enzyme of interest is fused to a coat protein of a filamentous phage and thereby displayed on the outside of the phage, where in principle it is able to retain enzymatic activity. Since the gene encoding the displayed protein is present in the phage particle, the gene of the mutant enzyme with the desired property is linked to its phenotype [71]. Other *in vitro* selection methods with a physical phenotype-genotype linkage are cell-surface display, ribosome display, plasmid display and mRNA-protein fusion [72]. Recently, a different approach was described to maintain a linkage between genotype and phenotype. *In vitro* compartmentalisation is a method in which compartments are formed as aqueous droplets in water-in-oil emulsions which contain

only one gene and a complete transcription/translation machinery [73]. These droplets mimic a bacterial cell by keeping the gene and its product together. The droplets containing an enzyme with the desired activity can be selected by FACS or, when the gene is physically bound to the substrate, by breaking the droplets and fishing out the desired product [74]. The advantages of *in vitro* over *in vivo* selection are the larger sample size of a mutant library and the larger amount of possible enzymatic activities to be tested. A drawback is that making the right water-in-oil emulsions with only one gene per droplet is tricky and *in vitro* translation can be a large problem

Screening for new enzymatic functions

Another way of finding the desired mutant enzyme is by screening. In screening methods all mutants have to be tested for the desired enzymatic reaction, even those that might not be active or accurately folded. The advantage is, however, that almost every enzymatic reaction can be tested, since the activity does not have to be dependent on growth rate or the formation of essential products. This can be done in a qualitative way by relatively simple visual screens such as the formation of coloured or fluorescent products or halos around a colony on a plate. For protein functions such as catalysis of a specific reaction or substrate specificity this is very difficult or even impossible. Quantitative methods are more suited to screen for these enzymatic activities, but are usually more elaborate. This implicates that in a normal time scale only small libraries can be tested (up to 10^5) or high throughput screening (HTS) has to be employed. HTS demands miniaturisation and automation of enzymatic assays. In the past decade a lot of research has been focussed on finding better, cheaper, quicker and more accurate HTS assays [98]. This has made HTS feasible for many laboratories all over the world now, resulting in a lot of smart enzymatic screening methods.

A screening method is used to find the best enzyme out of a large pool of mutants. The resulting amount of clones depends on the accuracy of the assay and should be optimised during the directed evolution experiment. For this optimisation one should consider the size of the library, the amount of (false) positive and false negative colonies that can be allowed, the costs of the assay and the possibility of performing a more accurate assay with the best mutants. Most screening assays are based on spectrophotometric methods in 96 or 384 wells plates [66]. Usually raw cell extracts are incubated with the substrate or an analogue thereof, which will give a coloured product. Another possibility is the use of a discontinuous assay in which the product, but not the substrate, is coloured by another chemical, like fluorescamine [99]. Yet another screening method based on coloured products is a spectroscopic colony screening technology. Colonies are grown on a membrane and transferred to the substrate in the presence of a colouring agent, resulting in blue spots. A digital camera can follow the colour formation as a function of time. Throughputs of ~50.000 micro colonies per membrane have been demonstrated [98]. More HTS methods are developed every day, both for specific enzymatic reactions and general applications, making screening the method of choice for many researchers.

Aims and outline of the thesis

This thesis describes the directed evolution of the cephalosporin acylase from *Pseudomonas* SY-77. The objective of this directed evolution experiment was to change the substrate specificity of the enzyme from hydrolysis of glutaryl-7-ACA towards hydrolysis of adipyl-7-ADCA or even CPC in order to obtain 7-ADCA and 7-ACA, which are key intermediates in the production of semi-synthetic β-lactam antibiotics. In order to become industrially relevant, the hydrolysis activity of the enzyme towards adipyl-7-ADCA or CPC should at least be the same as the activity towards glutaryl-7-ACA of the wild type enzyme.

To achieve this goal, we started with the randomisation of the β-subunit by epPCR to find amino acids important for substrate specificity. For this directed evolution experiment both a new selection strategy and a robotic screening assay had to be developed. The positions of the selected mutant residues were examined in the crystal structure of the highly homologous cephalosporin acylase from *P. diminuta* KAC-1. This pointed our attention towards single residues that are important for the substrate specificity of the enzyme. These amino acids were picked for a site-saturated mutagenesis approach in order to explore their contribution to substrate specificity. As a final approach, multiple mutants were constructed using both rational and random strategies: the best mutant from the α-subunit was combined with several improved mutants of the β-subunit, a library of two totally randomised residues was created and a total randomised library of all five important substrate specificity residues was constructed.

Throughout the thesis both rational and random mutagenesis strategies are used to find an industrial applicable adipyl- or CPC-acylase. The results presented here, point to the conclusion that this is the right way to improve biocatalysts.