CHAPTER I

INTRODUCTION

Since ages, biocatalysts have been used by humans to improve their daily products. The first biocatalysts have been used for food industries. Since a century, their use has been gradually extended to many fields (health, cosmetic, fine chemistry, environment, and bio-energy). In parallel to the diversification of their applications, biologists tried to understand their molecular mechanisms and to characterize them from a kinetic and biochemical point of view (optimal T°, pH, Aw...). Then, they overcame the climatic, geographical and politic risks of their production by learning how to produce them either in homologous or heterologous expression systems. Twenty years ago, the emergence of enzymatic engineering made it possible to adapt the performances of catalysts to industrial conditions. Indeed, these genetic techniques allowed the improvement of the activity and thermostability of enzymes, as well as to make them resistant to extremes pH or tolerant to organic solvent. Nevertheless, the area where these genetic techniques yet display all their huge potential still remains the improvement of enzyme selectivity. Two approaches can be used to improve enzymes. The first approach consists in a rational approach based on the study of the relations between the structure and the function of the biocatalyst. It enables targets to site-directed mutagenesis to be chosen, such as amino-acids in the active site of the enzyme. The second approach is an approach by directed evolution, which consists in selecting a biocatalyst with the improved properties among a library of enzyme variants randomly generated by genetic engineering. The first method requires the knowledge of the three-dimensional structure of the enzyme and the fine comprehension of the mechanisms involved at the molecular level. On the contrary the second method does not imply any structural or molecular knowledge pre-necessity but necessitates the development of an efficient high-throughput screening method able to rapidly test a large library of variants.

This thesis project modestly lies within this scope of study, focusing more particularly on lipases from *Candida rugosa* and *Yarrowia lipolytica*. *C. Rugosa* expresses 5 main lipases. Despite their high homology (superior to 70%), they display

different selectivities. They were greatly studied in the past and it was demonstrated that they are vey interesting enzymes in different application fields. However, the difficult to study them separately and to express them as heterologous proteins, because this yeast has an original non-universal codon usage, impeded their development in the last years. Today this difficulty is overcame by the possibility of synthesizing, for a reasonable price, genes adapted to the expression in a given host. In our opinion, lipases from *C. rugosa* are original tools and their comprehension should be deepened. Indeed, their active site was shown to be a long tunnel of 22 Å with the noteworthy characteristic that the catalytic triad is at the entrance of the tunnel and that the substrate, in particular fatty-acid, must be able to place in this tunnel to be catalyzed. It confers to this enzyme a remarkable interest to purify molecules of interest. For example, it would be possible to modify the geometry of the tunnel in order to control the chain length specificity of these enzymes and thus to purify fatty-acids of interest. The second yeast studied in this thesis, the oleaginous yeast *Y. lipolytica*, has been described relatively recently, contrarily to those from *C. rugosa*. It is one of the most active enzymes regarding the hydrolysis of long chain fatty-acids, such as oleic acid, a propriety exploited for the processing of waste treatment in fat industry. In addition, it is very stable at acidic pH, and this could thus be potentially interesting for the substitution of the human pancreatic lipase for patients displaying this deficiency (cystic fibrosis for instance); it is currently in clinical evaluation for such an application. Moreover, it presents interesting selectivity towards different substrates with high added value, such as 2-halogeno-arylacetic acids (precursor of chiral compounds used in the drug industry), esters of DHA (fatty-acid belonging to the family of omega-3) and esters of 2-ethylhexanol (plasticizer, lubricant...).

The objective of this thesis was to improve the properties of these lipases by enzymatic engineering using a rational approach based on the study of the relation between the structure and the function of these enzymes. The reaction of interest was the resolution of a racemic mixture of pharmaceutical interest, pertaining to the family of 2- halogeno aryl acetic acids. These compounds are intermediates in the synthetic pathways of many drugs, such as prostaglandin, prostacyclin, semi-synthetic penicillin or thiazolium salts, and analgesic. The use of racemic mixtures in drugs is

strongly regulated and thorough investigations are requested on the toxicological effects of each enantiomer before they are accepted in drug composition. The utilization in the process of the sole active enantiomer is thus often privileged whenever it can be obtained in a pure form. However, classical methods used to separate enantiomers, such as chemical asymmetric synthesis, stereoselective crystallization or chiral chromatography, are usually expensive. The use of enantioselective enzymes thus appears as an appealing alternative to separate enantiomers. However, to obtain purities compatible with pharmaceutical legislation, biocatalysts with high enantioselectivity are required (*E*-value $\geq$200). It was shown previously that the nature of the halogen largely influenced the enantioselectivity of lipase Lip2 from *Y. lipolytica*. The most important factor to explain this phenomenon is the obstruction of the atom. Indeed, a bromine atom, more bulky than fluorine or chlorine, enables the enantioselectivity to be maximized, whereas the nature of the ester part only influences the selectivity to a minor extend. For these reasons, and because a long chain ester has the practical advantage to reduce times of analysis, we consequently decided to work on a racemic mixture of 2-bromo phenyl acetic acid octyl esters.

For the individual study of *C. rugosa* lipases, the first step was to express and produce them in an effective host. We chose the yeast *Y. lipolytica* because many proteins from different microorganisms (Lip2 lipase for instance) were successfully expressed as recombinant proteins in this yeast. In addition, this yeast combines ease of use, low cost, post-translational modification efficiency and secretion abilities, whereas *Saccharomyces cerevisiae* leads to hyperglycosylation and presents low secretion efficiency and *Pichia pastoris* displays low transformation efficiency, varying copy numbers and incomplete cassette integration. Furthermore, it appeared necessary, upstream from this study of enzyme improvement, to validate a previously constructed strain of *Y. lipolytica* specifically designed for the screening of optimized enzymatic activities. The use of *Escherichia coli*, *S. cerevisiae* and *P. pastoris* enables the comparison of variant selectivity because *E*-values are ratios of initial rates; however, it is generally not possible to directly compare activities of variants from crude extracts because the control of the expression system is not sufficiently accurate. Causes of this lack of accuracy are numerous: marker gene conversion,

multiple integration or plasmids integrated in multi-copies, difficulty to control host growth and protein expression. As a consequence, it is often necessary to use long and fastidious purification steps to be able to compare the activities of generated enzyme variants. Recently, a new strain of *Y. lipolytica* dedicated to high-throughput screening procedures was developed. An expression system was constructed with long terminal repeats of retrotransposon Ylt1, termed zeta regions, flanking the expression cassette. In the French *Y. lipolytica* strain formerly used, which does not contain a zeta sequence, the expression cassette integration was done at random into the chromosome. A new strain, namely JMY1212, was constructed with integration into the genome of a "zeta docking platform" enabling the expression cassette insertion at a defined locus. Thanks to homologous insertion, the transformation efficiency was improved to 8000 transformants per µg DNA, compatible with the construction of a large variant library. Moreover, the rate of gene conversion of the marker used for the selection of the transformants was shown to be very low (0.23%) compared with *P. pastoris*, in which marker gene conversion represents about 10 to 50 % of the transformants. The first objective of this thesis project was to test the reproducibility of this strain, i.e. to evaluate its performances for the control of the expression level and thus the direct comparison of the activities of a large number of enzyme variants.

This thesis is organized in three chapters.

- The second chapter reviews the existing literature on the various fields of our study (general information on lipases, lipases from *C. rugosa* and *Y. lipolytica*, expression system).

- The third chapter describes the material and methods used in this study and come as a complement to the experimental part of the three publications.

- The fourth chapter is written in the form of three publications:

    - Publication1: Overexpression and characterization of *Candida rugosa* lipase isoenzymes in *Yarrowia lipolytica*

- Publication 2: A single amino acid located in the substrate binding site of *Candida rugosa* lipases Lip1, Lip3 and Lip4, direct their enantioselectivity towards 2-bromo phenyl acetic acid octyl esters
- Publication 3: Biodiesel production catalysed by immobilized *Candida rugosa* lipase

Finally, the general conclusion gives an overview of the results obtained during this thesis work and opens the way to prospects.