CHAPTER II

LITERATURE REVIEW

PART I: Lipases –powerful enzymes for industrial purposes

1. Generality on lipases

Lipases (triacylglycerol ester hydrolases, EC 3.1.1.3) are ubiquitous enzymes encountered in microbes, plants and animals, which catalyse in nature the hydrolysis of fats and oils. Nevertheless, for the benefit of biotechnologists, and according to the environment and reaction conditions, lipases are able to catalyse a very large range of synthesis reactions. Their robustness and their ability to recognise a large range of substrates, which can be very different from their natural substrates, confer to this enzyme class important potentiality in biotechnology. Lipases are currently used in several different industrial applications such as washing detergents, energy, brewery, agro-industries, depollution and production of fine chemicals and pharmaceuticals (Jaeger and Reetz, 1998; Jaeger and Eggert, 2002).

1.1. Catalysed reactions

1.1.1 Hydrolysis reaction

Lipases naturally catalyse the hydrolysis of tri-, di- and mono- glycerides into fatty-acids and glycerol (shown below). They are also active on many esters of fatty acids. In all the cases, the reaction is carried out at an interface in a biphasic system reaction. This biphasic system results from the presence of an immiscible organic phase to water, this phase being made up either of the substrate alone, or of the substrate dissolved in a solvent immiscible with water.

$$R_1COOR_2 + H_2O \underset{2}{\overset{1}{\rightleftharpoons}} R_1COOH + R_2OH$$

During the reaction of hydrolysis, one will be able to distinguish lipases by their regioselectivity, type-selectivity or enantioselectivity (see below, 3. Specificity of lipases).

1.1.2. Synthesis reactions

Lipases catalyse a large variety of synthesis reactions which are listed below. These synthesis reactions usually occur in media with low thermodynamic water activity, the thermodynamic activity being a measure of molecule availability in a solvent. The medium then consists in a free-solvent medium (molten medium) or in an organic solvent. In these conditions, lipases present high remarkable stability, in particular under their immobilized form. These reactions are widely used in industrial processes.

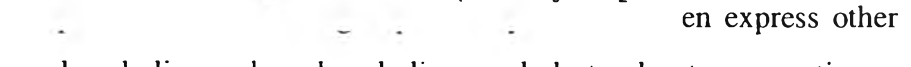
- esterification

- transesterification
  - alcoholysis

  - Interesterification

- thio-esterification:

- aminolysis or amidation:

Moreov  en express other activities such as phospholipase, lysophospholipase, cholesterol esterase, cutinase, amidase and other esterase type of activities (Svendsen, 2000).

1.2. Source of lipases

Some of the common sources of lipases are listed in Table I-1. Pancreatic lipase of porcine origin is one of the earliest characterised and is still the best known

lipase. The most important source of animal lipase is the pancreas of cattle, sheep, hogs and pigs. The disadvantage with pancreatic (animal) lipases is that they cannot be used in the processing of vegetarian or kosher food. In addition, these extracts contain components which have undesirable effect. For instance, the pig pancreatic extract contains trypsin, which produces bitter tasting amino acids. They are also likely to contain residual animal viruses, hormones, etc. Plant lipases are not used commercially whereas animal and microbial lipases are used extensively. Microbes are thus the major source of lipases. Yeast has earned acceptability since long and are considered natural. In addition, due their ease of handling and growth, they were often used to produce lipases used in food and other industries since ages (Arpigny and Jaeger 1999; Jaeger *et al.,* 1999). Some of the lipases currently produced by yeasts are reported in Table I-2.

Table I-1. Common mammalian, fungal and bacterial sources of lipases.

Source	Name
Mammalian	Human Pancreatic Lipase Horse Pancreatic Lipase Pig Pancreatic Lipase Guinea Pig Pancreatic Lipase
Fungal	*Rhizomucor meihei* *Pencillium camberti* *Humicula lanuginose* *Rhizopus oryzae* *Aspergillus niger* *Candida rugosa** *Candida Antarctica* Lipase A* *Candida Antarctica* Lipase B* *Geotrichium candidum**
Bacterial	*Chromobacterium viscosum* *Pseudomonas cepacia* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Pseudomonas fragi* *Bacillus thermocatenulatus* *Staphylococcus hyicus* *Staphylococcus aereus* *Staphylococcus epidermidis*

*indicate yeast lipases. (http://www.au-kbc.org/bata/bioproj2/sources.htm)

Table I-2. Reported lipase producing yeasts, their cellular localization and number of lipase isoenzymes (Vakhlu and Kour, 2006).

Source	Cellular - location	Isoforms	Reference
Arxulaadeni nivorans	Extra-cellular	1*	Boer et al. 2005
Candida albicans	Extra-cellular	10*	Hube et al. 2000
Candida antarctica	Extra-cellular	2*	Høegh et al. 1995 Rotticci et al. 2001
Candida ernobii	Extra-cellular	1	Pignede et al. 2000a Pignede et al. 2000b
Candida parapsilosis CBS 604	Cell bound	1*	Neugnot et al. 2002
Candida rugosa/cylindracea			
ATCC 14380	Extra-cellular	5*	Brocca et al. 1995
DMS 2031	Extra-cellular	3	Benjamin and Pandey, 2001
L 1754	Extra-cellular	2	Veeraragavan and Gibbs, 1989
Candida curvata	-	-	Lazar and Schroder, 1992
Candida tropicalis	-	-	Lazar and Schroder, 1992
Candida deformans CBS 2071	Extra-cellular	3*	Bigev et al. 2003
Geotrichum asteroids FKMF 144	Extra-cellular	2	Kazanina et al. 1981
Geotrichum candidum			
ATCC 34614	Extra-cellular	2*	Shimada et al. 1990
NRCC205002		2*	Bertolini et al. 1994
NRRL Y-552		2*	Bertolini et al. 1994
NRRL Y-533		2*	Bertolini et al. 1994
CMICC 335426		2	Charton et al .1992
ATCC 66592		2	Jacobsen and Poulsen, 1992
Geotrichum sp. FO401B	Extra-cellular	2	Ota et al. 2000
Kurtzmanomyces sp. I-11	Extra-cellular	1	Kakugawaet al. 2002
Kluyveromyces lactis	Extra-cellular	1*	Oishi et al.1999
Saccharomysis cerevisiae	-	-	Oishi et al.1999
Saccharomysi fibuligera	-	-	Pandey et al.1999
Trichosporona steroides	Extra-cellular	1	Dharmsthiti and Ammaranond, 1997
Trichosporon cutaneum	Extra-cellular	-	Chen et al.1993
Trichosporon fermentans WU-C12	Extra-cellular	2*	Chen et al. 1993 Chen et al. 1994 Arai et al. 1997
Yarrowia lipolytica	Extra-cellular and cell bound	2, 1*	Ota et al. 1982 Destain et al. 1997 Pignede et al. 2000a

1.3. Application of lipases

Lipases (mainly microbial lipases) have many advantages on a biotechnological aspect: i) ease of production, ii) stability in organic solvents, iii) no requirement for cofactors, iv) broad range of specificity v) active under ambient conditions and thus low energy expenditure requirement vi) great selectivity in the reactions implemented. These characteristics made them particularly interesting for industrial applications as biocatalysts. Here are some of their important applications (Pandey *et al.*, 1999):

- in detergents industry, as additives. Lipases have to be both active and stable at high temperatures, alkaline pH and in the presence of the other components of the detergent. Since the first marketed lipase (Lipolase; Novozyme), more powerful enzymes were proposed (Lipex- Novozyme).

- in food industry; in the industry of dairy products, for instance, they are used to increase the aromatic of cheeses, and to accelerate the maturation. In the manufacture of bread, they are used like emulsifier of the bread dough (Lipopan F - Novozyme). Lipids can also modified by *sn*-1,3 regiospecific lipases to produce structural triglycerides (cocoa butter equivalent, human milk substitute, DHA triglycerides...)

- in paper industry, to remove triglycerides and pulp waxes.

- in leather industry, to remove greases.

- in environmental processes, to treat the oil wastes.

- in the industry of perfumes and cosmetics; lipases are used in two major applications: production of surfactant (mono- and di-acylglycerols) and production of flavours.

- in chemical industries; to obtain optically active pure compounds (used in the pharmaceutical and phytosanitary industries). Many chiral drugs are synthesized with lipases starting from racemic mixtures. It is the case of prostaglandins, cephalosporines, non-steroid anti-inflamatory drugs (naproxene, ibuprofene, ketoprophene), hydantoins, penicillins.

- in the field of energy; they are used for obtaining bio-diesel starting from renewable resources such as plant oils.

Table I-3 reports some commercially available lipases and their industrial applications.

Table I-3. Some commercially available microbial lipases and their applications (Jaeger and Reetz, 1998).

Type	Source	Application	Producing company
Fungal	*C. rugosa*	Organic synthesis	Amano, Biocatalysts, Boehringer Mannheim, Fluka, Genzyme, Sigma
	C. antarctica	Organic synthesis	Boehringer Mannheim, Novo Nordisk
	T. lanuginosus	Detergent additive	Boehringer Mannheim, Novo Nordisk
	R. miehei	Food processing	Novo Nordisk, Biocatalysts, Amano
Bacterial	*Burkholderia cepacia*	Organic synthesis	Amano, Fluka, Boehringer Mannheim
	P. alcaligenes	Detergent additive	Genencor
	P. mendocina	Detergent additive	Genencor
	Ch. viscosum	Organic synthesis	Asahi, Biocatalysts

2. Structure of lipases

The first structures of lipases were elucidated with the beginning of 1990. These were the lipases from *Rhizomucor miehei* (Brady *et al.,* 1990) and human pancreatic lipase (Winkler *et al.,* 1990). Thereafter, many other lipases were crystallized, as shown in Table I-4, thus highlighting a common structural organization, the so-called α/β hydrolase fold (data concerning *Candida rugosa* lipases will be stated apart in paragraph II.2.). By the end of 2005 about 2000 non-redundant sequences of lipases and related enzymes were present in protein sequence databases. No specific sequence similarity is shared by all known lipases. On the contrary, they appear to be astonishingly variable. The lone consensus shared by all of them, containing the catalytic serine, is the pentapeptide Gly-X-Ser-X-Gly (with rare cases where glycines are substituted by other small residues. In the Lipase Engineering Database (LED), lipases are grouped in 16 superfamilies and 39 homologous families (Fischer and Pleiss, 2003). The classification of lipases in superfamilies and homologous families is shown in Table I-5.

Table I-4 Structural data existing for some extensively studied fungal lipases.

Lipase sources	References	Resolution	PDB code	Substrate/inhibitor
Rhizomucor miehei	Brady *et al.*, 1990	1.90 Å	1TGL	Close form
	Brady *et al.*, 1990	1.90 Å	3TGL	Close form
	Derewenda *et al.*, 1992	2.60 Å	4TGL	Open form + Diethyl phosphonate
	Brzozowski *et al.*, 1991	3.00 Å	5TGL	Open form + n-hexylphosphonate ethyl ester
Humicola (Thermomyces) lanuginosa	Brzozowski *et al.*, 1991	2.60 Å	3DT3	
	Yapoudjian *et al.*, 2002	2.35 Å	1DTE	
		3.00 Å	1EIN	Di-undecylphosphatidylcholine
		2.50 Å	1DU4	
		2.40 Å	1DT5	
		2.20 Å	1GT6	Open form + Oleic acid (serine catalytic mutated)
Rhizopus delemar (oryzae, nieveus)	Derewenda *et al.*, 1994	1.84 Å	1TIB	Close form
		2.60 Å	1TIC	1 Close form +1 Open form In the same crystal structure
Penicillium camembertii	Derewenda *et al.*, 1994	2.10 Å	1TIA	
Penicillium expansum	Bian *et al.*, 2009	1.30 Å	3G7N	

2.1. The Common α/β structural Fold

The first description of the above mentioned α/β fold, which is common to many hydrolytic enzymes including proteases, was done by Ollis *et al.* (1992). Almost of lipases belong to the super family of α/β hydrolases. The design of the canonical α/β hydrolase fold is based on a central, mostly parallel β-sheet of eight strands with only one antiparallel strand (β2). Strands β3 to β8 are connected by α-helices packed on both sides of the β-sheet (Figures I-1 - I-2). Variations from the canonical fold can affect the number of α-strands, the presence of insertions, and the architecture of the substrate binding subdomains (Figure I-3). The consensus motif previously mentioned forms a sharp turn (the nucleophile elbow) in a strand-turn-

helix motif in strand β5 which forces the nucleophile residue of the catalytic triad to adopt unusual main chain Φ and Ψ torsion angles. Due to its functional relevance, the nucleophile elbow is the most conserved feature of the fold. It can be note that lipases can have one or more disulphides bridges which can be important for the catalytic activity or the stability of the enzyme.

Table I-5. Classification of lipases in superfamilies and homologous families; The catalytic triad is in bold letters, oxyanion hole residues are underlined and number of Swiss-Prot entries is in parentheses

Superfamilies	Homologous families	Source	Oxyanion hole	Ser	Asp/Glu	His	Swiss-Prot
Acinetobacter calcoaceticus	Ac. esterase	*Ac. calcoaceticus*		147 GDSCG			EST_ACICA
Bacillus subtilis	B. lipase	*B. subtilis*		106 AHSMG			LIPA_BACSU
Candida antarctica	C. antarctica lipase B[1]	*C. antarctica*	61 LVPGTGT	128 TWSQG	210 ATDEI	247 IDHAG	LIPB_CANAR
Candida rugosa	C. rugosa lipase[2]	*C. rugosa* (5), *Geot. geotrichum* (2)	134 WIFGGGF	222 GESAG	354 NDEGT	462 TFHSN	LIP1_CANRU
Carboxylesterases	A. oxidans hydrolase	*A. oxidans*	106 WIHGGGL	186 GQSGG	365 RDEGT	400 AVHCI	PCD_ARTOX
	B. subtilis esterase	*B. subtilis*	102 WIHGGAF	187 GESAG	308 RDEGY	397 AFHAL	PNBA_BACSU
	Ca. elegans cytoplasmic esterase	*Ca. elegans*	121 YIHGGGY	205 GQSAG	328 EYEGL	445 AVHCT	EST2_CAEEL
	Ca. elegans gut esterase	*Ca. briggsae, Ca. elegans*	113 WVHGGGY	197 GYSAG	318 NSEGS	449 SPHAN	EST1_CAEBR
	Cu. pipiens esterase	*Cu. pipiens*	105 YIYGGGF	189 GHSAG	322 SEEGL	440 TAHAD	EST1_CULPI
	Dic. esterase	*Dic. discoideum* (2)	135 FIPGGAF	213 GESAG	338 QDEAI	441 VCHGT	CRYS_DICDI
	Hel. virescens esterase	*Hel. virescens*	136 FIHGGGF	218 GQSAG	349 SSECE	463 VGHIE	EST1_HELVI
	M. bile salt activated lipase[3]	*R. norvegicus, M. musculus* H. sapiens, Bos taurus	125 WIYGGAF	209 GESAG	338 DMDGH	453 ADHAD	BAL_RAT
	M. carboxylesterase	*R. norvegicus* (5), *M. musculus* (3), *O. cuniculus* (2), *Me. auratus, H. sapiens, An. platyrhynchos*	138 WIHGGGL	219 GESAG	338 KQEFG	451 GDHGD	EST1_RAT
	My. persicae esterase	*My. persicae*	131 FIHGGGY	212 GMSAG	337 QDEGL	461 FTHGD	ESTE_MYZPE
	Pseudoobscura esterase	*D. pseudoobscura* (3), *D. melanogaster* (2), *D. simulans, D. mauritiana, D. virilis*	124 QIHGGAF	205 GHSAG	336 TEDGG	465 TVHGD	ESTB_DROPS
	Cholinesterases[4]	*Tor. californica*	135 WIYGGGF	219 GESAG	346 KDEGS	459 VIHGY	ACES_TORCA
Cutinases	Fusarium solani[5]	*F. solani*	54 YARGSTE	134 GYSQG	189 VGDLV	202 APHLA	CUTL_FUSSO
	Colletotrichum gloeosporioides	*Co. gloeosporioides, Co. capsici, Ma. grisea, Al. brassicicola, As. rabiei, Asp. oryzae, Bo. cinerea*	53 FARASTE	134 GYSQG	189 IADAV	202 PAHFL	CUTL_COLGL
	Mycobacterium tuberculosis	*Myc. tuberculosis*	36 FARGTGE	108 GYSQG	175 PTDPI	189 SGHID	CUT3_MYCTU
Filamentous fungi	Rhizomucor lipase[6]	*Rh. miehei, Rh. delemar, Pen. camemberti, Hum. lanuginosa*	172 VFRGSSS	236 GHSLG	295 ERDIV	249 LDHLS	LIP
	Saccharomyces YJ77 lipase	*Sa. cerevisiae*	98 AFRGSTT	179 GHSLG	251 TGDYI	313 YEHRA	YJ77_YEAST
Haemophilus influenzae	Haemophilus lipase	*Haem. influenzae, E. coli*		117 GHSMG		264 SGHWV	Y193_HAEIN
Moraxella1	Moraxella lipase 1[7]	*Moraxella* sp., *S. exfoliatus*	59 ISPGETA	131 GHSMG	175 DGDTV	207 ASHFT	[10]
Moraxella2	Moraxella lipase 2	*Moraxella* sp.		237 GDSAG			LIP2_MORSP
Mycoplasma	Mycoplasma lipase	*Myp. pneumoniae* (3), *Myp. genitalium* (3)	26 FLHGFGS	95 GHSMG	210 SNDIV	239 VGHSP	ESL2_MYCPN
	Moraxella lipase 3	*Moraxella* sp., *Psy. immobilis*	72 LIHGFGG	140 GNSMG	262 DSDQV	290 VGHVP	LIP3_MORSP
	Non-haem peroxidase[8]	*P. pyrrocinia, S. liquidans, P. fluorescens, S. aureofaciens* (2)	26 FHHGWPL	94 GHSTG	226 EDDQI	255 YSHGM	PRXC_PSEPY
Pseudomonas	Cepacia lipase[9]	*P. aeruginosa, Vib. cholerae, P. fragi, P. cepacia, P. glumae*	*57 LVHGLSG*	*129 GHSQG*	*306 QNDGL*	*328 WNHLD*	*LIP_BURCE*
	Staphylococcus lipase	*St. epidermidis, St. hyicus, St. aureus*	315 LVHGFNG	416 GHSMG	607 ENDGL	646 WDHVD	LIP_STAEP
Pseudomonas fluorescens	Pseudomonas fluorescens lipase	*P. fluorescens* (2)		204 GHSLG			LIPA_PSEFL
Saccharomyces cerevisiae	Saccharomyces TGL2 lipase	*S. cerevisiae*		142 AHSMG			TGL2_YEAST
Yarrowia lipolytica	Yarrowia lipolytica lipase	*Y. lipolytica*	101 WIHGGGN	191 GESAG	301 IVDGT	390 CHHAV	LIP1_YARLI

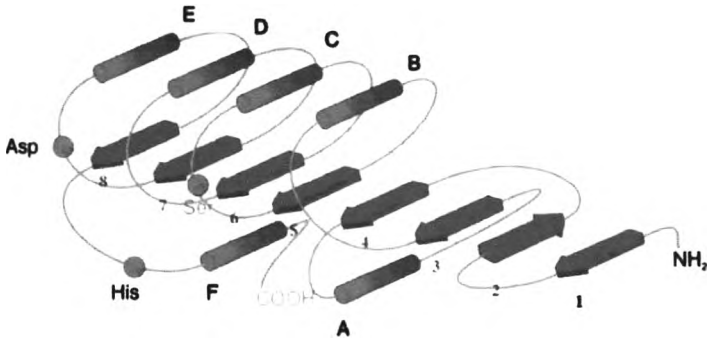


Figure I-1. Scheme of the 3D folding of α/β hydrolases. α Helices are indicated by *cylinders*, and β strands are indicated by *shaded arrows*. The topological position of the active-site residues is shown by a *solid circle*; the nucleophile is the residue after β5, the Asp/Glu residue is after β7, and the His residue is in the loop between β8 and α F (Jaeger *et al.*, 1999).

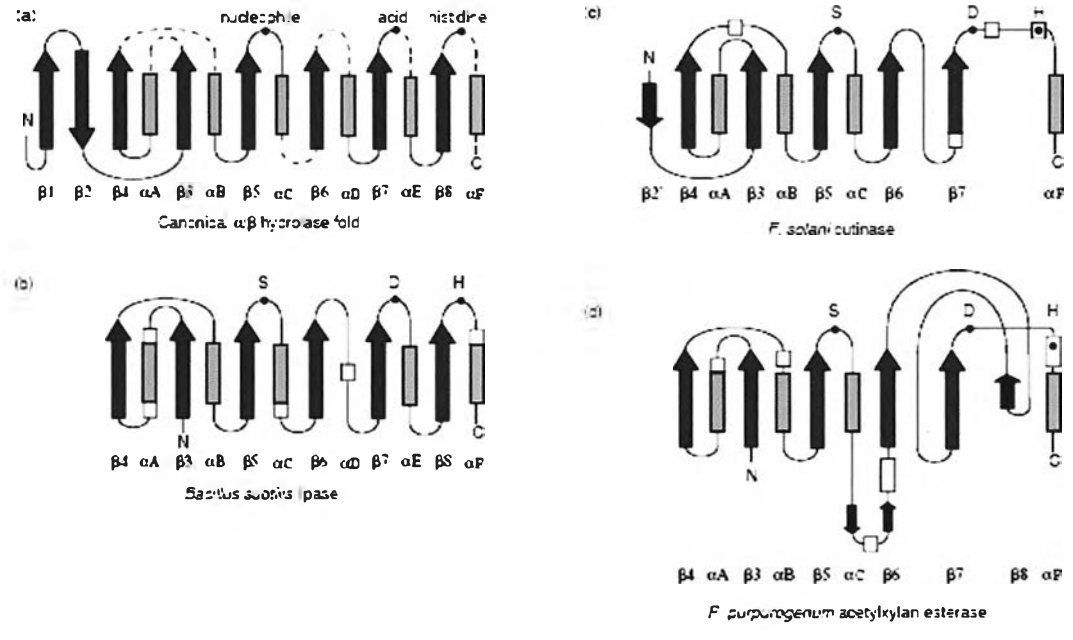


Figure I-2. Secondary structure topology of (a) the canonical α/β hydrolase fold, (b) *B. subtilis* lipase, (c) *F. solani* cutinase and (d) *P. purpurogenum* acetylxylan esterase. The broken lines in (a) indicate loops of variable length (van Pouderoyen *et al.*, 2001)

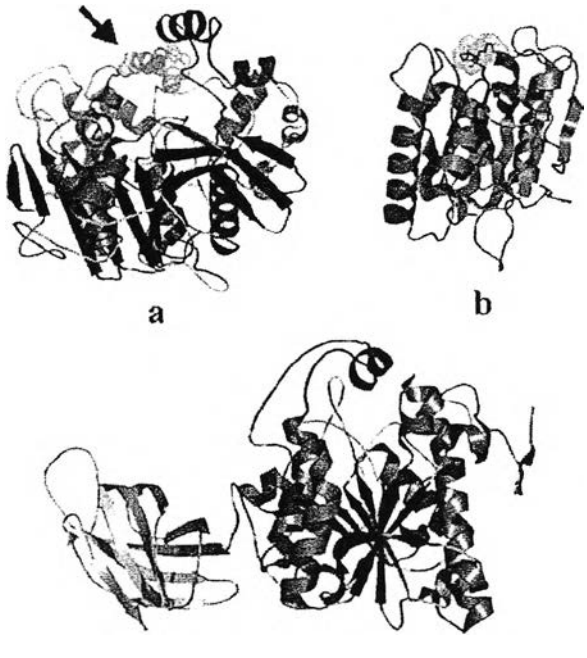


Figure I-3. Variation on the α/β hydrolase fold design in lipases of different complexity: (a) the *Candida rugosa* Lip1 (PDB 1LPO by Grochulski *et al.*, 1994). The arrow marks the oligosaccharide chain linked to Asn 351, (b) the mini-lipase from *Bacillus subtilis* distinguished by the lack of a lid structure and (c) the human pancreatic lipase with the colipase binding domain on the left side (Polaina, 2007).

2.2. Catalytic triad

In α/β hydrolases the active site consists of a catalytic triad comprising a nucleophile, an acidic residue and a histidine, reminiscent of that of serine proteases but with a different order in the sequence: nucleophile-acid-histidine (Ollis *et al.*, 1992). The lipase catalytic triad is very preserved in lipases. It is invariably compost of a serine, a histidine and an aspartic acid (or more rarely for *Geotrichum candidum* and *Candida rugosa* a glutamic acid). The active-site Ser is part of the Gly-Glu(His)-Ser-Ala-Gly-Ala/Gly conserved sequence mentioned above. The catalytic serine is always located on a nucleophile elbow " γ-turn " between the β5 strand and the following propeller α helix (Figure I-2).

2.3. Oxyanion hole

The tetrahedral intermediary formed during the catalytic reaction (see catalytic mechanism part) is stabilized by at least two hydrogen bonds formed with the amide groups of the carbon chain. These two amino-acids taking part in the stabilization of the tetrahedral intermediate form the oxyanion hole. The first residue of the oxyanion hole is located in the N-terminal of lipases and is well preserved inside the various families of lipases. It is located in loop between 3β and α-helix αA. Two types of oxyanion hole were identified by Pleiss and collaborators: GX and GGGX (Pleiss *et al.*, 2000) (Figure I-4).

The type of oxyanion hole seems to be correlated with the specificity of substrates of lipases. Indeed, lipases of the type GX will have a specificity marked for the substrates with medium and long carbon chains whereas the oxyanion holes of the type GGGX are rather found in carboxylesterases and lipases specific of the short carbon chains. Lipases of the fungal type have an oxyanion hole of type GX (X being a serine or a threonine). The team of Pleiss also highlighted, for lipases of the GX type, that another amino acid, generally aspartic acid or asparagine in lipases of the fungal type, stabilized the residue of the oxyanion hole through interaction with the lateral chain of this former.

Recently, a new class was identified, the class Y. The stabilization of the tetrahedral intermediate is not done by a connection with a proton amide of the principle chain (for types GX and GGGX) but with a side chain hydroxyl group of a tyrosin. This tyrosin is strictly preserved within these superfamilies. This class gathers primarily peptidases, endopeptidases and esterases (Fischer *et al.*, 2006).

The second residue of the oxyanion hole is the X2 residue of the sequence consensus G-X1-S-X2-G. For lipases of *Rhizomucor miehei* presenting a sequence consensus GHSLG, it is a leucine. It is positioned in the nucleophile elbow and its position is very preserved for lipases (Figure I-5). The oxyanion hole can be present in the closed formed of the lid. The interaction of amino acid which stabilizes the oxyanion hole does not take place, this interaction being only possible when lipase is in open form.

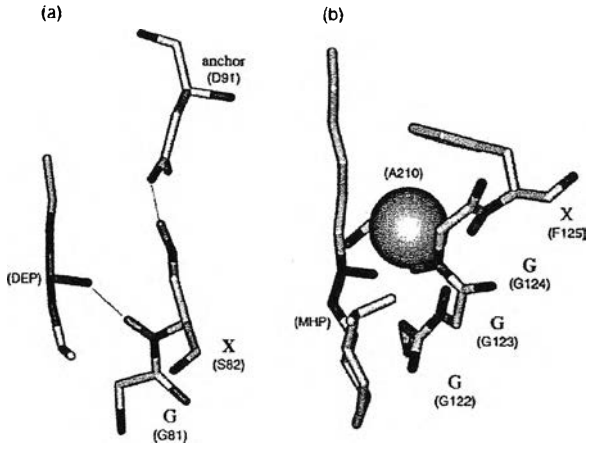


Figure I-4. Illustration of the two principal types of oxyanion holes. (a) GX type in RML (PDB entry 4TGL.): stabilization of substrate analogous inhibitor diethylphosophate DEP.by hydrogen bonding to first oxyanion hole residue X (S82); stabilization of S82 by hydrogen bond to the anchor residue D91. (b) GGGX type in *Candida rugosa* lipase, CRL (PDB entry 1LPM): stabilization of substrate analogous inhibitor (*R*)-menthyl hexyl phosphonate by the first oxyanion hole residue G (G124); stabilization of the oxyanion by locking side chain of A210 between G124 and the side chain of X (F125) (Pleiss *et al.*, 2000).

2.4. Catalytic mechanism

The catalytic mechanism of lipases (Cygler, *et al.*, 1994) is similar to the one observed with serine proteases. The first stage is an acylation. The transfer of proton between the aspartic acid (or glutamic acid), histidine and catalytic serine, induces an activation of the hydroxyl group of the catalytic serine. The nucleophile attack of this activated hydroxyl on the carbonyl of the substrate leads to a first tetrahedral intermediary (INT1) which carries a negative charge on the oxygen of the carbonyl group. The tetrahedral intermediate is stabilized by at least 5 hydrogen bonds, including at least two bonds between the negative charge of the substrate carbonyl oxygen and the amide proton of the oxyanion hole. By return of the doublet of oxygen and the transfer of proton of histidine, a molecule of water (R'=OH) or alcohol is released depending on the substrate. The enzyme-substrate complex thus formed is called acyl-enzyme. The second phase is deacylation and utilizes a nucleophile attack which can be exerted by a water molecule (hydrolysis) or by an alcohol (synthesis - esterification) on the carbonyl of the acyl-enzyme. A second tetrahedral intermediate

is formed (INT2) according to a process similar to the stage of acylation and leads to the formation of the second product of the reaction (carboxylic acid or ester) with regeneration of the native enzyme (Figure I-5).

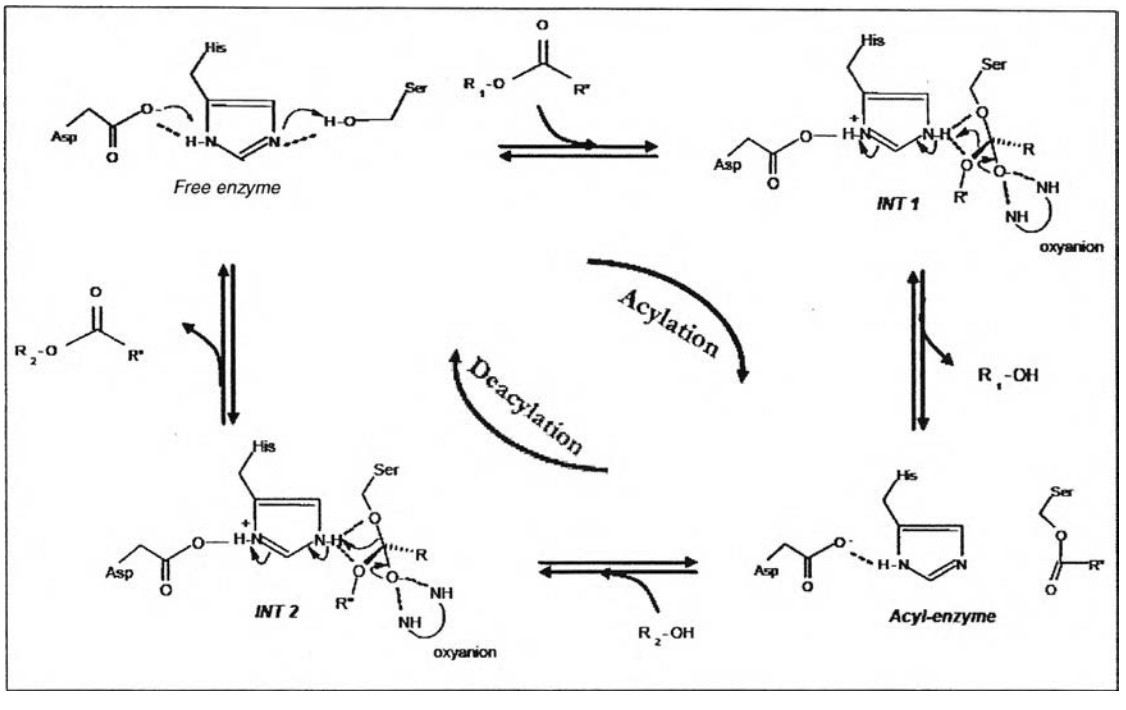


Figure I-5. Catalytic mechanism of lipases according to Cygler et al., 1994. INT1 and INT2 represent the first and the second tetrahedral intermediate, respectively.

2.5. The lid of lipases

One particularity of lipase catalyzed reactions is that they occur at an interface water/hydrophobic phase (substrate or organic solvent). In the presence of this interface, catalytic reaction rate are much higher than that observed in a monophasic system. This phenomenon is called interfacial activation (Figure I-6). It is linked to the presence of a mobile element composed of one or more α-helices, the so-called lid. The lid region is an amphipatic structure which covers the catalytic active site of most lipases and is involved in modulating activity and selectivity of lipases (Secundo *et al.*, 2006). When the interface is absent, the entrance of the active site is blocked

and the enzyme is inactive. In the presence of an interface, the lid moves, liberating the access to the active site (Figure I-7).

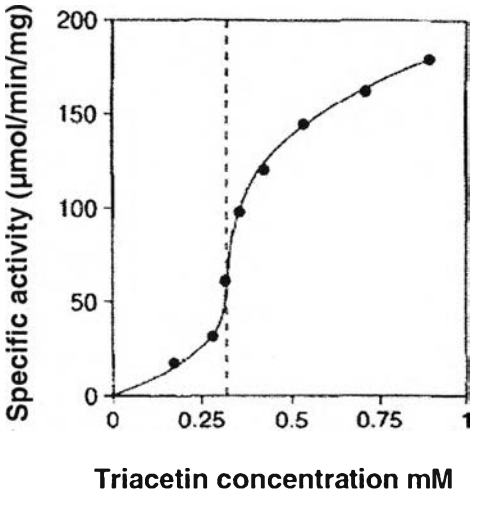


Figure I-6. Hydrolysis of triacetin catalyzed by the pancreatic lipase of pig according to water soluble substrate. The vertical line in dotted represents the limit of solubility of substrate used (reviewed in Verger 1997).

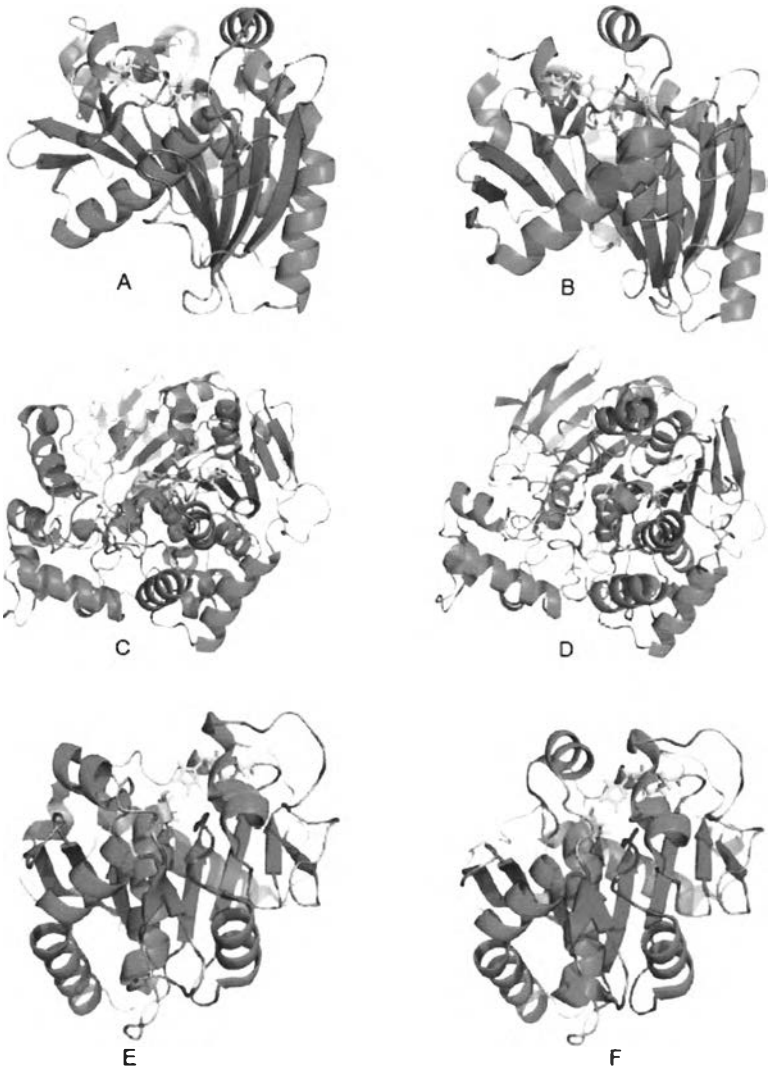


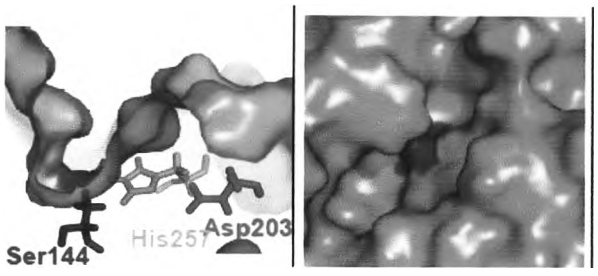
Figure I-7. Open conformations (A, C and E) and closed (B, D and F) of lipases from *Rhizomucor miehei, Candida rugosa* Lip1, and *Thermomyces lanuginosa*, respectively. For *R. miehei*, open structure (PDB: 4TGL) (A) and closed structure (PDB: 3TGL) (B). For *C. rugosa*, open structure (PDB: 1LPN) (C) and closed structure (PDB: 1TRH) (D). For *T. lanuginosa*, open structure (PDB: 1EIN) (E) and closed structure (PDB: 1DT3) (F). The catalytic residues are represented in yellow and the amphiphilic shutters in green (Derewenda and Derewenda 1991; Grochulski *et al.,* 1993; Grochulski *et al.,* 1994; Brzozowski *et al.,* 2000).

2.6. The Substrate binding site

The active site in lipases is located within the protein structure and substrate access through a binding site located in a pocket on the top of the central β-sheet. Although lipases share the same structural fold their substrate binding regions are considerably different in size, structure and physico-chemical features, in particular regarding the hydrophobicity of residues lining the pocket. The length, shape and hydrophobicity of the binding pocket have been related to chain length preference, obtaining good agreement with experimental results (Pleiss *et al.*, 1998).

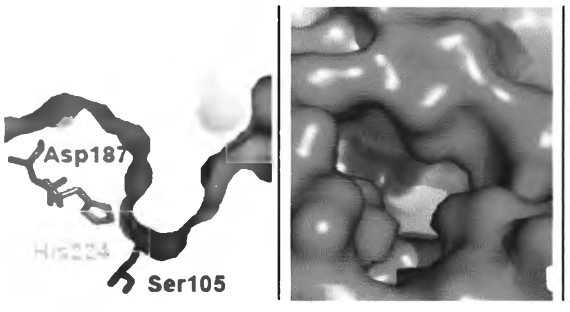
The structures of the lipases crystallized in presence and absence of analogues of lipids made it possible to identify the binding sites of the acyl (-R-C=O') and alkyl (-OR') group of the ester substrate. The binding site for the alkyl is similar in all lipases. It is a pocket containing two zones, i) a broad hydrophobic pocket upstream of the catalytic serine and ii) a small amino acid, usually a serine, downstream from the catalytic which would be responsible for the stereo-selectivity of lipases for secondary alcohols. On the other hand, the topology of the binding site of the acyl group varies considerably according to lipases. For example, in *Candida rugosa* lipases, this site corresponds to a long tunnel which can adapt to long chain of carbon, whereas in *Rhizomucor miehei* lipase, it is shorter and located on the surface of protein. *Burkholderia cepacia* lipase presents a rather much broader binding site than *Candida rugosa*. From these observations, three topologies of active sites were proposed by Pleiss and colleagues of which the side and top views are represented in Figure I-8.

a) Lipases with hydrophobic active site located close to the surface of protein: *Rhizomucor* and *Rhizopus*



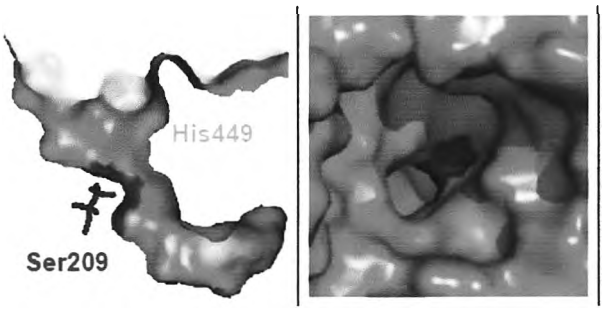
Rhizomucor miehei
(PDB : 5TGL)

b) Lipases with active site in funnel shape: *Burkholderia cepacia* and *Candida antarctica*



Candida antarctica
(PDB : 1TCA)

c) Lipases with active site in tunnel shape: *Candida rugosa*



Candida rugosa
(PDB : 1LPN)

Figure I-8. Three topologies of the active sites of lipases. The Surfaces of the active sites were modelled by using the crystallographic coordinates deposited in database PDB. The amino-acids of the catalytic triad are coloured (Catalytic Ser: blue, Asp or Glu: red and His: orange) (from Lafaquiere 2010).

3. Specificity of lipases

The specificity of lipases can be classified in three kinds. The type- selectivity is defined by the selectivity observed according to the length of the carbon chains, to the degree of non-saturation and substitution of the substrate. The type-selectivity is connected to the topology of the active site and will not be discussed further (see paragraph 2.6. for the description of the three topologies of active sites proposed by Pleiss and colleagues). The regio-selectivity, which can be of type *sn*-1, *sn*-2 or *sn*-3, corresponds to the ester bond preferentially hydrolysed on the triglycerides (Figure I-9). The results are fatty acid, di-acylglycerol and mono-acylglycerol. The Enantioselectivity translates the preference between two enantiomers of a racemic mixture.

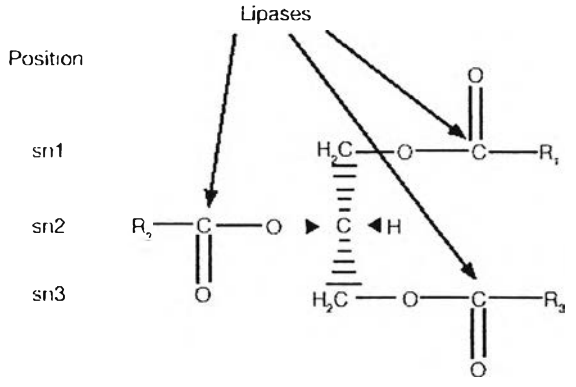


Figure I-9. Fischer representation of a triacylglycerol molecule. Identification of the ester bonds potentially hydrolysed by lipases (Egloff *et al.*, 1995).

3.1 Triglyceride regioselectivity of lipases

Most microbial lipases hydrolyze triacylglycerols at *sn*-1 or *sn*-3 position, and only few also at *sn*-2 position. However, the *sn*-1(3) stereopreference and stereoselectivity of microbial lipases toward triacylglycerols are varying. For example, the crystallization of the lipase from *Burkolderia cepacia* (Lang, *et al.*, 1998) reveals the presence of four pockets or cavities; one is the oxyanion hole, the three others accommodating the carbon chains of the fatty acids in position *sn*1, *sn*2 and *sn*3 (Figure I-10). The differences of size and hydrophobicity between the various

sites of stowing, determine the regioselectivity of lipase. The alcohol-binding pocket can be subdivided into a mixed hydrophilic/hydrophobic cleft for the *sn*-2 moiety of the substrate and a smaller hydrophobic groove for the *sn*-1 chain.

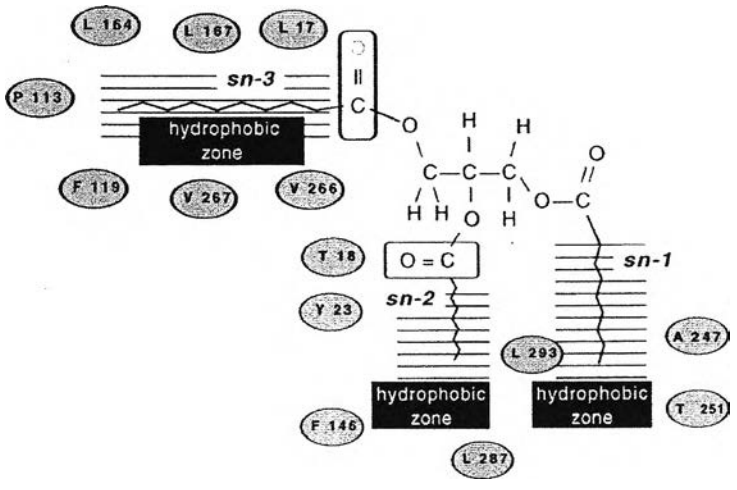


Figure I-10. Active site of *Burkholderia cepacia* lipase. The binding pockets for the *sn*-1, *sn*-2, and *sn*-3 moieties of the lipid substrate are indicated. The binding pocket residues lining are indicated (Jaeger *et al.*, 1999).

A study by Pleiss *et al.* (2000) highlighted a structural motif formed by the catalytic histidine and another amino acid which differs according to the different lipases, the so-called "histidine gap". The authors showed that the stereoselectivity of four lipases, *Rhizopus oryzae*, *Rhizomucor miehei*, *Candida rugosa*, and lipase B from *Candida Antarctica* (CALB) could be explained by the form of this motif and the flexibility of the substrate. The *sn*-2 substituent positions in the histidine gap and directly contact the second amino-acid of the motif. The stronger are steric interactions, the more the selectivity will be directed towards the *sn*-3 position. Thus, the stereoselectivity could be affected by the rigidity of the *sn*-2 substituent and by the shape of the histidine gap. A narrow histidine gap implies a *sn*-3 stereopreference for all types of substrates (Figure I-11a) whereas a histidine gap of intermediate size implies a *sn*-1 stereopreference for the flexible substrates and *sn*-3 for the rigid substrates (Figure I-11b); finally, a broad histidine gap implies a *sn*-1 stereopreference

(Figure I-11c) for all substrates. The form of the active site and in particular the positioning of some loops will also play a role in the stereopreference.

a) CAL B in complex with trioctanoin in *sn*-1 orientation. The functional ester group of the *sn*-2 substituent points toward the *His gap* (I189 and H224), the active site H224 on an α-helix between D223 and S230.

b) CRL in complex with trioctanoin in *sn*-1 orientation. The functional ester group of the *sn*-2 fatty acid points toward the *His gap* (F344 and H449). The non-hydrolyzed *sn*-3 fatty acid is located in the trench. The alcohol binding site is restricted by a long turn structure from G122 to P134 which does not contribute to stereoselectivity.

c) RML (light gray) in complex with trioctanoin in *sn*-1 orientation. The functional ester group of the *sn*-2 fatty acid points toward the *His gap* (H257 and L258). In dark gray, the G-elbow loop of ROL is highlighted.

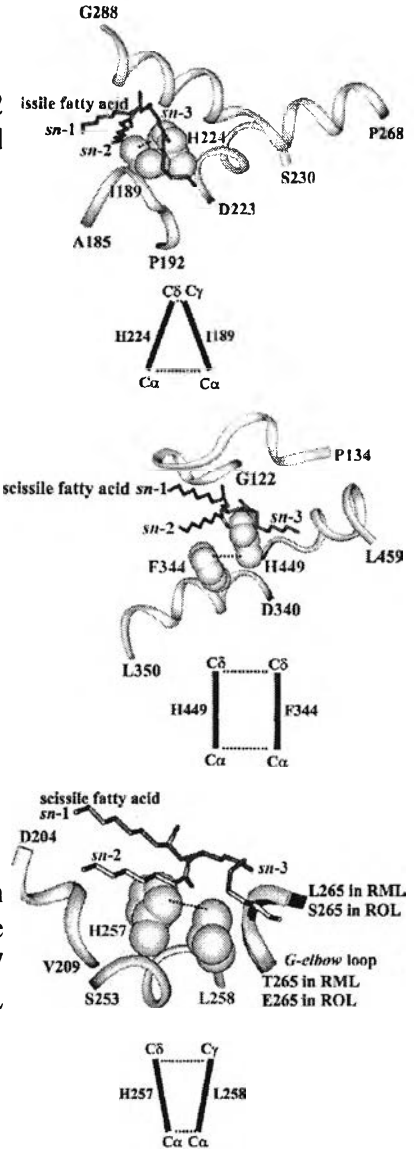


Figure I-11. Different histidine gaps presented by Pleiss *et al.* (2000).

3.2. Enantioselectivity

3.2.1. Chiral molecule

Chirality is the property of an object that is not super imposable with its image in a mirror. The 2 forms, R and S, which represent the object and its image, are called enantiomers (Figure I-12).

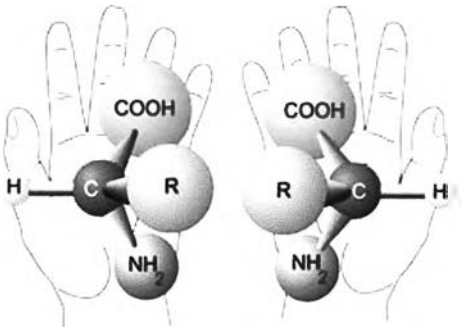


Figure I-12. Example of two chiral molecules.

Some molecules can comprise in their structure elements of chirality. Generally, it is asymmetrical carbon (carbon with four different substituents). The two enantiomers have very close properties, in particular their physical properties such as melting point, solubility in an achiral solvent, reactivity, etc.. On the other hand, they have different properties with respect to a dissymmetrical physical phenomenon like, for example, the deviation of the polarized light. An equimolar mixture of two enantiomers, inactive on polarized light, is called racemic. The molecules of pharmaceutical interest are often chiral ones. Moreover, in general only one enantiomer is required for therapeutic activity, while the other might be inactive or worse toxic. Consequently, the legislation gradually imposed a rigorous control of the use of racemic mixtures.

The American FDA (Food and Drug Administration) and The European EMEA (European Medecines Agency) do not impose the systematic use of pure enantiomer, but require a thorough biological study. Because of these new regulations, the production of pure enantiomers increased considerably. The sales of drugs made of pure enantiomeric were of 225 billion euros in 2005, thus accounting for 37% of the market of the drug company according to the provided figures by

Technology Catalyst International. In 2006, 75% of the pharmaceutical molecules of low molecular-weight approved by the FDA were pure enantiomers. It is a market in constant evolution (11% per year between 2000 and 2005) (Erb 2006). Within this framework, and because their enantioselectivity prevents from the use of expensive purification processes, lipases are used in the drug company and the industry of the pesticides to carry out the resolution of racemic mixtures. The enantioselectivity can be measured by the enantiomeric ratio E, which represents the ratio of initial speeds of conversion of the two enantiomers. For a pharmaceutical application, the ratio must be higher than 200 for an effective separation of two enantiomers, which means that one of the enantiomers reacts 200 times faster than the other.

3.2.2. Obtention of enantiomerically pure compounds

The enantiomerically pure molecules can be obtained either by chemical or biological ways. Three methods can be employed:

- Chemical synthesis starting from chiral precursors

Enantiomerically pure natural substances such as the amino-acids, terpenes, alkaloids, sugars and their derivatives can be used as precursors for the synthesis of derived enantiopure molecules.

- Asymmetrical chemical synthesis

The synthesis of enantiomerically pure starting from raw achiral material requires the use of a chiral auxiliary. The chiral auxiliary is either i) used in catalytic quantity, ii) associated with the substrate to create a stereogenic carbon in a stereoselective way , or iii) associated during the reaction with the prochiral carbon of the substrate. First the stereoselectivity of the reaction is not always easy to envisage. Moreover, the use of stoichiometric quantities of chiral auxiliary might be expensive. Finally, this method might require two additional stages of synthesis to bind and remove the chiral auxiliary. The asymmetrical synthesis can also be catalysed by enzymes or microorganisms.

- Resolution of racemic mixtures

The resolution of racemic mixtures is, by far, the method mostly used in the industrial sector for the preparation of enantiomerically pure compounds. The resolution gives

access to the two enantiomers, which represents an advantage for the drug company if the preclinical studies are to be conducted for the two forms. Techniques employed for the resolution of racemates are mainly divided into four groups.

• Preferential or selective crystallization: this process occurs only with some racemates called conglomerates. They have the characteristic to form crystals containing only one enantiomer making it easy to separate optical isomers. Approximately 10% of the whole of racemic have this property.

• Salt crystallization of diastereoisomers: this process includes a chemical reaction in the presence of an optically active acid (or base) called the solving agent, which leads to the formation of a mixture of diastereoisomers salts. Diastereoisomers can then be separated by simple crystallization. The best theoretical performance of this reaction is 50%.

• Separation by chromatography: this separation is based on the difference in interaction of the two enantiomers with a homochiral adsorbent. The high price of the stationary phases remains the major disadvantage of this technique.

• Kinetic resolution: this means of separation uses the difference in reaction rate of each enantiomer with an optically pure reagent. Thus, the chemical modification of only one out of the two enantiomers will make it possible to ensure the resolution of racemates by traditional methods of separation used in organic chemistry. Here again, the yield which cannot exceed 50% is the intrinsic limitation of this technique except in the case of a kinetic resolution concomitant to *in situ* racemization. The properties of the residual substrate and the product are sufficiently different so that a traditional separation by chromatography can then be carried out.

In parallel to the chemical ways, enzymatic technologies are appreciated for the separation of enantiomers because of the high specificity and selectivity of the reactions implemented by biocatalysts. Among the above mentioned techniques allowing the kinetic resolution of racemates, enzymes have multiple advantages compared to traditional catalysts of chemical industry. Indeed, they authorize less polluting processes, catalyse the reactions under mild conditions of pH and temperature, and have a great specificity compared to chemical catalysts. Hydrolases, which represent 75% of the enzymes used in industry are the enzymes more frequently employed in the processes of racemate resolution. Among these

hydrolases, lipases are the most widespread enzymes for the resolution of of racemic mixtures of alcohols, acids, and amines. However, in spite of the extended work relating to the enantioselective resolution catalyzed by these enzymes, the molecular factors responsible for the enantioselectivityare still badly understood and not easily predictable.

3.2.3. Determination of enantioselectivity

3.2.3.1. Enantiomeric excess (ee)

The kinetic resolution of a racemic mixture can be characterized by the determination of the enantiomeric excesses of the substrate (ee_s) (Equation 1) or of the product (ee_p) (Equation 2) defined in Figure I-13. However, the variation of these enantiomeric excesses is function of the conversion rate (Figure I-14). With low levels of conversion, only the best recognized enantiomer is transformed thus leading to a very high ee_p. Progressively, as the reaction progresses, the quantity of enantiomer best recognized decreases in the racemic mixture, with the profit of the other enantiomer. Finally, this latter might be transformed by the enzyme, involving a reduction in ee_p and an increase in ee_s.

$$\text{Enz} + R \underset{k_2}{\overset{k_1}{\rightleftharpoons}} (\text{Enz-}R) \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \text{Enz} + \text{P}$$

$$\text{Enz} + S \underset{k'_2}{\overset{k'_1}{\rightleftharpoons}} (\text{Enz-}S) \underset{k'_4}{\overset{k'_3}{\rightleftharpoons}} \text{Enz} + \text{Q}$$

$$\%\,ee_s = \frac{[R-S]_t}{[R+S]_t} \times 100 \quad (1) \qquad \%\,ee_p = \frac{[P-Q]_t}{[P+Q]_t} \times 100 \quad (2) \qquad \%\,c = 1 - \frac{[R+S]_t}{[R_0+S_0]_{t=0}} \times 100 \quad (3)$$

Figure I-13. Enzymatic resolution. [R] and [S] are concentrations represent of enantiomers substrates, [P] and [Q] concentrations of the produced enantiomers of the reaction. Enz = enzyme. (ee_s) enantiomeric excess of the substrate, (ee_p) enantiomeric excess of the product, (c) conversion rate.

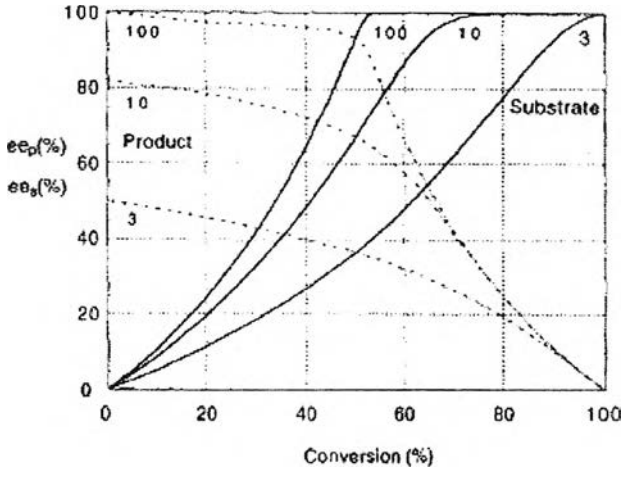


Figure I-14. Variation of enantiomeric excesses according to the conversion rate of reaction in the case of an irreversible reaction. Line 3, 10 and 100 represent the enantioselectivity of the enzymatic reaction (Anthonsen, 1997).

3.2.3.2. Enantiomeric ratio; E

In order to have a parameter making it possible to describe the degree of enantioselectivity of an enzyme, Chen and colleagues introduced the enantioselectivity (E), which is defined as a concept of "enantiomer ratio" catalytic effectiveness, k_{cat}/K_M, measured for each enantiomer (Chen C-S. 1982; Chen C-S. 1987). The E value thus defines the capacity of an enzyme to differentiate the enantiomers (R) and (S) under given reaction conditions. It is an intrinsic property of the enzyme which describes quantitatively its stereoselectivity. A non enantioselective enzyme has an E value = 1.

$$E = \frac{v_R}{v_S} = \frac{\left[\frac{k_{cat}}{K_m}\right]_R}{\left[\frac{k_{cat}}{K_m}\right]_S}$$

As indicated in Figure I-14, the profile of variation of enantiomeric excesses according to the conversion rate depends on the enantioselectivity of the reaction. For highly enantioselective enzymes, the reaction rate decreases brutally beyond 50% of conversion. The enantioselectivity can be calculated starting from the Equations (4)

and (5), given bellow, which utilize the conversion rate (3), and the enantiomeric excesses of the substrate (1) and the product (2). If the conversion rate is very weak or very high, Equation (6) must be utilised. These Equations apply only in the case of irreversible enzymatic reactions (as reactions in Figure I-13 where $k_4 = k'_4 = 0$) (Rakels *et al.*, 1993).

$$E = \frac{\ln\left[1-c(1+ee_p)\right]}{\ln\left[1-c(1-ee_p)\right]} \quad \textbf{(4)} \qquad \text{For product}$$

$$E = \frac{\ln\left[(1-c)(1-ee_s)\right]}{\ln\left[(1-c)(1+ee_s)\right]} \quad \textbf{(5)} \qquad \text{For substrate}$$

$$E = \frac{\ln\dfrac{\left[ee_p(1-ee_s)\right]}{(ee_p+ee_s)}}{\ln\dfrac{\left[ee_p(1+ee_s)\right]}{(ee_p+ee_s)}} \quad \textbf{(6)}$$

In some cases, the expression (7) which defines the enantioselectivity as the ratio initial rate of transformation of each enantiomer is preferred because it is much less complex, faster to estimate at the experimental level and is a good approximation of the enantioselectivity.

$$E = vi_R / vi_S \quad (7)$$

3.2.4. Thermodynamic analysis of enantioselectivity

Various parameters influence the enantioselectivity of an enzyme, and particulary the nature of the solvent, the water activity (a_w) or the temperature. Results presented in literature indicate that these factors can either increase, or reduce the enantioselectivity according to the considered cases (nature of the enzyme, substrate and type of reaction). Beyond the physico-chemical parameters influencing the enantioselectivity, the molecular recognition between the enzyme and the enantiomers was also identified like a key parameter responsible for the enantioselectivity of lipases. This can be apprehended through the study of energy parameters of the reaction, i.e. the free energy of activation, the enthalpy and the entropy. The relation between the difference in free energy of Gibbs of the states of transition for the enantiomers and the enantiomeric ratio (E) is indeed given by the theory of the state of transition (Equation 8).

$$\Delta_{R-S}\Delta G = -RT\ln E = \Delta_{R-S}\Delta H - T\Delta_{R-S}\Delta S \quad \textbf{(8)}$$

The enantioselectivity, corresponding to the difference in free energy of activation between the two enantiomers thus corresponds to the sum of two terms: an enthalpic term and an entropic term (Equation 9).

$$\ln E = -\frac{\Delta_{R-S}\Delta H}{R} x \frac{1}{T} + \frac{\Delta_{R-S}\Delta S}{R} \quad (9)$$

The difference in enthalpy of activation ($\Delta_{R-S}\Delta H$) translates in a certain manner the difference in complementary of each enantiomer in the state of transition, caused by steric and electrostatic interactions between the enzyme, its substrate and the components of the reaction (solvent). As the complementary between the enantiomer and the enzyme will be optimal the enthalpy of formation of the enzyme-enantiomer intermediate will be weak and the enantiomer will be preferred. The entropic term is dependent on several factors such as the freedom space of the substrate in the active site (hydrodynamic volume), the differences in degree of freedom of the protein residues, and the differences in solvatation between the two enantiomers (Figure I-15).

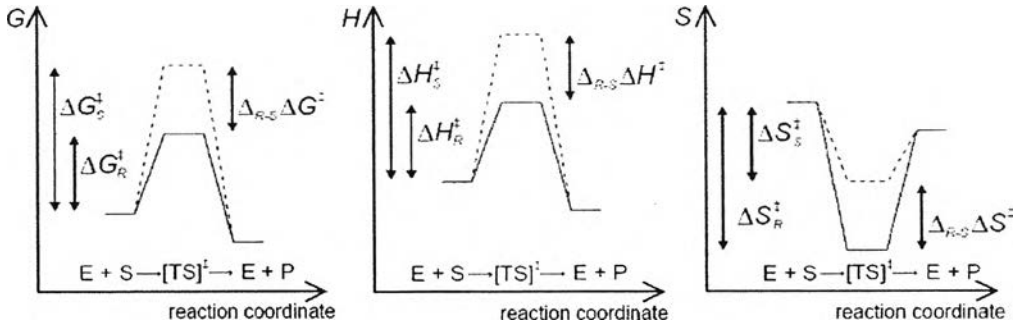


Figure I-15. Energy profiles for the enzymatic catalysis of a chiral substrate made of enantiomers (R) and (S), G: Gibbs' energy, H enthalpy and S: entropy), according to Ottosson *et al.*, 2001).

Considering the enantiomers are strongly constrained in the active site, the contribution of the entropy was often regarded as negligible. However, many examples of the literature highlighted the importance of the entropic term. The neglection of this term can be valid only if the complexes of the two states of transition have same rigidity and interact in the same way with the solvent. Ottosson and co-worker (2002) showed that in the resolution of the racemate of 3-hexanol by

lipase B of *C. antarctica* the entropic term represent 25 to 60% of the enthalpic component (Ottosson *et al.,* 2002). The enthalpic term ΔΔH can be easily approximate experimentally or by means of molecular modeling. However, the determination of the entropic term is much more delicate.

In the case of lipases, the system passes by various states of transition not easily detectable in experiments because of their very weak life spans (Figure I-16).

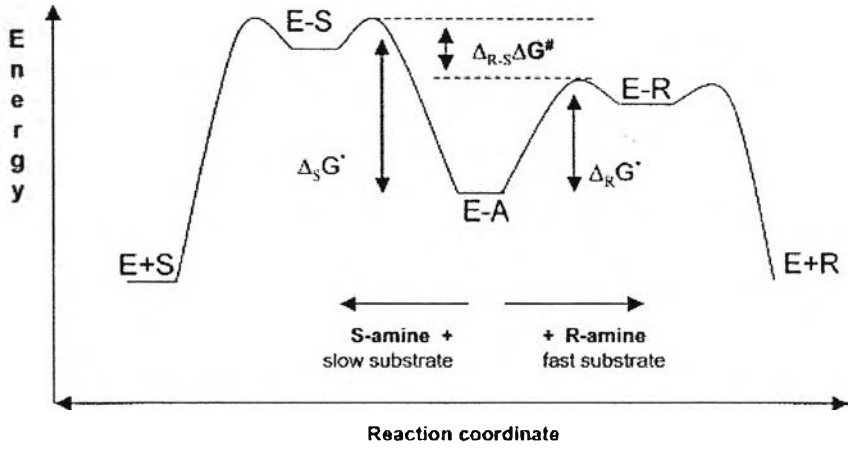


Figure I-16. Energy profile for the reaction of the lipase with both enantiomeric substrates proceeding via two diastereomeric TS discriminated by a strong $\Delta_{R-S}\Delta G^{\ddagger}$ difference. Although the enthalpic portion of $\Delta_{R-S}\Delta G^{\ddagger}$ dominates and strongly favors the fast-reacting substrate, the slow-reacting substrate is favored entropically (Bocola *et al.,* 2003).

3.2.5. Prediction of enantioselectivity, empirical rules

Before the accession of structural data on lipases, empirical rules were used to predict the enantioselectivity of lipases according to the geometry of the substrates.The first predictive model was proposed by Kazlauskas *et al.*: this empirical model resulted from the systematic sifting of many secondary alcohols and established that the enantiopreference was correlated with the difference in size between the two groups of the secondary alcohol and with their distinct positioning in the pockets of the lipases active site(Kazlauskas 1991). This rule is represented in figure I-17 and is applicable to reactions of hydrolysis (the substrate being an ester) and of trans-esterification (the substrate being alcohol). It supposes that the base of the enantioselectivity rests on the difference in size between the two substituents of a secondary alcohol. This rule implies the existence of pockets of different sizes around

the active site of the enzyme which will accommodate the various groups of the secondary alcohol (figure I-17).

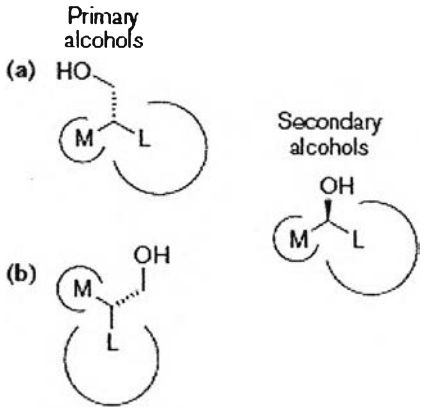


Figure I-17. Empirical rules to predict the fast-reacting enantiomer of secondary and primary alcohols with *Pseudomonas cepaciae* lipase PCL. The primary-alcohol rule is shown in two different orientations to suggest relative orientations of primary and secondary alcohols in the alcohol-binding site. (a) The large (L) substituent of both primary and secondary alcohols are in a similar place. (b) The large substituent is in different places in the primary and secondary alcohol models. Both suggestions assume that the hydrogen at the stereocenter, the medium (M) substituents, and the alcohol oxygen bind in similar places (Kazlauskas 2000).

3.2.6. Improvement of enantioselectivity

The evolution of biomolecular engineering during these last years made it possible to develop effective processes of biocatalysis by reorganizing enzymes, i.e. modifying and/or improving their natural capacities (Bornscheuer and Pohl 2001; Zhao *et al.*, 2002; Jaeger and Eggert 2004). Two strategies for such an approach can be considered: i) rational design, by directed mutagenesis if the structure of the enzyme is known and if the amino-acids implied in the selectivity are identified (Svendsen 2000), or ii) randomized techniques for directed evolution, which do not require structural knowledge of the enzyme (Reetz and Jaeger 2000) but a highly efficient screening method.

3.2.6.1. Rational and semi-rational engineering

Rational engineering consists in modifying the structure of an enzyme by changing amino-acids by means of directed mutagenesis. The targets of mutagenesis are selected according to the properties which one wishes to modify or improve. In the case of the modification or improvement of the enantioselectivity, the important positions are often close to the active site. These positions are defined starting from the structural analysis of protein, by studies of molecular modelling such as the stowing of the substrate in the active site or using chimiometric methods. For that, it is important to know: the three-dimensional structure of the enzyme, the reactional mechanism, the catalytic residues and the conformation of the active site. In this type of approach, only some amino-acids are targeted. The mutants are then expressed and the produced enzymes are characterized to determine improvement made from the mutation. Semi-rational engineering term is used to describe methods allowing the introduction of all the possible changes on one or more selected amino acids. The approach then consists in creating a library of variants obtained by saturation mutagenesis on selected positions and possibly in recombining the best mutations in a second phase. Various methods were proposed which make it possible to optimize the degeneration of library in order to limit the size of the libraries created. The important size of the generated libraries might require the use of high-throughput techniques of which several examples will be further presented.

3.2.6.2. Directed evolution

Directed evolution allows miming in the laboratory the natural evolution of proteins by random creation of genetic diversity followed by a selection process or screening of variants presenting the required improvement. The first stage thus consists in generating a library of mutants by specific methods (error prone PCR being the most widely used method nowadays). The likelihood of success in a directed evolution experiment is directly related to the total library size, as evaluating more mutants increases the chances of finding one with the desired properties. An iterative process, i.e. performing multiple rounds of evolution, is useful not only because a new library of mutants is created in each round, but because each new library uses better mutants as templates. The advantage of the directed evolution

approach is that the researcher needs not to understand the mechanism of the desired activity in order to improve it. Most directed evolution projects seek to evolve properties that are useful to humans in an agricultural, medical or industrial context (Otten and Quax 2005).

3.2.6.3. Selection and screening of improved variants

a) Selection

In directed evolution experiments, success often depends on the efficiency of screening or selection methods. Genetic selections have proven to be extremely valuable for evolving enzymes with improved catalytic activity, improved stability, or with altered substrate specificity. In contrast, enantioselectivity is a difficult parameter to select for. Boersma *et al.*, (2008) present a successful strategy that selects for enantioselectivity. Authors demonstrated the selection of *Bacillus subtilis* lipase A variants with inverted and improved enantioselectivity. They created a library of mutated lipases in an aspartate auxotroph *Escherichia coli* and the library was plated on minimal medium that was supplemented with the aspartate ester of the desired enantiomer (*S*)-(+)-1,2-O-isopropylidene-*sn*-glycerol. To inhibit growth of less enantioselective variants, a covalently binding phosphonate ester of the opposite enantiomer, the (*R*)-(-)-1,2-O-isopropylidene-*sn*-glycerol, was added as well. After three selection rounds in which the selection pressure was increased by raising the phosphonate ester concentration, a mutant was selected with an improved enantioselectivity increased from *ee* -29.6 % to +73.1 %. Interestingly, the author presenting an inversion of the enantioselectivity compared to the wild enzyme which prefers *R* enantiomer.

b) Screening

Whereas the principles of combinatorial chemistry are well established in pharmaceutical research, extension to the area of catalysis is not as advanced. One reason is that few general methods for high-throughput screening of heterogeneous and homogeneous catalysts have been devised. This applies all the more to enantioselective catalysts, as the methods usually used for the analysis of the

enantioselectivity are high performance liquid chromatography (HPLC) or gas chromatography (GC) using chiral stationary phases, which often require analysis extending from a few minutes to several minutes, making it impossible the screening of large libraries. Reetz (1999) has developed a screening system for the catalytic enantioselective hydrolysis of chiral *p*-nitrophenol esters in which the course of the reactions of the *R*- and *S*-configured substrates is monitored in a parallel manner by UV/Vis spectroscopy. With the use of such a technique, they were able to screen about 800 different enantioselective catalysts per day.

Jaeger and Reetz (2000) described another method based on electrospray ionization mass spectrometry (ESI-MS) which enables the determination of enantioselectivity in about 1000 biocatalysed asymmetric reactions per day. Two basically different stereochemical processes can be monitored by this approach, namely kinetic resolution of racemates and asymmetric transformation of substrates.

Conclusions

Despite their broad diffusion in biotransformation reactions the use of lipases (and of most enzymes) in industrial processes is still limited by intrinsic weaknesses of the biological catalyst, in particular low stability under operational conditions and low activity or specificity on particular or non-natural substrates. Knowledge of the molecular determinants of enzyme properties has accumulated allowing the rational choice of the catalyst for a given process. On the other hand, the cloning of genes encoding unknown enzymes from non-conventional sources and the modification of those already available by a combination of molecular techniques are very promising as potential sources of novel catalysts with improved or completely new properties.

PART II: Lipases from *Candida rugosa*

1. Genetic and biochemical data

C. rugosa (formerly, *Candida cylindracea*) is a non-sporogenic, pseudofilamentous, unicellular, and non-pathogenic yeast. It is one of the most extensively studied microorganism with respect to its lipase secretion. Indeed, *C. rugosa* secrets a mixture of enzymes, amongst which at least seven isoenzymes, encoded by a 'lipase minigene family'. Five of the encoding genes, namely Lip1 to Lip5, were sequenced and intensively studied, whereas only a partial sequence was obtained for Lip6 and Lip7 (Brocca *et al.*, 1995). Recently, a new gene was sequenced from *Candida rugosa* ATCC14830, namely LipJ08 (Xu *et al.*, 2009). A particulariry of these genes is that, as all *Candida* species, some of the universal codons for leucine, CTG, are translated into a serine. Nearly 20 serines encoding codons on Lip1 to Lip5 and LipJ08, including the catalytic Ser, are CTG. Their location on the translated proteins is given in Figure I-18.

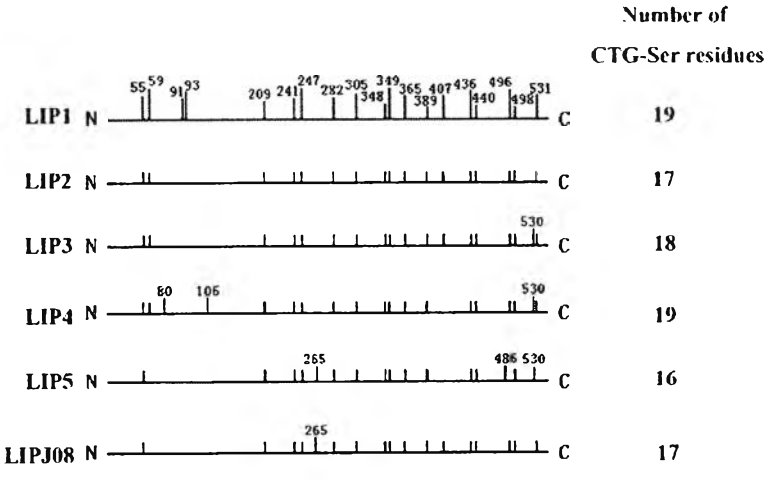


Figure I-18. Distribution of codon triplets, CTG, encoding the serine residues including the active site Ser 209, in five *C. rugosa* lipase isoforms. (Akoh *et al.*, 2004; Xu *et al.*, 2009)

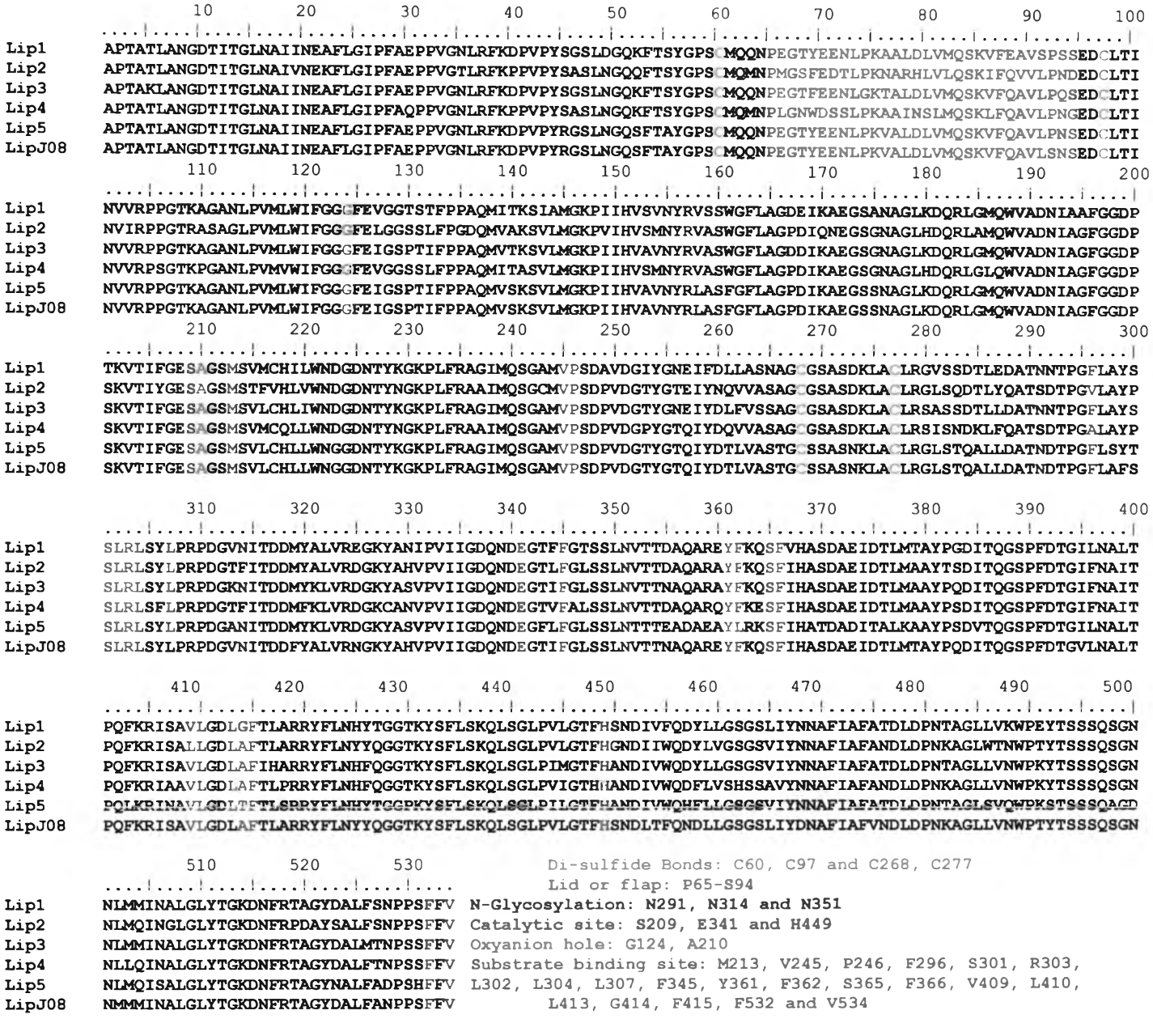


Figure I-19. Sequence alignment of LIP1, LIP2, LIP3, LIP4, LIP5 and LIPJ08 (swissprot respective accession number P20261, P32946, P32947, P3294, P32949 and FJ74369) showing di-sulphide bonds: C60, C97 and C268, C277, in light blue; lid: P65-S94, in pink; N-glycosylation site: N291, N314 and N351, in blue; Catalytic site: S209, E341 and H449, in red; oxyanion hole: G124, A210, in orange; and substrate binding site: M213, V245, P246, F296, S301, L302, R303, L304, L307, F345, Y361, F362, S365, F366, V409, L410, L413, G414, F415, F532 and V534, in green. Informations from Pleiss *et al.,* 1998; Akoh *et al.,* 2004 and Xu *et al.*, 2009; alignment using BioEdit.

Table I-6. Pair wise of percent identity of lipases from *C. rugosa* (Lotti *et al.*, 1994; Ferrer *et al.*, 2001).

	LIP1	LIP2	LIP3	LIP4	LIP5	LIPJ08
LIP1	100	—	—	—	—	—
LIP2	79	100	—	—	—	—
LIP3	88	82	100	—	—	—
LIP4	80	83	83	100	—	—
LIP5	81	77	86	78	100	—
LIPJ08	86	80	88	80	88	100

C. rugosa lipases (CRLs) belong to a large protein family that contains mostly esterases but also a few proteins devoid of esterase activity. Apart from CRLs, the only other lipases presently identified within this family are those from *Geotrichum candidum* and *Galactomyces geotrichum*. Other members of this family include acetyl and butyrylesterases, carboxylesterases, and cholesterol esterases. CRLs are closely related to *G. candidum* lipases (GCL) with which they share approximately 40% amino acid sequence identity. Their relationship to acetylcholinesterases is more distant, e.g. they show approximately 25% amino acid identity to *Torpedo californica* acetylcholinesterase. In the classification developed by Pleiss *et al.*, 1998, CRL belong to the type GGGX with X being a phenylalanine.

The five isoenzymes present the same disulfide bonds profile, involving amino acids C60, C97 and C268, C277, but they differ in N-glycosylation sites, isoelectric points, and hydrophobicity profiles (Table I-7). It should be noted that variations in glycosylation influence the biochemistry of the *CRLs* isoenzymes (Sanchez *et al.*, 1999; Ferrer *et al.*, 2001; Pernas *et al.*, 2001).

Glycosylation at position Asn351 is conserved in all isoenzymes (Longhi *et al.*, 1992; Lotti *et al.*, 1993; Benjamin and Pandey 1998; Ferrer *et al.*, 2001; Akoh *et al.*, 2004). It is also present in the case of *G. candidum* lipases, as Asn364 (Bertolini *et al.*, 1994). Brocca *et al.* (2000) reported the behaviour of different Lip1 mutants in which the glycosylation points were specifically modified. While mutations in the Asn-291 did not lead to special changes in activity, aw dependence, or resistance to

temperature, variations in the Asn314 and in the Asn351 turned out crucial, and a decrease of the activity was observed upon mutations of these residues. Notably, the Asn-351 is the most sensitive point, since it directly interacts with the lid movement. Finally, Tang *et al.* (2001) reported the catalytic and stability parameters of a mutated Lip4 with/without glycosylation points, focusing on the conserved Asn351. It was stated that the glycosylated enzyme was more thermostable (58 vs. 52 °C), and that some biocatalytical aspects varied from one mutant to another, e.g., the specificity towards long chain lipids.

Table I-7. Proteins predicted on the basis of nucleotide sequences as compared with proteins isolated from commercial preparations (Benjamin and Pandey, 1998).

Gene	Amino acid mature protein	MW (kDa)	pI	N-glycosylation sites	
LIP1	534	57.223	4.5	3	291, 314, 351
LIP2	534	57.744	4.9	1	351
LIP3	534	57.291	5.1	3	291, 314, 351
LIP4	534	57.051	5.7	1	351
LIP5	534	56.957	5.5	3	291, 314, 351

2. Structural data

3D structures are available for three CRL out of 5 isosymes as showed in table I-8.

Table I-8. Structural data of CRL existing in protein database.

Lipase sources	References	Resolution	Isoenzyme	PDB code	Substrate/inhibitor
Candida rugosa	Grochulski *et al.*, 1993	2.06 Å	Lip 1	1CRL	Open form
		2.10 Å	Lip 1	1TRH	Closed form
	Cygler *et al.*, 1994	2.18 Å	Lip 1	1LPM	Complex with (1*R*)-menthyl hexyl phosphonate
		2.18 Å	Lip 1	1LPS	Complex with (1*S*)-menthyl hexyl phosphonate
		2.18 Å	Lip 1	1LPN	
	Grochulski *et al.*, 1994	2.18 Å	Lip 1	1LPO	complexes with 1-hexadecanosulfonic acid
		2.18 Å	Lip 1	1LPP	
	Mancheno *et al.*, 2003	1.97 Å	Lip 2	1GZ7	
	Ghosh *et al.*, 1995	2.00 Å	Lip 3	1CLE	
	Pletnev *et al.*, 2003	1.40 Å	Lip 3	1LLF	

Additionally, some structural data for Lip4 (Tang *et al.*, 2001), as well as theoretical predictions for Lip5 (Mancheno *et al.*, 2003), were published.

All isoenzymes present the same catalytic triad (Ser209-Glu341-His449), which is stabilised by a hydrogen-bonding network, slightly different in the crystallised isoenzymes. It is noteworthy that the acid member of the triad in CRL is a glutamate instead of the usual aspartate. In these lipases, the residues of the catalytic triad are located at the top of the β-sheet near its centre (Figure I-20). The studies of complexes of CRL with substrate-like inhibitors indicated that, as in serine proteases, the oxyanion hole is formed by the main chain NH group of residues Gly 124 and Ala 210. It is also likely that a third NH group, that from Gly 123, also participates in tetrahedral intermediate stabilisation through hydrogen bond formation. These two glycines are located within a short loop near the nucleophile elbow. Another residue that is likely to be important for the catalytic events is Glu 208. Indeed, this side chain

is well conserved within the homologous family and provides a negative character to the immediate vicinity of the nucleophile.

The structures of the two crystallised lipases LIP1 and LIP3, sharing 88% amino acid sequence identity are very similar and the root-mean-square deviation for all 534 C^{α} atoms is 0.3 Å. However, they differ in their substrate specificity (see below paragraph on CRLs specificities).

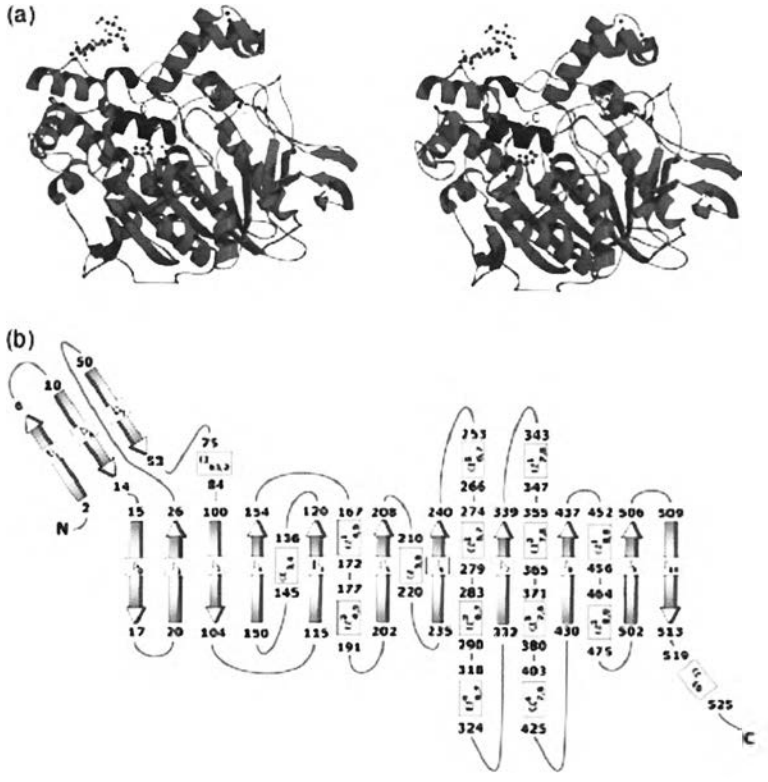


Figure I-20. Overall structure of Lip2 from *Candida rugosa* (PDB Code 1GZ7) (a) Ribbon representation with α-helices, β-strands and coils colored in red, green and gray, respectively. The helical and coil segments forming the flap region are shown in dark blue and orange, respectively. The catalytic triad residues (Ser209, Glu341 and His449), the disulfide bridges and the Asn-attached N-acetylglucosamine moieties are shown in ball-and-stick representation. The secondary structures were calculated with Dictionary of Secondary Structure of Proteins (DSSP). (b) A representation of the lipase topology with the secondary structure elements identified (b or β-strands; α-helix). The nomenclature follows that introduced by Cygler and team (Cygler *et al.*, 1993). Data from (Mancheno *et al.*, 2003).

The scaffold of the three-dimensional structure of CRLs is formed by an extended, 11 stranded mixed β-sheet. This sheet shows a strong twist; the first and last strands are nearly perpendicular to each other (Figure I-20 and I-21). The centre of the sheet contains parallel strands. The polypeptide chain extensions on the N-terminal side of the β-strands are rather short while those on the C-terminal sides are much longer. There are three α-helices lining the convex side of the sheet and two helices on the concave side. The helices are for the most part aligned along the direction of the strands. In the middle of one of the helices on the convex side of the sheet are a bend and a widening of one turn of the helix. A similar feature has also been observed in the acetylcholinesterase and lipase from *Geotrichum* structures and was attributed to the presence of a proline residue in the middle of the helix (Cygler and Schrag 1999). The distortion of the helix persists in CRL despite the fact that the proline residue is not conserved. In addition to this scaffold, there are eight helices and loops forming a cap located on the top (C-terminal side) of the β- sheet.

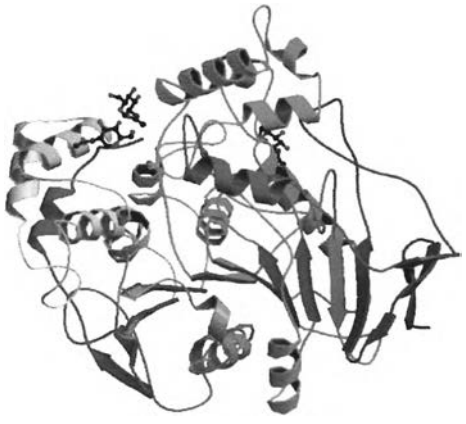


Figure I-21. Ribbon diagram of Lip1 from Candida rugosa (PDB code 1CRL), with open and closed states of the lid superimposed. The central mixed β-sheet is light blue and the smaller N-terminal β-sheet is dark blue. Helices that pack against the central β-sheet are dark green. The closed conformation of the lid is yellow and the open conformation is red. The residues forming the catalytic triad are shown in red (Grochulski *et al.,* 1993).

The binding pocket of CRL is exceptional and completely differs from those of most other lipases. It presents an internal cavity, found also in lipases from *Geotricum candidum* (Dominguez de Maria *et al.*, 2006). This tunnel starts near the nucleophile Ser 209, below the mobile loop, and extends inward, over the β-sheet and under the helical cap (Figures I-22, side view). The tunnel ends between Tyr 361 and Ser 365 having a total length of nearly 25 Å with a diameter of about 4 Å. It is not straight but rather has an L shape and is lined along its length mainly by hydrophobic side chains. In the open state of CRL the tunnel remains intact, except that one of its ends is now open to the solvent. The mouth of the tunnel is near the active site and close to the conserved Glu 208. This residue confers a negative charge on the vicinity of the active site.

The substrate binding site of CRL has been mapped by co-crystallization of these lipases with several substrate-like inhibitors containing a single fatty acyl chain. The complexes of LIP1 with phosphonate and sulfonate analogs of monoglycerides revealed the location of the scissile fatty acid chain. The substrate binding site is formed by G124 (oxyanion hole), F125, the catalytic S209, A210 (oxyanion hole), M213, V245, P246, F296, S301, L302, R303, L304, L307, F345, Y361, F362, S365, F366, V409, L410, L413, G414, F415, F532 and V534 (Pleiss *et al.*, 1998; Akoh *et al.*, 2004). The length of the hexadecyl moiety of the inhibitor, which has been crystallized with LIP1 CRL, corresponds to a fatty acid of chain length C17. The active site serine lies just behind the entrance to the tunnel (Figure I-22B), which is located near C3 and closely contacts the fatty acid chain up to its v-end. The most hydrophobic parts of the tunnel are located from C4 to C7 and C12 to C14. Without any adjustments this site can accommodate a C18 acyl chain and its L shape is particularly well suited for the oleoyl acyl chain. The part of the inhibitor that corresponds to the leaving group is located at the bottom of the depression created in the open state by the flipping of the lid. In contrast to other lipases, the alcohol binding site is not limited by a wall. The lid is located at the lower right-hand side (Figure I-22E, top view) and not at the right-hand wall (front view) like in the other lipases. The substrate accesses the binding pocket from the right-hand side (Figure I-22E, side view).

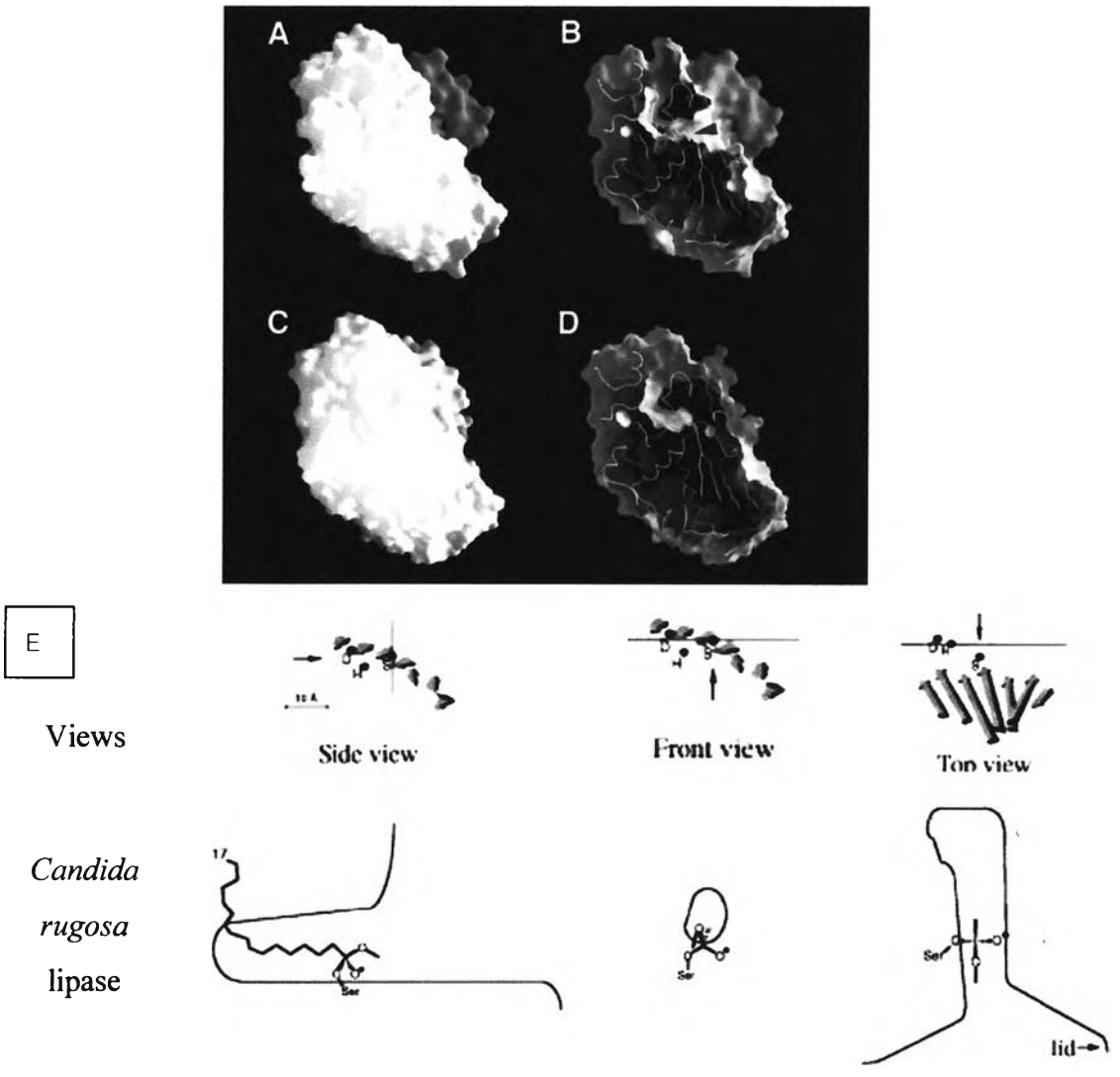


Figure I-22. Location of the tunnel in the Lip2 from *Candida rugosa*. (A) The surface representation of the open form with the residues of the lid colored magenta; (B) a cutaway view that shows the location of the tunnel with the entrance indicated by the arrow. Also shown are the backbone trace of the protein and the inhibitor hexadecane sulfonate (yellow). Several strands of the central β-sheet are seen below the tunnel; (C, D) corresponding views of the closed conformation. The internal cavity corresponding to the tunnel is apparent (Cygler and Schrag 1999). (E) Orientation of the cross-sections which are planes perpendicular to the paper plane and indicated by a straight line; Shape of the binding site of CRL. The direction of the view is indicated by an arrow (Pleiss *et al.*, 1998).

3. Specificities and applications of CRL

CRL have been widely used for the hydrolysis and synthesis of a wide range of esters of commercial interest as well as for the resolution of racemic mixtures (Dominguez de Maria *et al.*, 2006). Indeed, the most interesting biocatalytic applications of CRL, in the pharmaceutical and health food industries, are the preparation of chiral building blocks by kinetic resolution of racemic mixtures and the purification of valuable polyunsaturated fatty acids (PUFA) such as *cis*-5,8,11,14,17-eicosapentaenoic acid (EPA) and *cis*-4,7,10,13,16,19-docosahexaenoic acid (DHA) from concentrates containing both fatty acids through discriminative hydrolysis (Shahidi and Wanasundara 1998; Secundo *et al.*, 2004). These applications can be explained by either typo- region- or enantioselectivity of the isoforms.

3.1. Type and Regioselectivity

CRL are classified as nonpositional specific lipases, catalyzing hydrolysis or esterification at the *sn*-1, -2, and -3 positions of the glycerol molecule. They accept different saturated as well as unsaturated esters with varying chain-length of acid and alcohol moieties and other nonglycerol moieties such as phospholipids, sugars, and polyhydric alcohols (polyols).

Many studies were performed in order to assay CRL specificities. Nevertheless, the difference in sample preparation (purified from native strain or recombinant production) and reaction conditions make it difficult to compare accurately the 5 isoforms of CRL. Such a study was carried out by (Lopez *et al.*, 2004), where they compare Lip1, Lip2 and Lip3 obtained by purification of either commercial or home produced powders. According to the authors, Lip1 shows the strongest lipase behaviour (affinity for long-chain esters) while Lip2 and Lip3 perform better the hydrolysis of short-chain water-soluble esters (esterase behaviour). Interestingly, a high activity of Lip2 and Lip3 towards cholesteryl esters is also described. As a matter of fact, Lip3 is available commercially under the name cholesterol esterase (Sigma). This is in accordance with the results obtained by (Kontkanen *et al.*, 2004), whose team describe the hydrolysis of steryl esters by a commercial mixture of CRL. Such reactivity towards cholesterol esters was further patented in the quantification of cholesterol HDL (high density lipoprotein) for

diagnostic purposes (Dimagno and Arter 2005). The different substrate specificity displayed by lip1 and lip3 has been structurally analysed and several amino acid residues located mainly in the mouth of the hydrophobic tunnel of these lipases have been proposed as responsible for substrate recognition (Mancheno *et al.*, 2003).

These selectivities are of great significance for the applications of CRL in the purification of particularly valuable amino acids from waste oils and their incorporation in TAG. Indeed, the production of structured triacylglycerols (STAG) for clinical and nutritional purposes is a subject of interest. Most attention has been focused on triacylglycerols (TAG) with medium-chain fatty acids (M) located in positions 1 and 3 of the glycerol backbone and a functional health related long-chain polyunsaturated fatty acid (DHA and EPA). A prerequisite for such reaction is the obtention of high quantities of EPA and DHA. These can be obtained from fish oil wastes by enzymatic ways, and CRL were shown to be of great interest for such reactions. Such studies were performed by various teams. For instance, a research from Jonzo *et al.*, (2000) reported the use of purified isoforms of commercial powder of CRL, namely LipA (corresponding to Lip1 isoform) and LipB (corresponding to a mixture of Lip2 and Lip3 isoforms) for the enrichment of DHA via a selective esterification of free fatty acid from sardine oil with cholesterol. Lip A led to a 4- fold enrichment of DHA in 10 h., with a DHA recovering yield of 95 %, while selective esterification with Lip B gave a DHA recovery of 93.8 % after 24 h with 3.4- fold enrichment. Results are shown in figure I-23.

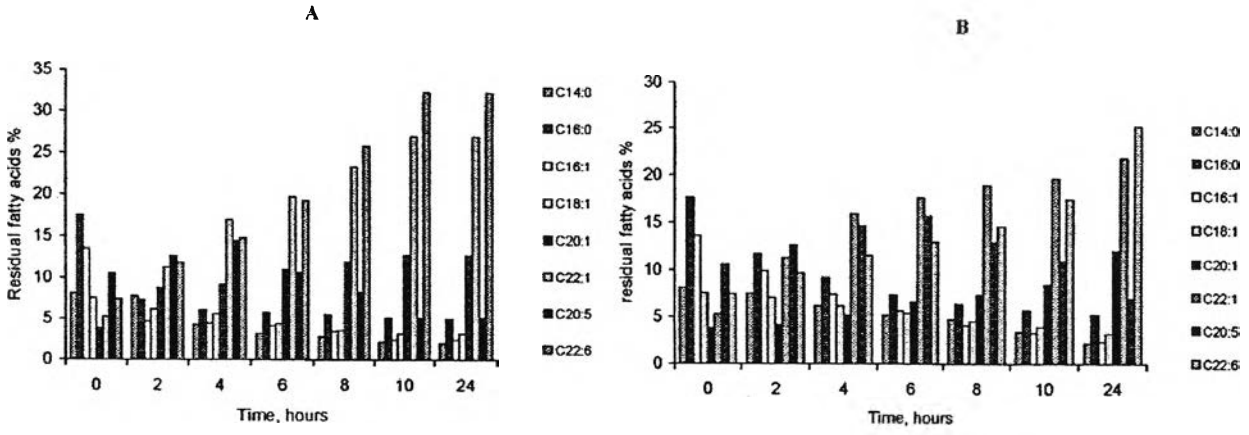


Figure I-23. Conversion of fatty acid from sardine oil by cholesterol with immobilized Lip A (A) or Lip B (B) of CRL at 40 °C in cyclohexane for 24h, C22:6 is DHA (Jonzo *et al.,* 2000).

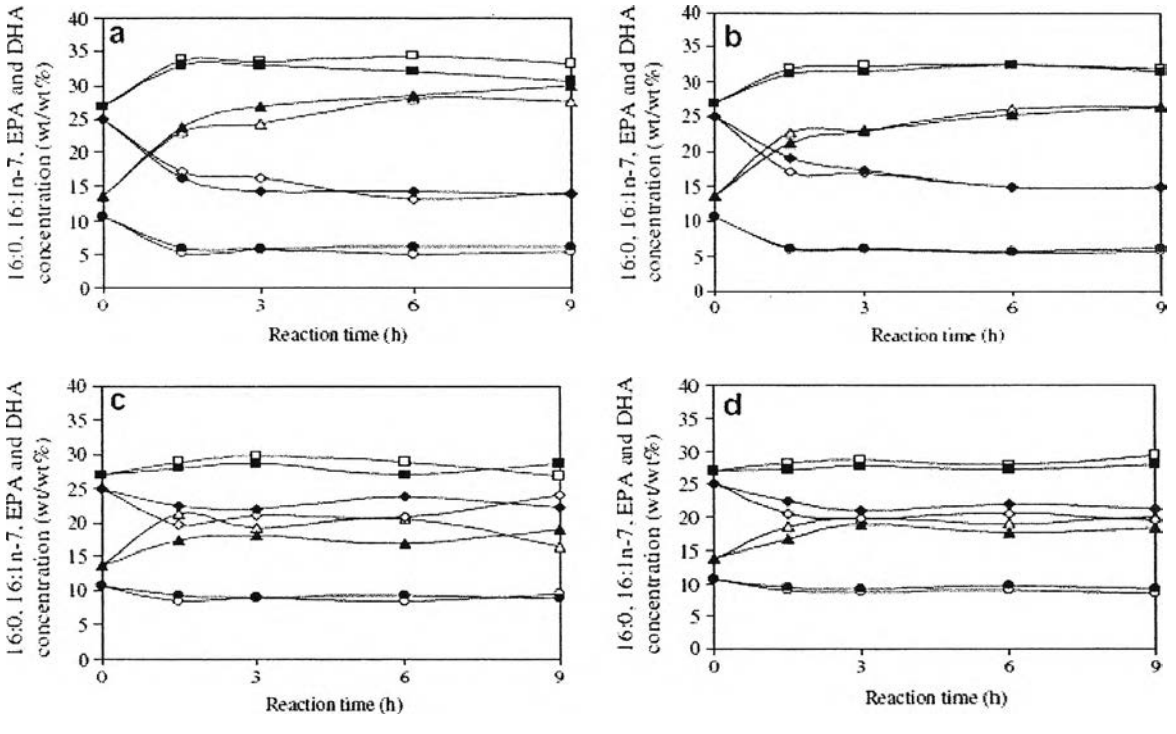


Figure I-24. Changes in 16:0, 16:1n- 7, EPA, and DHA concentration (wt/wt%) in final n- 3 PUFA concentrate with lipases from (a) CR, (b) CC, (c) MJ and (d) AN during hydrolysis at 37 ° C. ◇, 16:0 with 250 U; ◆, 16:0 with 500 U; ○, 16:1n - 7 with 250 U; ●, 16:1n- 7 with 500 U; □, EPA with 250 U; ■, EPA with 500 U; Δ, DHA with 250 U; ▲, DHA with 500 U; U = U/mL of enzyme activity (Okada and Morrissey 2007).

Okada and Morrissey (2007) studied the production of n-3 polyunsaturated fatty acid concentrate from sadine oil catalyzed commercial lipases: *C. rugosa* (CR, Sigma), *C. cylindracea* (CC, Fluka), *Mucor javanicus* (MJ, Aldrich) and *Aspergillus niger* (AN, Aldrich). Their results showed that fatty acids profile of DHA and EPA catalyzed by *C. rugosa* and *C. cylindracea* significantly increased after 1.5 h higher than compared to *M. javanicus* and *A. niger* (Figure I-24).

3.2. Enantiopreference of *C. rugosa* lipases

CRL have been widely used for kinetic resolution of racemic compounds through various reactions. As described in paragraph I-3.2.5, empirical rules based on the size of substituents at the stereocenter were used to predict enantioselectivity of lipases. Recent progresses enabled to further understand the enantiopreference of lipases by means of molecular modelling (modelling of the tetrahedral intermediate or docking of the substrate in the active site) or by X-Ray of substrate-enzyme complexes.

Astonishingly, Berglund *et al.*, (1998) even reported the reversion of the enantiopreference of CRL based on molecular modelling. They identified two different productive modes of binding of enantiomers of a 2-methyldecanoic acid ester in the active site of CRL (Figure I-25a) and designed a substrate analog of 2-methyldecanoic acid ester preventing the *S*-fast-reacting enantiomer to bind its preferential mode (Figure I-25b).

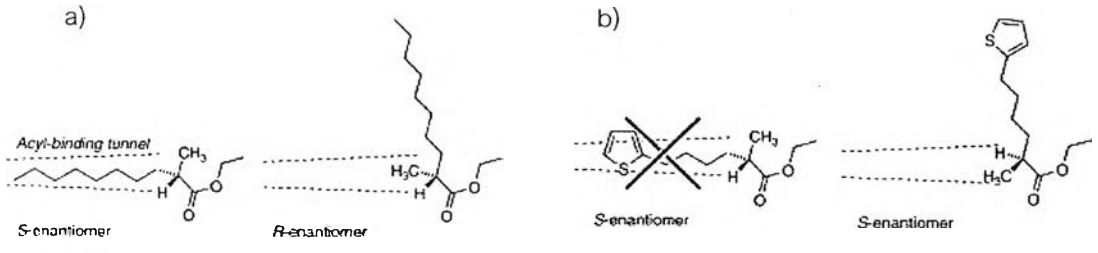


Figure I-25. Schematic view of the active site of CRL illustrating : a) The two productive modes of binding of enantiomers of a 2-methyldecanoic acid ester in the active site of CRL. b) Imposed binding mode of the designed *S*-enantiomer of 2-methylcarboxylic acid ester substrate in CRL. Molecular modelling suggests *R*-preference in this situation. Data from (Berglund *et al.*, 1998).

The 3-D structure of CRL also helped to understand the enantiopreference of lipases in many chiral reactions catalyzed by these enzymes. For instance, the structures of Lip1 in complex with a phosphonate inhibitor indicated that the orientation of the imidazole ring of the catalytic histidine would be distorted upon binding of the slow-reacting enantiomer (Figure I-26). This was found to be the most likely cause of the difference in reactivity for the two enantiomers.

ต้นฉบับ หน้าขาดหาย

presence in the mixture of several isoenzymes in different proportions. This factor greatly limits the application of crude CRL (mixed isoforms) in the production of specific compounds with high reproducibility. Due to the high sequence homology, it is difficult to purify each isoenzyme directly from cultures of *C. rugosa* on a preparative scale for industrial applications.

However, several procedures using chromatographic techniques were proposed either starting from commercial powder or personal fermentations (Pernas *et al.*, 2000 and 2002, Tenkanen *et al.*, 2002, Mancheno *et al.*, 2003, and Lopez *et al.*, 2004). In order to facilitate such strategies, the profile of isoenzymes proportion was also studied upon induction/repression by various reagents in the culture medium. Table I-9 is a summary of the composition in isoforms of CRL of various commercial powders and the conditions used to promote the production of these different isoforms.

Table I-9. Proportion of isoenzymes in different preparations of CRL; commercial (comparision of different suppliers) and lab scale production, effect of the carbon source (inducer) (adapted from Dominguez de Maria *et al.*, 2006).

Inducer/origin sample	Isoenzymatic profile (%)			References
	Lip 1	Lip 2	Lip 3	
Sigma (commercial)	73	8	19	Dominguez de Maria *et al.*, 2005
Sigma (commercial)	89	0	11	Lopez *et al.*, 2004
Roche (commercial)	66	13	21	Dominguez de Maria *et al.*, 2005
Amano (commercial)	74	0	26	Lopez *et al.*, 2004
Oleic acid (batch)	0	50	50	Sanchez *et al.*, 1999 Ferrer *et al.*, 2001
Oleic acid (batch)	0	60	40	Dominguez de Maria *et al.*, 2005
Olive oil (batch)	0	73	27	Dominguez de Maria *et al.*, 2005
Sunflower oil (batch)	0	73	27	Dominguez de Maria *et al.*, 2005
1-Dodeccanol (batch)	22	34	44	Dominguez de Maria *et al.*, 2005
Glycerol (batch)	<5	<5	<5	Dominguez de Maria *et al.*, 2005
Oleic acid (fed-batch low feed rate)	0	50	50	Sanchez *et al.*, 1999 Lopez *et al.*, 2004
Oleic acid (fed-batch high feed rate)	15	38	47	Lopez *et al.*, 2004

A way to counteract this problem was solved by recombinant protein expression. However, A great difficulty to realise heterologous expression of CRL is that, as mentioned in paragraph II.1., *Candida* species are unusual in the utilization of the genetic code. The production of CRL has been hampered for many years by these

difficulties in expressing active CRL. With the progresses in molecular biology of the past decade, different techniques were developed which enable the rapid mutation of multiple codons at the same time (overlap extension PCR strategy by Chang *et al.* 2006 on CRL for instance, Single-step overlap-primer-walk PCR by Kumar and Rajagopal (2008), for a review see Czar *et al.* 2009). This led to successful expression of active CRL and thus provided the required means for extensive mutagenesis studies. Table I-10 summarises the studied carried out on recombinant CRL expression and the production/activity obtained.

Table I-10. Expression system for *C. rugosa* lipases in previous reports.

CRL	Expression host	Enzyme produced (mg/L culture)	Lipolytic activity (U/mL enzyme)[a]	Activity assay	Reference
Lip 1	*S. cerevisiae*	10 -20	n.d.[b]	-	Fusetti *et al.*, 1996
Lip 1	*C. maltosa*	0.1	n.d.[b]	-	Mileto *et al.*, 1998
Lip 1	*P. pastoris* *S. cerevisiae*	-	150 5-7	*p*-nitrophenyl palmitate	Brocca *et al.*, 1998
Lip 1	*P. pastoris*	-	253.3	Tributyrin	Chang *et al*, 2005
Lip 1	*P. pastoris*	152	277.18 ± 13.5	Triacyl glycerol	Chang *et al.*, 2006
Lip 2	*P. pastoris*	2.3	n.d.[b]	-	Lee *et al*, 2002
Lip 2	*P. pastoris*	-	2.44	Lipase colorimetric (LIP Kit, Roche Diagnostics, Germany)	Ferrer *et al.*, 2009
Lip 3	*P. pastoris*	-	1.2	*p*-nitrophenyl butyrate	Chang *et al.*, 2006
Lip 4	*E. coli*	-	25.8 U/mg	Tributyrin	Tang *et al.*, 2000
Lip 4	*P. pastoris*	-	7 13	*p*-nitrophenyl laurate Tributyrin	Tang *et al.*, 2001
LipJ08	*P. pastoris*	-	4.7	*p*-nitrophenyl palmitate	Xu *et al.*, 2009

[a]One unit of enzyme will hydrolyze 1.0 µmol of substrate per min and mL of enzyme in assay condition.
[b]n.d. = not determined

Questions related to the role of various residues in the catalytic events and in substrate recognition were thus addressed by the use of third-generation CRL (Brocca *et al.*, 2000, Manetti *et al.*, 2000, Tang *et al.*, 2001, Schmitt *et al.*, 2002, Brocca *et al.*, 2003, Secundo *et al.*, 2004) and lead to a much better understanding of CRL.

Part III: Lipase Lip2 from *Yarrowia lipolytica*

Y. lipolytica is characterized by its capacity to secrete large amounts (up to g/L) of various proteins. These proteins include two proteases, one induced under alkaline conditions (Ogrydziak and Mortimer, 1977), the other with acidic pH (Yamada and Ogrydziak, 1983), a RNase (Cheng and Ogrydziak, 1986), a phosphatase (Moran *et al.*, 1989), and several lipases (Peters and Nelson, 1948). Lipase activities have been detected and analyzed since last two decades, for a review see Vakhlu and Kour (2006)

16 genes encoding lipases were found in the genome of Y. *lipolytica*. Among them, only three proteins, encoded by LIP2, LIP7 and LIP8 were characterised to date (Pignede *et al.*, 2000, Fickers *et al.*, 2005). These three proteins display around 30% sequence identity and were found either extra cellular or cell-bound. Lip2 was isolated first and subsequently extensively studied and will be introduced below.

1. Biochemical and structural data

Y. lipolytica lip2 is a glycosylated 301 amino acid protein of 38 kDa encoded by the LIP2 gene. It is produced as a *Pre-Pro*-enzyme of 334 amino acids. The first 33 amino acids consist of signal peptide, namely ***prepro***, a signal sequence with four motive X-Ala, X-Pro (substrates of diaminopeptidase) and apparently molecular weight of 38 kDa (Pignède *et al.*, 2000). In some report note that the presence of different isoformes, thus, three to four isoformes present different in isolelectric points from 5.0 to 5.4 were identified by IEF (Aloulou *et al.*, 2007; Destain, *et al.*, 1997). Lipase is active at low temperature (up to 5°C) and is quickly inactivated beyond 50°C, the optimum temperature is between 30°C and 40°C. It is active in a pH range from 4.0 to 8.0 with an optimum between pH 6.0 and 8.0 according to the substrates.

2. Application of Y. lipolytica lipase

Lip2 displays many industrial applications; these are summarised in table I-11. It is particularly promising regarding the resolution the racemic mixtures of 2-substituted carboxylic acids and this is the reason why we are studying this lipase in

the laboratory. These compounds are of great interest for drug companies because they are used as starting material for the synthesis of drugs (precursor of analgesics, prostaglandins and semi-synthetic penicillin). Recently, Lip2 was shown to display an activity and enantioselectivity higher than all the lipases previously tested for the resolution of racemates of 2-bromo-arylacetic acids derivatives (Guieysse, *et al.*, 2004). Moreover, the mutations of several positions surrounding the active site of Lip2 lead to a mutant, where the valine in position 232 was replaced by an alanine, displaying a tremendous increase in enantioselectivity (Bordes *et al.*, 2009). This newly constructed variant, which allowed meeting industrial expectations thanks to only a single mutation, opened even larger perspectives for the industrial use of Lip2.

Table I-11. Industrial applications of lip2 from *Y. lipolytica* (summarized by Beopoulos *et al.* 2009, in Handbook of Hydrocarbon and Lipid Microbiology, 2010).

Application propose	Utilization	References
Single cell oils	Accumulated lipid in the cell used as alternative source for produce polyunsaturated fatty acids (PUFA)	Athensteadt *et al.*, 2006
PUFA production	Engineered by introducing and expressing heterologous genes encoding the ω-3/ω-6 biosynthetic pathway in the oleaginous host	Damude *et al.*, 2006
Pharmaceutical	Resolution reaction of 2-halogeno carboxylic acids intermediates in drug industrials.	Guieysse *et al.*, 2004 Bordes *et al.*, 2007 Cancino *et al.*, 2008 Bordes *et al.*, 2009
Fine chemistry	Used for bioconversion processes for production of enzymes and secondary metabolites, such as: isocitrate ltase, citric acids, γ-dodecalactone and pyruvic acids	Thevenieau *et al.*, 2009
Bioremediation	Treatment of mineral oil pollution and plant oil waste.	Oswal *et al.*, 2002
	Used for bioremediation of sites contaminated with petroleum	Zinjarde and Pant 2002

3. Production of Lip2

Because of its potential industrial applications (table I-11), many approaches were attemped to optimize *Y. lipolytica* extracellular lipase production. Destain *et al.*, (1997) reported the selection of mutants with increased lipase secretion (up to 1000 U/mL) after chemical mutagenesis of the wild-type strain CBS6303. Other studied

from Nicaud *et al.*, (2002) described the construction and selection of overproducing mutants using genetic engineering of the producing strain (Fickers, *et al.*, 2003; Fickers, *et al.,* 2005). They also developed a set of expression vectors described in paragraph IV.3.2.1. Attemps to produce it in other organisms, or optimization of the medium and manufacturing control (Destain *et al.*, 2005; Fickers, *et al.*, 2005) were also reported. Lip2 is conventionally produced as homologous in *Y. lipolytica* modified to improve the production (Fickers, *et al*, 2003; Pignede, *et al.*, 2000). The wild-type strain secretes lip2 when using oil (olive oil or triolein) or fatty-acid (oleic acid) as carbon source in the culture medium. Fickers and coworker (Fickers *et al.*, 2003) isolated a strain, namely LgX64.81, created by chemical mutagenesis on the wild-type strain, which was able to produce great quantity of lipase on medium without oil and containing glucose. In this case, the system of regulation of lip2 seemed to have been modified (Fickers, *et al.,* 2005; Turki *et al.*, 2010). In recombinant strains, the production of lipase is dependent on the promoter used. Promoter POX2, inducible by oleic acid or methyl oleate, allows strong productions of lip2 in *Y. lipolytica* (Fickers *et al*, 2005).

Besides the medium composition, cultivation strategy affects the effectiveness of the process. When most lipase production were performed in batch culture, as it is simple and enables a short process time, Turki *et al.*, (2010) developed a fed-batch process for the production of lip2 and obtained a 2-fold increase in the volumetric lipase productivity.

PART IV: *Yarrowia lipolytica*: a powerful expression system

As the amount of protein necessary will depend on the applications, from a few milligrams of proteins for a functional characterization to kilograms for pharmaceutical applications, one will choose to either isolate its protein from natural sources or produce it in a recombinant way. Since the production of the first recombinant protein in 1977, namely the peptide hormone somatostatin, which was produced in *E. coli* (Itakura *et al.*, 1977), the use of recombinant proteins in biotechnology did not cease growing. As a matter of fact, to date, the majority of proteins marketed (drug proteins/enzymes) are produced in a recombinant way. Consequently, many systems of expression were developed depending on the desired application. The choice of an expression system will thus depend on the scope of the study, making a good compromise between the level of production necessary, the quality of protein produced and the cost of the process.

1. Microbial expression systems

From their facility of use, low cost of cultivation and the availability of many commercial vectors, the gram negative *E. coli* and gram positive *Bacillus subtilis* have became the most extensively used prokaryotic hosts for the production of recombinant proteins. A great quantity of proteins, from antimicrobial peptides to enzymes and challenging membrane proteins were produced in *E. coli* and *B. subtilis* (Ferguson *et al.*, 1998; Zhou *et al.*, 2010; Singh *et al.*, 2010). However, these bacteria present important limitations. First, the quality of the protein obtained, in particular when these are proteins of eukaryotic origin. Indeed, because of the differences between the prokaryotic and eukaryotic cell machinery, the recombinantly produced proteins can display some differences compared to the natively produced protein, mostly as regards to post-translation modifications. Indeed, as prokaryotic expression systems, *E. coli* and *B. subtilis* cannot perform post-translational modifications, such as N- and O-glycosylation, fatty acid acylation, phosphorylation, and disulfide-bond formation, that are often required for proper folding and thus functionality of the interest protein (Yin *et al.*, 2007).

In addition, proteins are often produced in *E. coli* as inclusion bodies (Idicula-Thomas and Balaji 2005), which entails the use of a denaturation and renaturation steps for obtention of soluble protein. Moreover, *E. coli* and *B. subtilis* produced proteins are rarely excreted, which complicates the purification.

2. Fungal expression systems

The development of new eukaryotic production systems (yeasts, insects and mammalian cells) makes it possible to mitigate these limitations. Nevertheless, the cultivation of insects and mammalian cells are more expensive and difficult to implement.

Within this framework, yeasts are particularly interesting in that they gather both the advantages of prokaryotic systems in term of simplicity of implementation and production costs, and the advantages of eukaryotic systems in term of protein quality. Moreover, yeasts are the sole eukaryotic expression system allowing high throughput screening of genetically engineered proteins. Finally, many types of yeast are classified as GRAS (Generally Recognised As Safe) by Food and Drug Administration, as their use prevents any kind of contamination by viruses compared with mammalian cells, which allow an industrial use of yeasts. Therefore, yeast will be the sole eukaryotic expression system described here.

2.1. *Saccharomyces cerevisiae* - the mostly used systems of expression

The yeast *S. cerevisiae* was the first eukaryotic expression system used. In 1984, this yeast enabled the production of the first recombinant vaccine against the hepatitis B virus (McAleer *et al.*, 1984). However, this yeast presents limitations: the protein is generally produced at low levels and not always secreted and a hyperglycosylation profile of proteins is often observed. Therefore, the use of nonconventional yeasts was strongly developed. Amongst them, the mostly used are *Pichia pastoris*, *Kluyveromyces lactis*, *Hansenula polymorpha* and *Yarrowia lipolytica* (Dominguez, et al., 1998).

2.2. Non conventional yeasts

2.2.1. Methylotrophic yeasts: *Pichia pastoris* and *Hansenula polymorpha*

These yeasts are able to grow on methanol as the only source of carbon. The regulation of methanol metabolizing enzymes involves accurately controlled strong promoters (promoter of AOX1 gene for alcohol oxydase in *P. pastoris* and promoter of MOX gene for methanol oxydase in *H. polymorpha*). These were used to overproduce recombinant proteins up to 30% of total proteins. *H. polymorpha* is regularly used in industrial fields because it is more thermo-tolerant than *P. pastoris* and its process of fermentation does not require methanol (MOX is replaced by glycerol). Nevertheless, *P. pastoris* is the most largely nonconventional yeast used amongst the scientific community due to available expression kits (Invitrogen).

2.2.2. *Kluyveromyces lactis*

K. lactis is currently used by industry in several countries (the United States, England and Japan) to industrially produce recombinant chymosine (one of the active components used in cheese dairy). Today more than 40 proteins of various origins were expressed successfully in this host (van Ooyen *et al.*, 2006). An advantage of *K. lactis* is the availability of an original cre-lox recombinase system which makes it possible to remove the selection marker once the gene was introduced, in order to re-used it for the introduction of other genes (Selten *et al.*, 1995). In addition, genetically engineered strains of *K. lactis* were developed to improve the secretion of produced proteins (Bartkeviciute and Sasnauskas 2003; Bartkeviciute and Sasnauskas 2004; Donnini *et al.*, 2004; Uccelletti *et al.*, 2004). A kit for protein expression in *K. lactis* can be purchased from New England Biolabs.

2.2.3. *Yarrowia lipolytica*

More than forty proteins of various origins were successfully expressed in this yeast. On the one hand, the production of proteins can be carried out using an original expression system based on non-homologous multiple genomic integration of the interest gene, leading to a better stability of the strains obtained (Madzak *et al.*, 2004). On the other hand, homologous integration in *Y. lipolytica* enables a targeted integration of the gene in the yeast genome (Gellissen *et al.*, 2005), enabling the

screening of few transformants to select mutant with a correct integration. Moreover, as in *K. lactis*, the cre-lox recombinase system allows the re-use of the selection markers. Finally, a *Y. lipolytica* strain specifically dedicated to high throughput screening was recently developed (Bordes *et al.*, 2007) enabling the reproducible 96-well plate production of protein and activity assays.

A commercial kit for protein expression in *Y. lipolytica* was recently developed by Yeastern biotech.

Few studies compare the available expression systems; a study by (Muller *et al.*, 1998) compared various alternate yeasts with the yeast *S. cerevisiae*. The comparison was carried out on 6 different enzymes and 5 expression hosts (*polymorpha, K. lactis, pombe, Y. lipolytica* and *S. cerevisiae*). The authors studied various parameters: quantity of secreted enzyme, glycosylation profile, effectiveness of transformation, and plasmid stability. Amongst the different systems studied, *Y. lipolytica* was shown to be the most effective as it exhibited a good transformation yield, stable plasmids and correct growth. In addition, no hyperglycosylation problem was raised and *Y. lipolytica* displayed a particularly high secretion capacity.

Finally, Table I-12 gathers the reported advantages and drawbacks of the different eukaryotic and prokaryotic recombinant expression system.

Table I-12. Advantages and disadvantages of recombinant expression systems.

Characteristics	Bacteria	Yeast	Insect cells	Mammalian cells
Cell Growth	Rapid	Rapid	Slow	Slow
Complexity of Growth Medium	Minimum	Minimum	Complex	Complex
Cost of Growth Medium	Low	Low	High	High
Expression Level	High	Low - High	Low - High	Low - Moderate
Extracellular Expression	Secretion to Periplasm	Secretion to Medium	Secretion to Medium	Secretion to Medium
Protein Folding	Refolding Usually Required	Refolding May Be Required	Proper Folding	Proper Folding
N-linked Glycosylation	None	High Mannose	Simple, No Sialic Acid	Complex
Other Post translational	No	Yes	Yes	Yes

As highlighted by table I-12, *Y. lipolytica* seems to be a particularly effective expression system for the production and directed evolution of enzymes and will thus be more accurately described in next pararagraph.

3. *Yarrowia lipolytica* expression system

As shown previously, this yeast offers a number of advantages as expression system. Its retains the advantages of microbial expression systems in its ease of manipulation and growth capacity, but, in contrast to these latter, its also possesses an eukaryotic subcellular organization able to perform post-translational processing of eukaryotic proteins. Moreover, it was shown to be a very attractive alternative host, especially in terms of performance reproducibility.

3.1. Characteristics

Y. lipolytica (formerly, *Candida, Mycotorula, Endomycopsis* or *Saccharomyces lipolytica*) is a non conventional, non pathogenic yeast (Holzschu *et al.*, 1979) which was classified as GRAS (generally regarded as safe) by the Food and

Drug Administration. It is naturally found in high lipid containing products such as cheeses, dairy products, plant oils and sewages.

Y. lipolytica is an ascomycota of the Saccharomycetales order. It was initially classified as *Candida lipolytica*, then as *Endomycopsis lipolytica* (Wickerham *et al.*, 1970) *Saccharomycopsis lipolytica*, (Yarrow 1972), and finally ***Yarrowia lipolytica*** (van der Walt and von Arx 1980). It is a dimorphic that grows as budding cells, hyphae or pseudohyphae depending on growth conditions. *Y. lipolytica* can use glucose (but not sucrose because it does not have invertase), alcohols, acetate and hydrophobic substrates (such as alkanes, fatty acids and oils) as carbon sources (Barth and Gaillardin 1997). It is strictly aerobic and grows at temperatures ranging from 5°C to 30°C.

The genome of type E150 *Y. lipolytica* was sequenced at the Génolevures Consortium in 2000 (Casaregola *et al.*, 2000; Dujon *et al.*, 2004). The choice of this yeast was based on its taxonomic distance to *S. cerevisiae*, its original lipid metabolism and it high secretion capacity. Its genome is organized in 6 chromosomes, representing a total of 20.5 Mb (much more than *S. cerevisiae* and *S. pombe*), with size varying between 2.6 and 4.9 Mb. No plasmidic DNA was identified. Several observations suggested that it diverged considerably from other ascomycota in that it displays: i) high GC content (approximately 50% compared with about 38% for the others ascomycota), ii) high gene density (3.3 Kb compared to 2 Kb for *S. cerevisiae*), iii) high intron frequency (13% of genes have at least 1 intron, compared to 4.8 % for *S. cerevisiae*), iv) ribosomal DNA is distributed in seven loci located in the subtelomeric areas (only one locus in *S. cerevisiae*), v) low level of gene similarity with the other yeasts, vi) high number of transfer RNA.

In addition, it display a high number of genes encoding protein involved in the metabolism of hydrophobic substrates (De Hertogh *et al.*, 2006; Thevenieau *et al.*, 2007), as well as a great diversity in transposable elements (Neuveglise *et al.*, 2002; Neuveglise *et al.*, 2005). Furthermore, many differences were observed between the isolated strains, which differ by the size of the chromosomes (Casaregola, *et al.*, 1997), the number of ribosomal DNA (Baker and Platt 1986) and the presence of the Ylt1 retrotransposon (Schmid-Berger *et al.*, 1994).

3.2. Genetic tools

Many genetic tools were developed for the expression of recombinant proteins in *Y. lipolytica* (Barth and Gaillardin 1997).

3.2.1 Expression vectors

Vectors used are shuttles allowing at the same time replication in bacterial hosts and expression in yeast. The yeast part of the vector contains the expression cassette (promoter – signal sequence - gene - termination), a selection marker, and all the elements necessary for maintenance in yeast. The list of these elements is recapitulated in Table I-13 according to (Madzak, *et al.*, 2004). Two types of shuttle vectors can be used, differing on their mode of maintenance into yeast cells.

- Replicative vectors:

Y. lipolytica does not naturally contain replicative plasmids. Replicative synthetic vectors could be built, requiring an origin of replication and a centromer (Baker *et al.*, 1993; Vernis *et al.*, 1997). Unfortunately, the number of vectors per cell is limited (1 to 3) and their maintenance requires a constant selection pressure. Their use is thus limited.

- Integrative vectors:

Homologous integration can be carried out by simple crossing-over (within the site LEU2, URA3, in the ribosomal DNA or the platform of chromosomal integration). The strains have the advantage of being stable during at least 100 generations without needing a pressure of selection. Random integration in the genome can also be carried out by non-homologous insertion (Hamsa and Chattoo 1994), due zeta zones (the long terminal repeat zeta of the *Y. lipolytica* retrotransposon Ylt1) flanking the expression cassette. The use of these vectors makes it possible to get rid of the bacterial part of the plasmid (Nicaud *et al.*, 1998; Pignede *et al.*, 2000).

A technique of transformation by lithium acetate method is possible to obtain a good effectiveness of transformation (Xuan *et al.*, 1988).

Table I-13. Components available for the construction of *Y. lipolytica* expression vectors (most commonly used and most interesting items are underlined). (Madzak et al., 2004)

Component[a]	Characteristics
Marker genes	
LEU2, *URA3*, *LYS5*, *ADE1*	Auxotrophy complementation
ura3d4	Item + copy number amplification (cf. text)
Phleo^R, *hph* (*E. coli*)	Antibiotic resistance (respectively, to phleomycin and hygromycin B)
SUC2 (*S. cerevisiae*)	Sugar utilization
Promoters (source)	
p*LEU2* (β-isopropylmalate dehydrogenase)	Inducible by leucine precursor
p*XPR2* (alkaline extracellular protease)	Inducible by peptones
p*POX2*, p*POT1* (respectively, acyl-CoA oxidase 2, 3-oxo-acyl-CoA thiolase)	Inducible by fatty acids and derivatives, and alkanes
p*ICL1* (isocitrate lyase)	Inducible by fatty acids and derivatives, alkanes, ethanol and acetate
p*POX1*, p*POX5* (acyl-CoA oxidases 1 and 5)	Weakly inducible by alkanes
p*G3P* (glycerol-3-phosphate dehydrogenase)	Inducible by glycerol
p*MTP* (bidirectional: metallothioneins 1 and 2)	Inducible by metallic salts
hp4d (hybrid promoter derived from p*XPR2*)	Growth-phase-dependent
p*TEF*, p*RPS7* (respectively: translation elongation factor-1α, ribosomal protein S7)	Constitutive
Secretion signals[b]	
Native	Frequently efficient in *Y. lipolytica*
XPR2 prepro	13 aa pre 10 XA or XP dipeptides 122 aa pro KR cleavage site
XPR2 pre + dipeptide	13 aa pre 5–10 XA or XP dipeptides
XPR2 pre	13 aa pre LA cleavage site
LIP2 prepro	13 aa pre 4 XA or XP dipeptides 10 aa pro KR cleavage site
Terminators	
XPR2t, *LIP2t*, *PHO5t*	Respectively, 430, 150 and 320 bp fragments
Minimal *XPR2t*	PCR-synthesized 100 bp fragment with added restriction sites
Elements for maintenance in yeast cells	
ARS18, *ARS68*	Autonomously replicative vectors can maintain only 1–3 copies/cell[c]
Homology to genome	Homologous integration (in *LEU2*, *URA3*, *XPR2* terminator, rDNA, or, when present, in zeta or pBR322 docking platform)
Zeta (Ylt1 LTR)	Non-homologous integration in Ylt1-devoid strains (cf. text)

3.2.2. Selection markers

A gene of resistance to kanamycine, ampicilline or tetracycline enables bacterial selection. *Y. lipolytica* is only sensitive to few antibiotics, such as bleomycine/phleomycine (Gaillardin and Ribet 1987)and hygromycine B (Cordero Otero and Gaillardin 1996). Therefore, the selection by complementation of auxotroph strains (mainly for LEU2 and URA3) remains the mostly used (Barth and Gaillardin, 1997). Very recently two other markers of selections, namely GUT2 and ADE2 were developed at the Laboratory of Microbiology of Grignon.

3.2.3. Promoters

The strong promoter of the XPR2 gene, encoding the extracellular alkaline protease of *Y. lipolytica* was undoubtedly the mostly used promoter. Nevertheless, its regulation is complex (pH> 6, obligatory presence of peptones in the medium) and renders its industrial use complicated. Therefore, starting from a sequence activating

(UAS1) (Blanchin-Roland *et al.,* 1994) slightly affected by culture medium composition, a hybrid promoter was obtained: hp4d (hybrid promoter 4 direct repeat) corresponds to 4 copies of UAS1 (Madzak *et al.,* 2000). This promoter is considered as semi-constitutive, because its induction is dependent on growth phase, but the elements controlling its regulation are still unclear. Two other strong constitutive promoters were thus afforded. TEF (translation elongation Factor-1 alpha) and RPS7 (ribosomal protein S7) were described by Muller *et al.* (1998). Inducible and strong promoters were also described. The promoters of POX2 gene (encoding for acyl-CoA-oxydase 2) (Pignede, *et al.*, 2000), ICL1 gene (Isocitrate lyase) (Juretzek *et al.,* 2001), and POT1 gene (3-oxo-acyl-CoA thiolase). These promoters activate the expression of enzymes involved in the metabolism of hydrophobic substrates and are thus inducible by alkanes and fatty-acids.

3.2.4. Secretion signals

Signal sequences of the principal proteins secreted by wild type *Y. lipolytica* are used for secretion of recombinant proteins. The most commonly used are the prepro sequence of the XPR2 gene, which encodes for the extracellular alkaline protease and the prepro sequence the LIP2 gene. Sometimes the proper signal sequences of the interest proteins were also used successfully. It was the case for six fungal proteins expressed in the study of Muller *et al.*, (1998) but also for proteins of other origins such as the α- amylase of rice (Park *et al.,* 1997), for instance, where secretion and maturation of the protein are more efficient with its own signal sequence than with signal sequences from *Y. lipolytica.* To date, no general rule makes it possible to predict which signal sequence must be used for a given protein.

3.2.5. Amplification systems

Several amplification systems were developed. Those are based on the use of a defective selection marker, a mutated allele of the *URA*3 gene in which part of the promoter was deleted: ura3d4 (Le Dall *et al.,* 1994). For auxotroph complementation, the gene thus has to be present in several copies. Two major strategies of amplification were used for *Yarrowia lipolytica*:

- Multiple homologous integration: repeated sequences can be targeted, such as clusters of ribosomal DNA or the above mentioned zeta zones (Juretzek, *et al.*,

2001). Single sequences, such as XPR2, can also be targeted. In all cases, tandem integration of the expression cassette and a great number of integrations are observed (60-100 copies). However, many copies are lost in conditions of induction and the number of ura3d4 copies often stabilizes around 10 to 13 (Le Dall, *et al.*, 1994).

- Nonhomologous integration: when zeta zones are used for integration in wild type strains without zeta zones, one observes a nonhomologous, multiple and dispersed recombination. This type of integration, even if it does not allow such a great number of copies, avoids the formation of tandem and thus enables the construction of very stable strains. Strains with up to 16 copies and stable on 120 generations were observed (Pignede, *et al.*, 2000).

Other ways were developed to improve the production of proteins. Indeed, overproducing strains were obtained by chemical mutagenesis and can be used as host for the production of heterologous proteins (Fickers, *et al.*, 2003; Fickers, *et al.*, 2005). In addition, an original strategy was developed to improve the production of proteins. As *Y. lipolytica* is strictly aerobic, its growth can be limited by oxygen, in particular when cells density is important. Thus, Bhave and Chattoo, (2003) expressed a haemoglobin of Vitreoscilla in *Y. lipolytica*; this allowed a better use of oxygen, an improved growth and therefore an increase in the quantity of secreted extracellular proteins.

3.2.6. Production strains

Strains used as host can be an important parameter in the level of recombinant proteins secretion. Indeed, De Baetselier *et al.* (1991) showed that the levels of secretion were different (up to a factor 100) for the expression of a glucose oxydase from *Aspergillus Niger* according to the strain of *S. cerevisiae* used. These differences could not be explained by any particular genetic marker.

Such studies were not conducted on the secretion by *Yarrowia lipolytica* different strains. However, several elements should be taken into account when choosing an expression host. The presence of zeta zones in the host strain will allow two different types of integration when the expression cassette is bordered by those latter: a homologous and multicopies integration in the American strain CBS6142-2, where the zeta zone are found in great number, or a nonhomologous integration in the

French strain W29 and German strain H222, which do not have zones zeta. The strain po1d is a derivative of the French strain W29 where the extracellular alkaline protease was deleted to avoid the hydrolysis of secreted proteins of interest and an invertase was added, which enable the strain to grow on saccharose, and abundant and not very expensive substrates such as molasses (Nicaud, *et al.*, 1989). This strain might thus also be interesting for protein production.

4. A dedicated expression system for directed evolution

An expression system for high-throughput screening must satisfy several points. First, the transformation yield must be sufficiently high to be able to generate large libraries; secondly, the screening of the enzymatic activity must be achievable as fast as possible; and the growth of variants and the protein expression must be identical between the variants in order to allow fast identification of potentially interesting proteins.

Nowadays, although displaying many limitations, the expression host that was mostly used for such purpose is *E. coli*. However, the protein to evolve might not always be functional when expressed in a prokaryotic host, and might also be found intracellular. Then, even if the transformation effectiveness is high with such host, the major hurdles to obtain a sufficient number of transformants by plate are the effectiveness of the ligation step and the low level of transformation of ligated products. Therefore, yeasts seem to be an alternative host for this type of application. Indeed, as discussed above, it gathers the advantages of prokaryotic systems in term of facility of use, effectiveness of transformation, and numerous genetic tools available and the advantages of eukaryotic systems in terms of post-translation modifications and protein secretion.

At the beginning, only *S. cerevisiae* (Cherry *et al.*, 1999; Bulter and Alcalde 2003) and *P. pastoris* (Boettner *et al.*, 2002)had been used to evolve enzymes. *S. cerevisiae* has high transformation effectiveness but a tendency to hyperglycosylation proteins and it often presents problems of secretion. *P. pastoris* presents several disadvantages, such as low transformation effectiveness, strong conversion rate of the selection marker without integration of the expression cassette (thus 10 to 50% of transformants will not express the protein of interest because of the recombination of

the sole selection marker) and a variable number of copies. Another advantage of yeasts, with regard to directed evolution, is the possibility of introducing genetic diversity directly by "*in vivo*" recombination (Pompon and Nicolas 1989; Cherry *et al.*, 1999; Abecassis *et al.*, 2000).

Previous work by Bordes *et al.*, (2007) developed a *Y. lipolytica* strain specifically dedicated to high through-put screening of enzymes improved through directed evolution. Yeast transformation was rendered reproducible by integration of a zeta docking platform in the LEU2 locus of the genome. Thus, integration of the expression cassette could occur at a targeted locus, entailing the homologous and monocopy integration always at the same position in the genome. This newly constructed host was tested and showed really good reproducibility for the entire screening process, i.e. from transformation of the strains to enzymatic assay, compared to the formerly used strain of *Y. lipolytica*.

PART V: Biodiesel production by immobilized lipases

1. Production of biodiesel

Previous studies have shown that triglycerides hold promise as alternative diesel engine fuels. Some natural glycerides contain higher levels of unsaturated and saturated fatty acids. They can not be used as fuel in a diesel engine in their original from. The high viscosity, acid composition, and free fatty acid content of such oil, as well as gum formation due to oxidation and polymerization during storage and combustion, carbon deposits, and lubricating oil thickening are some of the more obvious problems (Darnoko *et al.*, 2000, Komers *et al.*, 2001 and Demirbas A., 2003). Consequently, considerable effort has gone into developing vegetable oil derivatives that approximate the properties and performance of hydrocarbon-based diesel fuels. Problems encountered in substituting triglycerides for diesel fuels are mostly associated with their high viscosity, low volatility and polyunsaturated character. There are various ways to produce biodiesel (Demirbas A., 2009) as follows:

1. **Direct use and blending**: The blending of vegetable oil with diesel fuel were experimented successfully by various researchers.
2. **Pyrolysis**: Refers to chemical change caused by application of heat to get simpler compounds from a complex compound. The process is also known as cracking.
3. **Microemulsion:** To solve the problem of high viscosity of vegetable oil, micro emulsions with solvents such as methanol, ethanol (Demirbas A., 2009), propanol (Ranganathan et al., 2008) and butanol have been used.
4. **Transesterification**: The most popular method of producing biodiesel is the transesterification of vegetable oils. Biodiesel obtained by transesterification process is a mixture of mono-alkyl esters of higher fatty acids. Transesterification is the alcoholysis of triglyceric esters resulting in a mixture of mono-alkyl esters and glycerol and the sequence of processes. The high viscosity component, glycerol, is removed and hence the product has low viscosity like the fossil fuels. The mixture of these mono-alkyl esters can therefore be used as a substitute for fossil fuels. The

transesterification process can be done in a number of ways such as using an alkali catalyst, acid catalyst, biocatalyst, heterogeneous catalyst or using alcohols in their supercritical state.

It has been recently found that enzymes such as lipase can be used to catalyze transesterification process by immobilizing them in a suitable support. The advantage of immobilization is that the enzyme can be reused without separation. Also, the operating temperature of the process is low (approximately 50 °C) compared to other techniques as show in Table I-14.

Table I-14 Comparison between alkali-catalyzed and lipase-catalyzed methods for biodiesel production (Fukuda et al., 2001).

	Alkali-catalyzed process	**Lipase-catalyzed process**
Reaction temperature	60-70 °C	30-40 °C
Free fatty acids in raw material	Saponified products	Methyl esters
Water in raw materials	Interference with the reaction	No influence
Yield of methyl esters	Normal	Higher
Recovery of glycerol	Difficult	Easy
Purification of methyl esters	Repeated washing	None
Production cost of catalyst	Cheap	Relatively expensive

2. Immobilization of lipases

Methods of enzyme immobilization

Enzyme immobilization technology may be an effective means to perform enzyme reuse and to improve its activity and stability (Chang et al., 2008). A number of method for immobilization of enzymes have been reported in the literatures review such as adsorption onto an insoluble material, covalent linking to an insoluble carrier, entrapping enzymes within the matrix and encapsulation in gel bead. The relative merits each are discussed briefly below;

2.1. Adsorption

Immobilization by adsorption is the simplest method and involves reversible surface interactions between enzyme/cell and support material. The forces involved

are mostly electrostatic, such as Van der Waals interactions, hydrophobic interactions, hydrogen bonds, ionic bonds (Villeneuve *et al.*, 2000).

2.2. Covalent linkage

This method of immobilization involves the formation of a covalent bond between the enzyme and support material. The bond is normally formed between functional groups present on the surface of the support and functional groups belonging to amino acid residues on the surface of the enzyme. A number of amino acid functional groups are suitable for participation in covalent bond formation. Those that are most often involved are the amino group (NH_2) of lysine or arginine, the carboxyl group (CO_2H) of aspartic acid or glutamic acid, the hydroxyl group (OH) of serine or threonine, and the sulfydryl group (SH) of cysteine.

2.3. **Encapsulation**

Encapsulation of enzyme and or cell can be achieved by enveloping the biological components within various forms of semi permeable membranes. It is similar to entrapment in that the enzymes and cell are free in solution, but restricted in space. Large proteins or enzymes cannot pass out of or into the capsule, but small substrates and products can pass freely across the semipermeable membrane. Many materials have been used to construct microcapsules varying from 10-100 μm in diameter; for example, nylon and cellulose nitrate have proven popular. The problems associated with diffusion are more acute and many result in rupture of the membrane if products from a reaction accumulate rapidly. A further problem is that the immobilized cell or enzyme particle may have a density fairly similar to that of the bulk solution with consequent problems in reactor configuration, flow dynamics, and so on.

Since most enzymes applied for industrial uses need to be immobilized in order to be reusable and therefore reduce the cost of the operation, there are many methods to immobilize enzyme. Each method provides different advantages and disadvantages depending upon its nature as shown in Table I-15.

Table I-15. Comparative methods for the immobilization of enzyme (http://www.org /ippagele/ipdate/2004/05/file/e200405-1101.pdf).

Characteristics	Adsorption	Covalent binding	Entrapment	Membrane confinement
Preparation	Simple	Difficult	Simple	Difficult
Cost	Low	High	Moderate	High
Binding force	Variable	Strong	Weak	Strong
Enzyme leakage	Yes	No	Yes	No
Applicability	Wide	Selective	Wide	Vary wide
Running problem	High	Low	High	High
Matrix effects	Yes	Yes	Yes	No
Large diffusional barriers	No	No	Yes	Yes
Microbial protection	No	No	Yes	Yes

3. Lipase immobilization

From previous literature, studies on different support materials and protocols for lipase immobilization and the effects of operational conditions on the enzyme activity and enantioselectivity are also explored.

Takaç and Bakkal immobilized *Candida rugosa* lipase on Amberlite XAD 7 and immobilization under the best reaction conditions in achieving high activity. The authors found that pH 6 and 45°C were obtained for immobilization *C.rugosa* lipase adsorbed on Amberlite XAD 7 was used in the hydrolysis of racemic Naproxen methyl ester (Takaç and Bakkal, 2007).

Huang and Cheng immobilized alkali lipase from *Penicillium expansum* on biomodal ceramic foam and determined on the preferable immobilization conditions. The optimal conditions for immobilization of lipase were found at pH 8, 12 g lipase/g support, 4 hr immobilizing time and 20 °C immobilizing temperature (Huang and Cheng, 2008).

Feresti and Ferreira studied the immobilized lipase from *Candida rugosa, Pseudomonas fluorescens* and *Candida antarctica* B onto chitosan and glutaraldehyde pretreated chitosan powder. The prepared biocatalysts were assayed in the direct esterification of oleic acid and ethanol to produce the ethyl oleate. The immobilization of lipase from *Candida antarctica* B onto untreated chitosan powder led to 75% conversion of the fatty acid in 24 hr of reaction.They concluded that *Candida antarctica* B was the most active (Feresti and Ferreira, 2007).

In Table I-16 showed summary of immobilization of lipase for biodiesel production.

Table I-16. Summary of transesterification catalyzed by lipase.

References	Oil/enzyme	Acyl acceptor	Conversion (%)	Technique employed
Watanabe *et al.* (2000)	Vegetable oil, Novozyme 435	Methanol	90-93	Stepwise addition of methanol
Samukawa *et al.* (2000)	Soybean oil, Novozyme 435	Methanol	97	Stepwise addition methanol and preincubation of enzyme in methyl oleate and soyabean oil
Ban *et al.* (2001)	Vegetable oil, *Rhizopus oryzae*	Methanol	90	Stepwise addition of methanol and application of glutaraldehyde for stability of enzyme
Iso *et al.* (2001)	Triolein, *Pseudomonas flourescens*	Butanol	90	Butanol was used as an acyl acceptor and no solvent was used
Shimada *et al.* (2002)	Waste cooking oil, Novozyme 435	Methanol	90	Stepwise addition of methanol
Bako *et al.* (2002)	Sunflower oil, Novozyme 435	Methanol	97	Stepwise addition of methanol and removal ofnglycerol by dialysis
Du *et al.* (2004)	Soy bean oil, Novozyme 435	Methyl acetate	92	A novel acyl acceptor, methyl acetate which had no inhibitory effects was use
Xu et al. (2004)	Soy bean oil, Novozyme 435	Methanol	98	Stepwise addition of methanol and removal of glycerol using the solvent, iso-propanol
Royon et al. (2007)	Cotton seed oil, Novozyme 435	Methanol	97	tert-Butanol was used as a solvent
Dizge and Keskinler (2009)	Canola oil, Thermomyces lanuginosus	Methanol	90	Stepwise addition of methanol
Gao et al. (2009)	Chinese tallow kernel oil, lipase	Methanol	93	