



HAL
open science

Approches multiples d'ingénierie pour l'utilisation d'enzymes hydrolytiques comme outils de synthèse

Julien Durand

► **To cite this version:**

Julien Durand. Approches multiples d'ingénierie pour l'utilisation d'enzymes hydrolytiques comme outils de synthèse. Biochimie, Biologie Moléculaire. INSA de Toulouse, 2017. Français. NNT : 2017ISAT0029 . tel-02011218

HAL Id: tel-02011218

<https://theses.hal.science/tel-02011218>

Submitted on 7 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Cotutelle internationale avec :

Présentée et soutenue par :
DURAND Julien

Le vendredi 1 décembre 2017

Titre :

Approches multiples d'ingénierie pour l'utilisation d'enzymes hydrolytiques
comme outils de synthèse

ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :

LISBP, UMR CNRS 5504, UMR INRA 792, INSA TOULOUSE

Directeur(s) de Thèse :

Dr. Michael O'DONOHUE

Dr. Sophie BOZONNET

Rapporteurs :

Pr. Uwe BORNSCHEUER, Université de Greifswald, Allemagne

Dr. Jean-Guy BERRIN, Biodiversité et Biotechnologie Fongiques, Marseille, France

Autre(s) membre(s) du jury :

Pr. Charles TELLIER, Université de Nantes, France

Dr. Sylvie ARMAND, Université de Grenoble, France

Dr. Régis FAURÉ, Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés,
Toulouse, France

THESE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue par :

Julien Durand

Titre :

Approches multiples d'ingénierie pour l'utilisation d'enzymes hydrolytiques comme outils de synthèse

Ecole doctorale et discipline ou spécialité :

ED SEVAB : Ingénieries microbienne et enzymatique

Directeurs de Thèse :

Dr. Michael O'Donohue

Dr. Sophie Bozonnet

Directeur de recherche, INRA

Ingénieur de Recherche, INRA

LISBP, INSA Toulouse

LISBP, INSA Toulouse

Unité de recherche :

LISBP, UMR CNRS 5504, UMR INRA 792, INSA TOULOUSE

"On n'est pas vieux tant qu'on cherche."

Jean Rostand - *Carnet d'un biologiste*

RÉSUMÉ

La Chimie Verte s'engage entre autres i) à mettre au point des procédés plus respectueux de l'environnement et viables économiquement, et ii) à essayer d'émanciper de la pétrochimie les filières de fabrication de produits, en promouvant l'utilisation de ressources renouvelables. Dans ce contexte, les enzymes représentent des alternatives aux catalyseurs chimiques toxiques et polluants et sont depuis longtemps appréciées pour leur extraordinaire efficacité et sélectivité pour les transformations biologiques et chimiques. Elles constituent notamment des outils de choix pour réaliser des réactions de synthèse et la panoplie de méthodes d'ingénierie de protéines disponibles à l'heure actuelle permet d'élargir encore plus le répertoire d'enzymes utilisables pour la synthèse spécifique et écoresponsable de molécules organiques à forte valeur ajoutée à partir de la biomasse végétale.

L'endoglycocéramidase II de *Rhodococcus* sp. M-777, une glycoside hydrolase à rétention de configuration, a ainsi été la cible d'un travail de remodelage du site actif afin de réorienter son activité vers la synthèse de polyglucosides d'alkyles, de potentiels biosurfactants. Ainsi, grâce à une stratégie de design semi-rationnel, ciblant des résidus impliqués dans la reconnaissance de la molécule acceptrice, une transglycosylase efficace permettant d'atteindre des rendements de production de cellobioside d'alkyle de plus de 70% a été obtenue. La détermination des paramètres cinétiques de la réaction et la modélisation de la mutation découverte permettent de proposer des pistes sur les raisons de cette inversion du ratio transglycosylation/hydrolyse (T/H).

Dans un deuxième temps, une stratégie d'évolution dirigée multi-étapes a été appliquée à la féruloyle-estérase A d'*Aspergillus niger* pour la rendre plus résistante aux chocs thermiques et à la présence de solvants, deux propriétés requises pour pouvoir envisager d'utiliser cette enzyme pour des réactions de transfert dans des conditions industrielles thermodynamiquement favorables. Une méthodologie de criblage haut-débit, s'appuyant sur l'utilisation de substrats chromogéniques originaux, a été développée et a permis l'exploration efficace d'une diversité génétique de plus de 10 000 clones. Au terme de cette étude, un catalogue d'enzymes améliorées, soit pour leur thermostabilité, soit pour leur résistance aux solvants, soit pour les deux propriétés simultanément, a été obtenu, et donne accès à un ensemble de mutations, analysées au regard de leur effet sur l'activité de l'estérase.

Une dernière partie, prenant en compte l'analyse des résultats obtenus, propose des perspectives expérimentales pour aller plus loin dans la compréhension fondamentale des mécanismes moléculaires régissant la modulation du ratio T/H ainsi que la thermostabilité ou la résistance à la présence de solvant. L'accumulation de ces connaissances permettra de pouvoir plus efficacement rationaliser le design de biocatalyseurs répondant aux attentes de la chimie verte.

Mots clés : Chimie verte, biocatalyseur enzymatique, évolution dirigée, synthèse de molécules

SUMMARY

Green chemistry promotes, among other things i) the development of more environmentally friendly processes, which need to remain economically viable, and (ii) the ending of polluting petrochemical industries by promoting the use of renewable resources. In this context, enzymes represent interesting alternatives to toxic and polluting chemical catalysts, and are appreciated for their extraordinary efficiency and selectivity in biological and chemical transformations. Notably, they constitute tools of choice for synthesis reactions, and the range of protein engineering methods available nowadays allows enlarging even more the repertoire of enzymes for the selective and eco-friendly synthesis of organic molecules with high added value from plant biomass.

Endoglycoceramidase II from *Rhodococcus* sp. M-777, a retaining glycoside hydrolase, was subjected to active-site remodelling in order to reorient its activity towards the synthesis of alkyl-polyglucosides, molecules with potential biosurfactant properties. Thus, thanks to a semi-rational design strategy, targeting residues involved in the recognition of the acceptor molecule, an efficient transglycosylase able to reach production yield of more than 70% of alkyl-cellobiosides was obtained. A complete kinetic characterization associated to a modelling study of the identified improved mutant help to identify the determinants of this complete reversion of the transfer / hydrolysis ratio (T/H).

In a second part, a multi-step directed evolution strategy was applied to *Aspergillus niger* feruloyl-esterase A, in order to make it more resistant to heat shocks and to the presence of solvents, two prerequisites to use this enzyme for transfer reactions under thermodynamically favourable industrial conditions. A high-throughput screening methodology, relying on the use of original chromogenic substrates, was developed and allowed to efficiently explore a genetic diversity of more than 10,000 clones. This led to the establishment of a catalog of enzymes optimized either for their thermostability, or for their solvent resistance, or even for both properties at the same time, and to a list of mutations, possibly involved in resistance mechanisms.

Finally, considering all obtained results, in the last chapter, we propose experimental perspectives to go further in the fundamental understanding of the molecular mechanisms governing the modulation of the T/H ratio, or the temperature / solvent resistance of these enzymes. This will pave the way towards a more efficient way to rationally design biocatalysts meeting the expectations of green chemistry.

Key words: Green chemistry, enzymatic biocatalyst, directed evolution, molecule synthesis

Publications scientifiques

Publiées

Durand, J., Biarnés, X., Watterlot, L., Bonzom, C., Borsenberger, V., Planas, A., Bozonnet, S., O'Donohue, M.J., and Fauré, R. (2016). A Single Point Mutation Alters the Transglycosylation/Hydrolysis Partition, Significantly Enhancing the Synthetic Capability of an endo -Glycoceramidase. *ACS Catal.* 6, 8264–8275.

Gherbovet, O., Fauré, R., Ferreira, F., **Durand, J.**, Ragon, M., Hostyn, G., Record, E., Bozonnet, S., and O'Donohue, M.J. (2016). Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities. *J. Mol. Catal. B Enzym.* 126, 24–31.

Bissaro, B., **Durand, J.**, Biarnés, X., Planas, A., Monsan, P., O'Donohue, M.J., and Fauré, R. (2015). Molecular Design of Non-Leloir Furanose-Transferring Enzymes from an α -L -Arabinofuranosidase: A Rationale for the Engineering of Evolved Transglycosylases. *ACS Catal.* 5, 4598–4611.

En préparation

Durand, J., Chaumeton, V., Gherbovet, O., Ragon, M., Jütten, P., L.A. Bras, J., I.P.D. Guerreiro, C., M.G.A. Fontes, C., Piechot, A., Nicaud, JM., Fauré, R., O'Donohue, MJ., Bozonnet, S. Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance.

Gherbovet, O., Ferreira, F., Ragon, M., **Durand, J.**, Bozonnet, S., O'Donohue, M.J., and Fauré, R. Regioselective chemoenzymatic syntheses of chromogenic substrates for feruloyl esterases

Communications orales

Colloque du Club Biocatalyse en Synthèse Organique - CBSO (7-10 juin **2016**, Evian-les-Bains) Dualité enzymatique : comment faire basculer une enzyme hydrolitique vers une activité de synthèse. B. Bissaro, **J. Durand**, X. Biarnés, T. Tandrup, C. Dumon, P. Monsan, L. Lo Leggio, A. Planas, S. Bozonnet, M.J. O'Donohue, R. Fauré

Journées du Groupe Français des Glycosciences - GFG (23-27 mai **2016**, Aussois) How to tip the balance from hydrolysis toward transglycosylation: molecular basis in retaining GHs. B. Bissaro, **J. Durand**, X. Biarnés, T. Tandrup, C. Dumon, P. Monsan, L. Lo Leggio, A. Planas, S. Bozonnet, M.J. O'Donohue, R. Fauré

European Carbohydrate Symposium (2-6 août **2015**, Moscou, Russie) New insight into the design of transglycosylases in retaining glycoside hydrolases. B. Bissaro, **J. Durand**, X. Biarnés, P. Monsan, A. Planas, S. Bozonnet, R. Fauré, M.J. O'Donohue

International Conference on Novel Enzymes (14-17 octobre **2014**, Gand, Belgique) Towards non-Leloir transarabinofuranosidases using a molecular evolution strategy applied to a GH51. B. Bissaro, **J. Durand**, X. Biarnés, P. Monsan, A. Planas, R. Fauré, M.J. O'Donohue

Communications par poster

Durand, J., Watterlot, L., Bonzom, C., Borsenberger, V., Bozonnet, S., Fauré, R., and O'Donohue, M.J. Enzymatic synthesis of cellobiose lipids by engineering of a type II endo glycoceramidase Background and challenges. Novel Enzymes 2014, Ghent, Belgique.

Durand, J., Chaumeton, V., Gherbovet, O., Ragon, M., Jütten, P., LA Bras, J., IPD Guerreiro, C., MGA Fontes, C., Piechot, A., Fauré, R., Bozonnet, S., O'Donohue, MJ. Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance. BIOCAT 2016, Hambourg, Allemagne.

J. Durand, V. Chaumeton, O. Gherbovet, M. Ragon, P. Jütten, J.L.A. Bras, C.I.P.D. Guerreiro, C.M.G.A. Fontes, A. Piechot, R. Fauré, S. Bozonnet, M.J. O'Donohue. Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance. Carbohydrate Bioengineering Meeting 2017, Vienne, Autriche.

J. Durand, V. Chaumeton, O. Gherbovet, M. Ragon, P. Jütten, J.L.A. Bras, C.I.P.D. Guerreiro, C.M.G.A. Fontes, A. Piechot, R. Fauré, S. Bozonnet, M.J. O'Donohue. Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance. International Conference on Novel Enzymes 2016, Groningue, Pays-Bas.

Remerciements

Ces quelques lignes ont pour objet de remercier toutes les personnes avec qui j'ai travaillé, collaboré, déprimé parfois, rit beaucoup ou refait le monde autour d'une bière, bref toutes les personnes avec qui j'ai traversé ces années si pleines et intenses de la vie de thésard.

Je voudrais en premier lieu remercier les membres du jury, Uwe Bornscheuer, Jean-Guy Berrin, Charles Tellier, Sylvie Armand et Régis Fauré qui m'ont fait l'honneur d'évaluer ce travail.

Pendant ces années de thèse j'ai eu la chance d'évoluer dans l'équipe pluridisciplinaire, professionnelle et attachante des Lignos, dirigée jusqu'à récemment par Michael O'Donohue. Michael, vous êtes un manager incroyable, respecté par tous pour vos connaissances, votre professionnalisme et votre éthique irréprochable. Nous avons eu une chance énorme de vous avoir à nos côtés et je suis heureux d'avoir pu bénéficier de votre vision si globale et aiguisée des biotechnologies.

Merci mille fois à toi Sophie, pour m'avoir fait confiance et pour tout ce que tu m'as appris. Ton intégrité et ta rigueur scientifique resteront mes modèles pour la suite. Je suis très fier d'avoir été ton premier thésard ! Régis, tu as été le premier à m'accorder ta confiance lorsque je suis arrivé au laboratoire pour mon stage de M2. Depuis j'ai eu la chance de travailler avec toi tout au long de cette thèse et de profiter de tes lumières sur les obscurs mystères de la chimie des sucres !

Un grand merci à l'équipe des permanents Lignos, Cédric, ta rigueur scientifique n'a d'égal que tes jeux de mots lors de nos réunions ligno. Claire, nouvelle boss en charge de manager l'équipe, vous côtoyer fut un vrai plaisir.

Passons à l'équipe des *capullos* qui ont jalonné la lignocellulase team durant cette thèse : Bastiani, Pablito, Greg *kikikiii*, Thomas, Pablo *le break danseur*, Justin. Vous avez été ma plus grande surprise lors de cette thèse. Je m'attendais à trouver des collègues j'ai trouvé des amis. Combien de fois nos discussions "what's app" m'ont fait pleurer de rire. Vous êtes des gars bien et j'espère que l'on continuera à bien se marrer ! Sayani merci pour ta gentillesse et tes supers soirées de « recipes exchange » qui nous ont fait voyager les papilles. Merci mille fois à Olga, je sais que nous resterons en contact (et je sais tout ce que je te dois !). Un grand merci à Eleni, Roose, Barbara, Louise, Marjorie, Cyrielle, Laurie, Hayang, Jiao, Franck, Zhongpeng et Victor. Tiens Victor, "last but not the least", j'ai eu beaucoup de chance

de t'avoir avec moi ces quelques mois, tu as fourni un super boulot et je te souhaite de prendre beaucoup de plaisir dans la suite de ton parcours.

Merci à l'équipe du CEMES : Marlène, Akli, Mounir, Clément, Virginie, Sayani, Alvaro, Jeremy, David... notre petit break quotidien était toujours un moment de détente et de convivialité. Merci aux gens qui ont travaillé dans notre petit bureau et qui ont contribué à en faire un lieu toujours agréable : Yannick, Virginie, Mégane, Romain, Lisa, Marion S., Manon M., Justin, Victor...

Je suis reconnaissant à tous les voisins et voisines de paillasse qui ont contribué à faire de ces 4 années un formidable souvenir : Seydou, Alvaro, Marlène, Yvonne, Vinciane, Alexandra, Marion C, Davide, Marco polo, Pauline, Angeline, Les Pibracais Etienne et Laurence, Les belles blondes Manon M et Laure, Mathilde, Sandra, Marie, Betty...Je pourrais écrire un paragraphe sur chacun de vous mais vous comprendrez qu'après ces quelques pages de manuscrit je préfère largement vous dire combien je vous kiffe de vive voix autour d'une bière par exemple ! Une pensée particulière à Akli et Yannick que j'ai suivi plus volontiers dans les poses clope que pour les footings. Nos discussions étaient de vrais moments de plaisir.

Je souhaite exprimer toute ma gratitude et mon immense respect à Monsieur Satgé et, à travers vous, à tous ces enseignants militants qui donne une seconde chance à des adultes passionnés de science. Vous avez contribué à changer ma vie et celle de bien d'autres.

Il est important pour moi de profiter de ces quelques lignes pour remercier aussi ma famille, votre soutien a été plus que précieux. Un énorme merci à mes tantes Joelle et Martine, si différentes mais avec la même bienveillance. Vanessa à qui je dois énormément et pour qui j'ai un immense respect. Papa, merci pour tout, j'essaierai d'être un aussi bon père et papi que tu l'es pour nous tous. JB, Manou, ces dernières années m'ont confirmé que j'avais un frère et une sœur en or même si je n'en doutai pas. Je vais finir en remerciant mes enfants, Rose et Raphaël, vous êtes formidables ! Vous lirez sûrement ces lignes, dans quelques années, peut être en tombant par hasard sur ce manuscrit. Rose, tu es arrivée au début de cette thèse et ta façon de vivre chaque jour comme une fête fait de toi l'être le plus attachant du monde. Quant à toi Raphaël, je ne serais pas surpris que ta curiosité naturelle sur le monde qui t'entoure te conduise devant un microscope ou un télescope dans quelques années !

Enfin, je dédie cette thèse à ma mère, ton courage et ton humour resteront mes plus belles leçons de vie.

Table des matières

Introduction	17
Chapitre I : Introduction bibliographique.....	21
1. La chimie verte et ses 12 principes	24
a. La catalyse	25
b. La catalyse chimique	25
c. La biocatalyse	26
2. Les enzymes, nouveaux outils industriels verts	26
a. Classification des enzymes	28
b. Promiscuité enzymatique	28
c. Ingénierie Enzymatique	30
d. Evolution dirigée.....	30
e. Evolution rationnelle ou semi-rationnelle.....	41
f. Production, criblage et sélection des enzymes.....	43
3. Sources de carbone renouvelable pour la chimie verte.....	51
a. Bilan carbone	51
b. Origine du carbone renouvelable	52
c. Déconstruction de la biomasse	57
4. Synthèse enzymatique de molécules à forte valeur ajoutée issues de la biomasse.....	58
a. Les transférases.....	58
b. Les hydrolases pour la synthèse de molécules.....	62
5. Objectifs de la thèse.....	68
Références :	70
Chapitre II : Modification du ratio transglycosylation/hydrolyse d'une endoglycocéramidase afin d'améliorer son activité de synthèse.....	83
1. Postulat pour une approche rationnelle de la conversion de GH en TG	85
2. Exemple de la <i>TxAbf</i>	86
3. Propriétés, applications et production des polyglycosides d'alkyles (APG)	87

4.	Conversion d'une endoglycocéramidase de type II (EGC II) en transglycosylase.....	88
a.	Trois familles d'EGC et leurs spécificités de substrats.....	89
b.	Choix de l'enzyme modèle	90
5.	Stratégie utilisée pour la conversion de l'EGC II en TG	91
	Références :	94
	A Single Point Mutation Alters the Transglycosylation/Hydrolysis Partition, Significantly Enhancing the Synthetic Capability of an Endo-Glycoceramidase	97
	ABSTRACT:	99
	INTRODUCTION	101
	EXPERIMENTAL SECTION	104
	RESULTS	112
	DISCUSSION	120
	CONCLUSION.....	123
	REFERENCES	124
	Supporting Information.....	128
	Chapitre III : Evolution dirigée d'une féruloyle estérase pour améliorer sa thermostabilité et sa résistance aux solvants	137
1.	Température et solvants pour la biocatalyse enzymatique.....	139
a.	Enzymes et solvants.....	139
b.	Enzymes et température.....	142
c.	Lien entre thermostabilité et résistance aux solvants ?	144
2.	Les feruloyle estérases dans les biotechnologies.....	144
a.	Mesure d'activité des féruloyle estérases	146
b.	La féruloyle estérase A d' <i>Aspergillus niger</i> (<i>AnFaeA</i>).....	147
3.	Stratégie utilisée pour l'évolution de <i>AnFaeA</i>	147
	Références :	149
	Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance	135
	ABSTRACT:	157

INTRODUCTION	158
EXPERIMENTAL SECTION.....	160
RESULTS	167
DISCUSSION.....	173
REFERENCES	180
Supporting Information.....	185
Chapitre IV : Conclusions-Perspectives.....	153
1. Une nouvelle transglycosylase pour la synthèse de polyglucosides d'alkyle.....	193
2. Une feruloyle estérase plus résistante à la température et aux solvants.....	196
3. Vers une féruloyle estérase améliorée pour la synthèse de molécules d'intérêt ?	199
Références :	203
ANNEXE 1 :	189
Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities	189
ANNEXE 2 :	241
Regioselective chemoenzymatic syntheses of chromogenic substrates for feruloyl esterases	241
(en préparation)	241

Définitions et Abréviations :

Définitions :

Biotechnologie blanche : procédés de fabrication de molécules chimiques par l'intermédiaire de microorganismes ou d'enzymes. Ce domaine trouve ses racines dans l'Antiquité, avec l'utilisation des levures pour l'élaboration de bières et de vins.

Céramide : sphingolipide résultant de la combinaison entre un acide gras et une sphingosine, alcool gras de type aminodiol éthylénique, via une liaison amide.

Chiralité : une molécule est dite chirale quand elle n'a aucun plan de symétrie ; se dit d'une molécule, non superposable à son image dans un miroir.

Enantiosélectivité : propriété d'une réaction au cours de laquelle sont obtenus, à partir d'un précurseur achiral*, deux énantiomères d'un produit chiral* en quantités inégales. *(voir *Chiralité*).

Extrêmophile : se dit d'un organisme dont les conditions de vie normales - extrêmes - sont mortelles pour la plupart des autres organismes : température optimale de croissance comprise entre 75 et 100 °C (hyperthermophile), entre 50 et 75 °C (thermophile) ou inférieures à 0 °C (psychrophile), pression exceptionnelle (barophile), milieu très chargé en sel (halophile), milieu très acide (acidophile) ou hyper-alcalin (alcalophile).

Gangliosides : glycoconjugués constitués d'un céramide lié par une liaison éther à un motif saccharidique chargé. Les gangliosides contiennent au minimum deux résidus osidiques dont un, au minimum, est lié à une molécule d'acide sialique (ou acide *N*-acétylneuraminique) par une liaison éther.

Glycoconjugués : composés constitués de glucides liés de manière covalente avec d'autres types moléculaires : polypeptides, protéines ou lipides.

Holistique : qui relève de l'holisme, qui s'intéresse à son objet comme constituant un tout.

Homodimère : en biologie des protéines, ce terme fait référence à une protéine dimérique composée de deux unités protéiques identiques.

Lyoprotecteur : molécule ayant un rôle protecteur lors de la lyophilisation.

Métagénomique : étude du contenu génétique d'échantillons issus d'environnements complexes prélevés dans la nature. La métagénomique vient après la génomique en étudiant les organismes directement dans leur environnement sans passer par une étape de culture en laboratoire.

Régiosélectivité : qualité d'une réaction qui agit sur un site particulier d'une molécule.

Stéréosélectivité : qualité d'une réaction qui se caractérise par la formation prédominante d'un produit à configuration stéréochimique particulière au dépend des autres.

Transition : en biologie moléculaire, remplacement d'une purine (Adénine ou Guanine) par une purine ou d'une pyrimidine (Thymine ou Cytosine) par une pyrimidine.

Transversion : en biologie moléculaire, remplacement d'une purine par une pyrimidine ou d'une pyrimidine par une purine.

Abréviations :

APG : Alkylpolyglycoside

CAZy : Carbohydrate-Active enZymes database

CCM : Chromatographie sur Couche Mince

E_a : Energie d'activation

EGC : Endoglycocéramidase

epPCR : error prone Polymerase Chain Reaction

FACS : Fluorescence-Activated Cell Sorting

FRET : Fluorescence Resonance Energy Transfer

GC : Gaz Chromatography (Chromatographie gaz)

GH : Glycoside Hydrolase

GSL : Glycosphingolipides

HPLC : High Performance Liquid Chromatography (Chromatographie en phase liquide à haute performance)

HTS : High Throughput Screening

ITCHY : Incremental Truncation for the Creation of Hybrid enzyme

RMN : Résonance Magnétique Nucléaire

SCOPE : Structure based COmbinatorial Protein Engineering

StEP : Staggered Extension Process

T/H : Ratio transglycosylation/hydrolyse

TG : Transglycosidase

T_m : melting temperature (température de fusion)

Introduction

La **Chimie Verte** se propose de relever un des défis majeurs de notre époque : réconcilier la Nature avec les procédés chimiques industriels. Ce concept a été clairement explicité en 1998 par les chimistes américains Paul Anastas et John C. Warner et décliné en 12 principes majeurs (Anastas and Warner, 2000), dont les objectifs sont la réduction -voire l'élimination- de l'utilisation et de la génération de substances dangereuses pour l'homme ou l'environnement, au cours des procédés de production des produits chimiques, dans un souci de diminution des risques sanitaires à la source. Inspirée du concept de « **développement durable** », elle vise un équilibre économique, social et environnemental du milieu dans lequel elle s'insère. Sont concernés les domaines de la chimie, de la biochimie, de la production ou transformation des matériaux et de l'énergie, pour lesquels il faudra alors, dans cette même optique, veiller à **substituer tout ou partie de la ressource fossile utilisée par une alternative renouvelable** (Colonna, 2012). De nouvelles solutions seront donc nécessaires pour **économiser à la fois l'énergie**, par exemple à travers le développement de technologies vertes (énergie éolienne, solaire, hydraulique), **et les matières premières carbonées**, notamment à travers l'exploitation de la biomasse, et en particulier de la biomasse végétale.

Dans ce contexte, l'utilisation de **réactions catalysées**, c'est-à-dire qui bénéficient de la présence d'un élément, en faible concentration dans le milieu, qui permet d'augmenter la vitesse d'une réaction chimique ou biochimique pour la rendre possible sans être modifié lui-même, représente un domaine de recherche particulièrement pertinent. A ce titre, les **enzymes** constituent des alternatives aux catalyseurs chimiques toxiques et polluants et sont depuis longtemps reconnues comme des **outils extraordinairement efficaces pour les transformations biologiques et chimiques**. C'est dans ce contexte que s'inscrivent les travaux décrits dans cette thèse, axés sur **l'ingénierie d'outils enzymatiques plus performants** à destination de procédés plus respectueux de l'environnement et économiquement viables.

Le chapitre I de ce manuscrit propose un tour d'horizon bibliographique de l'état actuel des connaissances afin de situer la problématique des travaux développés, notamment dans le contexte de ce concept de Chimie Verte, abordé dans la première section. Une deuxième partie est consacrée à l'utilisation des enzymes et des options disponibles pour l'optimisation de leurs performances dans le cadre de la transformation de la biomasse végétale, une source de carbone renouvelable pour la chimie verte dont la diversité est détaillée ensuite. Enfin, dans la mesure où les travaux de cette thèse concernent l'optimisation d'enzymes pour la **synthèse** de molécules, une section est consacrée aux enzymes spécifiquement dédiées à ces fonctions.

Le chapitre II concerne la conversion d'une glycoside hydrolase (GH), l'endoglycocéramidase II (EGC II) de *Rhodococcus* sp. en biocatalyseur pour la synthèse de potentiels nouveaux biosurfactants. Dans le cadre du projet européen BIOSURF (Novel Production Strategies for Biosurfactants) qui

visait à trouver de nouvelles voies de synthèse enzymatique de molécules tensioactives, alternatives aux méthodes issues de l'industrie pétrochimique, nous avons travaillé à la conversion de l'activité hydrolytique de cette enzyme en activité de transfert permettant la synthèse de polyglycosides d'alkyles (APG). Nous avons choisi d'adapter une méthode développée par l'équipe de Charles Tellier et adaptée précédemment avec succès au laboratoire sur une arabinofuranosidase (*TxAbf*). Cette stratégie, qui consiste à modifier les barrières énergétiques de la réaction au profit d'une activité de synthèse par des changements subtils du site actif a donc été utilisée sur une nouvelle famille de GH avec l'objectif de valider cette approche d'une part et bien sûr, de créer ainsi un mutant optimisé pour la synthèse d'APG.

Dans le **chapitre III**, nous avons utilisé une approche d'évolution dirigée dans le but d'augmenter la résistance à la température et aux solvants de la féruloyle estérase A d'*Aspergillus niger*, des préalables nécessaires pour pouvoir envisager d'utiliser cette enzyme pour des réactions de synthèse de molécules féruloylées ayant potentiellement des propriétés antioxydantes, un des objectifs affichés du projet européen OPTIBIOCAT (OPTImized esterase BIOCATalysts for cost-effective industrial production) dans lequel s'est inscrit cette recherche. Cette démarche participe ici aussi à la mise en place de bioprocédés efficaces, avec un impact environnemental plus doux que les procédés chimiques actuellement utilisés.

Enfin, cette thèse se conclut sur le **chapitre IV** qui fait le bilan des expériences menées et des résultats clés obtenus sur chaque projet, et propose des pistes de discussion et des perspectives pour compléter cette étude.

Les **Annexes** regroupent deux publications, qui concernent la conception, la synthèse, la caractérisation et la mise en œuvre de substrats chromogéniques pour la détection et la mesure d'activité féruloyle estérase.

Références

- Anastas, P.T., Warner, J.C., 2000. Green chemistry: theory and practice, 1. paperback. ed. Oxford Univ. Press, Oxford.
- Colonna, P., 2012. Le carbone renouvelable dans les systèmes alimentaires, énergétiques et chimiques : Leçon inaugurale prononcée le jeudi 15 décembre 2011. Collège de France.

Chapitre I : Introduction bibliographique

La fin du 20^{ème} siècle a été marquée par le double constat que les ressources fossiles s'épuisent peu à peu car non renouvelables et que leur utilisation démesurée est devenue une source considérable de pollution pour l'Homme et pour la nature. Outre leur épuisement programmé à court terme, l'industrie pétrolière provoque des dégâts considérables sur les sols, l'eau et l'air par des rejets de molécules toxiques et de gaz à effets de serre. Les risques d'accidents sont aussi importants : explosions, marées noires, ruptures d'oléoducs, incendies, etc.

Dans ces conditions, notre société doit trouver des solutions pour réduire sa dépendance vis-à-vis des ressources fossiles et principalement du pétrole. Pour trouver des procédés industriels offrant un moindre impact environnemental, de gros efforts d'innovation et de recherche ont été réalisés depuis quelques années. Dans cette optique, la biomasse végétale issue de la photosynthèse focalise beaucoup d'attentions car elle constitue une immense ressource de carbone renouvelable qui pourrait offrir une alternative -au moins partielle- à l'utilisation des ressources fossiles. Cette biomasse carbonée peut être utilisée pour la production d'énergie (Cotana et al., 2016; Valdivia et al., 2016) et pour la synthèse de nombreux produits qui sont à ce jour principalement issus de la pétrochimie tels que les tensioactifs ou les lubrifiants. Le secteur des biomatériaux se tourne également vers l'utilisation du carbone renouvelable (Ten and Vermerris, 2013; Snell et al., 2015) : ainsi la loi du 17 août 2015 relative à la transition énergétique pour la croissance verte, dit dans son article 5 que "l'utilisation des matériaux biosourcés concourt significativement au stockage de carbone atmosphérique et à la préservation des ressources naturelles". En effet, seule une faible partie des cultures alimentaires est réellement commercialisée, le reste des plantes est souvent considéré comme du déchet et reste inutilisé. Cette biomasse agricole ainsi que les végétaux au sens large comme les cultures dédiées à la production de matière organique non alimentaire, les résidus de l'industrie du bois, ou encore les déchets organiques issus des ordures ménagères représentent une ressource majeure de carbone renouvelable (Ragauskas et al., 2006). Des millions de tonnes de déchets végétaux pourraient ainsi être exploités chaque année sans porter préjudice à la biodiversité et sans épuiser les sols.

Cette biomasse renouvelable composée principalement d'hydrates de carbone possède une structure bien plus complexe (diversité de structures, de liaisons et d'interactions) (Cosgrove, 2005) que les hydrocarbures fossilisés. Mais cette complexité et cette diversité de structures offrent aussi un atout majeur pour la synthèse de nombreuses molécules complexes à haute valeur ajoutée, tels que des prébiotiques ou des biomatériaux. Ainsi, les industries cosmétique (Talmaciu et al., 2015), chimique (Bugg and Rahmanpour, 2015) et pharmaceutique (Pachau, 2015) y trouvent autant une source d'innovation qu'une alternative à la pétrochimie.

Dans un même temps, l'utilisation de cette nouvelle source de carbone s'accompagne dans nos sociétés d'une prise de conscience sur la nécessité d'utiliser des processus de fabrication plus respectueux de l'environnement. C'est dans ce cadre qu'est né dans les années 1990 le concept de chimie verte ou chimie renouvelable ou encore chimie écologique.

1. La chimie verte et ses 12 principes

La chimie verte vise à réduire au maximum les composés présentant un danger pour l'homme et pour l'environnement. Elle s'appuie sur 12 principes fondamentaux (Figure 1), qui synthétisent les propositions d'un livre intitulé "Green chemistry: theory and practice" (Anastas and Warner, 2000).



Figure 1 : Les douze principes de la chimie verte

Ces 12 principes reflètent les préoccupations modernes que les biochimistes doivent s'imposer tout au long de la mise en place de nouveaux procédés industriels visant à la synthèse de molécules. Trois de ces principes semblent essentiels pour répondre au défi d'une chimie plus responsable et soucieuse de son impact environnemental :

Principe 3. *Une conception de synthèses moins dangereuses grâce à l'utilisation de conditions douces et la préparation de produits peu ou pas toxiques pour l'homme et l'environnement.*

En effet il n'est plus raisonnable de se soucier uniquement des voies de synthèse et des rendements finaux. Les réactions doivent générer un minimum de déchets toxiques pour l'homme et l'environnement.

Principe 7. *Utiliser une ressource naturelle ou une matière première renouvelable plutôt que des produits fossiles, dans la mesure où la technique et l'économie le permettent.*

Une grande majorité des atomes de carbone utilisés industriellement par les chimistes organiciens provient actuellement du pétrole et est donc épuisable. En revanche, le carbone renouvelable offert par les organismes vivants tels que les végétaux est une option qui doit être privilégiée.

Principe 9. *L'utilisation d'agents catalytiques (aussi sélectifs que possible) est préférable à celle de procédés stœchiométriques.*

L'utilisation des procédés catalytiques où le catalyseur n'est pas modifié et est donc réutilisable "indéfiniment", est donc à privilégier face aux procédés stœchiométriques où une molécule est consommée par réaction, ce qui génère des produits secondaires à retraiter. L'ibuprofène, par exemple, un analgésique très répandu, était produit industriellement depuis les années 1960 à partir de 6 réactions stœchiométriques. Ce procédé a permis de générer annuellement des milliers de tonnes d'ibuprofène mais, également une quantité encore plus importante de sous-produits non recyclés qu'il a fallu détruire ou retraiter. Cette gabegie a duré jusque dans les années 1990 où la société BHC proposa une synthèse en trois étapes catalysées. En plus d'être un procédé plus rapide et moins onéreux, l'unique sous-produit, l'acide éthanoïque, étant réutilisé, la quantité de déchets à retraiter est devenue nulle. Il est donc pertinent de bien maîtriser les différentes méthodes qui s'offrent à nous pour catalyser une réaction.

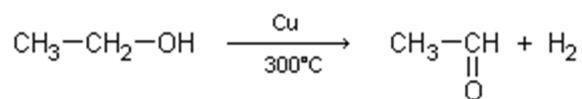
a. La catalyse

La catalyse fait référence à la modification, généralement à l'accélération de la cinétique d'une réaction chimique par un catalyseur. La plupart des réactions, pour avoir lieu, nécessitent un apport d'énergie appelé : 'énergie d'activation' (E_a). Le catalyseur peut être solide, liquide ou gazeux. Il n'apparaît pas dans l'équation globale de la réaction (contrairement aux réactifs et aux produits) mais est impliqué dans au moins une étape de la réaction pour en modifier la vitesse. Les réactions catalytiques sont dites 'homogènes' si le catalyseur est soluble dans le milieu ou, 'hétérogènes' si le milieu présente deux phases différentes. Comme explicité ci-après, une réaction peut être catalysée par un catalyseur chimique ou biologique. Le choix du catalyseur dépend des propriétés de la réaction et des composés (Ager, 2012). Il est donc important de comparer les différents types de catalyseurs disponibles pour utiliser le catalyseur le mieux approprié.

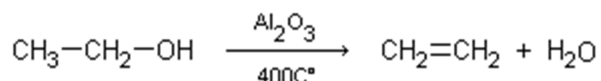
b. La catalyse chimique

La catalyse chimique consiste à utiliser une espèce chimique pour augmenter la vitesse d'une réaction. En effet, la présence d'acide ou d'ions métalliques dans un milieu peut fortement faire diminuer l' E_a nécessaire à la réaction et donc augmenter la vitesse de la réaction. Une réaction peut évoluer différemment selon le catalyseur utilisé, qui peut ainsi orienter une transformation chimique. Par

exemple l'évolution de l'éthanol lors de son chauffage dépend du catalyseur utilisé. En présence de cuivre, l'éthanol subit une déshydrogénation pour donner de l'éthanal,



alors qu'en présence d'alumine (oxyde d'aluminium Al_2O_3) il subit une déshydratation pour donner de l'éthylène.



Les catalyseurs chimiques sont devenus courants car ils sont bien appropriés pour la synthèse de molécules simples issues de la pétrochimie. Par contre, la complexité des molécules provenant de la biomasse végétale nécessite des catalyseurs plus spécifiques qui peuvent se trouver parmi les organismes vivants qui construisent ou déconstruisent ces structures complexes.

c. La biocatalyse

Le terme de biocatalyse est employé lorsque le catalyseur est, ou provient d'un organisme vivant tel qu'une enzyme (biocatalyseur). Dans le passé, la plupart des réactions de synthèse étaient réalisées par des chimistes peu sensibilisés à la biocatalyse. L'utilisation de biocatalyseurs était donc la solution choisie en dernier recours. Pourtant, les Hommes utilisent depuis des siècles des organismes naturels pour catalyser des réactions : fermentation alcoolique des sucres utilisée par les Sumériens en Mésopotamie, panification des farines apparue en Égypte antique, etc. (Tonnac, 2010).

La biocatalyse utilise donc des ressources du vivant capables d'abaisser l'énergie d'activation d'une réaction chimique et va indéniablement dans le sens du développement durable ; de nombreux industriels se tournent donc vers elle car elle offre des procédés à la fois compétitifs et plus respectueux de l'environnement. Un des avantages majeurs à utiliser des biocatalyseurs pour les industriels réside dans le fait qu'il n'est souvent plus nécessaire d'utiliser de hautes températures pour réaliser les réactions, ces catalyseurs naturels étant généralement actifs à température ambiante. Par exemple, la plupart des lessives contenant des enzymes sont à présent très efficaces à basses ou moyennes températures (30 °C). L'industriel peut ainsi réduire le coût énergétique et donc les coûts de production. De plus, l'utilisation de solvant est très significativement réduite car la plupart des micro-organismes sont actifs en milieu aqueux.

2. Les enzymes, nouveaux outils industriels verts

Le mot enzyme vient du Grec $\acute{\epsilon}\nu\zeta\upsilon\mu\omicron$ (inzymo) ($\acute{\epsilon}\nu$ en, dans, et $\zeta\upsilon\mu\eta$ levure, levain) "dans la levure", il est proposé par Kühne en 1878 pour faire la distinction entre les micro-organismes entiers et les

substances excrétées par les micro-organismes. Le principe de la catalyse enzymatique s'appuie sur le modèle "clé-serrure" proposée en 1894 par E. Fischer : l'enzyme catalyse une réaction sur un substrat spécifique (la clé) qui se positionne dans l'enzyme (la serrure) avec une complémentarité stérique. Le substrat est alors modifié et conduit au produit (Kobayashi and Ohmae, 2006). Ce n'est que vers la fin des années 1920 que les enzymes sont identifiées comme étant des protéines.

Ces protéines catalytiques peuvent accélérer fabuleusement la vitesse d'une réaction, jusqu'à 10^{19} fois (Bar-Even et al., 2011). Un autre de leurs avantages est qu'elles peuvent catalyser des réactions régio- et stéréo-sélectives. Cette sélectivité permet de s'affranchir des étapes de "protection/déprotection" souvent nécessaires dans les procédés chimiques traditionnels. Dans un contexte de synthèse chimio-enzymatique, l'emploi d'enzymes réduit ainsi le nombre d'étapes nécessaires, limite les déchets et réduit le coût énergétique du procédé. Ainsi, dans le cas concret des sucres, l'incroyable diversité des monosaccharides est due aux carbones asymétriques qui les composent : plus la chaîne carbonée est longue, plus le nombre de stéréoisomères est important. La complexité s'accroît encore pour les polysaccharides en prenant en compte la diversité des types de liaisons (stéréo et régio-spécifiques). Un hexasaccharide réducteur possède ainsi plus de 10^{12} isomères différents (Laine, 1994). Il faut donc utiliser des outils adaptés à cette complexité et seules certaines enzymes peuvent fournir des outils avec la spécificité nécessaire pour la synthèse ou pour l'hydrolyse "contrôlée" de ce type de substrat.

Depuis les années 1970, les énormes progrès concernant la production d'enzymes recombinantes en fermenteur a fortement fait baisser leurs coûts de production, qui sont donc devenus de plus en plus compétitifs pour des réactions à l'échelle industrielle (Reetz, 2013). Pour ces raisons, la catalyse enzymatique en chimie organique est en plein essor. Un des exemples les plus significatifs est peut-être la synthèse de près de 10 000 tonnes de pénicilline par an par voie enzymatique. La penicillin acylase est active en milieu aqueux et à température ambiante, son intégration dans le processus de synthèse a permis d'éviter une étape qui combinait l'utilisation d'un groupement protecteur (triméthylsilyle) et une réaction se déroulant à -40 °C (Batchelor et al., 1959). Les enzymes sont à ce jour principalement utilisées dans l'industrie des détergents, de l'amidon, de l'agroalimentaire humaine et animale, de la chimie fine ou encore de la santé (Combes and Monsan, 2009).

Ces dernières années, grâce au potentiel énorme que représentent ces biocatalyseurs, le nombre d'enzymes disponibles s'est accru exponentiellement et par voie de conséquence, la diversité des réactions disponibles aussi, ceci grâce d'une part à la découverte de nouvelles enzymes lors de criblage à haut-débit de banques métagénomiques par exemple (Uchiyama and Miyazaki, 2009) et d'autre part à l'ingénierie enzymatique dont nous allons présenter différentes approches plus loin. Il faut toutefois garder à l'esprit que l'énorme majorité des enzymes industrielles est d'origine microbienne, majoritairement fongique, et que seule une infime fraction des micro-organismes vivants sur notre planète a pour l'instant été identifiée (National Research Council (U.S.) and National Academies Press (U.S.), 2007). Le potentiel d'outils catalytiques qui reste à découvrir au sein de la

biodiversité est donc immense. Néanmoins, il est important de pouvoir classifier toutes ces enzymes connues et à venir.

a. Classification des enzymes

L'Enzyme Commission numbers (EC) est une classification numérique basée sur les réactions chimiques que catalysent les enzymes, et qui régit la nomenclature appliquée à celles-ci. Pour identifier une enzyme, le sigle EC est suivi de quatre nombres (EC X.X.X.X.). Ces quatre nombres donnent chacun une information de plus en plus spécifique sur la réaction catalysée. Le premier nombre, indique l'appartenance de l'enzyme à l'une des six grandes familles de réaction.

EC 1	Oxidoréductases
EC 2	Transférases
EC 3	Hydrolases
EC 4	Lyases
EC 5	Isomérases
EC 6	Ligases

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)
<http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2/>

Le second indique la famille du substrat ou le groupement fonctionnel sur lequel l'enzyme est active. Le troisième indique spécifiquement le type de réaction et le quatrième le numéro de série propre à l'enzyme. Cette classification simple est devenue la référence pour nommer les enzymes. L'inconvénient de ce type de classification est qu'il ne tient pas compte des similarités de séquence primaire ou de la structure des enzymes. Cela implique que deux enzymes structurellement très proches mais dont l'évolution aura modifié la spécificité de substrat respectif appartiendront à des familles différentes. Inversement, des enzymes très éloignées phylogénétiquement se retrouvent dans une même famille. Cette classification pose aussi la question des enzymes à substrats multiples. Pour ces raisons, plusieurs autres bases de données ont été créées utilisant des critères de classification différents. Chacune de ces classifications permet d'obtenir des informations différentes et complémentaires sur une enzyme.

b. Promiscuité enzymatique

Lorsqu'une enzyme est active sur plusieurs substrats, on parle alors de promiscuité de substrat. Il est ici important de faire la différence avec les deux autres grands types de promiscuité enzymatique proposés par Hult et Berglund : i) la promiscuité de conditions de réactions (activité enzymatique dans diverses conditions réactionnelles différentes de celles naturelles), et ii) la promiscuité catalytique qui est basée sur la capacité d'un site actif unique à catalyser plusieurs transformations chimiques (Bornscheuer and Kazlauskas, 2004), comme les pyruvate décarboxylases de certaines levures (*Saccharomyces*, *Candida*), qui non seulement décarboxylent le pyruvate mais lient également l'acétaldéhyde et le benzaldéhyde (par une activité lyase) pour former le (R)-phénylacétylcarbinol,

précurseur pour la synthèse de l'éphédrine (Ward and Singh, 2000). Cette dernière est à nouveau scindée en deux catégories. La promiscuité catalytique "accidentelle", présente sur des enzymes naturelles et la promiscuité catalytique dite "induite" qui apparait suite à des manipulations délibérées pour modifier le type de réaction catalysée par l'enzyme sauvage (Hult and Berglund, 2007).

Dans certains cas, la promiscuité de substrat des enzymes permet l'utilisation de molécules non naturelles qui miment les substrats naturels, principalement au niveau du site actif de l'enzyme (Figure 2). Les substrats synthétiques sont parfois nécessaires pour quantifier et caractériser la capacité d'une enzyme à catalyser une réaction, également lorsque le substrat naturel est instable, inconnu ou encore pour la synthèse de nouvelles molécules (Kobayashi and Makino, 2009). L'efficacité catalytique d'une enzyme peut varier fortement avec des substrats différents et les résultats sont alors difficilement comparables. Le substrat permettant l'activité optimale de l'enzyme n'est pas forcément le mieux adapté pour une réaction, il faut considérer sa disponibilité, sa stabilité, sa solubilité et, si nécessaire, l'accessibilité de la réaction à une méthode de détection pour faciliter le suivi d'une réaction (Bisswanger, 2014). Il est évident que pour la comparaison des résultats, le même substrat doit toujours être utilisé.

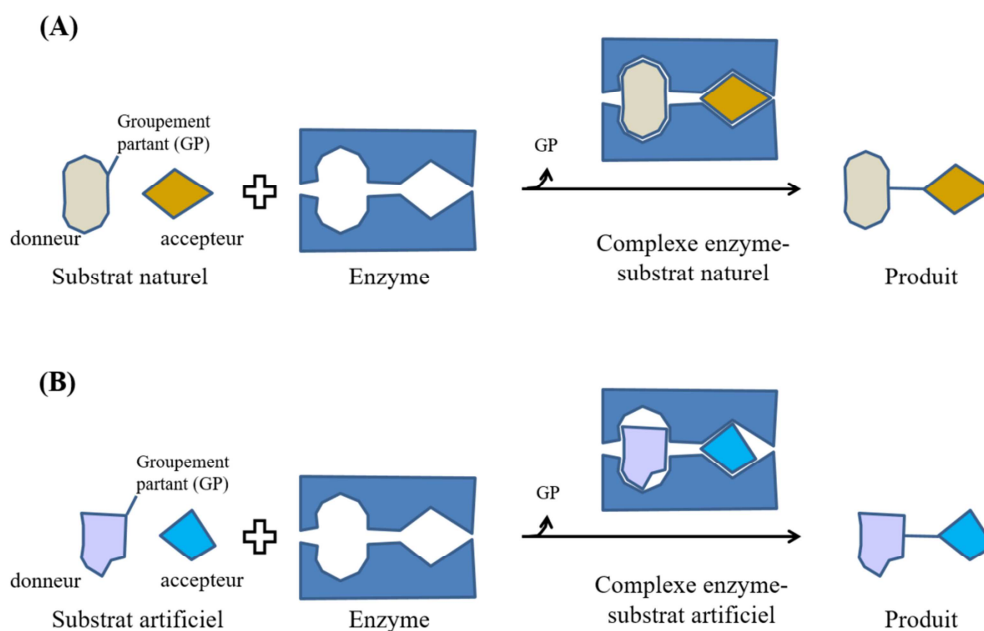


Figure 2 : **Diversité des complexes enzyme-substrat pour des réactions de synthèse enzymatiques** : (A) modèle d'une réaction obéissant à un modèle "clé – serrure" (B) réaction *in vitro* impliquant un complexe enzyme-substrat artificiel menant à un produit avec formation de liaisons, le substrat artificiel possède une structure semblable à celle du substrat naturel au voisinage du site actif. D'après (Kobayashi and Makino, 2009)

Hormis pour les organismes extrémophiles, les enzymes ont généralement évolué pour avoir une activité optimale dans des conditions modérées, c'est à dire à température ambiante, dans des milieux

aqueux, à pH neutre... Cependant, certains procédés requièrent l'utilisation de conditions plus extrêmes parfois non adaptées aux enzymes. Pour répondre à cette attente, des techniques d'évolution dirigée d'enzymes se perfectionnent depuis les années 1990 et permettent aujourd'hui d'obtenir des enzymes améliorées capables de résister à des conditions beaucoup plus drastiques, ce qui permet d'augmenter significativement les rendements réactionnels ainsi que leurs stabilités (Lutz, 2010; Dalby, 2011; Bornscheuer et al., 2012; Kumar and Singh, 2013; Packer and Liu, 2015; Molina-Espeja et al., 2016). Lorsque l'on souhaite créer ou modifier les propriétés d'une enzyme, plusieurs techniques sont aujourd'hui disponibles au travers de l'ingénierie enzymatique.

c. Ingénierie Enzymatique

L'ingénierie enzymatique vise dans un premier temps à comprendre les mécanismes fondamentaux qui régissent au niveau moléculaire les réactions catalysées par ces protéines. Ces connaissances peuvent ensuite être mises à profit pour améliorer ou adapter ces biocatalyseurs à certaines fonctions ou conditions et répondre ainsi à des besoins précis. La fin du XX^e siècle a marqué l'explosion de l'ingénierie enzymatique, principalement à partir des années 1980 avec l'invention de la réaction de polymérisation en chaîne (PCR) qui permet de dupliquer et amplifier des fragments d'ADN (Saiki et al., 1985). La décennie suivante (1990) a vu apparaître de nombreuses techniques de biologie moléculaire et d'évolution dirigée qui ont permis d'obtenir des résultats spectaculaires (Willem P. Stemmer, 1994; W. P. Stemmer, 1994; Lutz, 2010; Kumar and Singh, 2013). Aujourd'hui, de nombreuses méthodes, généralement complémentaires, sont utilisées en laboratoire.

d. Evolution dirigée

L'évolution dirigée repose sur le principe Darwinien de l'évolution des espèces. Charles Darwin donne en 1859 cette définition de l'évolution naturelle : "*Transformation d'un organisme au cours des générations suite à l'altération du génotype et maintien du phénotype le mieux adapté par sélection naturelle.*" Autrement dit, des mutations se produisent naturellement au niveau génétique lors de la réplication de l'ADN et génèrent des variations de phénotypes. Certains nouveaux phénotypes, mieux adaptés à leur environnement, vont se transmettre à la descendance au dépend des anciens. C'est par ce mécanisme (très lent à l'échelle humaine) de sélection naturelle que les espèces s'adaptent à leur environnement. Ce processus d'évolution peut être reproduit de façon artificielle, en laboratoire, sur des micro-organismes ou sur une séquence d'ADN codant pour une protéine isolée en quelques étapes qui sont schématisées sur la Figure 3.

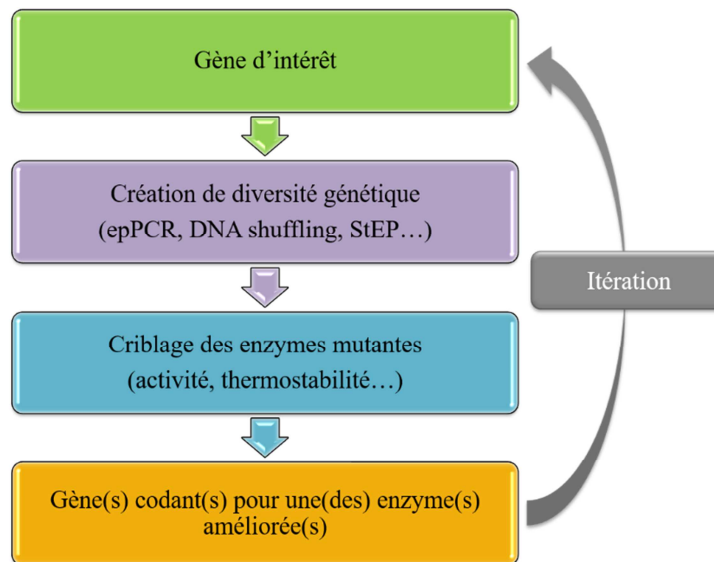


Figure 3 : **Principales étapes du processus d'évolution dirigée**

L'évolution dirigée repose toujours sur le principe de cycles répétés, composés de trois étapes clés : mutagenèse, criblage/sélection et choix du mutant amélioré, constituant le point de départ du cycle suivant (Arnold et al., 2001; Bloom and Arnold, 2009). Les travaux utilisant une approche d'évolution dirigée en laboratoire ont réellement débuté à la fin des années 1990 et la Figure 4 illustre clairement l'évolution au cours du temps du nombre de publications contenant les mots "évolution dirigée" dans le titre de l'article (figure réalisée en utilisant la base de données PubMed).

De façon plus détaillée, le principe repose sur le fait de générer artificiellement et de façon beaucoup plus fréquente que dans la nature, des mutations sur une séquence génétique. Cette diversité est alors soumise à une pression de sélection ou à un criblage afin de sélectionner un nouveau phénotype mieux adapté à la pression de sélection que l'enzyme de départ. Les conditions de laboratoire permettent de soumettre cette diversité à des conditions non présentes dans la nature et de conférer ainsi de nouvelles fonctionnalités qui n'auraient pas pu être mises à jour dans un milieu naturel (Molina-Espeja et al., 2016).

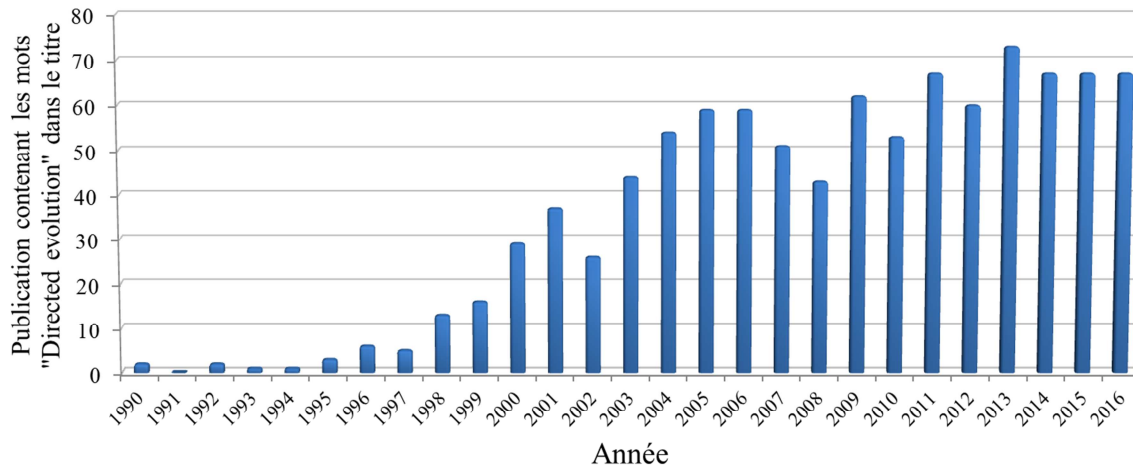


Figure 4 : **Evolution du nombre de publications contenant dans le titre les mots "directed evolution" dans la base de données PubMed (Modifié d'après (Pourmir and Johannes, 2012))**

Création aléatoire de diversité génétique

Différentes techniques de laboratoire permettent de générer artificiellement et de manière très rapide, une importante diversité génétique. Certaines de ces techniques sont applicables sur tout le génome d'un micro-organisme (*in vivo*), dans ce cas, l'organisme est soumis à des conditions mutagènes :

- liées à la présence d'agents chimiques tels que le méthanesulfonate d'éthyle, le méthyl-nitro-nitrosoguanidine, la *N*-nitroso-*N*-éthylurée qui provoque une méthylation sur les nucléotides guanine ou thymine, ou la methoxylamine qui se lie covalamment aux niveaux des résidus d'ADN endommagés. Ces modifications changent peu la conformation de l'ADN ce qui rend les modifications difficilement détectables par les systèmes de réparation. L'ADN polymérase par contre aura beaucoup plus tendance à commettre un mésappariement sur ces résidus modifiés lors de la réplication et donc à générer des mutations (Hoffmann, 1980; Gdaniec et al., 1996; Wyatt and Pittman, 2006).

- dues à une irradiation des micro-organismes par des rayons UV qui vont entraîner la formation de dimères de pyrimidine et par conséquent une torsion de la double hélice d'ADN qui va gêner la réplication et provoquer des mésappariements (Wong et al., 2006).

- ou permises par l'utilisation d'une souche commerciale génétiquement modifiée d'*E. coli*, nommée XL1-Red, délétée de trois voies principales de réparation de l'ADN (*mutS*, *mutT*, et *mutD*); cette souche va permettre un taux de mutations environ 5 000 fois plus important que sur la souche sauvage (Greener et al., 1997). Le taux de mutations souhaité peut être régulé par le nombre de cycles de croissance mais ce taux de mutations reste relativement faible comparé à d'autres méthodes, même après 24 heures de culture (Rasila et al., 2009). Cette souche a démontré son efficacité dans de

nombreux travaux d'évolution dirigée (Bornscheuer et al., 1998; Henke and Bornscheuer, 1999; Lu et al., 2001).

Les micro-organismes ainsi mutés sont cultivés dans un environnement comprenant une pression de sélection : pression, température ou pH extrême ou dans un milieu contenant un élément normalement toxique à la cellule, etc. Le micro-organisme va alors conserver les mutations, où qu'elles soient sur le génome, qui lui permettent de supporter ces conditions extrêmes. On peut noter que ces approches sont aussi utilisées pour l'évolution dirigée d'enzymes spécifiques. Ainsi, les plasmides contenant le gène d'intérêt vont se dupliquer dans la souche et des mutations vont potentiellement se répartir sur toute la longueur du plasmide. Les plasmides sont ensuite purifiés et utilisés pour transformer une souche de production et ainsi générer une banque d'enzymes mutantes (Greener et al., 1997).

De manière plus ciblée, des méthodes équivalentes sont aussi disponibles pour générer de la diversité aléatoire sur un fragment d'ADN précis, une séquence codant pour une enzyme par exemple. La principale technique pour obtenir de la diversité est la PCR à erreur (error-prone PCR, epPCR). L'epPCR mime des erreurs de réplication qui se produisent lors de l'évolution naturelle des protéines. Elle se base sur l'amplification d'une séquence codante comprise entre deux amorces nucléotidiques en incorporant, au cours des différents cycles de la PCR, des mutations aléatoires (contrairement à une PCR fidèle) qui se traduiront potentiellement par des variations d'acides aminés sur la séquence primaire de la protéine. Pour cela, l'ADN polymérase de *Thermophilus aquaticus* ou *Taq* polymérase, qui est dépourvue d'activité exonucléase et ne peut donc pas réparer les mésappariements, est généralement utilisée. Plusieurs paramètres permettent de contrôler le taux de mutations comme la quantité de matrice initiale par exemple ou le nombre de cycles d'amplification; il est aussi possible d'utiliser le produit d'une epPCR comme matrice pour une seconde réaction ou encore de modifier la composition du milieu réactionnel en déséquilibrant les concentrations des différents dNTP ou en utilisant des sels de manganèse au lieu du magnésium ou encore en utilisant des analogues nucléotidiques (Zaccolo et al., 1996).

Plus la diversité génétique de la banque de séquences codantes est importante, plus la probabilité de trouver des mutants intéressants est grande. Une des limites de cette approche est qu'il est impossible de générer et de cribler toutes les combinaisons possibles. En effet, l'espace de séquences pour une protéine de 100 acides aminés est de $20^{100} = 10^{130}$ possibilités en utilisant les 20 acides aminés naturels, quantité inconcevable à cribler (Dalby, 2011). Concrètement, la probabilité d'introduire par mutagenèse aléatoire plus d'une mutation au sein d'un même codon est quasi nulle. Un acide aminé donné ne peut donc être muté qu'en un nombre limité d'autres résidus (Table 1).

Table 1 : **Possibilités de substitutions d'acides aminés sur la base d'une mutation par codon.**

Acide aminé	Possibles acides aminés
Phe	Ser, Tyr, Cys, Leu, Ile, Val
Leu	Ser, Trp, Ile, Val, Phe, Pro, His, Gln, Arg
Ile	Phe, Leu, Val, Thr, Asn, Ser
Met	Phe, Leu, Val, Thr, Lys, Arg
Val	Phe, Leu, Ile, Met, Ala, Asp, Glu, Gly
Ser	Phe, Leu, Tyr, Cys, Pro, Thr, Ala, Asn, Ile, Arg, Gly
Pro	Leu, His, Gln, Arg, Ser, Thr, Ala
Thr	Ile, Met, Asn, Lys, Ser, Arg, Pro, Ala
Ala	Val, Asp, Glu, Gly, Ser, Pro, Thr
Tyr	Phe, Ser, Cys, His, Asn, Asp
His	Tyr, Asn, Asp, Leu, Pro, Arg
Gln	Leu, Pro, Arg, Lys, Glu
Asn	Ile, Thr, Ser, Tyr, His, Asp, Cys
Lys	Ile, Met, Thr, Arg, Gln, Glu, Asn
Asp	Val, Ala, Gly, Tyr, His, Asn
Glu	Val, Ala, Gly, Gln, Lys
Cys	Phe, Ser, Tyr, Arg, Gly
Trp	Leu, Ser, Arg, Gly
Arg	Leu, Pro, His, Gln, Met, Thr, Lys, Cys, Trp, Ser, Gly
Gly	Val, Ala, Asp, Glu, Cys, Trp, Arg, Ser

Parce que l'effort de criblage constitue actuellement le goulot d'étranglement d'un projet d'évolution dirigée, il est primordial d'optimiser la qualité de la banque pour réduire le nombre de clones à cribler. Il est ainsi convenu que les caractéristiques d'une banque offrant le plus de probabilités pour générer des enzymes améliorées sont :

- un nombre d'acides aminés mutés compris entre 2 et 5 par gène, ce qui permet généralement de conserver une bonne proportion de mutants actifs (Goldsmith and Tawfik, 2013),
- des mutations nucléotidiques sans biais entre transition et transversion. Une mutagenèse non biaisée devrait donc présenter une moyenne de 66 % de transversions contre 33 % de transitions car chacun des quatre nucléotides possède 2 possibilités de transversion contre une seule pour une transition. Par conséquent, il faut veiller à bien maintenir un taux de transversions important car, du fait de la dégénérescence du code génétique, les transitions sont moins susceptibles de conduire à des substitutions d'acides aminés, et sont donc plus susceptibles de produire des mutations silencieuses.
- un nombre limité voire nul d'insertions ou de délétions pour éviter le risque de modifier le cadre de lecture, ce qui affecterait trop grandement l'intégrité structurale des enzymes et donc leur activité.

Rasila et al., ont comparé en 2009 les caractéristiques de différentes méthodes de mutagénèse. Les résultats obtenus lors d'epPCR avec le kit Mutazyme II présentent le moins de biais en ce qui concerne la distribution des types de mutations. Les pourcentages de transition et de transversion constatés sont sensiblement identiques alors que le taux de transitions est largement majoritaire dans toutes les autres techniques de mutagénèse testées. L'utilisation du kit Mutazyme II ne permet donc pas d'obtenir une répartition totalement aléatoire et non biaisée concernant le type de mutations mais constitue la méthode s'en approchant le plus (Rasila et al., 2009). Les résultats obtenus par epPCR démontrent toutefois que cela reste un processus hautement stochastique (Hanson-Manful and Patrick, 2013).

Recombinaison de séquences codantes

Recombinaison de séquences homologues

La recombinaison de séquences génétiques permet de générer de la diversité génétique, non plus par l'incorporation de mutations aléatoires mais par la recombinaison aléatoire de plusieurs séquences codantes. Cette approche est envisagée dans deux principaux cas. Le premier se présente lorsque, à la suite d'un criblage (voir paragraphe suivant), plusieurs gènes modifiés comportant différentes mutations sont détectés comme améliorant la fonction souhaitée. Une recombinaison suivie d'un nouveau criblage permet de sélectionner les mutations intéressantes et éliminer les mutations sans intérêt. Il est fréquent de constater des effets additifs voire synergiques lors de la recombinaison de mutations sur une même séquence (Zhang et al., 1997). Le deuxième cas concerne la possibilité de recombiner deux ou plusieurs séquences codant pour des enzymes homologues afin d'associer des particularités phénotypiques intéressantes de chaque enzyme sur un mutant chimère (Li et al., 2015).

En 1994, Stemmer propose pour cela une technique qui est, aujourd'hui encore, largement utilisée dans les laboratoires. Cette approche nommée DNA shuffling consiste à fragmenter différentes séquences géniques et à les recombiner aléatoirement pour créer une librairie de séquences qui peut être criblée (W. P. Stemmer, 1994). Les séquences sont fragmentées par enzyme de restriction ou à la DNaseI puis les fragments sont soit ligués avec une DNA ligase ou par PCR (Stemmer et al., 1995) respectivement (Figure 5). Pour que ces recombinaisons puissent être réalisées, une identité de séquences au moins égale à 60 % est nécessaire. Cette approche a permis par exemple d'augmenter 540 fois la résistance au moxalactam (antibiotique) d'une souche d'*E. coli* exprimant une protéine chimère obtenue par recombinaison de 4 séquences géniques de céphalosporinase provenant de quatre micro-organismes différents. Ces résultats prometteurs ont été obtenus après un seul tour d'évolution (1 criblage) et le criblage de 5.10^4 clones. Cette protéine chimère optimisée comprend des fragments de 3 des 4 séquences parents utilisées, mais présente également 33 résidus sur les 381 acides aminés de la séquence finale qui ne proviennent d'aucune des 4 séquences parents. Ces mutations sont donc apparues lors de la PCR d'assemblage des différents fragments (Cramer et al., 1998). Cette étude a été

l'une des premières à montrer la possibilité de recombinaison de séquences provenant d'espèces différentes.

Une approche alternative nommée RACHITT pour "RANDOM CHimeragenesis on Transient Templates" a été mise au point peu de temps après. Cette méthode présente l'avantage de générer un nombre de recombinaisons par gène supérieur à celui obtenu avec du DNA shuffling. L'approche repose sur l'hybridation aléatoire de différents fragments d'ADN parentaux, le rognage et la ligation sur une séquence polynucléotidique transitoire comme représentée sur la Figure 5. Cette technique a permis d'une part d'augmenter la vitesse de réaction globale mais aussi d'élargir la spécificité de substrat d'une protéine chimère issue de deux dibenzothiophène monooxygénases (DBT-MO) provenant de deux souches bactériennes distinctes (Coco et al., 2001).

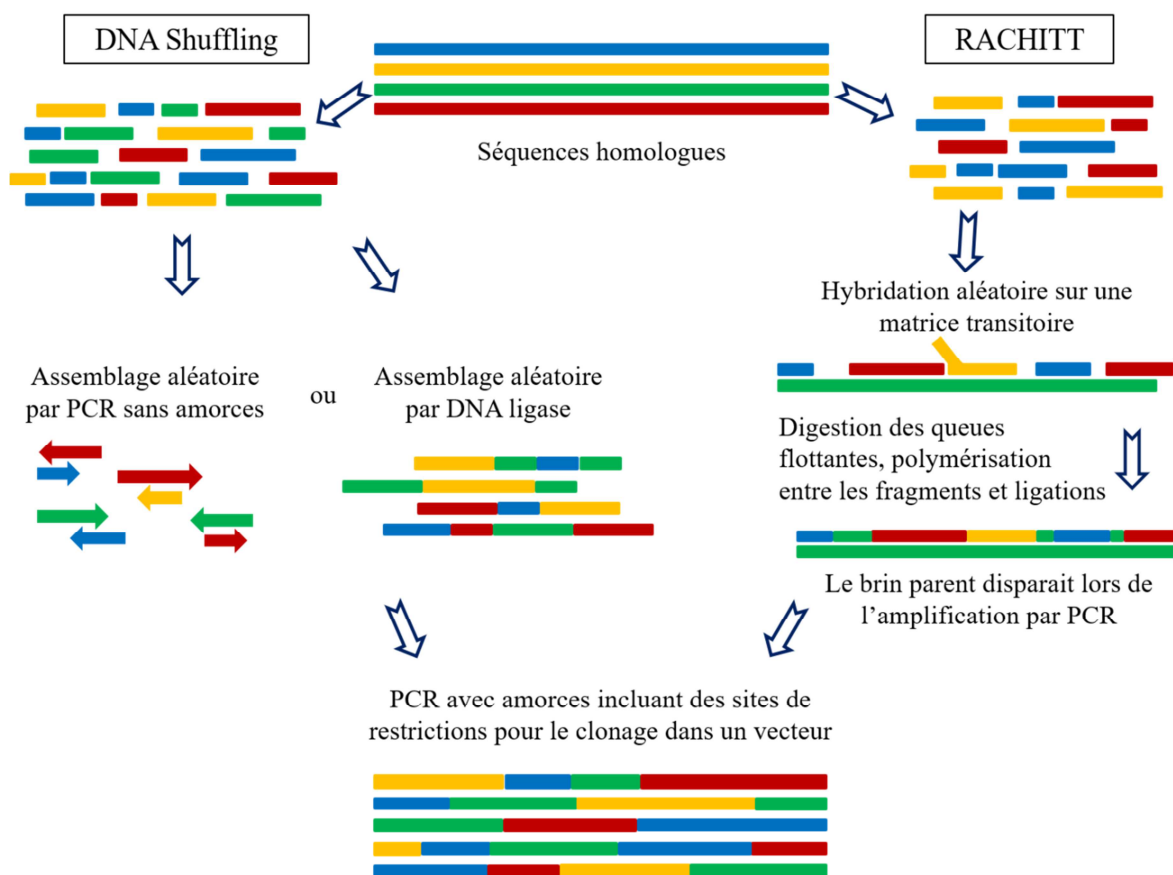


Figure 5 : Principes de la recombinaison par DNA shuffling (Willem P. Stemmer, 1994) et RACHITT (Coco et al., 2001)

En 2002, un nouveau procédé est expérimenté par Ness et al., qui permet à chaque acide aminé d'une protéine d'être modifié indépendamment (Figure 6). La méthode, appelée "Synthetic Shuffling", repose sur l'utilisation d'amorces nucléotidiques synthétiques partiellement dégénérées d'environ 60 pb. Les amorces se chevauchant, la globalité de la séquence est reconstituée au cours d'une PCR. Dans cette étude, des mutations préalablement sélectionnées sur 15 gènes mutants de la subtilisine (protéase

à sérine) ont été ainsi aléatoirement redistribuées sur différentes séquences. Après criblage, quatre variants présentent une activité améliorée par rapport à la Savinase® commercialisée, deux d'entre eux avec une activité deux à quatre fois supérieure après un choc thermique, et tous plus thermostables que n'importe lequel des parents (Ness et al., 2002). Dans cette approche, il n'est pas nécessaire d'avoir la séquence codante de l'enzyme pour générer de la diversité mais l'avantage majeur est qu'elle permet d'accéder à un espace de séquence plus grand et mieux contrôlé qu'avec les approches précédemment citées (Valetti and Gilardi, 2004). En effet, d'habitude les recombinaisons de fragments homologues se font préférentiellement entre des zones à forts taux d'identités, alors qu'ici de courtes séquences d'acides aminés sont directement ciblées, soit en y incluant une (des) mutation(s) précise(s), soit en utilisant des amorces dégénérées pour augmenter l'espace de séquences à cribler.

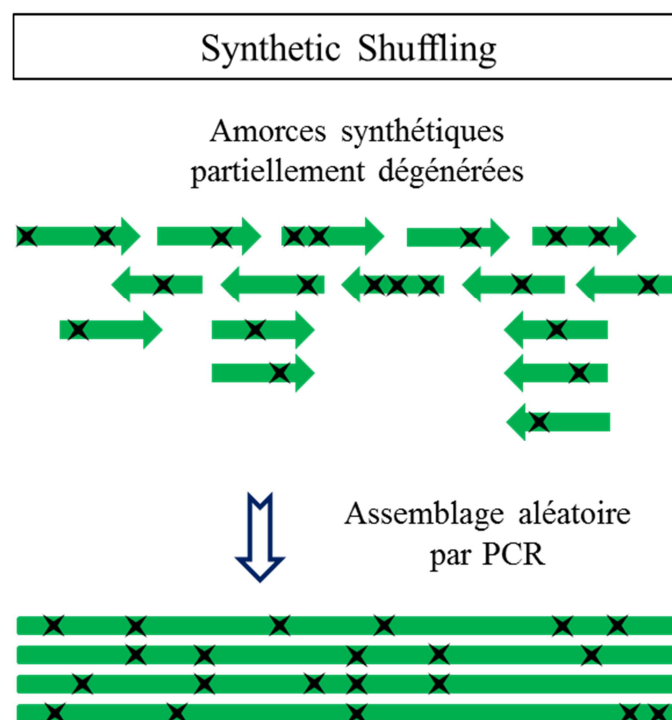


Figure 6 : **Principe de recombinaison par Synthetic Shuffling** (Ness et al., 2002)

Pour conclure cette liste non exhaustive de techniques utilisées pour la recombinaison de séquences homologues, Zhao *et al.*, proposent en 1998 une technique de recombinaison nommée StEP pour "Staggered Extension Process" (Zhao et al., 1998). Elle repose sur une simple PCR qui utilise comme matrice, soit la diversité obtenue artificiellement lors d'un premier cycle d'évolution soit différentes séquences homologues à recombinaison. La particularité de cette PCR est d'utiliser un temps d'élongation très court (1 ou 2 sec.). Toutes les séquences à recombinaison sont utilisées dans la réaction de PCR pour servir de matrices. Lors de chaque cycle, les amorces s'hybrident aléatoirement sur une des séquences matrices. Le temps d'élongation étant très court, la polymérase va greffer une dizaine de nucléotides sur l'amplicon avant de se déshybrider et de s'hybrider à nouveau sur une nouvelle séquence matrice

au cycle suivant (Figure 7). La preuve de concept a été faite sur une librairie de mutants obtenus à partir de 5 variants de la subtilisine E. Les 5 séquences de départ comprenant de 1 à 5 mutations sont issues d'un premier tour d'évolution effectué par PCR à erreurs. Ces 5 mutants améliorent la demi-vie de la protéase à 65 °C d'un facteur compris entre 3 et 8 comparé à la subtilisine E sauvage. Le criblage de cette nouvelle diversité a mis au jour 3 nouveaux mutants possédant une demi-vie à 65 °C multipliée d'un facteur 25 à 50, comparé à l'enzyme sauvage (Zhao et al., 1998). Cette méthode ingénieuse et simple est particulièrement adaptée pour la recombinaison de plusieurs mutants améliorés d'une même séquence codante (Zhao, 2004; Zhao and Zha, 2006).

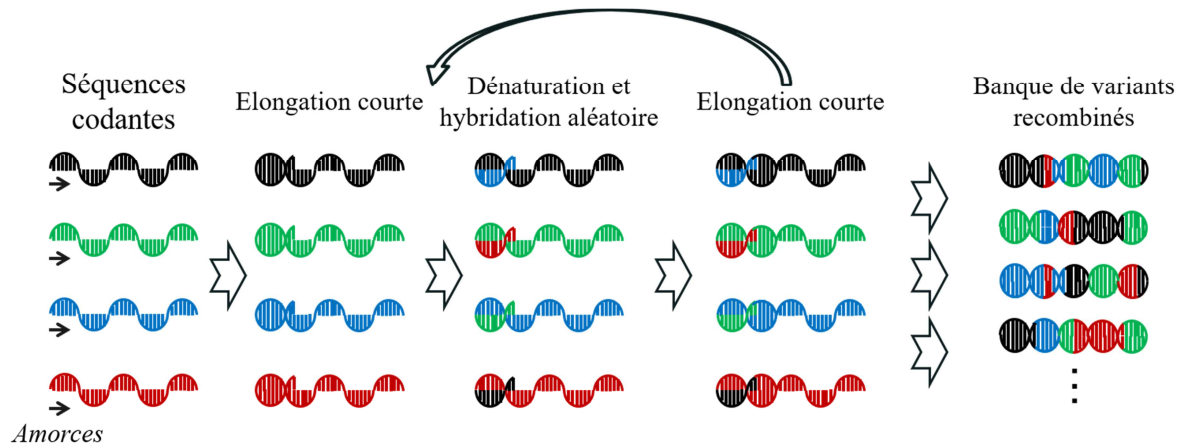


Figure 7 : Cycles schématisés de la réaction de recombinaison par StEP (modifié d'après (Packer and Liu, 2015)).

Recombinaison de séquences non homologues

Parce que la recombinaison de gènes faiblement homologues demeure intéressante, des techniques alternatives et plus spécifiques ont été développées pour ces cas particuliers. L'objectif ici est de s'affranchir des étapes de recombinaison homologue ou d'hybridation -qui seraient virtuellement impossibles du fait de la faible homologie entre les séquences ciblées- et d'utiliser en remplacement des ligations de fragments d'ADN à bouts francs pour réassembler les séquences. Cela permet de recombiner de manière combinatoire des séquences géniques de plus faible homologie, voire totalement différentes. La méthode ITCHY (Incremental Truncation for the Creation of Hybrid enZymes) permet une chimérisation aléatoire entre deux gènes indépendamment de l'identité entre les séquences. Cependant, l'inconvénient du procédé est que les membres de ces bibliothèques ne contiennent qu'un seul évènement de recombinaison par séquence (Ostermeier et al., 1999) (Figure 8). Pour remédier à cet inconvénient, la méthode Scratchy combine les deux méthodes que sont l'ITCHY et le DNA shuffling (Figure 8) : elle permet l'introduction de recombinaisons multiples entre les séquences d'intérêts indépendamment de leurs pourcentages d'identités (Lutz et al., 2001). Griswold et al., ont ainsi pu produire des chimères entre deux glutathion S-transférases (GST), l'une provenant

d'un rat et l'autre humaine, possédant 54,3 % d'identité. Un criblage à haut-débit par cytométrie en flux utilisant un substrat fluorogénique a permis l'isolement d'un variant qui a échangé les hélices 4 et 5 de de la GST humaine avec les séquences correspondantes de la GST de rongeur. Cette enzyme chimère possède une efficacité catalytique (k_{cat}) améliorée pour la conjugaison de la 7-amino-4-chloromethyl coumarine (CMAC) sur le glutathion, et plus surprenant, présente également une activité de conjugaison du glutathion sur l'acide étacrynique, une activité qui n'est pourtant retrouvée dans aucune des deux enzymes parentales. Ces résultats soulignent le pouvoir de la méthode qui peut aussi fournir un moyen d'ingénierie d'enzymes humanisées qui pourrait être utile en thérapeutique clinique (Griswold et al., 2005).

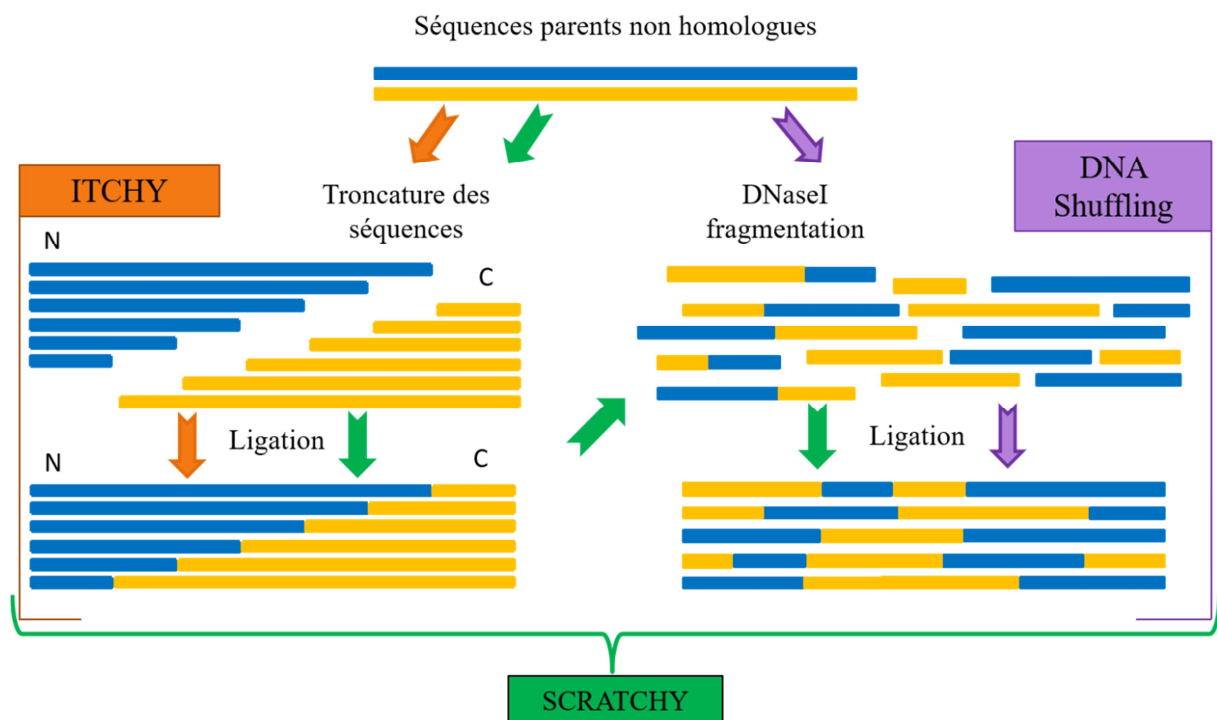


Figure 8 : **Principe des méthodes de recombinaison ITCHY** (Ostermeier et al., 1999) et **SCRATCHY** (Lutz et al., 2001).

Pour conclure cette liste des principales techniques utilisées pour la recombinaison de séquences non-homologues, on peut également citer la méthode SHIPREC, acronyme pour Sequence Homology-Independent Protein RECombination, qui est utilisée pour créer des mutants à partir de deux enzymes partageant une homologie de structure indépendamment de leur similarité des séquences (Figure 9). Les mutants ainsi produits posséderont le côté N-terminal d'une protéine parent et le côté C-terminal de la seconde enzyme parent, avec une jonction entre les deux séquences aléatoirement positionnées. La démonstration de la méthode a là aussi permis de créer une chimère entre une enzyme humaine et une enzyme bactérienne. L'objectif atteint était de rendre plus soluble, tout en conservant son activité, une enzyme membranaire humaine cytochrome de type P450, en la couplant avec une autre enzyme

cytochrome P450 bactérienne, présentant une bonne solubilité. Les deux séquences protéiques présentent un taux d'identité de seulement 16 %, et un mutant identifié lors de cette étude (RC2), en plus de conserver une bonne efficacité catalytique, montre une activité enzymatique dans la fraction cytosolique 8 fois supérieure à l'enzyme P450 humaine. De plus, les enzymes présentes dans la fraction membranaire sont plus facilement re-solubilisées. Il faut toutefois constater que toutes les chimères actives, donc correctement repliées, possèdent des recombinaisons aux extrémités N- ou C-terminale, soit des enzymes présentant une très grande majorité de l'une ou l'autre des enzymes parents. Ce phénomène représente une limite importante en termes de diversité pour cette méthode (Sieber et al., 2001).

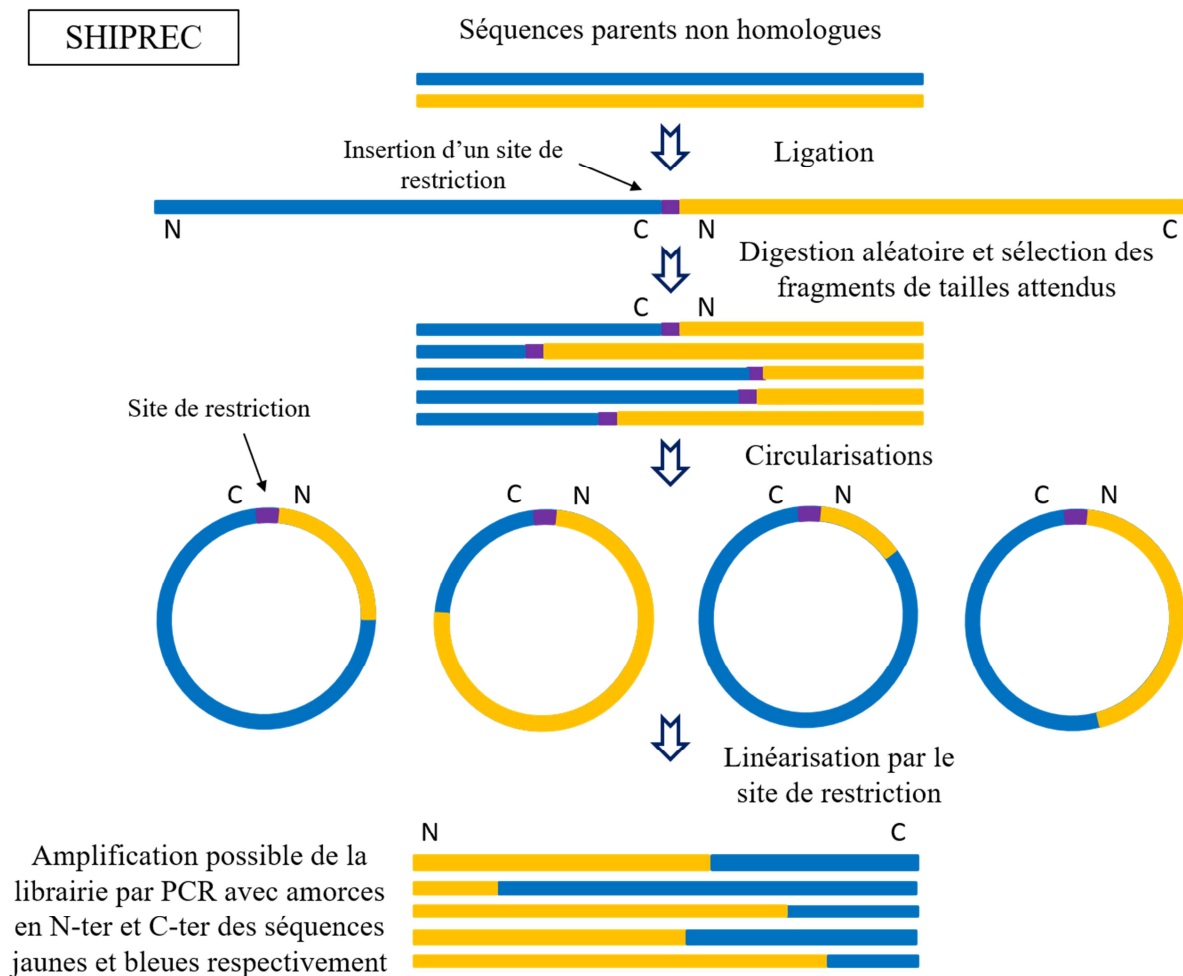


Figure 9 : **Principe de la méthode SHIPREC** (Sequence Homology–Independent Protein REcombination)

Afin de couvrir le plus large espace de séquences possible, l'évolution dirigée repose sur un processus itératif. Le gène du mutant sélectionné lors d'un premier tour d'évolution peut servir de matrice pour créer une nouvelle diversité génétique. C'est souvent après plusieurs tours d'évolution (création de

diversité génétique + criblage) que les résultats obtenus deviennent très significatifs en comparaison avec l'enzyme de départ. Les 126 tours d'évolution réalisés sur la Cre recombinase pour modifier ses caractéristiques de reconnaissance de substrat et générer un variant qui excise efficacement et sélectivement l'ADN proviral du VIH intégré du génome humain en est un exemple extrême (Sarkar et al., 2007).

e. Evolution rationnelle ou semi-rationnelle

Une approche rationnelle nécessite, contrairement aux techniques d'évolution dirigée, un minimum de connaissances des liens séquence-structure-fonction de l'enzyme. Ainsi, sur la base d'hypothèses formulées, des résidus sont précisément ciblés pour mutation afin de modifier les propriétés de l'enzyme. Les principales étapes représentées sur la Figure 10 reposent ici sur le choix des résidus à muter puis sur la confirmation des prédictions par des données expérimentales.

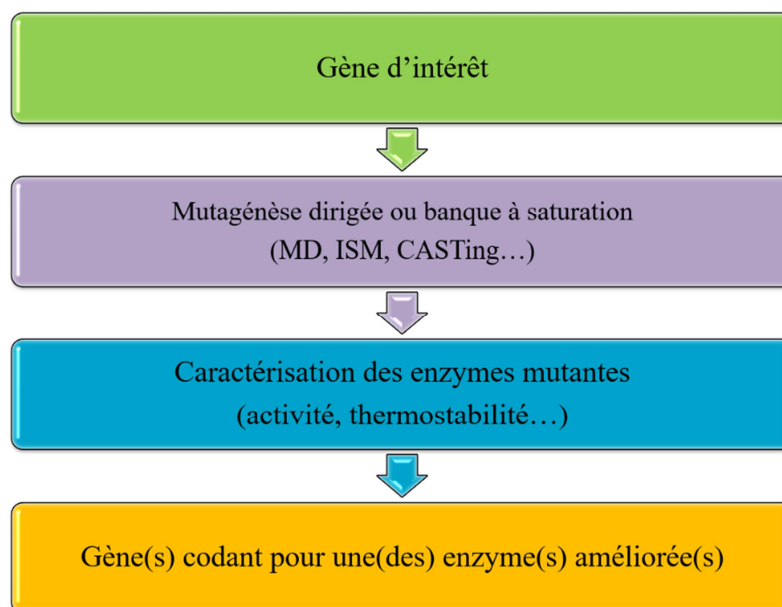


Figure 10 : Principales étapes du processus d'évolution rationnelle

Un exemple de méthode rationnelle pour améliorer la stabilité des enzymes est appelé le "consensus concept" (Pantoliano et al., 1989). Elle consiste, suite à un alignement multiple des séquences homologues d'une même enzyme, à déterminer une séquence consensus, c'est-à-dire une séquence composée des résidus majoritairement présents sur chaque position tout au long de la séquence. L'idée repose sur le fait que l'acide aminé consensus contribue plus à la stabilité de la protéine que les acides aminés non majoritairement représentés (Lehmann et al., 2000b). Même si ce postulat est loin d'être systématiquement confirmé (Wang et al., 1999), un résultat assez spectaculaire a été obtenu par Lehmann et al. En effet, sur la base d'un alignement de 13 séquences de phytases fongiques, une enzyme consensus a été construite, qui présente une augmentation de sa température de fusion (T_m) de

15 à 22 °C par rapport à chacun de ses parents. De plus, malgré cet important gain de thermostabilité, son activité catalytique à 37 °C n'est pas affectée (Lehmann et al., 2000a).

De nombreux logiciels utilisant les structures tridimensionnelles des enzymes ou les séquences protéiques tels que RONN (Yang et al., 2005), PoPMuSiC (Dehouck et al., 2009) ou encore FireProt (Bednar et al., 2015) permettent de faire des prédictions à propos des résidus potentiels à muter afin d'obtenir des protéines plus stables. Ces mutations ciblées sont alors réalisées par mutagenèse dirigée puis le mutant est produit et testé pour confirmer les prédictions faites *in silico*. Malgré l'évolution incontestable de ces outils de prédiction lors de cette dernière décennie (Jiang et al., 2008), les résultats expérimentaux sont loin de confirmer systématiquement les prédictions faites *in silico*.

Située à mi-chemin entre l'évolution dirigée et la mutagenèse dirigée, la stratégie semi-rationnelle permet de combiner la puissance des deux approches (Chica et al., 2005). Des approches telles que l'ISM (Iterative Site saturation Mutagenesis) ou le CAST (Combinatorial Active site Saturation Test) permettent, après identification de quelques zones clés, identifiées généralement grâce à la structure de l'enzyme, de réduire considérablement la taille des banques de mutants, et donc l'effort de criblage qui va avec, en ne générant de la diversité qu'à quelques positions précisément identifiées. Le meilleur mutant issu de chaque tour de criblage est alors utilisé comme matrice pour générer une nouvelle diversité au niveau de chaque région ou acide aminé ciblés. Ces approches semi-rationnelles composées de petites banques 'intelligentes' ont montré une grande efficacité pour faire évoluer les enzymes (Reetz et al., 2005, 2006a, 2006b; Reetz and Carballeira, 2007).

Un exemple de ces approches mixtes est le SCOPE pour "Structure based COmbinatorial Protein Engineering". Cette technique de recombinaison de séquences non homologues permet d'associer des domaines structuraux propres à plusieurs enzymes, non pas aléatoirement, mais en se basant sur des informations concernant la conformation des domaines structuraux présents dans les deux protéines qui vont être recombinées. Ainsi deux ADN polymérases, une de rat (Pol β) et une du virus de la peste porcine africaine (Pol χ) qui diffèrent grandement en taille et en activité mais partagent des éléments de structure ont été soumises à ce procédé. Un criblage par sélection des polymérases chimères a mis en évidence des mutants permettant une croissance de la souche hôte à 42 °C significativement plus rapide que pour les cellules contenant les enzymes parentales Pol β ou Pol χ (O'Maille et al., 2002). Ce type d'approches mixtes se développe et est très prometteur pour accroître l'efficacité de l'ingénierie enzymatique. Grâce aux techniques de la biologie moléculaire, combinées à celles de la biologie structurale et de la modélisation *in silico*, nous pouvons aujourd'hui d'une part, améliorer significativement l'efficacité des enzymes, mais également concevoir des biocatalyseurs totalement originaux, capables de réaliser de nouvelles réactions.

Une fois la diversité génétique obtenue, il faut déterminer le protocole de criblage ou de sélection le plus judicieux. Il doit permettre d'isoler spécifiquement les catalyseurs améliorés. Là aussi, les

alternatives sont multiples et doivent être judicieusement choisies en fonction des conditions propres à chaque situation (Bornscheuer et al., 2012).

f. Production, criblage et sélection des enzymes

La mise au point du protocole de criblage est certainement l'étape clé d'un travail d'évolution dirigée. "You get what you screen for !" (On ne trouve que ce que l'on cherche) signifie que chaque protocole de criblage conditionne la qualité de la sélection de mutants isolés. Si le protocole ne répond pas précisément à la question, les enzymes isolées ne seront pas optimisées pour le critère initialement choisi. Cet effort de mise au point du criblage doit donc être adapté pour chaque enzyme -il n'existe pas de protocole de criblage 'type'- et doit donc être spécifique de la modification souhaitée (augmenter l'activité, modifier la spécificité de substrat, améliorer la stabilité en présence de solvants, augmenter la résistance aux hautes températures, accroître l'activité de transfert chez des hydrolases, résister à un antibiotique, augmenter l'énantiosélectivité, etc.). Dans certains cas où le criblage direct est impossible, une réaction couplée à l'activité que l'on souhaite évaluer doit être mise au point pour permettre une détection. Il faut donc dans tous les cas, lors de l'élaboration du procédé, prendre en compte tous les paramètres (température de criblage, variation du pH lors de la réaction, effet de la concentration des substrats, choix judicieux de l'organisme de production et des substrats, etc.).

Production des enzymes

En théorie, tout organisme pourrait servir d'hôte de production car le mécanisme ribosomal de synthèse des protéines est un processus universellement conservé, ce qui permettrait par exemple à une bactérie de produire des protéines humaines (Braun et al., 2002). Cependant, seulement un petit nombre d'organismes génétiquement modifiés sont adaptés aux travaux d'évolution dirigée et aux contraintes de criblages à hauts-débits (Packer and Liu, 2015). Une publication de 2012 rapporte que près de 86 % des études contenant l'expression "évolution dirigée" dans le titre de la publication ont porté sur la bactérie *Escherichia coli* comme organisme de production. A peine plus de 10 % ont utilisé des organismes eucaryotes tels que les levures *Saccharomyces cerevisiae* (8,7 %) ou *Pichia pastoris* (1,6 %) et à peine plus de 1 % de ces travaux ont utilisé des cellules de mammifères ou d'insectes pour surexprimer les enzymes (Pourmir and Johannes, 2012). La prédominance d'*E. coli* et, dans une moindre mesure de *S. cerevisiae*, dans ce type d'étude tient à leur bonne efficacité de transformation, leur vitesse rapide de croissance -avec un temps de génération de 20 min et 2 heures pour *E. coli* et *S. cerevisiae* respectivement-, aux protocoles de manipulations génétiques bien établis et à leur aptitude à maintenir de façon stable des plasmides (Pourmir and Johannes, 2012). Une large sélection de lignées cellulaires d'*E. coli*, modifiées génétiquement, existe et permet de s'adapter aux spécificités de nombreuses enzymes (Rosano and Ceccarelli, 2014). La souche BL21 par exemple, largement utilisée pour la production d'enzymes recombinantes, est déficiente pour la protéase Lon, qui dégrade les protéines exogènes (Gottesman, 1996) ainsi que pour le gène codant pour la protéase

membranaire OmpT, dont la fonction est de dégrader les protéines extracellulaires (Grodberg and Dunn, 1988). Cette protéase présente l'inconvénient de dégrader les protéines produites suite à la lyse cellulaire. *E. coli* reste donc sans aucun doute l'organisme procaryote de référence pour la production de protéines, même si aujourd'hui, certaines bactéries telles que *Bacillus subtilis* ou *Lactococcus lactis* par exemple, sont capables de sécréter des protéines recombinantes (Westers et al., 2004; Morello et al., 2007).

Toutefois, lorsque l'enzyme d'intérêt nécessite d'être glycosylée, l'utilisation de levures est alors nécessaire (Wildt and Gerngross, 2005). Dans ce cas, les séquences codant les protéines recombinantes peuvent se transmettre à la descendance soit de façon épisomale, après l'insertion de la séquence codante dans un plasmide, soit en intégrant la séquence codante dans le génome de la levure hôte. Une étude démontre que l'intégration des séquences codantes dans le génome de la levure permet une expression plus homogène des enzymes, ce qui est primordial dans les procédés de criblages (Jensen et al., 2014). Certaines lignées cellulaires ont été génétiquement modifiées afin de permettre une recombinaison homologue ciblée afin d'obtenir un taux de transcription équivalent pour chaque clone (Bordes et al., 2007).

Enfin, la production hétérologue de certaines protéines nécessite une optimisation de la séquence nucléotidique afin qu'elle soit adaptée à l'organisme hôte (Gustafsson et al., 2004). En effet, la séquence codante de la protéine d'intérêt peut contenir des codons plus rares chez l'organisme hôte : très souvent, les micro-organismes expriment des différences assez marquées dans l'utilisation de codons synonymes, ainsi par exemple, le codon AUA qui code l'isoleucine est largement évité chez *E. coli*, par rapport aux deux autres codons synonymes AUU et AUC. Cette préférence d'usage des codons est très variable d'un organisme à l'autre et la variation de la vitesse de traduction peut perturber le repliement de l'enzyme.

Criblage et sélection des enzymes

Une fois la banque de mutants constituée, plusieurs stratégies peuvent s'envisager pour isoler les mutants optimisés : on peut distinguer les méthodes dite de "sélection" et les méthodes de "criblage" (Packer and Liu, 2015; Xiao et al., 2015). Les principales options possibles pour identifier de nouveaux biocatalyseurs enzymatiques améliorés au sein d'une librairie de mutants sont schématisées sur la Figure 11.

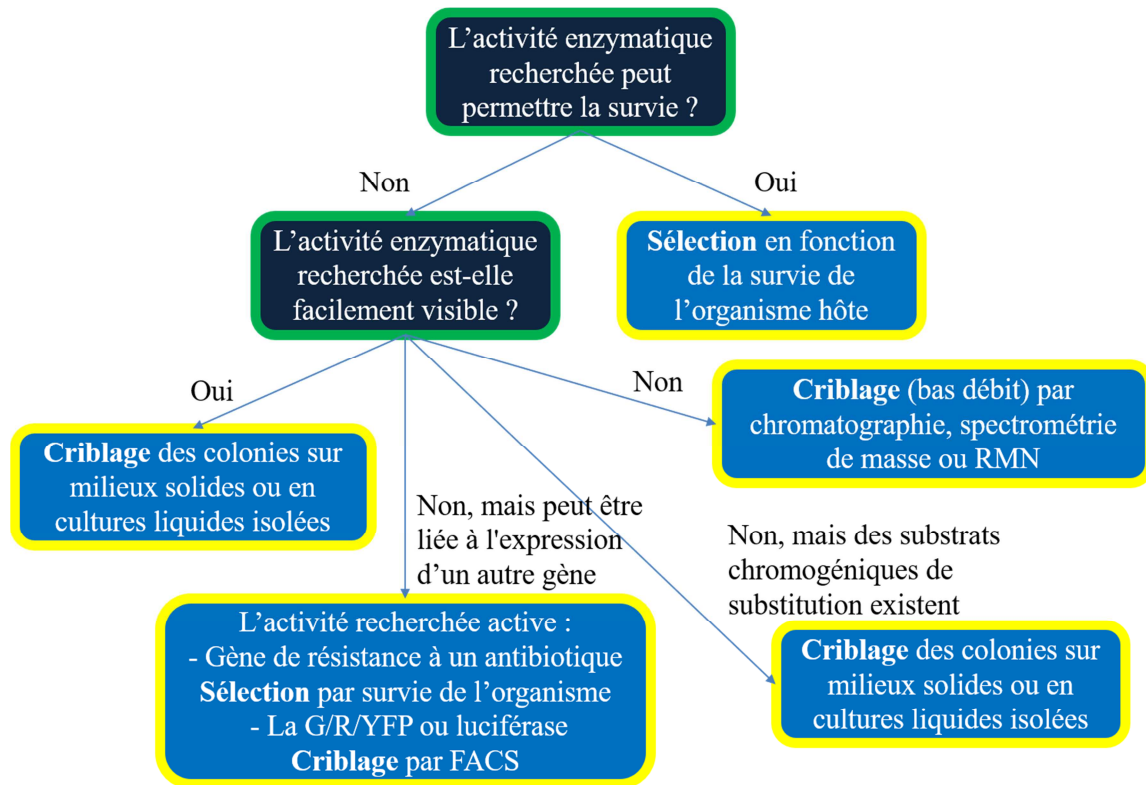


Figure 11 : **Différentes options de sélection/criblage pour l'évolution dirigée d'enzymes.** Modifié d'après (Packer and Liu, 2015)

Sélection

Les méthodes de sélection offrent un débit de criblage inégalable car elles permettent de tester tous les variants dans un seul et même pool, et en un minimum de procédure. La sélection *in vitro* est conçue pour la recherche de chimères ayant une affinité pour une autre protéine ou un ligand. Les méthodes de sélection *in vivo* reposent sur la recherche d'un phénotype particulier qui permettra à la cellule de survivre et croître dans un milieu qui est normalement incompatible avec la survie du phénotype de départ (Boersma et al., 2007).

Les méthodes de recherches de mutants améliorés par sélection peuvent donc être séparées en deux groupes (Figure 12) :

- les méthodes de "display", *in vitro*, où un ensemble de mutants d'une librairie est soumis en même temps et dans un même échantillon à un test d'affinité. Ces méthodes nécessitent donc une association entre la protéine chimère et sa séquence codante. Une très grande quantité de mutants est cultivée simultanément puis soumise à la procédure de sélection par affinité pendant laquelle seuls les mutants ayant une affinité pour les molécules ou protéines cibles immobilisées vont être retenus. Pour identifier la séquence correspondant au mutant sélectionné, il faut donc que la séquence codante correspondante puisse être récupérée dans le même temps : soit en étant produite à la surface de l'hôte

(phage display (Smith, 1985)), soit en permettant que la protéine produite reste covalamment liée à la séquence codante dont elle est issue (ARN display (Wilson et al., 2001), ribosome display (Hanes and Plückthun, 1997), plasmide display (Speight et al., 2001)). Ces approches permettent de sélectionner des mutants sur la base de leur affinité mais sont rarement déployées pour des criblages d'amélioration d'activité enzymatique (Fa et al., 2004).

- les méthodes *in vivo*, de sélection pour une activité enzymatique. Ces méthodes comportent une première étape de transformation d'un organisme hôte. La sélection repose ensuite sur la capacité potentielle d'un des mutants à se développer dans un milieu sélectif grâce à une activité enzymatique recherchée. Certaines souches hôtes peuvent avoir subi la délétion d'un gène codant pour une protéine indispensable d'une voie métabolique, la sélection consiste alors à rechercher une enzyme mutante capable de rétablir cette activité et donc la croissance de la souche en question. Cette approche est appelée sélection par complémentation (Boersma et al., 2007), dont un bel exemple a été réalisé par l'équipe dirigée par Donald Hilvert, sur la chorismate mutase provenant de l'archée thermophile *Methanococcus jannaschii* (MjCM). Cette enzyme, sous forme d'homodimère, catalyse la conversion du chorismate en préphénate dans la voie de biosynthèse de la tyrosine (Tyr) et de la phénylalanine (Phe). L'objectif est de modifier structurellement cette enzyme pour qu'elle puisse conserver une bonne activité sous forme monomérique (MacBeath, 1998). Pour cela, une combinaison aléatoire de six acides aminés a été insérée au milieu d'une hélice α pour permettre un repliement à 180° de cette même hélice α et former une enzyme active sous forme monomérique. La même équipe avait préalablement réalisé la délétion dans une souche d'*E.coli* de l'activité de la chorismate mutase endogène, permettant ainsi une sélection par complémentation sur milieu non complémenté en Tyr et Phe (Kast et al., 1996). Seules les cellules exprimant un mutant actif, rétablissant la prototrophie de la souche, c'est-à-dire capable de restaurer la synthèse de ces deux acides aminés essentiels, poussent sur un milieu non complémenté en Tyr et Phe. Cette sélection à haut-débit a permis de tester une diversité de près de 10^8 transformants pour ne sélectionner que 26 mutants permettant la viabilité des cellules. La structure exclusivement monomérique ou la présence d'un équilibre entre les deux formes ont pu être vérifiées dans un deuxième temps sur ce petit nombre de mutants par une chromatographie d'exclusion sur colonne, ce qui a permis d'isoler un mutant monomérique conservant sensiblement les mêmes paramètres cinétiques que ceux obtenus avec l'enzyme sous forme d'homodimère.

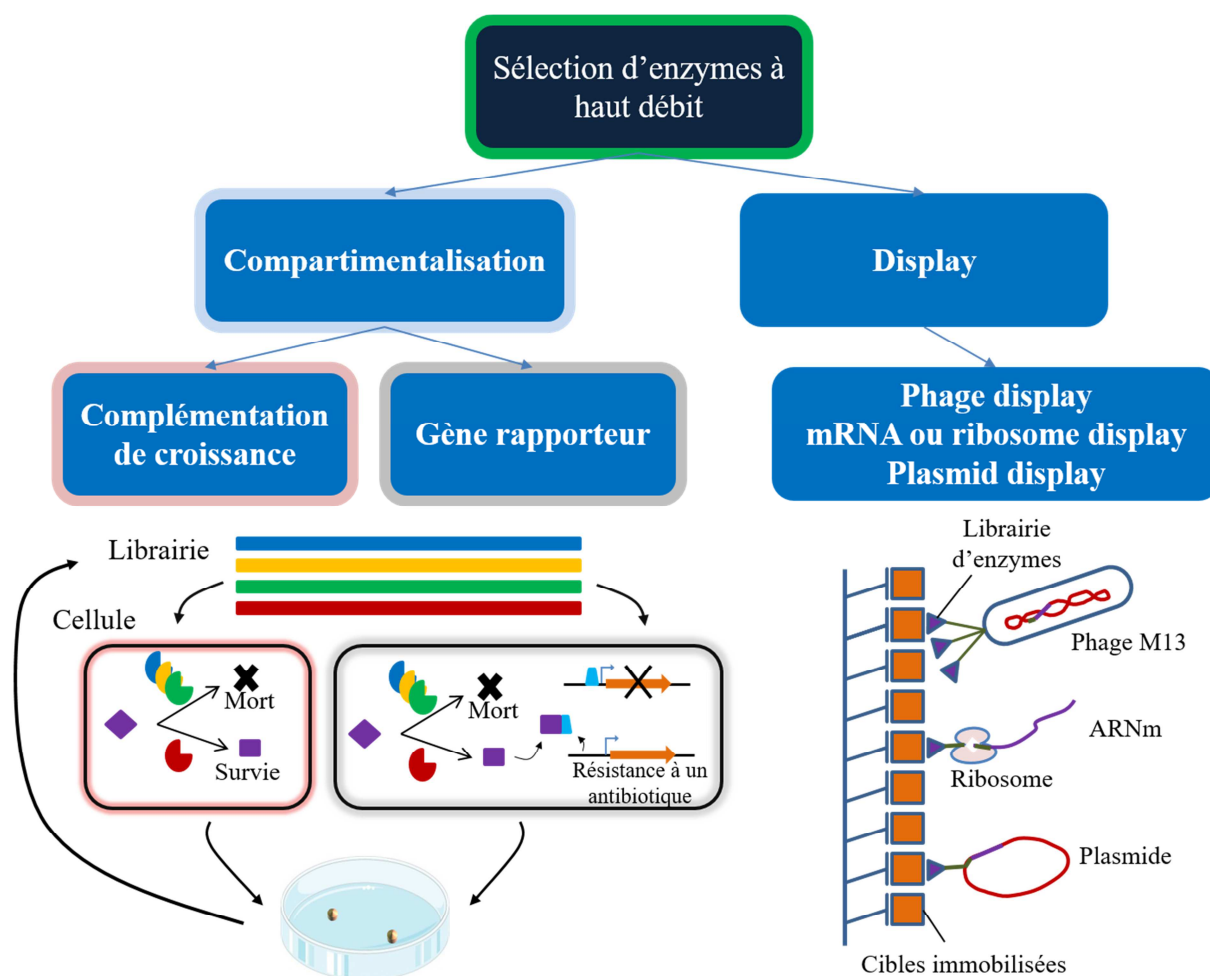


Figure 12 : **Représentation schématique des principales approches utilisées pour la sélection d'enzymes mutantes.**

Une autre option de sélection consiste à utiliser un gène rapporteur (Figure 12). Cette technique et ses applications potentielles ont été décrites en 2013 (van Rossum et al., 2013). La sélection basée sur l'activité d'un gène rapporteur permet de sélectionner des fonctions enzymatiques plus diverses par rapport à la complémentation de croissance (Dietrich et al., 2010). La stratégie de sélection *in vivo* la plus employée est la régulation transcriptionnelle du gène rapporteur par le produit de la protéine testée. L'activité de l'enzyme est évaluée indirectement par l'activation ou la levée de la répression transcriptionnelle d'un gène rapporteur par un régulateur transcriptionnel. La revue de (Dietrich et al., 2010) détaille les différents profils de réponses mais indique que cette méthode est encore relativement peu utilisée pour l'évolution de nouveaux biocatalyseurs. Les deux principales limitations proviennent d'une part de la complexité et du temps nécessaire à la mise au point du procédé (Carreño and de Lorenzo, 2010) et d'autre part, *in fine*, à la sélection de nombreux faux positifs, due aux mécanismes de contournement générés par la cellule.

Les méthodes de sélection, bien que très puissantes en termes d'espace de séquence exploré ne sont pas applicables à toutes les activités enzymatiques. Lorsqu'il n'est pas possible de corrélérer,

directement ou indirectement, l'activité recherchée à la viabilité de la cellule hôte par exemple, un procédé de criblage permettant l'évaluation de l'activité enzymatique de chaque mutant est donc à mettre en place. Il est à noter que les résultats exploitables obtenus par des approches de sélection sont souvent moins riches d'enseignement que ceux générés par des méthodes de criblage (Xiao et al., 2015). En effet, les méthodes de sélection écartent systématiquement tous les mutants inactifs et donc bloquent l'accès aux informations génétiques responsables de ces pertes d'efficacité.

Criblage

Lors d'un criblage, tous les mutants enzymatiques de la banque sont testés individuellement pour l'activité d'intérêt. Les techniques de criblage se perfectionnent et se diversifient constamment grâce aux avancées de la robotique et à l'imagination des scientifiques. Elles nécessitent une adaptation pour chaque projet d'ingénierie enzymatique, car il n'existe pas de protocole standard préétabli applicable à n'importe quelle enzyme ; la conception pertinente du criblage va déterminer sa robustesse et, généralement, la qualité des résultats obtenus. La taille des banques criblées varie généralement entre 10^2 et 10^5 mais peut atteindre 10^9 mutants, tout dépend de la cadence de tests que permet la méthode retenue (Figure 13).

- Fluorescence et très haut débit :

L'utilisation de cellules vivantes transformées produisant des mutants enzymatiques dont la réaction produit un intermédiaire fluorescent peut offrir, de loin, le plus haut-débit de criblage. Le tri des cellules par cytométrie de flux (FACS) (Julius et al., 1972; Watson and Dive, 1994) permet de suivre l'apparition de la fluorescence emprisonnée dans la cellule *in vivo* en triant jusqu'à plus de 10^9 cellules par heure. Un exemple de l'efficacité de cette technique concerne l'amélioration de l'efficacité catalytique de la protéase à serine d'*E. coli*, OmpT (EC 3.4.21.87). Le substrat utilisé pour le criblage des banques de mutants aléatoires permet le transfert d'énergie entre molécules fluorescentes (FRET). Il est composé d'un peptide comportant un site de coupure spécifique de la protéase OmpT et d'un couple de fluorophores (tétraméthylrhodamine et bore-dipyrométhène) liés covalamment au peptide de chaque côté du site de coupure. La proximité géographique de ces deux fluorophores dans la molécule native bloque l'émission de fluorescence mais celle-ci est rétablie lors de l'hydrolyse du peptide par l'enzyme. Après hydrolyse, la tétraméthylrhodamine fluoresce et reste liée à la surface de la cellule grâce à sa liaison avec trois acides aminés chargés positivement qui forment des liaisons ioniques avec la surface des cellules chargées négativement. Le tri des cellules par FACS a permis d'isoler, parmi les 1.9×10^6 cellules criblées en 24 minutes, 352 clones ayant l'intensité de fluorescence la plus élevée. Après un second criblage robotisé en microplaques, 3 mutants ont été purifiés et montrent une efficacité catalytique d'hydrolyse sur un substrat naturel entre 9 et 60 fois supérieure à l'enzyme sauvage (Olsen et al., 2000). Bien que OmpT soit une enzyme de surface, la stratégie mise au point lors de cette étude est transposable à une grande variété d'enzymes hétérologues en utilisant des

méthodes bien établies pour l'adressage des protéines à la surface d'*E. coli*, de *S. cerevisiae*, ou encore *Streptococcus* spp (Georgiou et al., 1997). Le criblage par FACS a aussi été utilisé avec succès pour l'amélioration de l'activité de transfert de sucres. En utilisant des sucres accepteurs couplé à des fluorophores, une glycosyltransférase (GT) du pathogène humain *Campylobacter jejuni* a ainsi vu son efficacité catalytique de synthèse augmenter 400 fois (Aharoni et al., 2006).

- méthodes visuelles/colorimétriques :

Les substrats chromogéniques peuvent mimer les substrats naturels (Puchart and Biely, 2007; Borsenberger et al., 2012); même si l'activité enzymatique sur ces substrats peut varier par rapport à l'activité sur substrat naturel, ils permettent souvent de bien retranscrire la spécificité des enzymes (Borsenberger et al., 2014).

Dans un premier temps, ces substrats peuvent être utiles pour un criblage en milieu solide (Reymond and Babiak, 2007), autorisant ainsi des cadences de criblage élevées (Bastien et al., 2013). Des logiciels d'imagerie permettent d'évaluer ou de comparer des activités enzymatiques via des variations d'intensité de coloration des colonies (Joo et al., 1999; Kone et al., 2009) mais ne permettent pas de quantifier précisément l'activité ou de déterminer les paramètres cinétiques des enzymes. Un cas particulier concerne le cas du criblage par visualisation de halos de dégradation : ici, il ne s'agit pas d'utiliser un substrat artificiellement coloré, mais d'observer la dégradation d'un substrat naturel rendant le milieu dans lequel il est dissous opaque. Cette méthode a, par exemple, été utilisée pour la détection d'une protéase qui génère un halo sur un milieu contenant du lait écrémé. Cette activité enzymatique a été détectée suite au criblage de 70 000 clones provenant d'une banque métagénomique issue de la peau de chèvre (Pushpam et al., 2011).

En milieu liquide, les cadences de criblage se réduisent naturellement mais les méthodes spectrophotométriques demeurent intéressantes car elles permettent une quantification précise des activités et la détermination de paramètres cinétiques. Outre la variation ou l'apparition de coloration liée à l'utilisation d'un substrat chromogénique (Bissaro et al., 2015a; Bøjstrup et al., 2015), il est également possible de suivre l'évolution d'une réaction par des changements de pH (Bornscheuer et al., 1998; Champion et al., 2010), un suivi de la turbidité de la solution (Huber et al., 2011) ou même les changements de température accompagnant la réaction (Cohen et al., 2001). Pour être réalisé à haut-débit, le criblage liquide en micro-plaques nécessite le plus souvent un appareillage spécialisé et précis (spectrophotomètre pour micro-plaques, stations de pipetage automatisées, etc.).

- Méthodes analytiques :

Ces méthodes à plus bas débit, sont généralement les plus sensibles et permettent d'utiliser directement les substrats naturels des enzymes.

* La chromatographie liquide à haute pression (HPLC) permet de séparer chaque produit d'une réaction enzymatique en fonction de ses propriétés physico-chimiques. Si le produit de réaction

est disponible pour obtenir une gamme étalon, la quantification de chaque produit est possible. Une étude d'évolution dirigée sur la mannitol déshydrogénase (MDH) a permis de cribler 10 000 enzymes mutantes par HPLC, en suivant leur capacité de conversion du substrat ribitol en ribose. L'énantiosélectivité des produits de réactions a ensuite été évaluée par électrophorèse capillaire (CE). Cette méthode a permis d'isoler un mutant possédant une énantiosélectivité supérieure à 99 % pour la synthèse de L-ribose, ce qui représente le seuil nécessaire pour que cette enzyme puisse être utilisée comme biocatalyseur dans des procédés industriels de synthèse de L-ribose (Sun et al., 2013). Aujourd'hui, les criblages par HPLC sont souvent couplés avec un spectromètre de masse (LC-MS) afin d'identifier plus précisément la nature des produits de réactions (Rodriguez-Aller et al., 2013).

* La chromatographie sur couche mince (CCM) permet de séparer les différents composés d'une réaction à l'aide d'une phase mobile (liquide) qui migre par capillarité à travers une phase stationnaire, généralement un gel de silice, sur laquelle est déposé l'échantillon. C'est une méthode simple qui demande peu d'investissement matériel (Pittet and Royer, 2002) où les composés sont identifiés en fonction de leurs distances de migration. Les molécules n'étant pas altérées par la méthode, elles peuvent être récupérées pour confirmer leur nature par d'autres méthodes d'analyse (Poole, 2003). Il est toutefois reproché à la CCM un manque de sensibilité et de résolution. La bio-autographie se définit comme la CCM associée à une détection biologique (Marston et al., 2002). Elle offre une alternative rapide et à haut-débit pour le criblage de métabolites secondaires enzymatique et pour de potentielles activités inhibitrices des enzymes (Choma and Grzelak, 2011; Bräm and Wolfram, 2017).

* La résonance magnétique nucléaire (RMN) est probablement la plus onéreuse des méthodes de criblage citées ici, mais aussi une des plus sensibles et informatives (Irague et al., 2011). Elle permet, d'une part de suivre l'évolution des différents produits de la réaction au cours du temps et d'autre part, de déterminer avec précision la conformation (sites de liaisons, énantiosélectivité, anomérie) des produits de synthèse (Reetz et al., 2002), informations qui ne sont pas toujours accessibles avec les autres méthodes de criblage.

Quel que soit la méthode utilisée, l'expérimentateur doit pouvoir lier le phénotype de l'enzyme améliorée à sa séquence génétique, sans quoi il est impossible d'identifier les mutations qui améliorent cette activité ou cette propriété.

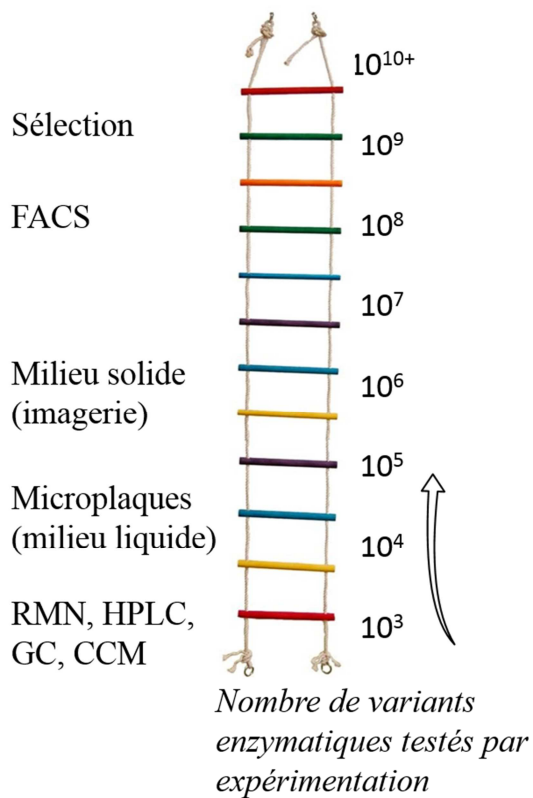


Figure 13 : Taille des banques qu'il est possible de cribler en fonction de la méthode utilisée. Modifié d'après (Dietrich et al., 2010).

Du fait de leur utilisation dans de nombreuses applications industrielles, les enzymes actives sur la biomasse végétale sont des candidates particulièrement intéressantes pour ces approches d'ingénierie enzymatique. L'origine des ressources en carbone renouvelable issues de la biomasse est très diverse. Bien que peu onéreuse et abondante, elle peut varier fortement en termes de structure.

3. Sources de carbone renouvelable pour la chimie verte

a. Bilan carbone

Si l'on regarde d'une manière globale la répartition des stocks et des flux de carbone à l'échelle du système planétaire, on constate que le bilan global des flux de carbone échangés avec l'atmosphère est positif (Figure 14). Cela signifie que l'atmosphère s'enrichit progressivement en carbone, principalement sous forme de dioxyde de carbone (CO_2). Même si le flux émis dans l'atmosphère, dû à l'exploitation par l'homme du carbone fossile, est faible en comparaison des échanges de carbone entre l'atmosphère et les océans ou les organismes vivants, il suffit à rendre le solde de ces "échanges de carbone" positif (Colonna, 2012). L'étude de l'évolution climatique montre que l'enrichissement de l'atmosphère en CO_2 participe au réchauffement climatique, et ce réchauffement est donc

principalement d'origine anthropique (Huber and Knutti, 2011), principalement à cause de l'exploitation du carbone fossile par l'homme et de sa transformation en CO₂ (Solomon et al., 2009). Il est à noter que sans l'action de l'homme, l'atmosphère verrait sa teneur en CO₂ diminuer ; la stabilisation de la concentration atmosphérique en CO₂ repose donc sur une maîtrise de ce flux.

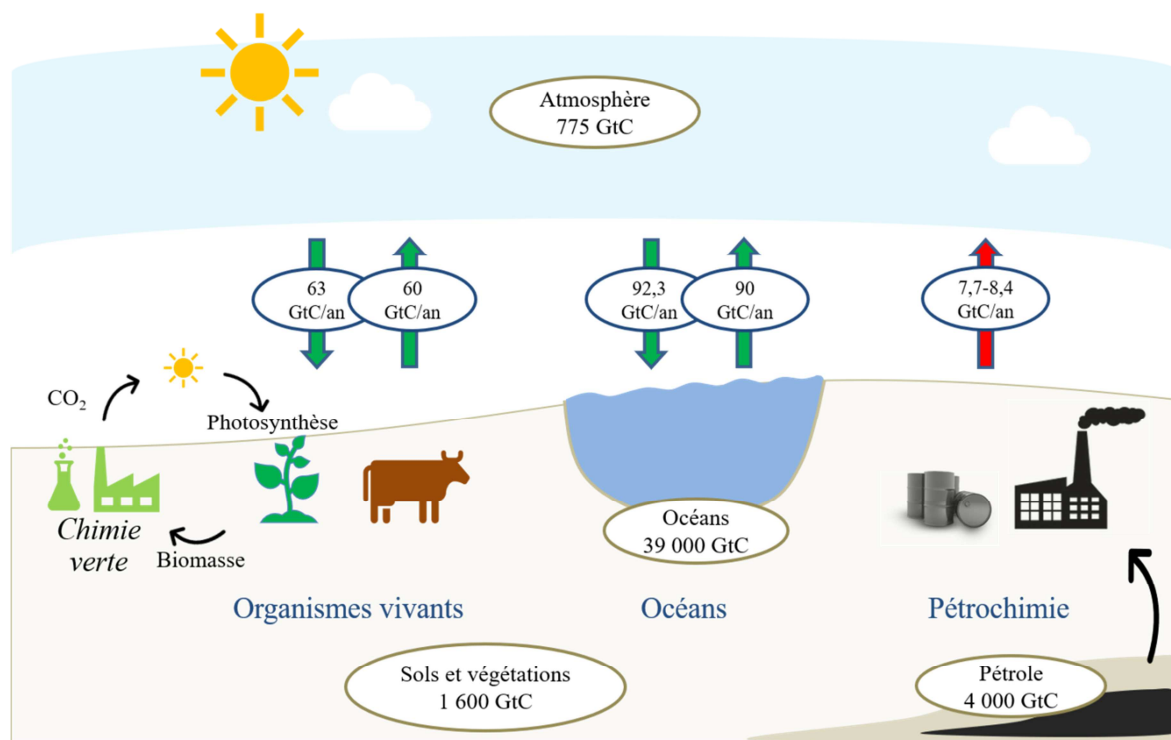


Figure 14 : Flux et répartition des stocks de carbone terrestre. Modifié d'après (Colonna, 2012). Valeurs données en milliards de tonnes de carbone (GtC), et (GtC/an) pour les flux.

Il est complexe de substituer directement le carbone fossile des procédés actuels par le carbone stocké sous forme de CO₂ dans l'atmosphère. Peu de procédés industriels utilisent le CO₂ comme matière première car une grande quantité d'énergie est nécessaire pour transformer ce carbone sous sa forme la plus oxydée. Les réactions utilisant du CO₂ comme matière première n'améliorent pas nécessairement l'effet de serre car la transformation du CO₂ nécessite tellement d'énergie que les rejets de CO₂ sont souvent supérieurs aux quantités de CO₂ consommé (Sakakura et al., 2007), malgré les développements récents dans le domaine de l'hydrogénation catalytique du CO₂ (Wang et al., 2011).

b. Origine du carbone renouvelable

La biomasse représente *a priori* une ressource carbonée abondante, renouvelable, plus facilement exploitable que le CO₂, dans le respect des principes de la chimie verte. La biomasse se définit comme la fraction biodégradable des produits et déchets provenant de l'agriculture, des substances végétales et animales issues de la terre et de la mer, de la sylviculture et des industries connexes, ainsi que la

fraction biodégradable des déchets industriels et ménagers (loi de programme n°2005-781 du 13 juillet 2005). La biomasse végétale résulte de la photosynthèse, processus bioénergétique qui transforme, grâce à l'énergie solaire, le CO₂ présent dans l'atmosphère et l'eau présente dans les plantes en diverses molécules complexes qui contribuent à la structure et aux réserves de carbone des végétaux (Figure 15).

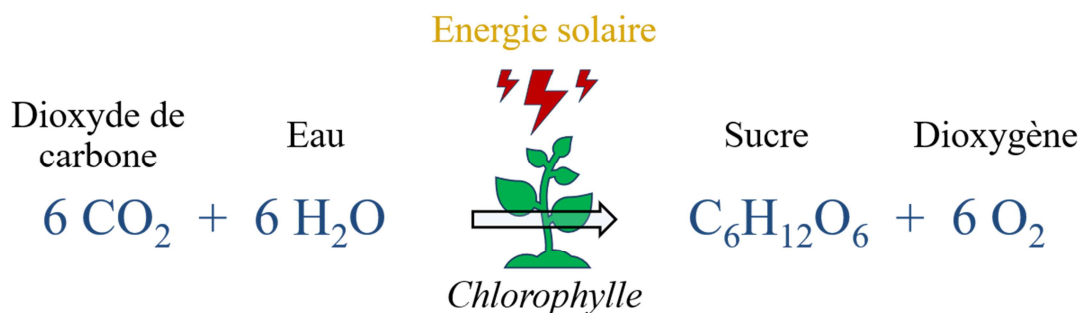


Figure 15 : **Réaction simplifiée de la photosynthèse oxygénique.**

La biomasse comprend quatre familles biochimiques de polymères carbonés pouvant répondre aux attentes de la chimie verte (Colonna, 2012) :

(1) **La biomasse lignocellulosique** représente le type de biomasse le plus abondant sur la surface terrestre. Sa production annuelle a été estimée à 10¹⁰ Mt dans le monde (Sánchez and Cardona, 2008). Elle se compose de cellulose (homopolymère d'unités D-glucosyles liées en β-1,4), d'hémicelluloses (polymère complexe ramifié composé de différents types de sucres tels que du D-glucose, du D-xylose, du D-mannose, du D-galactose, du L-rhamnose ou du L-arabinose) et de lignine (biopolymère de monolignols constituant un réseau tridimensionnel hydrophobe complexe) qui sont les constituants majeurs des parois des cellules végétales. Leurs proportions respectives dans les parois végétales varient en fonction des espèces et du stade de développement de la plante (Figure 16). Cette biomasse, dont la diversité de structures permet de nombreuses applications en chimie verte (Isikgor and Becer, 2015) peut être utilisée de deux manières différentes :

- Soit comme une ressource naturelle de structures polymériques organisées et directement exploitables pour leurs propriétés fonctionnelles avérées, comme biomatériaux ou pour l'habillement par exemple. C'est le cas du bois, du chanvre ou encore du coton qui sont composés quasi-exclusivement de cellulose. La cellulose, de par sa structure simple et son abondance naturelle, est également largement exploitée dans divers domaines comme le papier ou les biomatériaux (Dhepe and Fukuoka, 2008). Plus récemment, de nouvelles applications à base de cellulose sous sa forme cristalline ont vu le jour (Pachau, 2015). Mahmoud et al., ont par exemple couplé des nano-cristaux de cellulose à des nanoparticules d'or pour créer des

supports d'immobilisation de protéines. Une cyclodextrine glycosyltransférase et une alcool oxydase ont été alors utilisées comme enzymes modèles avec de très bons rendements de récupération (Mahmoud et al., 2009).

- Soit comme un réservoir de polymères, d'oligomères et de molécules, qui peuvent servir de précurseur dans diverses applications de la chimie ou comme molécules fermentescibles pour la production de bioéthanol (Alvira et al., 2010). La décomposition de cette biomasse lignocellulosique en monomères fermentescibles ouvre la voie aux conversions biotechnologiques et la diversité des molécules présentes dans ce type de biomasse élargit considérablement le spectre des molécules biosourcées disponibles (Kikuzaki et al., 2002; Zakzeski et al., 2010; Dumon et al., 2012). Parmi la variété des produits qu'il est possible d'obtenir à partir du D-glucose extrait de la cellulose, l'acide lactique, l'éthanol, l'acide acétique et l'acide lévulinique, obtenus par voies microbienne ou chimique sont des composés particulièrement intéressants pour de nombreuses industries (Kamm and Kamm, 2004). L'exploitation, longtemps plus marginale, des pentoses, présents uniquement dans les hémicelluloses, est en voie de développement. Ce retard est dû à une complexité et une diversité plus grande des structures polysaccharidiques présentes dans l'hémicellulose. Par voie de conséquence, cette complexité nécessite un arsenal enzymatique large, et de nombreuses études se consacrent à la découverte de ces biocatalyseurs ou à leur optimisation (Dumon et al., 2012). La valorisation récente des pentoses est aussi passée par la mise au point de procédés de synthèse enzymatique de xylo-oligosaccharides qui possèdent un intérêt nutritionnel, entre autres comme prébiotiques (Grootaert et al., 2007). Un autre composant de l'hémicellulose, l'acide férulique possède des propriétés antioxydantes (Graf, 1992) intéressantes pour l'industrie cosmétique, alimentaire et pharmaceutique (Kikuzaki et al., 2002; Ou and Kwok, 2004; Horbury et al., 2016) et qui peut également être utilisé comme précurseur dans la synthèse de vanilline (Gallage et al., 2014).

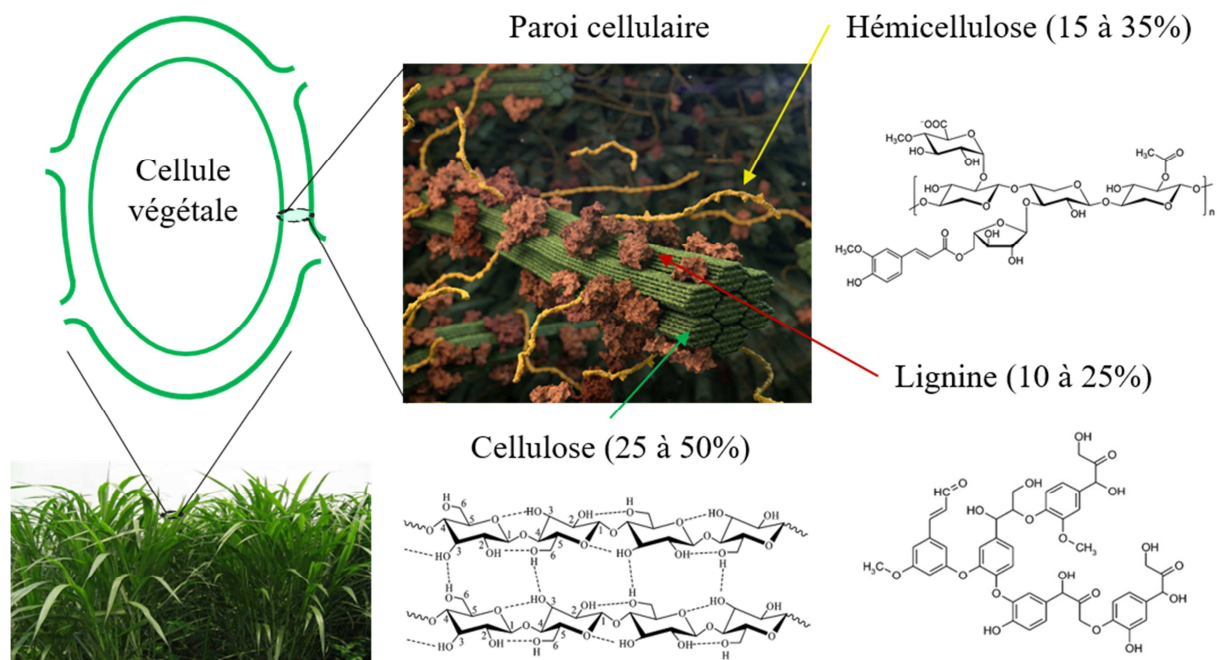


Figure 16 : Composition de la biomasse lignocellulosique et pourcentage moyen de chacun des trois composants de la lignocellulose (valeurs en % variable en fonction des végétaux, d'après (Isikgor and Becer, 2015)).

(2) **L'amidon** se compose de deux homopolymères, l'amylose ($\approx 25\%$) et l'amylopectine ($\approx 75\%$) (French, 1973), deux homopolymères de glucose (comme la cellulose), qui se différencient par leur degré de ramification et leur degré de polymérisation. L'amidon se présente sous la forme de petits granules dans les graines, les fruits, les feuilles, les tubercules et les bulbes de certaines plantes, et sert de réserve énergétique. Dans les tubercules de pomme de terre, l'endosperme de maïs, la patate douce ou les racines de manioc et d'igname par exemple, la teneur en amidon varie entre 65 et 90 % du poids sec total (Preiss, 2009). La majorité des 2 500 millions de tonnes estimées de récoltes annuelles d'amidon est destinée à l'alimentation humaine ou animale, mais une partie de plus en plus conséquente est utilisée pour la production de molécules non comestibles (Zeeman et al., 2010), telles que le bioéthanol par exemple, dont les différentes étapes du procédé ont été largement optimisées (Bothast and Schlicher, 2005).

(3) **Le saccharose** ou l' α -D-glucopyranosyl-(1 \leftrightarrow 2)- β -D-fructofuranoside est un diholoside formé d'une molécule de D-glucose et d'une molécule de D-fructose. Le saccharose est principalement extrait de la canne à sucre et de la betterave sucrière. Une fois raffiné, il est l'un des ingrédients les plus utilisés dans l'industrie agro-alimentaire pour son pouvoir sucrant (sucre de table). Le saccharose est l'un des substrats de fermentation les plus classiques car un grand nombre de micro-organismes peuvent le consommer. Le saccharose permet aussi d'obtenir par voie enzymatique des biopolymères de type dextrans, polysaccharides de D-glucoses liés par des liaisons α -(1,6), linéaires ou ramifiés, qui sont

très utilisés entre autre en pharmacologie (Wang et al., 2016). Sous sa forme estérifiée avec un acide gras, le saccharose est employé dans l'industrie cosmétique et suscite depuis peu un grand intérêt dans le domaine pharmaceutique. En effet, ces esters sont des excipients biodégradables affichant un comportement émulsifiant et solubilisant avéré (Szűts and Szabó-Révész, 2012). Au vu de l'intérêt que représente le saccharose, des modifications de génomes ont été réalisées sur certaines plantes dans le but d'augmenter leur production et donc le rendement d'extraction du saccharose. Ces travaux, bien que prometteurs dans un premier temps, se sont trouvés confrontés à des phénomènes d'évitement par contournement des voies métaboliques et les effets s'atténuent pour redevenir identiques aux rendements des plantes d'origine (Patrick et al., 2013).

(4) **Les lipides.** Les acides gras sont essentiels à toutes formes de vies et se retrouvent dans divers types de lipides qui, en milieu aqueux, s'organisent en réseaux bidimensionnels structurant toutes les membranes biologiques. Les acides gras libres ou acides carboxyliques aliphatiques, se composent d'un groupe carboxyle et d'une chaîne carbonée de 4 à 28 atomes de carbone, généralement en nombre pair. Un acide gras insaturé comporte une ou plusieurs doubles liaisons carbone-carbone contrairement aux acides gras saturés qui ne possèdent aucune double liaison. Les acide gras ou dérivés sont extraits directement de certains végétaux de manière industrielle (Merrien et al., 2012). Ces lipides issus de végétaux sont à la base de nombreuses applications comme par exemple, la production de polyuréthane nécessaire pour la synthèse de thermoplastique, et offrent donc ainsi une alternative sérieuse à la synthèse de plastiques issus de la pétrochimie (Maisonneuve et al., 2016). De plus, la France impose depuis 2015 la présence de 10 % de biocarburant, le plus souvent des esters méthyliques ou éthyliques issus d'huile de Colza ou de tournesol, dans les carburants fossiles. L'augmentation de la consommation de lipides que nécessite la production de biocarburants ne peut être durable sans améliorer les performances quantitatives, qualitatives et écologiques des différentes variétés d'oléagineux (Merrien et al., 2012). Pour répondre aux défis imposés par cette surexploitation, des programmes scientifiques ont permis d'optimiser différentes souches de colza, par croisement ou par génie génétique, en fonction des attentes spécifiques à chaque industrie (composition en acides gras spécifiques pour les huiles alimentaires, fort rendement pour les biocarburants, etc.) (Abadi and Leckband, 2011). Une alternative réside dans la production industrielle d'acides gras par des microorganismes à partir de matières premières renouvelables (Steen et al., 2010; Zhou et al., 2016), et s'accompagne d'un accès nouveau à toute une diversité moléculaire méconnue (présence de doubles liaisons, de fonctions époxydes) dont le potentiel est immense (Adrio, 2017; Chew et al., 2017).

Ces quatre sources de carbone renouvelable peuvent donc se substituer au pétrole dans un grand nombre d'applications (production de biopolymères, détergents, cosmétiques, biocarburants, etc.) - sous réserve que l'outil enzymatique adéquat soit disponible- et s'inscrivent pleinement dans le cahier des charges de la chimie verte et de la bio-économie. Il reste toutefois à préciser que contrairement à la

lignocellulose, l'amidon, le saccharose, ou les lipides provenant des oléagineux font partie des ressources alimentaires exploitées par l'homme. Leur exploitation peut générer une augmentation du prix des denrées alimentaires qu'il est important de maîtriser (O'Donohue, 2013 Colloque biotechnologie : relever le défi du carbone renouvelable).

c. Déconstruction de la biomasse

Pour exploiter la ressource végétale carbonée et synthétiser de nouvelles molécules biosourcées, le premier objectif consiste à déconstruire ces structures, principalement polysaccharidiques, afin d'en collecter les briques élémentaires. Or, les parois des cellules végétales ont évolué pour se défendre contre les agressions externes qui peuvent être mécaniques, thermiques, chimiques ou biologiques (Silveira et al., 2013). A l'état naturel, la cellulose comme les polymères polyphénoliques de la lignine présentent une structure cristalline, qui rend difficile l'accès des enzymes lignocellulolytiques à leurs substrats. Pour la production de bioéthanol issu de la lignocellulose, des prétraitements thermo-chimiques permettent de faciliter l'accessibilité des enzymes hydrolytiques aux fractions polysaccharidiques. La déconstruction peut alors se faire avec des dosages enzymatiques plus faibles et sur des temps de réaction considérablement réduits (Alvira et al., 2010). Lors du projet européen BIOCORE (BIO-COMmodity REfining) -focalisé sur la faisabilité industrielle d'un concept de bioraffinerie pour la conversion de sous-produits céréaliers (paille notamment), de résidus forestiers ou de cultures ligneuses en un large éventail de produits (biocarburants, polymères, matériaux)-, la société CIMV SA (Levallois-Perret, France) a mis au point une méthode permettant une séparation en trois fractions distinctes et exploitables des trois composants majeurs de cette matière première, sous des formes non-dénaturées et donc potentiellement valorisables: la cellulose, l'hémicellulose et la lignine. Cette technologie a été validée sur plusieurs types de ressources lignocellulosiques, pour démontrer l'adaptabilité de ce procédé (Delmas, 2008).

Une fois l'accessibilité aux enzymes améliorée par des prétraitements, la dépolymérisation complète des différents types de lignocellulose nécessite un arsenal d'enzymes très variées. La Figure 17 montre un exemple de la diversité des enzymes nécessaires à l'hydrolyse des différentes unités présentes sur une chaîne d'arabinoxylane qui n'est qu'une des nombreuses structures possibles de l'hémicellulose. Une des stratégies pour diminuer le coût de l'hydrolyse est d'élaborer des cocktails enzymatiques sur mesure, contenant toutes les activités nécessaires, spécifiques à chaque type de biomasse (Zhou et al., 2009; Gao et al., 2010).

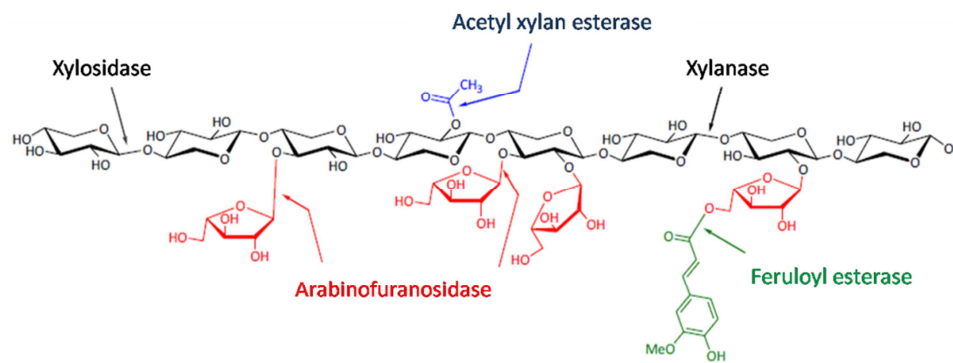


Figure 17 : **Structure d'une chaîne d'arabinoxylane**. Les flèches indiquent les sites de clivage pour chaque enzyme impliquée dans l'hydrolyse des différentes unités.

Pour identifier des enzymes permettant de dégrader efficacement la biomasse végétale, il suffit d'observer attentivement la Nature qui offre une très grande diversité de micro-organismes (bactéries, levures, champignons), capables de dégrader ces différentes variétés de biomasse végétale et de les utiliser comme source de carbone pour leur croissance. Ces micro-organismes peuvent se développer soit en milieu naturel, au contact direct des différentes biomasses (sols, humus, plantes) ou dans les systèmes digestifs d'insectes ou de mammifères qui se nourrissent de végétaux (termites, vers, bovins...) (Ufarte et al., 2015). Ces micro-organismes possèdent donc des enzymes capables de dégrader ces structures polysaccharidiques (Gao et al., 2010) ou lipidique (Michely et al., 2013) pour les rendre digestibles.

4. Synthèse enzymatique de molécules à forte valeur ajoutée issues de la biomasse

Comme indiqué précédemment, les enzymes sont classées dans l'Enzyme Commission number (EC number) selon le type de réaction qu'elles catalysent. Si l'on recherche des enzymes dédiées aux réactions de synthèse, il est logique de s'intéresser en premier lieu aux enzymes dont l'activité première est de catalyser des réactions de transfert entre molécules. Ces enzymes appartiennent à la famille des transférases. Étonnamment, certaines enzymes appartenant à la famille des hydrolases vont se révéler tout aussi pertinentes pour la synthèse de molécules issues de la biomasse.

a. Les transférases

Les transférases regroupent des enzymes qui sont directement dédiées à la catalyse de réactions de synthèse entre deux molécules organiques. Elles utilisent un substrat 'donneur' qui comprend le groupement fonctionnel à transférer, et le substrat qui reçoit ce groupement est appelé 'accepteur'. Le

code EC correspondant aux transférases commence par 2, le second nombre de 1 à 10 dépend de la nature du groupement transféré :

Table 2 : Classification des transférases dans l'Enzyme Commission number (EC number)

EC 2.1	enzymes transférant un groupe à 1 carbone (méthyltransférase)
EC 2.2	enzymes transférant un groupe carbonyle (aldéhyde ou cétone)
EC 2.3	acyltransférases
EC 2.4	glycosyltransférases
EC 2.5	enzymes transférant un groupe alkyle ou aryle autre que méthyle
EC 2.6	enzymes transférant un groupe azoté (transaminase)
EC 2.7	enzymes transférant un groupe phosphoré (phosphotransférase, mais aussi polymérase et kinase)
EC 2.8	enzymes transférant un groupe sulfuré (sulfurtransférase et sulfotransférase)
EC 2.9	enzymes transférant un groupe contenant du sélénium
EC 2.10	enzymes transférant un groupe contenant du molybdène ou du tungstène.

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2/>

Glycosyltransferases :

Dans les voies de biosynthèse des polysaccharides naturels, la formation des liaisons glycosidiques est principalement catalysée par des glycosyltransférases (GT), dont les principaux substrats donneurs sont des monosaccharides liés à un groupement partant nucléotidique de type sucres mono- ou di-phospho-nucléotidiques autrement appelés : donneurs de Leloir (Figure 18 A). Les GT utilisant ces substrats sont alors appelées enzymes de Leloir. Moins fréquemment, on trouve parmi les substrats possibles des sucres activés par des groupements phosphate ou polyprénol phosphate (Figure 18 B) (Lairson et al., 2008; Breton et al., 2012). Par exemple, la cellulose naturelle est synthétisée *in vivo* à partir d'uridine-5'-diphosphoglucose (UDP- α -D-Glcp) via la cellulose synthase (EC 2.4.1.12) (Scheible and Pauly, 2004; Somerville, 2006).

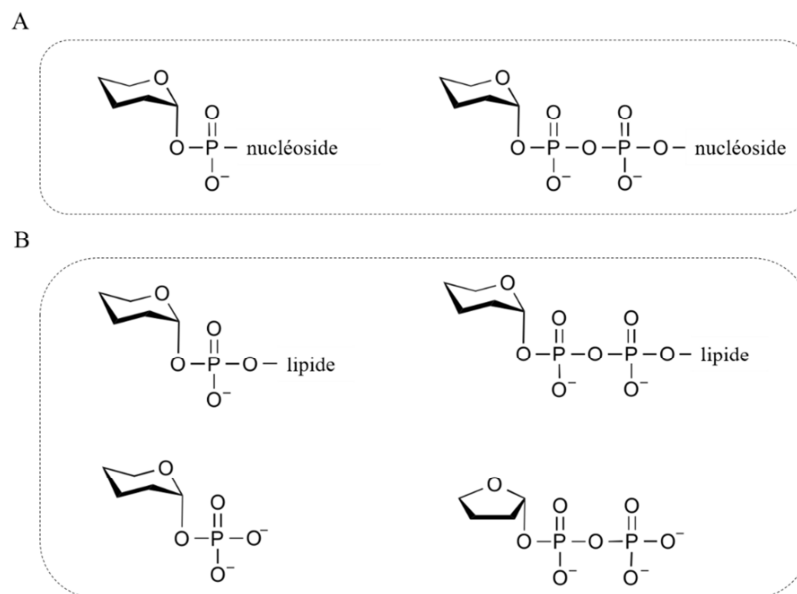


Figure 18 : **Substrats naturels des GTs.** (A) Sucres mono- ou di-phospho-nucléotidiques (donneurs de Leloir); (B) Donneurs non-nucléotidiques (non-Leloir), ils sont de type phosphate de polyprénol, pyrophosphate de polyprénol, phosphate de glycosyle ou 1-pyrophosphate de glycosyle.

La synthèse enzymatique de molécules glycosylées complexes *in vitro* a débuté dans les années 1980 à l'aide de GT (Wong et al., 1982).

Bungaruang et al. proposent en 2013 un procédé de synthèse enzymatique de la nothofagine, un antioxydant de type flavonoïde présent dans le thé rouge, par 3'-C- β -D-glycosylation parfaitement sélective de la phlorétine, une molécule végétale abondante. Dans ce procédé multi-étapes, la saccharose synthase de *Glycine max* (soja) permet d'abord de produire de l'UDP-D-Glcp en utilisant du saccharose qui sert de donneur de glycosyles; simultanément, la C-glycosyltransférase d'*Oryza sativa* (riz) glycosyle la phlorétine à partir de l'UDP-D-Glcp produit. Cette réaction génère du D-fructose comme produit secondaire mais présente un rendement de conversion de la phlorétine supérieur à 90 % lorsque le saccharose est en large excès (Bungaruang et al., 2013).

Les GT présentent l'avantage de catalyser des réactions de glycosylation irréversibles, car elles ne peuvent pas hydrolyser les produits formés (Frankova and Fry, 2013). Néanmoins, elles ont généralement besoin de substrats donneurs activés onéreux et/ou difficiles à synthétiser (Figure 18 A), le nombre d'enzymes caractérisées disponibles est très limité et l'activité de ces enzymes est souvent faible (même si une faible activité n'est pas incompatible avec de très bons rendements de synthèse). En effet, ces enzymes sont produites par les végétaux pour la synthèse de leurs parois cellulaires, qui est un processus lent et la sélection naturelle n'a pas forcément optimisé ces enzymes sur leur efficacité catalytique. Les GT sont généralement des protéines de type transmembranaire, ce qui rend leur production difficile. De plus, leurs substrats sont monosaccharidiques, ils ne permettent pas le

transfert de plusieurs sucres en une étape (Lairson et al., 2008). Le plus grand défi restant dans ce domaine est la caractérisation fonctionnelle des GT car la spécificité de donneur et d'accepteur n'est pas encore connue pour la grande majorité des 360 000 séquences codantes répertoriées (octobre 2017) -le plus souvent putatives- de cette classe d'enzymes (Breton et al., 2012). En outre, les mécanismes catalytiques de certaines GT restent encore non totalement définis, toutes ces raisons présentent des freins aux travaux d'évolution de cette classe d'enzymes (Lairson et al., 2008; Breton et al., 2012).

Féruylotransférases :

Dans la nature, les féruylotransférases sont impliquées dans la synthèse de la subérine (Molina et al., 2009), un hétéropolymère cireux constitué de domaines poly-aliphatiques et poly-aromatiques, qui sert de barrière lipophile à la paroi cellulaire primaire des végétaux. La subérine permet de contrôler les flux d'eau et de solutés, de contribuer à la résistance de la paroi cellulaire et constitue une barrière naturelle contre les pathogènes (Nawrath, 2002). Lors de la synthèse de la subérine, les féruylotransférases catalysent le transfert du féruylate à partir de féruyl-CoA sur des acides gras ω -hydroxylés et des alcools gras (Serra et al., 2010). Les féruylotransférases sont aussi impliquées dans la synthèse de l'hémicellulose et de la pectine via le transfert d'acide féruylque sur des résidus L-arabinose ou D-galactose des polysaccharides pectiques et hémicellulosiques. Ces groupements féruyls sont alors la principale cause de formation de réticulations internes dans ces polymères structuraux de la paroi cellulaire (Lenucci et al., 2009).

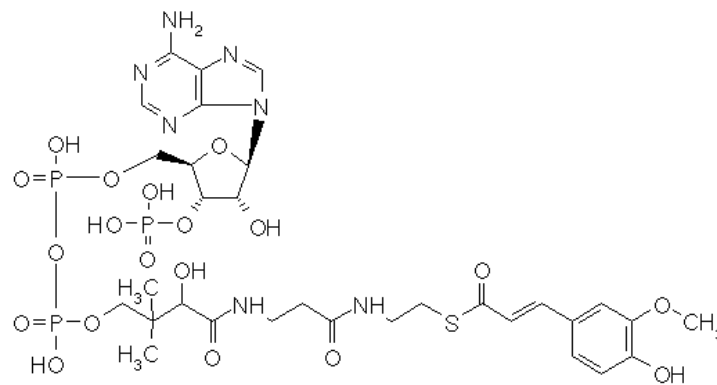


Figure 19 : **Représentation du féruylotransférases, substrat naturel des féruylotransférases.**

De même que pour les GT, peu de féruylotransférases sont à ce jour caractérisées et leur efficacité catalytique est faible. Le seul exemple d'application relevé dans la littérature utilisant une possible féruylotransférase (expérimentation réalisée sur extraits bruts enzymatiques) concerne le transfert d'acide caféique ou féruylque sur un flavonoïde glucosylé (la quercitrine ou quercétine 3-O- β -D-glucopyranoside) pour un rendement optimisé de 38 % en ajoutant à l'extrait cellulaire, 10 μ M d'acide caféique comme donneur, 10 μ M d'ATP et 10 μ M de coenzyme A. Le rendement obtenu est 4 fois plus faible avec l'acide féruylque (Nobuyoshi et al., 2000). Les freins à l'utilisation des féruyls

transférases pour la synthèse de molécules féruloylées sont très similaires à ceux rencontrés pour les GT. En effet, ces enzymes nécessitent des substrats activés de type féruloyl coenzyme A onéreux (Figure 19) (Ferrer, 2005; Mastrangelo et al., 2009).

b. Les hydrolases pour la synthèse de molécules

Enzymes hydrolytiques.

Le mot hydrolyse vient du grec *hydro* (eau) et *luein* (briser), et correspond à une réaction chimique qui consiste à rompre une liaison covalente par action d'une molécule d'eau. Les enzymes hydrolytiques ou hydrolases, qui catalysent cette réaction, sont réparties en 13 sous-groupes, en fonction du type de liaison hydrolysée (Table 3). Certaines enzymes hydrolytiques sont difficilement classables car elles présentent une large promiscuité de substrat pouvant correspondre à plusieurs sous-groupes, voire une promiscuité catalytique pouvant correspondre à d'autres familles d'enzymes (Faber, 2011).

Table 3 : Classification des hydrolases dans l'Enzyme Commission number (EC number)

	Hydrolyse de liaisons :
EC 3.1	ester
EC 3.2	impliquant un sucre
EC 3.3	éther
EC 3.4	peptidique (peptidases)
EC 3.5	carbone-azote, autre que les liaisons peptidiques
EC 3.6	regroupe les enzymes actives sur les anhydrides d'acide
EC 3.7	carbone-carbone
EC 3.8	carbone-halogène
EC 3.9	azote-phosphore
EC 3.10	soufre-azote
EC 3.11	carbone-phosphate
EC 3.12	soufre-soufre
EC 3.13	carbone-soufre

Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2/>

Dans les organismes vivants, les polymères, tels que les protéines, les nucléotides, les polysaccharides ou encore les graisses sont catabolisés par cette classe d'enzymes. Parmi les hydrolases, les plus communes sont les lipases, les glycoside hydrolases (GH) ou les protéases qui hydrolysent les graisses, les sucres et les protéines, respectivement.

Les principales caractéristiques qui ont fait des hydrolases la classe enzymatique la plus utilisée par les biochimistes au cours des deux dernières décennies sont leur indépendance vis à vis des cofacteurs, le fait que beaucoup affichent une promiscuité de substrat très large et le grand nombre d'enzymes

disponibles (Faber, 2011). Les enzymes lignocellulolitiques sont généralement produites par des micro-organismes (levures, champignons ou bactéries) qui se nourrissent des sucres lignocellulosiques. Chaque micro-organisme a développé et adapté ses propres cocktails d'enzymes ou des "collaborations symbiotiques" au sein d'une communauté microbienne permettant de dégrader cette structure lignocellulosique complexes (Wongwilaiwalin et al., 2010). Ces enzymes de déconstruction de la lignocellulose se sont donc adaptées non seulement à la spécificité de structure de cette biomasse à dégrader mais aussi aux conditions naturelles environnementales de l'organisme qui les produit (température, pH, etc.) (Kumar et al., 2008). Les hydrolases sont donc bien plus abondantes et présentent une grande variété de spécificité physico-chimique.

C'est bien entendu principalement pour leur activité hydrolytique que cette catégorie d'enzymes est utilisée, mais certaines enzymes classées parmi les hydrolases sont aussi capables de former des liaisons pour la synthèse de molécules. Pour ces raisons, ce n'est donc pas vers les transférases que les industriels se tournent majoritairement mais plutôt vers les hydrolases pour réaliser des réactions de synthèse en les faisant fonctionner en mode synthèse.

Les transglycosidases.

Même si la plupart des GH ont comme fonction principale l'hydrolyse de liaisons glycosidiques - particulièrement intéressante de nos jours notamment dans le domaine de la déconstruction de la biomasse lignocellulosique (Xu, 2011; Guerriero et al., 2016)-, certaines enzymes présentent une activité principale de transfert et sont appelées transglycosidases (TG). Parmi elles, on trouve les glucane-saccharases (Monsan et al., 2010) ou les xylogluco-endotransglycosylases (Rose, 2002) qui catalysent quasi exclusivement des réactions de transglycosylation.

Classées au sein de mêmes familles, GH et TG s'appuient sur des mécanismes réactionnels identiques, qui ont été décrits pour ces hydrolases pour la première fois en 1953 par Daniel Koshland (Koshland, 1953).

Le mécanisme dit de **rétenition** du centre anomérique (rGH) conserve l'anomérie α ou β du donneur de glycosyle (Figure 20). Il se compose de deux étapes successives d'inversion de configuration et passe par la formation d'un intermédiaire covalent glycosyl-enzyme. Ce mécanisme enzymatique met en jeu deux acides aminés catalytiques portant des fonctions acides carboxyliques et distants en moyenne de 5,5 Å. Ici, la fonction carboxylate du résidu catalytique nucléophile attaque le carbone anomérique du sucre de manière concomitante à la protonation de l'oxygène inter-glycosidique participant à la liaison qui va être rompue par l'autre résidu catalytique acide/base jouant le rôle d'acide, facilitant le départ de la partie aglycone du substrat et menant ainsi à la formation de l'intermédiaire covalent glycosyl-enzyme. C'est l'étape dite de glycosylation de l'enzyme. Dans un second temps, le résidu catalytique acide/base (jouant alors le rôle de base) active une molécule d'eau permettant son attaque nucléophile sur le carbone anomérique de l'intermédiaire glycosyl-enzyme. Il en résulte une hydrolyse avec

rétenion de l'anomérie de départ du substrat (Withers, 2001). C'est l'étape de déglycosylation de l'enzyme.

La réaction est appelée **transglycosidation** lorsque la fonction alcool d'une molécule acceptrice, attaque l'intermédiaire glycosyl-enzyme à la place d'une molécule d'eau. On peut également parler de réaction de condensation lorsque le donneur (substrat de départ) et l'accepteur sont deux molécules différentes, et dans le cas particulier où le donneur joue en même temps le rôle d'accepteur, on parle alors d'auto-condensation (Rather et al., 2013). Les trois espèces (le donneur, l'accepteur et l'eau) entrent alors en compétition pour attaquer l'intermédiaire covalent formé et en fonction de l'espèce qui prend le dessus, il y a apparition de produits d'hydrolyse, de condensation (transglycosidation) et/ou d'autocondensation avec des proportions variables (Mangas-Sánchez and Adlercreutz, 2015). Toutes les transglycosidases naturelles connues à ce jour possèdent un mécanisme catalytique de rétenion de configuration.

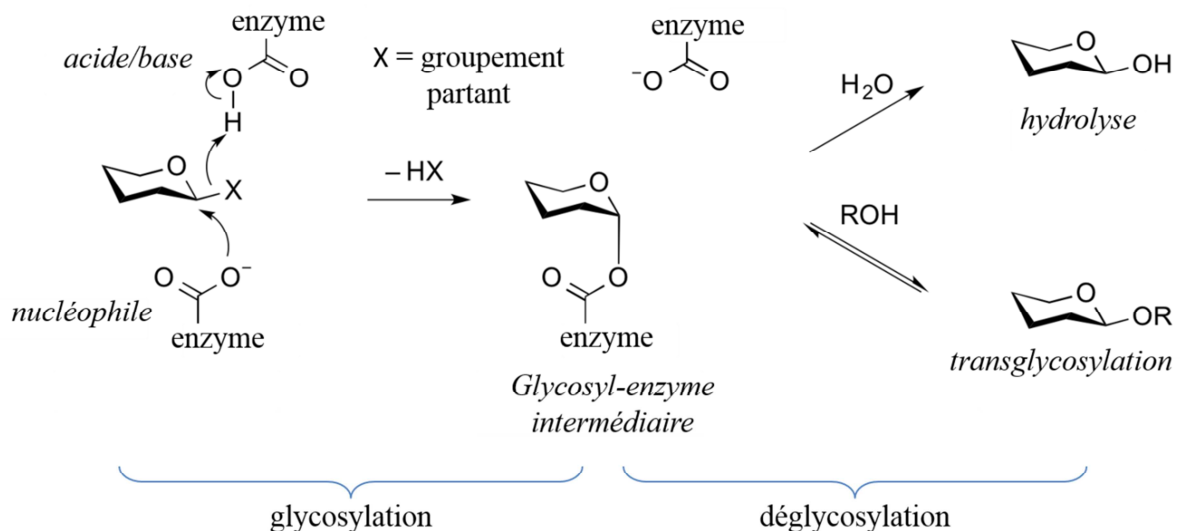


Figure 20 : Mécanisme catalytique dit de "rétenion de configuration" des rGH. La première étape génère un intermédiaire covalent glycosyl-enzyme et la deuxième étape détermine si la réaction est hydrolytique par ajout d'une molécule d'eau, ou de transglycosylation par liaison sur une fonction hydroxyle d'une molécule acceptrice (d'après <https://www.cazypedia.org/index.php/Transglycosylases>).

La comparaison entre ces TG naturelles et leurs GH équivalentes -c'est-à-dire synthétisant ou dégradant les mêmes molécules respectivement- montre des structures pouvant être très proches, mais les déterminants précis qui régissent la partition des activités hydrolytique ou de synthèse sont trop subtils pour être détectés par une simple comparaison de leurs structures moléculaire (Bissaro et al., 2015b).

Le second mécanisme décrit pour l'hydrolyse de molécules glycosylées est dit d'inversion car la réaction inverse l'anomérie α ou β du substrat donneur. Il implique deux acides aminés portant des fonctions acides carboxyliques agissant par un mécanisme acide/base. Ce mécanisme se déroule en une étape et la distance entre les deux résidus catalytiques est de 7 à 10 Å. Le résidu catalytique basique active une molécule d'eau qui attaque le carbone anomérique du sucre donneur, simultanément à la protonation de l'oxygène de la liaison glycosidique scissile par le résidu catalytique acide de l'enzyme (Withers, 2001). Ce mécanisme reste à ce jour spécifique des hydrolases, dans la mesure où aucune transglycosydase naturelle à inversion de configuration n'a été encore décrite.

Cas particulier des glycosynthases.

En 1998, une nouvelle stratégie d'ingénierie enzymatique pour la synthèse de sucres est proposée à partir de certaines rGH donnant naissance au concept de "glycosynthases" (Mackenzie et al., 1998; Malet and Planas, 1998). Les glycosynthases sont des enzymes dans lesquelles le résidu catalytique nucléophile est muté, généralement en alanine, en serine ou en glycine, de telle sorte que l'enzyme n'est plus capable de former l'intermédiaire glycosyl-enzyme. Pour restaurer une activité catalytique à cette protéine, il faut alors utiliser des substrats saccharidiques possédant un atome de fluor comme groupement partant dans une configuration anomérique inversée par rapport au substrat naturel. L'atome de fluor porté par le donneur mime l'intermédiaire covalent glycosyl-enzyme mais accroît la barrière énergétique de l'étape de "déglycosylation", ainsi seules des interactions avec la molécule acceptrice peuvent provoquer la réaction de synthèse (Ducros et al., 2003; Hidaka et al., 2010). Même si les glycosynthases ont une activité relativement faible et qu'elles nécessitent des substrats fluorés, elles peuvent être optimisées par ingénierie et sont utilisées dans de nombreuses réactions de synthèses (Wang and Huang, 2009). Depuis 2006, quelques exemples de glycosynthases basées sur des GH à mécanisme d'inversion de configuration ont été décrites (Honda and Kitaoka, 2006) ce qui élargit encore les enzymes candidates à leur conversion en glycosynthase (Wang and Huang, 2009).

Conversion de GH en TG par ingénierie.

Nous avons défini les TG naturelles comme des enzymes possédant une activité de transfert d'unités glycosyles prépondérante par rapport à leur activité hydrolytique. Ceci concerne une minorité de protéines au sein des familles de GH. Pour autant, de nombreuses rGH présentent une activité de transglycosylation, mais généralement très faible. Cette dualité s'évalue par le rapport Transglycosylation/hydrolyse (T/H) qui représente le rapport de l'activité de transfert sur l'activité d'hydrolyse pour une enzyme et un couple de substrats donneur/accepteur. Il est possible, par des modifications subtiles des propriétés physico-chimiques du site catalytique d'augmenter ce ratio de manière assez spectaculaire voire de convertir une GH en TG. C'est un exemple de cette approche qui va faire l'objet du second chapitre (Durand et al., 2016).

Les estérases et la transestérification

La plupart des estérases appartiennent à la famille des carboxylestérases (CE) de la classification CAZy, qui, contrairement aux GH qui présentent une grande diversité de structures (Davies and Henrissat, 1995), possèdent toutes un repliement de type α/β hydrolase (domaine Pfam PF00561) (Punta et al., 2012). Ces enzymes hydrolysent des liaisons esters pour générer un acide et un alcool, et partagent le même mécanisme catalytique, impliquant une triade catalytique (Ser-His-Asp) qui se déroule en deux étapes : d'abord l'acylation d'un résidu sérine nucléophile formant un intermédiaire covalent acyl-enzyme suivie par une réaction de désacylation (Blow, 1990; Brady et al., 1990). Dans cette deuxième étape, soit une molécule d'eau soit une molécule hydroxylée peut attaquer l'intermédiaire covalent et provoquer une réaction d'hydrolyse ou de transestérification sur la molécule acceptrice (Figure 21).

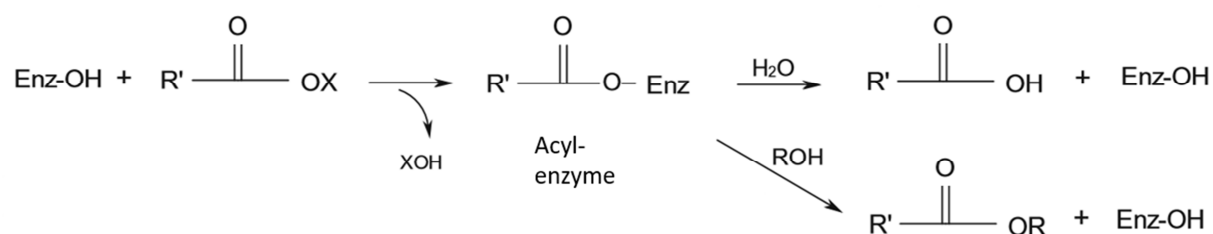


Figure 21 : **Mécanisme catalytique des carboxylestérases.** La première étape génère un intermédiaire covalent acyl-enzyme et la deuxième étape détermine si la réaction est hydrolytique par attaque d'une molécule d'eau ou si c'est une réaction de transestérification par liaison de l'acide sur une fonction hydroxyle d'une molécule acceptrice (Adlercreutz, 2017).

Par conséquent, ce mécanisme catalytique -autorisant la synthèse d'esters- associé à une diversité de spécificité de substrats très large dans cette famille d'enzymes (où l'on retrouve des protéases, lipases, féruloyle estérases, acétyl-xylane estérase,...) confèrent donc à cette classe d'enzymes, au même titre que les GH, un intérêt indéniable dans le domaine de la biotransformation enzymatique (Bornscheuer and Kazlauskas, 2006).

Les lipases sont ainsi largement employées pour la synthèse de molécules par transestérification, à partir d'acides gras contenus dans les huiles végétales. Ces réactions, souvent réalisées en présence de solvants organiques, consistent généralement à coupler des acides gras, soit à des alcools pour former des esters d'alkyles pour l'industrie cosmétique ou utilisés comme lubrifiant (Steinke et al., 2000; Soumanou and Bornscheuer, 2003), soit à des mono-, oligo-, et poly-saccharides afin d'estérifier des sucres pour l'industrie des détergents, des cosmétiques, alimentaire, etc. (Ganske and Bornscheuer, 2005; Šabeder et al., 2006). Les lipases ou les acylases peuvent également catalyser des réactions diverses et inattendues telles que des liaisons carbone-carbone, carbone-hétéroatome, hétéroatome-

hétéroatome, ou encore des réactions d'oxydation, d'époxydation (Busto et al., 2010). Par exemple, Branney et al., ont démontré que la lipase B de *Candida antarctica* (CAL-B, EC 3.1.1.3) peut catalyser une réaction d'aldolisation (formation d'une liaison carbone-carbone) et ont ensuite réussi à augmenter d'un facteur 4 l'activité spécifique de cette enzyme après mutation de la sérine nucléophile du site actif par mutagenèse dirigée (Branney et al., 2003).

Optimisation de l'activité de transestérification par ingénierie.

Les performances de synthèse d'une enzyme pour une réaction donnée ne sont pas toujours suffisantes pour une application dans un processus industriel, il est alors possible d'une part de modifier les conditions de la réaction : immobilisation, contrôle de l'activité de l'eau, choix de solvants ou de substrats mieux adaptés, etc. mais il y a aussi la possibilité, par les méthodes d'ingénierie rationnelle ou d'évolution dirigée de trouver des mutants enzymatiques mieux adaptés à une réaction de synthèse donnée (Bornscheuer et al., 2002). Ces différentes approches ne sont pas exclusives, une enzyme améliorée par ingénierie moléculaire peut présenter des caractéristiques très différentes dans d'autres conditions réactionnelles. Les facteurs déterminant la partition entre l'activité de synthèse et d'hydrolyse des lipases n'étant pas encore bien identifiés, la prédiction des mutations positives est une tâche difficile. Cependant, l'information dérivée des approches d'évolution dirigée contribuera certainement à la compréhension des relations structure-fonction.

Toutefois, une étude comparative des séquences et des structures de 3 lipases ayant des comportements différents, a permis de localiser un résidu clé dans la partition des activités acyltransférase/hydrolyse. La lipase *CduLac* de *Candida dubliniensis* nouvellement découverte montre un potentiel de transestérification plus élevé que les lipases CAL-A de *Pseudozyma antarctica* ou *CpLIP2* de *Candida parapsilosis*. Le résidu apolaire alanine A366 de *CduLac* situé à moins de 5 Å de l'histidine catalytique et son équivalent polaire S369 dans *CpLIP2* ou chargé négativement E370 dans CAL-A ont été ciblés comme potentiellement responsables des différences d'activité de transfert, et soumis à une permutation circulaire. Les mutants obtenus ont montré des changements significatifs dans l'activité acyltransférase. Alors que la mutation S369E entraîne une augmentation de l'activité hydrolytique de *CpLIP2*, S369A conduit à une alcoololyse plus forte. De même, la seule mutation E370A dans CAL-A abaisse d'un facteur 6 la concentration de méthanol pour laquelle les vitesses initiales d'alcoololyse et d'hydrolyse de l'oléate d'éthyle sont égales (Jan et al., 2016).

Remarque sur l'utilisation d'enzymes hydrolytiques comme outil synthèse.

Une donnée importante à considérer lors de réactions de synthèse avec des enzymes hydrolytiques est la possible dégradation des produits de la réaction par hydrolyse secondaire. En effet, le produit de synthèse peut devenir un substrat pour l'enzyme et en fonction de son affinité avec l'enzyme, celui-ci peut être hydrolysé et disparaître aussi rapidement qu'il est formé. Même si les déterminants régissant cette réaction non souhaitée ne sont pas clairement identifiés, il est possible de l'atténuer ou de

l'annuler par ingénierie du site actif de l'enzyme ou en modifiant les conditions de réactions de l'enzyme (Hansson and Adlercreutz, 2001; Bissaro et al., 2015a).

5. Objectifs de la thèse

La mise au point de procédés plus respectueux de l'environnement mais qui restent viables économiquement est devenue un enjeu majeur à l'ère de la chimie verte, une chimie qui notamment s'émancipe de la pétrochimie en utilisant des ressources renouvelables. La conception de nouvelles enzymes, qui permettront une synthèse efficace et plus écoresponsable de molécules organiques à forte valeur ajoutée, nécessite un effort de recherche et de développements permanent, et bien sûr un accroissement de la compréhension fine des différents mécanismes qui régissent le fonctionnement des enzymes.

Les travaux de cette thèse concernent la conception de tels outils, en mettant à profit des méthodes modernes d'ingénierie enzymatique et des connaissances récemment acquises sur le mécanisme catalytique de ces enzymes. Compte-tenu des avantages conséquents de l'utilisation d'enzymes hydrolytiques pour réaliser des réactions de synthèse, nos travaux ont porté sur deux hydrolases, une de type acyl- et l'autre de type glycosyl- fonctionnant par un mécanisme de rétention. Ces deux enzymes aux mécanismes différents partagent une réaction catalytique en deux étapes où la deuxième étape finalise une réaction hydrolytique, si ajout d'une molécule d'eau, ou de transfert, si création d'une liaison covalente entre le donneur et une fonction hydroxyle d'une molécule acceptrice différente de l'eau (Adlercreutz, 2017) (Figure 22).

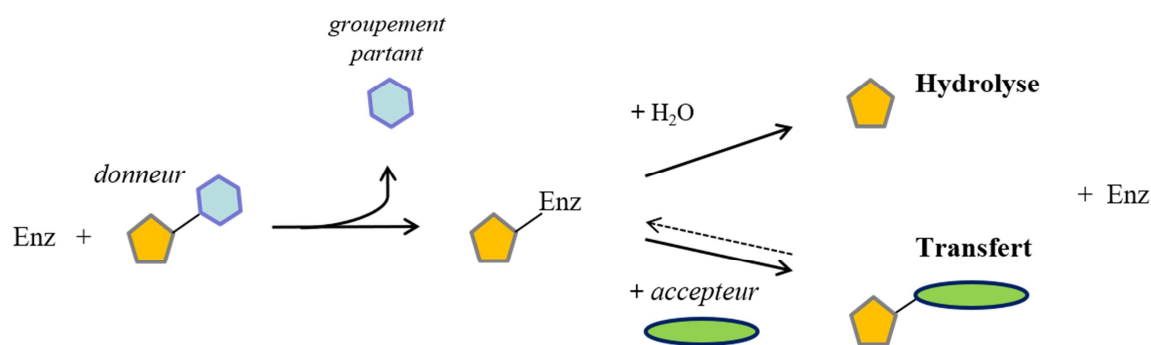


Figure 22 : schéma du mécanisme réactionnel en deux étapes d'une réaction d'hydrolyse ou de transfert. Mécanisme réactionnel schématisé d'enzymes hydrolytiques impliquant l'établissement d'un complexe covalent donneur-enzyme mettant en évidence les 2 voies possibles qui sont le transfert ou l'hydrolyse.

Premièrement, dans le cadre du projet européen BIOSURF (Novel Production Strategies for Biosurfactants) qui visait à trouver de nouvelles voies de synthèse enzymatique de surfactants, alternatives aux méthodes issues de l'industrie pétrochimique, nous avons travaillé à la conversion de

l'endoglycocéramidase II (EGC II) de *Rhodococcus* sp. (rGH 5) en transglycosylase permettant la synthèse d'alkyl de polyglycosides (APG). Cette enzyme a été choisie pour la dualité des substrats naturels (glycosphingolipides) qu'elle hydrolyse, libérant ainsi une partie hydrophobe (céramide) et une partie hydrophile (polysaccharide). Cette particularité a été mise à profit pour la synthèse d'APG qui présentent cette même nature amphiphile et donc potentiellement des propriétés tensioactives. Nous avons choisi d'adapter la méthode développée par l'équipe de Charles Tellier (Feng, 2005a; Osanjo et al., 2007; Kone et al., 2009) et adaptée au laboratoire (Arab-Jaziri et al., 2013, 2015; Bissaro et al., 2015a) qui consiste dans un premier temps à contrarier artificiellement l'activité hydrolytique de l'enzyme, soit en générant des mutants moins actifs par évolution dirigée, soit en travaillant avec un substrat difficilement hydrolysable par l'enzyme. La seconde étape vise à optimiser la reconnaissance de l'accepteur par le sous-site accepteur de l'enzyme. L'objectif final était d'inverser le ratio T/H de l'enzyme et de créer ainsi un mutant optimisé pour la synthèse d'APG, venant compléter la boîte à outils enzymatique avec une nouvelle transglycosidase.

Le deuxième axe de recherche développé dans cette thèse s'est inscrit dans le cadre du projet européen OPTIBIOCAT (OPTimized esterase BIOCATalysts for cost-effective industrial production) qui avait pour objectif de développer de nouvelles estérases pour la synthèse de molécules féruloylées ou glucuronoylées, ayant des propriétés antioxydantes. Nous avons utilisé une approche d'évolution dirigée dans le but d'augmenter la résistance à la température et aux solvants de la féruloyle estérase A d'*Aspergillus niger*, des préalables nécessaires pour pouvoir envisager d'utiliser cette enzyme pour des réactions de transfert, dans des conditions thermodynamiquement favorables, c'est-à-dire en présence de solvants et à température élevée. Cette démarche participe donc ici aussi à la mise en place de bioprocédés efficaces, avec un impact environnemental plus doux que les procédés chimiques actuellement utilisés.

Outre un intérêt applicatif concret, piloté par deux projets différents, cette thèse a pour objectifs, d'une part d'élaborer et de mettre en œuvre des méthodes de criblage innovantes et, plus important encore, de générer de nouvelles informations fonctionnelles sur les déterminants essentiels des mécanismes catalytiques de ces protéines. Une attention particulière sera accordée aux différents facteurs qui modulent le ratio T/H, ou associés à une meilleure stabilité à la chaleur et à une meilleure résistance aux solvants des clones obtenus.

Références :

- Abbadi, A., Leckband, G., 2011. Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Technol.* 113, 1198–1206.
- Adlercreutz, P., 2017. Comparison of lipases and glycoside hydrolases as catalysts in synthesis reactions. *Appl. Microbiol. Biotechnol.* 101, 513–519.
- Adrio, J.L., 2017. Oleaginous yeasts: Promising platforms for the production of oleochemicals and biofuels: Microbial Production of Oleochemicals and Biofuels. *Biotechnol. Bioeng.* 114, 1915–1920.
- Ager, D., 2012. Final Analysis: Biological or Chemical Catalysis? *Platin. Met. Rev.* 56, 140–142.
- Aharoni, A., Thieme, K., Chiu, C.P.C., Buchini, S., Lairson, L.L., Chen, H., Strynadka, N.C.J., Wakarchuk, W.W., Withers, S.G., 2006. High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nat. Methods* 3, 609–614.
- Alvira, P., Tomás-Pejó, E., Ballesteros, M., Negro, M.J., 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. *Bioresour. Technol.* 101, 4851–4861.
- Anastas, P.T., Warner, J.C., 2000. *Green chemistry: theory and practice*, 1. paperback. ed. Oxford Univ. Press, Oxford.
- Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., Fauré, R., O'Donohue, M.J., 2013. Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.* 30, 536–544.
- Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R., O'Donohue, M.J., 2015. Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α -L-arabinofuranosidase. *Carbohydr. Res.* 401, 64–72.
- Arnold, F.H., Wintrode, P.L., Miyazaki, K., Gershenson, A., 2001. How enzymes adapt: lessons from directed evolution. *Trends Biochem. Sci.* 26, 100–106.
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* 50, 4402–4410.
- Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., Noirot, C., Dumon, C., O'Donohue, M., 2013. Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol. Biofuels* 6, 78.
- Batchelor, F., Doyle, F., Nayler, J.H.C., Rolinson, G.N., 1959. Synthesis of penicillin: 6-aminopenicillanic acid in penicillin fermentation. *Nature* 183, 257–258.
- Bednar, D., Beerens, K., Sebestova, E., Bendl, J., Khare, S., Chaloupkova, R., Prokop, Z., Brezovsky, J., Baker, D., Damborsky, J., 2015. FireProt: Energy-and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.* 11, e1004556.
- Bissaro, B., Durand, J., Biarnés, X., Planas, A., Monsan, P., O'Donohue, M.J., Fauré, R., 2015a. Molecular Design of Non-Leloir Furanose-Transferring Enzymes from an α -L-Arabinofuranosidase: A Rationale for the Engineering of Evolved Transglycosylases. *ACS Catal.* 5, 4598–4611.
- Bissaro, B., Monsan, P., Fauré, R., O'Donohue, M.J., 2015b. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* 467, 17–35.
- Bisswanger, H., 2014. Enzyme assays. *Perspect. Sci.* 1, 41–55.

- Bloom, J.D., Arnold, F.H., 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci.* 106, 9995–10000.
- Blow, D., 1990. Enzymology. More of the catalytic triad. *Nature* 343, 694–695.
- Boersma, Y.L., Dröge, M.J., Quax, W.J., 2007. Selection strategies for improved biocatalysts: Selecting improved biocatalysts. *FEBS J.* 274, 2181–2195.
- Bøjstrup, M., Marri, L., Lok, F., Hindsgaul, O., 2015. A Chromogenic Assay Suitable for High-Throughput Determination of Limit Dextrinase Activity in Barley Malt Extracts. *J. Agric. Food Chem.* 63, 10873–10878.
- Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* 70, 493–502.
- Bornscheuer, U.T., Altenbuchner, J., Meyer, H.H., 1998. Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnol. Bioeng.* 58, 554–559.
- Bornscheuer, U.T., Bessler, C., Srinivas, R., Hari Krishna, S., 2002. Optimizing lipases and related enzymes for efficient application. *Trends Biotechnol.* 20, 433–437.
- Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., Robins, K., 2012. Engineering the third wave of biocatalysis. *Nature* 485, 185–194.
- Bornscheuer, U.T., Kazlauskas, R.J., 2006. *Hydrolases in organic synthesis: regio- and stereoselective biotransformations*. Wiley-VCH; John Wiley, Weinheim; Chichester.
- Bornscheuer, U.T., Kazlauskas, R.J., 2004. Catalytic Promiscuity in Biocatalysis: Using Old Enzymes to Form New Bonds and Follow New Pathways. *Angew. Chem. Int. Ed.* 43, 6032–6040.
- Borsenberger, V., Dornez, E., Desrousseaux, M.-L., Massou, S., Tenkanen, M., Courtin, C.M., Dumon, C., O'Donohue, M.J., Fauré, R., 2014. A ¹H NMR study of the specificity of α -L-arabinofuranosidases on natural and unnatural substrates. *Biochim. Biophys. Acta* 1840, 3106–3114.
- Borsenberger, V., Ferreira, F., Pollet, A., Dornez, E., Desrousseaux, M.-L., Massou, S., Courtin, C.M., O'Donohue, M.J., Fauré, R., 2012. A versatile and colorful screening tool for the identification of arabinofuranose-acting enzymes. *Chembiochem Eur. J. Chem. Biol.* 13, 1885–1888.
- Bothast, R.J., Schlicher, M.A., 2005. Biotechnological processes for conversion of corn into ethanol. *Appl. Microbiol. Biotechnol.* 67, 19–25.
- Brady, L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Høge-Jensen, B., Nørskov, L., Thim, L., Menge, U., 1990. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Lett. Nat.* 343, 767–770.
- Bräm, S., Wolfram, E., 2017. Recent Advances in Effect-directed Enzyme Assays based on Thin-layer Chromatography: Recent advances in effect-directed enzyme assays based on TLC. *Phytochem. Anal.* 28, 74–86.
- Branneby, C., Carlqvist, P., Magnusson, A., Hult, K., Brinck, T., Berglund, P., 2003. Carbon–Carbon Bonds by Hydrolytic Enzymes. *J. Am. Chem. Soc.* 125, 874–875.
- Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E., LaBaer, J., 2002. Proteome-scale purification of human proteins from bacteria. *Proc. Natl. Acad. Sci.* 99, 2654–2659.
- Breton, C., Fournel-Gigleux, S., Palcic, M.M., 2012. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* 22, 540–549.
- Bugg, T.D., Rahmanpour, R., 2015. Enzymatic conversion of lignin into renewable chemicals. *Curr. Opin. Chem. Biol.* 29, 10–17.

- Bungaruang, L., Gutmann, A., Nidetzky, B., 2013. Leloir Glycosyltransferases and Natural Product Glycosylation: Biocatalytic Synthesis of the C-Glucoside Nothofagin, a Major Antioxidant of Redbush Herbal Tea. *Adv. Synth. Catal.* 355, 2757–2763.
- Busto, E., Gotor-Fernández, V., Gotor, V., 2010. Hydrolases: catalytically promiscuous enzymes for non-conventional reactions in organic synthesis. *Chem. Soc. Rev.* 39, 4504.
- Carreño, C.A., de Lorenzo, V., 2010. Genetic Traps for Surveying New Catalysts in (Meta) Genomic DNA, in: Timmis, K.N. (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 4563–4579.
- Champion, E., Moulis, C., Morel, S., Mulard, L.A., Monsan, P., Remaud-Siméon, M., André, I., 2010. A pH-Based High-Throughput Screening of Sucrose-Utilizing Transglucosidases for the Development of Enzymatic Glucosylation Tools. *ChemCatChem* 2, 969–975.
- Chew, K.W., Yap, J.Y., Show, P.L., Suan, N.H., Juan, J.C., Ling, T.C., Lee, D.-J., Chang, J.-S., 2017. Microalgae biorefinery: High value products perspectives. *Bioresour. Technol.* 229, 53–62.
- Chica, R.A., Doucet, N., Pelletier, J.N., 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* 16, 378–384.
- Choma, I.M., Grzelak, E.M., 2011. Bioautography detection in thin-layer chromatography. *J. Chromatogr. A* 1218, 2684–2691.
- Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., Monticello, D.J., 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* 19, 354–359.
- Cohen, N., Abramov, S., Dror, Y., Freeman, A., 2001. In vitro enzyme evolution: the screening challenge of isolating the one in a million. *Trends Biotechnol.* 19, 507–510.
- Colonna, P., 2012. Le carbone renouvelable dans les systèmes alimentaires, énergétiques et chimiques: Leçon inaugurale prononcée le jeudi 15 décembre 2011. Collège de France.
- Combes, D., Monsan, P., 2009. Biocatalyse ou catalyse enzymatique. *Procédés Chim. - Bio- Agro Chim. Verte, Techniques de l'ingénieur* 1–18.
- Cosgrove, D.J., 2005. Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6, 850–861.
- Cotana, F., Cavalaglio, G., Coccia, V., Petrozzi, A., 2016. Energy Opportunities from Lignocellulosic Biomass for a Biorefinery Case Study. *Energies* 9, 748.
- Crameri, A., Raillard, S.-A., Bermudez, E., Stemmer, W.P., 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 391, 288–291.
- Dalby, P.A., 2011. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* 21, 473–480.
- Davies, G., Henrissat, B., 1995. Structures and mechanisms of glycosyl hydrolases. *Structure, Current Biology* 3, 853–859.
- Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., Rooman, M., 2009. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25, 2537–2543.
- Delmas, M., 2008. Vegetal Refining and Agrochemistry. *Chem. Eng. Technol.* 31, 792–797.
- Dhepe, P.L., Fukuoka, A., 2008. Cellulose conversion under heterogeneous catalysis. *ChemSusChem* 1, 969–975.
- Dietrich, J.A., McKee, A.E., Keasling, J.D., 2010. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu. Rev. Biochem.* 79, 563–590.

- Ducros, V.M.-A., Tarling, C.A., Zechel, D.L., Brzozowski, A.M., Frandsen, T.P., von Ossowski, I., Schülein, M., Withers, S.G., Davies, G.J., 2003. Anatomy of glycosynthesis: structure and kinetics of the *Humicola insolens* Cel7B E197A and E197S glycosynthase mutants. *Chem. Biol.* 10, 619–628.
- Dumon, C., Song, L., Bozonnet, S., Fauré, R., O'Donohue, M.J., 2012. Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.* 47, 346–357.
- Durand, J., Biarnés, X., Watterlot, L., Bonzom, C., Borsenberger, V., Planas, A., Bozonnet, S., O'Donohue, M.J., Fauré, R., 2016. A single point mutation alters the Transglycosylation/Hydrolysis partition, significantly enhancing the synthetic capability of an *endo*-glycosceramidase. *ACS Catal.* 8264–8275
- Fa, M., Radeghieri, A., Henry, A.A., Romesberg, F.E., 2004. Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.* 126, 1748–1754.
- Faber, K., 2011. *Biotransformations in organic chemistry: a textbook*, 6., and corr. ed. ed. Springer, Berlin.
- Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C., Dion, M., 2005. Converting a β -glycosidase into a β -transglycosidase by directed evolution. *J. Biol. Chem.* 280, 37088–37097.
- Ferrer, J.-L., zubietta, C., Dixon, R.-A., Noel J.-P., 2005. Crystal Structures of Alfalfa Caffeoyle Coenzyme A 3-O-Methyltransferase. *PLANT Physiol.* 137, 1009–1017.
- Frankova, L., Fry, S.C., 2013. Biochemistry and physiological roles of enzymes that “cut and paste” plant cell-wall polysaccharides. *J. Exp. Bot.* 64, 3519–3550.
- French, D., 1973. Chemical and physical properties of starch. *J. Anim. Sci.* 37, 1048.
- Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M., Møller, B.L., 2014. Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nat. Commun.* 5.
- Ganske, F., Bornscheuer, U.T., 2005. Optimization of lipase-catalyzed glucose fatty acid ester synthesis in a two-phase system containing ionic liquids and *t*-BuOH. *J. Mol. Catal. B Enzym.* 36, 40–42.
- Gao, D., Chundawat, S.P.S., Liu, T., Hermanson, S., Gowda, K., Brumm, P., Dale, B.E., Balan, V., 2010. Strategy for identification of novel fungal and bacterial glycosyl hydrolase hybrid mixtures that can efficiently saccharify pretreated lignocellulosic biomass. *BioEnergy Res.* 3, 67–81.
- Gdaniec, Z., Ban, B., Sowers, L.C., Fazakerley, G.V., 1996. Methoxyamine-Induced Mutagenesis of Nucleic Acids. A Proton NMR Study of oligonucleotides containing N4-methoxycytosine paired with adenine or guanine. *Eur. J. Biochem.* 242, 271–279.
- Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss, R., 1997. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol.* 15, 29–34.
- Goldsmith, M., Tawfik, D.S., 2013. Enzyme engineering by targeted libraries, in: *Methods in Enzymology*. Elsevier, pp. 257–283.
- Gottesman, S., 1996. Proteases and their targets in *Escherichia coli*. *Annu. Rev. Genet.* 30, 465–506.
- Graf, E., 1992. Antioxidant potential of ferulic acid. *Free Radic. Biol. Med.* 13, 435–448.
- Greener, A., Callahan, M., Jerpseth, B., 1997. An efficient random mutagenesis technique using an *E.coli* mutator strain. *Mol. Biotechnol.* 7, 189–195.

- Griswold, K.E., Kawarasaki, Y., Ghoneim, N., Benkovic, S.J., Iverson, B.L., Georgiou, G., 2005. Evolution of highly active enzymes by homology-independent recombination. *Proc. Natl. Acad. Sci.* 102, 10082–10087.
- Grodberg, J., Dunn, J.J., 1988. *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* 170, 1245–1253.
- Grootaert, C., Delcour, J.A., Courtin, C.M., Broekaert, W.F., Verstraete, W., Van de Wiele, T., 2007. Microbial metabolism and prebiotic potency of arabinoxylan oligosaccharides in the human intestine. *Trends Food Sci. Technol.* 18, 64–71.
- Guerriero, G., Hausman, J.-F., Strauss, J., Ertan, H., Siddiqui, K.S., 2016. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. *Eng. Life Sci.* 16, 1–16.
- Gustafsson, C., Govindarajan, S., Minshull, J., 2004. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22, 346–353.
- Hanes, J., Plückthun, A., 1997. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4937–4942.
- Hanson-Manful, P., Patrick, W.M., 2013. Construction and analysis of randomized protein-encoding libraries using error-prone PCR, in: Gerrard, J.A. (Ed.), *Protein Nanotechnology*. Humana Press, Totowa, NJ, pp. 251–267.
- Hansson, T., Adlercreutz, P., 2001. Enhanced transglucosylation/hydrolysis ratio of mutants of *Pyrococcus furiosus* β -glucosidase: Effects of donor concentration, water content, and temperature on activity and selectivity in hexanol. *Biotechnol. Bioeng.* 75, 656–665.
- Henke, E., Bornscheuer, U.T., 1999. Directed evolution of an esterase from *Pseudomonas fluorescens*. Random mutagenesis by error-prone PCR or a mutator strain and identification of mutants showing enhanced enantioselectivity by a resorufin-based fluorescence assay. *Biol. Chem.* 380, 1029–1033.
- Hidaka, M., Fushinobu, S., Honda, Y., Wakagi, T., Shoun, H., Kitaoka, M., 2010. Structural explanation for the acquisition of glycosynthase activity. *J. Biochem.* 147, 237–244.
- Hoffmann, G.R., 1980. Genetic effects of dimethyl sulfate, diethyl sulfate, and related compounds. *Mutat. Res. Genet. Toxicol.* 75, 63–129.
- Honda, Y., Kitaoka, M., 2006. The first glycosynthase derived from an inverting glycoside hydrolase. *J. Biol. Chem.* 281, 1426–1431.
- Horbury, M.D., Baker, L.A., Quan, W.-D., Greenough, S.E., Stavros, V.G., 2016. Photodynamics of potent antioxidants: ferulic and caffeic acids. *Phys. Chem. Chem. Phys.* 18, 17691–17697.
- Huber, M., Knutti, R., 2011. Anthropogenic and natural warming inferred from changes in Earth's energy balance. *Nat. Geosci.* 5, 31–36.
- Huber, R., Wulfhorst, H., Maksym, L., Stehr, R., Pöhnlein, M., Jäger, G., Spieß, A.C., Büchs, J., 2011. Screening for enzyme activity in turbid suspensions with scattered light. *Biotechnol. Prog.* 27, 555–561.
- Hult, K., Berglund, P., 2007. Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* 25, 231–238.
- Irague, R., Massou, S., Moulis, C., Saurel, O., Milon, A., Monsan, P., Remaud-Siméon, M., Portais, J.-C., Potocki-Véronèse, G., 2011. NMR-based structural glycomics for high-throughput screening of carbohydrate-active enzyme specificity. *Anal. Chem.* 83, 1202–1206.
- Isikgor, F.H., Becer, C.R., 2015. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym. Chem.* 6, 4497–4559.
- Jan, A.-H., Subileau, M., Deyrieux, C., Perrier, V., Dubreucq, É., 2016. Elucidation of a key position for acyltransfer activity in *Candida parapsilosis* lipase/acyltransferase (CpLIP2) and in

- Pseudozyma antarctica* lipase A (CAL-A) by rational design. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1864, 187–194.
- Jensen, N.B., Strucko, T., Kildegaard, K.R., David, F., Maury, J., Mortensen, U.H., Forster, J., Nielsen, J., Borodina, I., 2014. Easy Clone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 14, 238–248.
- Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., Hilvert, D., Houk, K.N., Stoddard, B.L., Baker, D., 2008. De novo computational design of retro-aldol enzymes. *Science* 319, 1387–1391.
- Joo, H., Arisawa, A., Lin, Z., Arnold, F.H., 1999. A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases. *Chem. Biol.* 6, 699–706.
- Julius, M.H., Masuda, T., Herzenberg, L.A., 1972. Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter. *Proc. Natl. Acad. Sci.* 69, 1934–1938.
- Kamm, B., Kamm, M., 2004. Principles of biorefineries. *Appl. Microbiol. Biotechnol.* 64, 137–145.
- Kast, P., Asif-Ullah, M., Jiang, N., Hilvert, D., 1996. Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis. *Proc. Natl. Acad. Sci.* 93, 5043–5048.
- Kikuzaki, H., Hisamoto, M., Hirose, K., Akiyama, K., Taniguchi, H., 2002. Antioxidant properties of ferulic acid and its related compounds. *J. Agric. Food Chem.* 50, 2161–2168.
- Kobayashi, S., Makino, A., 2009. Enzymatic polymer synthesis: an opportunity for green polymer Chemistry. *Chem. Rev.* 109, 5288–5353.
- Kobayashi, S., Ohmae, M., 2006. Enzymatic polymerization to polysaccharides, in: Kobayashi, S., Ritter, H., Kaplan, D. (Eds.), *Enzyme-Catalyzed Synthesis of Polymers*. Springer-Verlag, Berlin/Heidelberg, pp. 159–210.
- Kone, F.M.T., Le Behec, M., Sine, J.-P., Dion, M., Tellier, C., 2009. Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.* 22, 37–44.
- Koshland, D.E., 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* 28, 416–436.
- Kumar, A., Singh, S., 2013. Directed evolution: tailoring biocatalysts for industrial applications. *Crit. Rev. Biotechnol.* 33, 365–378.
- Kumar, R., Singh, S., Singh, O.V., 2008. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J. Ind. Microbiol. Biotechnol.* 35, 377–391.
- Laine, R.A., 1994. A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 4, 759–767.
- Lairson, L.L., Henrissat, B., Davies, G.J., Withers, S.G., 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* 77, 521–555.
- Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L., van Loon, A.P.G.M., 2000a. From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng. Des. Sel.* 13, 49–57.
- Lehmann, M., Pasamontes, L., Lassen, S.F., Wyss, M., 2000b. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* 1543, 408–415.
- Lenucci, M.S., Piro, G., Dalessandro, G., 2009. In muro feruloylation and oxidative coupling in monocots. *Plant Signal. Behav.* 4, 228–230.

- Li, J.-J., Pei, X.-Q., Zhang, S.-B., Wu, Z.-L., 2015. Improving the thermostability of feruloyl esterase by DNA shuffling and site-directed mutagenesis. *Process Biochem.* 50, 1783–1787.
- Lu, X., Hirata, H., Yamaji, Y., Ugaki, M., Namba, S., 2001. Random mutagenesis in a plant viral genome using a DNA repair-deficient mutator *Escherichia coli* strain. *J. Virol. Methods* 94, 37–43.
- Lutz, S., 2010. Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* 21, 734–743.
- Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., Benkovic, S.J., 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci.* 98, 11248–11253.
- MacBeath, G., 1998. Redesigning enzyme topology by directed evolution. *Science* 279, 1958–1961.
- Mackenzie, L.F., Wang, Q., Warren, R.A.J., Withers, S.G., 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* 120, 5583–5584.
- Mahmoud, K.A., Male, K.B., Hrapovic, S., Luong, J.H.T., 2009. Cellulose nanocrystal/gold nanoparticle composite as a matrix for enzyme immobilization. *ACS Appl. Mater. Interfaces* 1, 1383–1386.
- Maisonneuve, L., Chollet, G., Grau, E., Cramail, H., 2016. Vegetable oils: a source of polyols for polyurethane materials. *OCL* 23, D508.
- Malet, C., Planas, A., 1998. From β -glucanase to β -glucansynthase: glycosyl transfer to α -glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* 440, 208–212.
- Mangas-Sánchez, J., Adlercreutz, P., 2015. Enzymatic preparation of oligosaccharides by transglycosylation: A comparative study of glucosidases. *J. Mol. Catal. B Enzym.* 122, 51–55.
- Marston, A., Kissling, J., Hostettmann, K., 2002. A rapid TLC bioautographic method for the detection of acetylcholinesterase and butyrylcholinesterase inhibitors in plants. *Phytochem. Anal. PCA* 13, 51–54.
- Mastrangelo, L.I., Lenucci, M.S., Piro, G., Dalessandro, G., 2009. Evidence for intra- and extra-protoplasmic feruloylation and cross-linking in wheat seedling roots. *Planta* 229, 343–355.
- Merrien, A., Carre, P., Quinsac, A., 2012. Des ressources oléagineuses variées potentiellement au service du développement de la chimie verte. *Ol. Corps Gras Lipides* 19, 6–9.
- Michely, S., Gaillardin, C., Nicaud, J.-M., Neuvéglise, C., 2013. Comparative physiology of oleaginous species from the *Yarrowia* clade. *PLoS ONE* 8, e63356.
- Molina, I., Li-Beisson, Y., Beisson, F., Ohlrogge, J.B., Pollard, M., 2009. Identification of an *arabidopsis* feruloyl-coenzyme A transferase required for suberin synthesis. *PLANT Physiol.* 151, 1317–1328.
- Molina-Espeja, P., Viña-Gonzalez, J., Gomez-Fernandez, B.J., Martin-Diaz, J., Garcia-Ruiz, E., Alcalde, M., 2016. Beyond the outer limits of nature by directed evolution. *Biotechnol. Adv.* 34, 754–767.
- Monsan, P., Remaud-Siméon, M., André, I., 2010. Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. *Curr. Opin. Microbiol.* 13, 293–300.
- Morello, E., Bermúdez-Humarán, L.G., Llull, D., Solé, V., Miraglio, N., Langella, P., Poquet, I., 2007. *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* 14, 48–58.
- National Research Council (U.S.), National Academies Press (U.S.) (Eds.), 2007. The new science of metagenomics: revealing the secrets of our microbial planet. National Academies Press, Washington, DC.

- Nawrath, C., 2002. The biopolymers cutin and suberin. Arab. Book 1, The American Society of Plant Biologists e0021.
- Ness, J.E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T.V., Govindarajan, S., Mundorff, E.C., Minshull, J., 2002. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* 20, 1251–1255.
- Nobuyoshi, N., Kohji, I., Hiroki, H., Sei-Ichirou, K., Tsutomu, F., 2000. Regioselective acylation of flavonoid glucoside with aromatic acid by an enzymatic reaction system from cultured cells of *Ipomoea batatas*. *J. Biosci. Bioeng.* 90, 347–349.
- Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., Iverson, B.L., 2000. Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol.* 18, 1071–1074.
- O’Maille, P.E., Bakhtina, M., Tsai, M.-D., 2002. Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol.* 321, 677–691.
- Osanjo, G., Dion, M., Drone, J., Solleux, C., Tran, V., Rabiller, C., Tellier, C., 2007. Directed evolution of the α -L-Fucosidase from *Thermotoga maritima* into an α -L-Transfucosidase [†]. *Biochemistry* 46, 1022–1033.
- Ostermeier, M., Shim, J.H., Benkovic, S.J., 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 17, 1205–1209.
- Ou, S., Kwok, K.-C., 2004. Ferulic acid: pharmaceutical functions, preparation and applications in foods. *J. Sci. Food Agric.* 84, 1261–1269.
- Pachua, L., 2015. A mini review on plant-based nanocellulose: production, sources, modifications and its potential in drug delivery applications. *Mini-Rev. Med. Chem.* 15, 543–552.
- Packer, M.S., Liu, D.R., 2015. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394.
- Pantoliano, M.W., Whitlow, M., Wood, J.F., Dodd, S.W., Hardman, K.D., Rollence, M.L., Bryan, P.N., 1989. Large increases in general stability for subtilisin BPN’ through incremental changes in the free energy of unfolding. *Biochemistry* 28, 7205–7213.
- Patrick, J., C. Botha, F., G. Birch, R., 2013. Metabolic engineering of sugars and simple sugar derivatives in plants. *Plant Biotechnol. J.* 11, 142–156.
- Pittet, A., Royer, D., 2002. Rapid, low cost thin-layer chromatographic screening method for the detection of ochratoxin A in green coffee at a control level of 10 μ g/kg. *J. Agric. Food Chem.* 50, 243–247.
- Poole, C.F., 2003. Thin-layer chromatography: challenges and opportunities. *J. Chromatogr. A* 1000, 963–984.
- Pourmir, A., Johannes, T.W., 2012. Directed evolution: selection of the host organism. *Comput. Struct. Biotechnol. J.* 2, 1–7.
- Preiss, J., 2009. Biochemistry and molecular biology of starch biosynthesis, in: *Starch - Chemistry and Technology*, pp. 83–148.
- Puchart, V., Biely, P., 2007. A simple enzymatic synthesis of 4-nitrophenyl β -1,4-d-xylobioside, a chromogenic substrate for assay and differentiation of endoxylanases. *J. Biotechnol.* 128, 576–586.
- Punta, M., Coghill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D., 2012. The Pfam protein families database. *Nucleic Acids Res.* 40, D290-301.
- Pushpam, P., Rajesh, T., Gunasekaran, P., 2011. Identification and characterization of alkaline serine protease from goat skin surface metagenome. *AMB Express* 1, 3.

- Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Hallett, J.P., Leak, D.J., Liotta, C.L., others, 2006. The path forward for biofuels and biomaterials. *science* 311, 484–489.
- Rasila, T.S., Pajunen, M.I., Savilahti, H., 2009. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal. Biochem.* 388, 71–80.
- Rather, M.Y., Mishra, S., Aravinda, S., 2013. Exploring the synthetic potential of cell bound β -glycosidase of *Pichia etchellsii*. *J. Biotechnol.* 165, 63–68.
- Reetz, M.T., 2013. Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.* 135, 12480–12496.
- Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D., Vogel, A., 2005. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed.* 44, 4192–4196.
- Reetz, M.T., Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* 2, 891–903.
- Reetz, M.T., Carballeira, J.D., Vogel, A., 2006a. Iterative Saturation Mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.* 45, 7745–7751.
- Reetz, M.T., Eipper, A., Tielmann, P., Mynott, R., 2002. A practical NMR-based high-throughput Assay for screening enantioselective catalysts and biocatalysts. *Adv. Synth. Catal.* 344, 1008–1016.
- Reetz, M.T., Wang, L.-W., Bocola, M., 2006b. Directed evolution of enantioselective enzymes: Iterative Cycles of CASTing for Probing Protein-Sequence Space. *Angew. Chem.* 118, 1258–1263.
- Reymond, J.-L., Babiak, P., 2007. Screening Systems, in: Ulber, R., Sell, D. (Eds.), *White Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 31–58.
- Rodriguez-Aller, M., Gurny, R., Veuthey, J.-L., Guillarme, D., 2013. Coupling ultra high-pressure liquid chromatography with mass spectrometry: Constraints and possible applications. *J. Chromatogr. A* 1292, 2–18.
- Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5.
- Rose, J.K.C., 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant Cell Physiol.* 43, 1421–1435.
- Šabeder, S., Habulin, M., Knez, Ž., 2006. Lipase-catalyzed synthesis of fatty acid fructose esters. *J. Food Eng.* 77, 880–886.
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of β -Globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.
- Sakakura, T., Choi, J.-C., Yasuda, H., 2007. Transformation of carbon dioxide. *Chem. Rev.* 107, 2365–2387.
- Sánchez, Ó.J., Cardona, C.A., 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresour. Technol.* 99, 5270–5295.
- Sarkar, I., Hauber, I., Hauber, J., Buchholz, F., 2007. HIV-1 proviral DNA excision using an evolved recombinase. *science* 316, 1912–1915.

- Scheible, W.-R., Pauly, M., 2004. Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr. Opin. Plant Biol.* 7, 285–295.
- Serra, O., Hohn, C., Franke, R., Prat, S., Molinas, M., Figueras, M., 2010. A feruloyl transferase involved in the biosynthesis of suberin and suberin-associated wax is required for maturation and sealing properties of *potato periderm*: FHT function in potato periderm. *Plant J.* 62, 277–290.
- Sieber, V., Martinez, C.A., Arnold, F.H., 2001. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* 19, 456–460.
- Silveira, R.L., Stoyanov, S.R., Gusarov, S., Skaf, M.S., Kovalenko, A., 2013. Plant biomass recalcitrance: effect of hemicellulose composition on nanoscale forces that control cell wall strength. *J. Am. Chem. Soc.* 135, 19048–19051.
- Smith, G., 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228, 1315–1317.
- Snell, K.D., Singh, V., Brumbley, S.M., 2015. Production of novel biopolymers in plants: recent technological advances and future prospects. *Curr. Opin. Biotechnol.* 32, 68–75.
- Solomon, S., Plattner, G.-K., Knutti, R., Friedlingstein, P., 2009. Irreversible climate change due to carbon dioxide emissions. *Proc. Natl. Acad. Sci.* 106, 1704–1709.
- Somerville, C., 2006. Cellulose Synthesis in Higher Plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78.
- Soumanou, M.M., Bornscheuer, U.T., 2003. Improvement in lipase-catalyzed synthesis of fatty acid methyl esters from sunflower oil. *Enzyme Microb. Technol.* 33, 97–103.
- Speight, R.E., Hart, D.J., Sutherland, J.D., Blackburn, J.M., 2001. A new plasmid display technology for the in vitro selection of functional phenotype-genotype linked proteins. *Chem. Biol.* 8, 951–965.
- Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S.B., Keasling, J.D., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559–562.
- Steinke, G., Kirchhoff, R., Mukherjee, K.D., 2000. Lipase-catalyzed alcoholysis of crambe oil and camelina oil for the preparation of long-chain esters. *J. Am. Oil Chem. Soc.* 77, 361–366.
- Stemmer, W.P., 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci.* 91, 10747–10751.
- Stemmer, W.P., 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370, 389–391.
- Stemmer, W.P., Cramer, A., Ha, K.D., Brennan, T.M., Heyneker, H.L., 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164, 49–53.
- Sun, B., Miller, G., Lee, W.Y., Ho, K., Crowe, M.A., Partridge, L., 2013. Analytical method development for directed enzyme evolution research: a high throughput high-performance liquid chromatography method for analysis of ribose and ribitol and a capillary electrophoresis method for the separation of ribose enantiomers. *J. Chromatogr. A* 1271, 163–169.
- Szűts, A., Szabó-Révész, P., 2012. Sucrose esters as natural surfactants in drug delivery systems—A mini-review. *Int. J. Pharm.* 433, 1–9.
- Talmaciu, A.I., Volf, I., Popa, V.I., 2015. A comparative analysis of the "Green" Techniques" applied for polyphenols extraction from bioresources. *Chem. Biodivers.* 12, 1635–1651.
- Ten, E., Vermerris, W., 2013. Functionalized polymers from lignocellulosic biomass: state of the art. *Polymers* 5, 600–642.
- Tonnac, J.-P. de (Ed.), 2010. *Dictionnaire universel du pain*, Bouquins. Robert Laffont, Paris.

- Uchiyama, T., Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr. Opin. Biotechnol.* 20, 616–622.
- Ufarte, L., Potocki-Veronese, G., Laville, E., 2015. Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. *Front. Microbiol.* 6.
- Valdivia, M., Galan, J.L., Laffarga, J., Ramos, J.-L., 2016. Biofuels 2020: Biorefineries based on lignocellulosic materials. *Microb. Biotechnol.* 9, 585–594.
- Valetti, F., Gilardi, G., 2004. Directed evolution of enzymes for product chemistry. *Nat. Prod. Rep.* 21, 490.
- van Rossum, T., Kengen, S.W.M., van der Oost, J., 2013. Reporter-based screening and selection of enzymes. *FEBS J.* 280, 2979–2996.
- Wang, L.-X., Huang, W., 2009. Enzymatic transglycosylation for glycoconjugate synthesis. *Curr. Opin. Chem. Biol.* 13, 592–600.
- Wang, Q., Buckle, A.M., Foster, N.W., Johnson, C.M., Fersht, A.R., 1999. Design of highly stable functional GroEL minichaperones. *Protein Sci. Publ. Protein Soc.* 8, 2186–2193.
- Wang, R., Dijkstra, P.J., Karperien, M., 2016. Dextran, in: Neves, N.M., Reis, R.L. (Eds.), *Biomaterials from Nature for advanced devices and therapies*. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 307–319.
- Wang, W., Wang, S., Ma, X., Gong, J., 2011. Recent advances in catalytic hydrogenation of carbon dioxide. *Chem. Soc. Rev.* 40, 3703.
- Ward, O.P., Singh, A., 2000. Enzymatic asymmetric synthesis by decarboxylases. *Curr. Opin. Biotechnol.* 11, 520–526.
- Watson, J.V., Dive, C., 1994. Enzyme kinetics. *Methods Cell Biol.* 41, 469–507.
- Westers, L., Westers, H., Quax, W.J., 2004. *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1694, 299–310.
- Wildt, S., Gerngross, T.U., 2005. The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* 3, 119–128.
- Wilson, D.S., Keefe, A.D., Szostak, J.W., 2001. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.* 98, 3750–3755.
- Withers, S.G., 2001. Mechanisms of glycosyl transferases and hydrolases. *Carbohydr. Polym.* 44, 325–337.
- Wong, C.-H., Haynie, S.L., Whitesides, G.M., 1982. Enzyme-Catalyzed Synthesis of N-acetyllactosamine with in situ regeneration of Uridine 5'-diphosphate glucose and Uridine 5'-diphosphate galactose. *ACS Catal.* 47, 5416–5418
- Wong, T.S., Zhurina, D., Schwaneberg, U., 2006. The diversity challenge in directed protein evolution. *Comb. Chem. High Throughput Screen.* 9, 271–288.
- Wongwilaiwalin, S., Rattanachomsri, U., Laothanachareon, T., Eurwilaichitr, L., Igarashi, Y., Champreda, V., 2010. Analysis of a thermophilic lignocellulose degrading microbial consortium and multi-species lignocellulolytic enzyme system. *Enzyme Microb. Technol.* 47, 283–290.
- Wyatt, M.D., Pittman, D.L., 2006. Methylating agents and DNA repair responses: Methylated Bases and Sources of Strand Breaks. *Chem. Res. Toxicol.* 19, 1580–1594.
- Xiao, H., Bao, Z., Zhao, H., 2015. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* 54, 4011–4020.

- Xu, F., 2011. Enzymatic Degradation of Lignocellulosic Biomass, in: Tao, J.A., Kazlauskas, R. (Eds.), *Biocatalysis for Green Chemistry and Chemical Process Development*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 361–390.
- Yang, Z.R., Thomson, R., McNeil, P., Esnouf, R.M., 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21, 3369–3376.
- Zaccolo, M., Williams, D.M., Brown, D.M., Gherardi, E., 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* 255, 589–603.
- Zakzeski, J., Bruijninx, P.C.A., Jongerius, A.L., Weckhuysen, B.M., 2010. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem. Rev.* 110, 3552–3599.
- Zeeman, S.C., Kossmann, J., Smith, A.M., 2010. Starch: its metabolism, evolution, and biotechnological modification in Plants. *Annu. Rev. Plant Biol.* 61, 209–234.
- Zhang, J.-H., Dawes, G., Stemmer, W.P.C., 1997. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci.* 94, 4504–4509.
- Zhao, H., 2004. Staggered extension process in vitro DNA recombination. *Methods Enzymol.* 388, 42–49.
- Zhao, H., Giver, L., Shao, Z., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* 16, 258–261.
- Zhao, H., Zha, W., 2006. In vitro “sexual” evolution through the PCR-based staggered extension process (StEP). *Nat. Protoc.* 1, 1865–1871.
- Zhou, J., Wang, Y.-H., Chu, J., Luo, L.-Z., Zhuang, Y.-P., Zhang, S.-L., 2009. Optimization of cellulase mixture for efficient hydrolysis of steam-exploded corn stover by statistically designed experiments. *Bioresour. Technol.* 100, 819–825.
- Zhou, Y.J., Buijs, N.A., Zhu, Z., Qin, J., Siewers, V., Nielsen, J., 2016. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* 7, 11709.

Chapitre II : Modification du ratio transglycosylation/hydrolyse
d'une endoglycocéramidase afin d'améliorer son activité de synthèse

Ce premier chapitre décrit comment, par une approche semi-rationnelle, nous sommes parvenus à convertir une glycoside hydrolase (GH) hydrolytique, une endoglycocéramidase, en une transglycosylase (TG) efficace pour la synthèse de dérivés de type alkylpolyglycosides (APG). La stratégie choisie ici s'appuie sur les nombreux travaux publiés à ce jour visant à comprendre les déterminants qui régissent la modulation du ratio transglycosylation/hydrolyse (T/H) de nombreuses GH. Ces études utilisent soit des approches d'évolution dirigée (Feng, 2005b; Pennec et al., 2015), soit des approches d'ingénierie rationnelle ou semi-rationnelle (Kuriki et al., 1996; Oh et al., 2008; Dilokpimol et al., 2011a; Ochs et al., 2011; Teze et al., 2014, 2015; Bissaro et al., 2015a) et ont fait l'objet d'une revue critique de la littérature en 2015 par Bissaro et al. (Bissaro et al., 2015b). Au regard de ces résultats expérimentaux, l'objectif est à présent de déterminer une éventuelle approche rationnelle ou semi-rationnelle pour la conversion de GH en TG.

1. Postulat pour une approche rationnelle de la conversion de GH en TG

Première constatation, il s'avère que les TG naturelles comparées à leurs homologues hydrolytiques ont une efficacité catalytique (k_{cat}/K_M) beaucoup plus faible, plus de mille fois pour certaines sucrases de la famille GH13 si l'on compare par exemple l'amylosaccharase de *Neisseria polysaccharea* qui est une glucosyltransferase (Potocki de Montalk et al., 2000) à la saccharose hydrolase de *Xanthomonas axonopodis* pv. (Kim et al., 2008). Cette observation est soutenue par plusieurs exemples de conversion par des méthodes d'évolution dirigée de GH en TG qui montrent une forte diminution de l'efficacité catalytique globale des cibles étudiées sur leurs substrats naturels, pour les mutants présentant une meilleure activité de synthèse (Kuriki et al., 1996; Bissaro et al., 2015a), comme par exemple le mutant N282T-F401S de la β -glycosidase de *T. thermophilus* (*Tt*- β -Gly) qui ne conserve que 0,008 % de son activité sur β -D-glucopyranoside de 4-nitrophenyle (*p*NP- β -D-Glcp) par rapport à l'enzyme sauvage (Feng, 2005b). Ce constat nous a permis de faire le postulat que pour générer une TG "efficace", il est nécessaire que son activité hydrolytique pour le substrat choisi soit faible. En effet, limiter autant que possible l'accessibilité des molécules d'eau au site actif (David et al., 2017) et/ou déstabiliser la déglycosylation par une molécule d'eau tout en favorisant le transfert vers une molécule acceptrice (Bissaro et al., 2015a) semblent être des prérequis cohérents pour des réactions qui se déroulent en milieux aqueux où les molécules d'eau sont omniprésentes (55,5 M) par rapport aux molécules acceptrices (présentes à des concentrations de l'ordre de la dizaine à la centaine de mM le plus souvent). La Figure 1 représente le diagramme énergétique de la réaction enzymatique en deux étapes des GH à rétention de configuration. Un mutant ayant une activité hydrolytique déstabilisée verra la barrière énergétique nécessaire pour franchir l'état de transition [TS][‡] des étapes de glycosylation et de déglycosylation augmentée (représentée en pointillés rouges sur la Figure 1).

Le second prérequis est d'utiliser un accepteur qui soit le plus affin possible avec le sous-site accepteur de l'enzyme, mais sans engendrer d'inhibition, pour diminuer la barrière énergétique de l'étape de déglycosylation médiée par l'accepteur (représentée en pointillés verts sur la Figure 1). Dans ces conditions, l'accepteur, bien que statistiquement défavorable, peut alors entrer en compétition avec l'eau, présente en large excès, et rendre énergétiquement favorable la seconde étape de déglycosylation en faveur de la transglycosylation (l'objectif étant d'augmenter autant que possible la barrière énergétique $[TS2]^\ddagger_{\text{eau}}$ et diminuer celle pour $[TS2]^\ddagger_{\text{accepteur}}$, Figure 1).

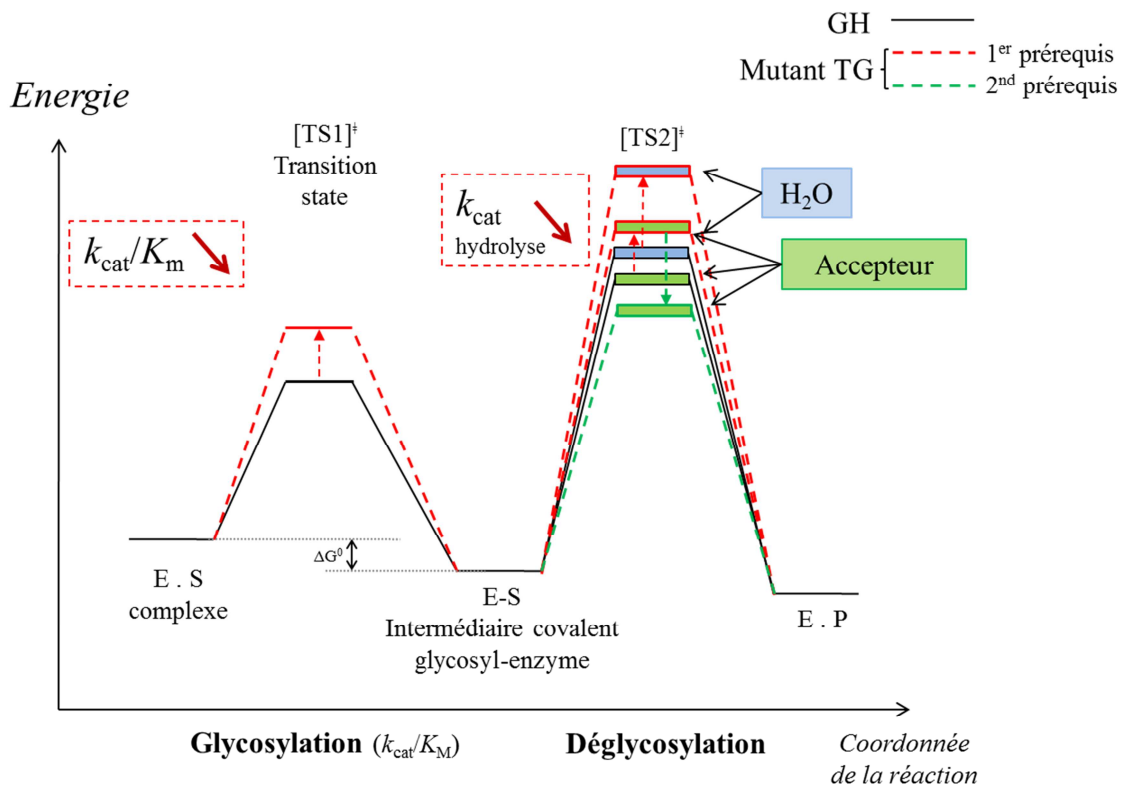


Figure 1 : **Diagramme énergétique de la réaction enzymatique en deux étapes des glycoside hydrolases à mécanisme de rétention de configuration (rGH)** (d'après Bissaro et al, 2015b). Le trait noir représente les barrières énergétiques classiques d'une rGH. Les pointillés rouges représentent un variant de rGH ayant une activité hydrolytique déstabilisée. Les rectangles bleus et verts représentent respectivement soit la barrière énergétique nécessaire à l'hydrolyse soit au transfert sur un accepteur (transglycosylation) lors de l'étape de déglycosylation. Les pointillés verts représentent une diminution de l'énergie d'activation lors du $[TS2]^\ddagger$ nécessaire pour le transfert de la partie glycone sur un accepteur.

2. Exemple de la TxAbf

Ces notions de contrôle des barrières énergétiques sont joliment exemplifiées dans les travaux de conversion de l' α -L-arabinofuranosidase de *Thermobacillus xylanilyticus* (TxAbf), une GH hydrolytique possédant une faible activité de transglycosylation, en trans-arabinofuranosylase (Bissaro

et al., 2015a). Par une approche d'évolution dirigée, 40 000 clones ont été dans un premier temps criblés sur un milieu solide contenant un substrat chromogène, l' α -L-arabinofuranoside de 5-bromo-4-chloro-3-indolyle (X- α -L-Araf), ce qui a permis de sélectionner 368 colonies dont les intensités de coloration étaient plus pâles que les colonies produisant l'enzyme de départ. Des travaux antérieurs permettent de considérer que ces enzymes qui présentent une hydrolyse altérée, sont des candidates pour l'identification d'une activité de transglycosylation accrue (Feng, 2005b; Kone et al., 2009; Arab-Jaziri et al., 2013, 2015; Bissaro et al., 2015a). Un criblage secondaire a ensuite été effectué sur ces clones afin de déterminer leurs activités relatives sur l' α -L-arabinofuranoside de 4-nitrophenyle (*p*NP- α -L-Araf), en présence ou en l'absence d'accepteur (xylotriose). Alors que la plupart des clones affichaient un comportement semblable à l'enzyme sauvage, à savoir une inhibition en présence d'accepteur, trois clones positifs présentaient une activité en présence de xylotriose multipliée jusqu'à neuf fois. Le séquençage de l'ADN des 3 clones a révélé que chacun d'entre eux contenait, en plus d'autres mutations, la substitution R69H, caractérisée par la suite comme responsable du phénotype mesuré. L'analyse du variant R69H purifié a démontré que cette mutation seule est suffisante pour modifier le rapport T/H. Cette Arginine en position 69 figure parmi les huit résidus conservés qui caractérisent le clan des GH-A (Durand et al., 1997), elle est située à 2,86 Å du résidu catalytique nucléophile (E298) avec lequel l'arginine forme une liaison hydrogène. La mutation de l'arginine équivalente de la β -glucosidase de *Thermus thermophilus* (R75A) qui appartient également au clan A, confère à l'enzyme une activité de synthèse également améliorée d'un facteur 2 ainsi qu'une forte diminution de son efficacité catalytique (Teze et al., 2014), également constatée sur la TxAbf. Il semble probable que la mutation R69H provoque une modification de la distribution des charges au niveau du résidu nucléophile de la TxAbf (E298) qui affecte les états de transitions, réduisant ainsi l'efficacité de la déglycosylation médiée par l'eau (Bissaro et al., 2015a). Cette mutation, R69H, combinée avec deux autres L352M et N216W, a conduit à un triple mutant avec un phénotype quasi-exclusivement dédié à la transglycosylation.

Au vu des résultats obtenus sur la TxAbf, nous avons voulu conforter cette stratégie en l'appliquant à une nouvelle GH hydrolytique appartenant à une autre famille (rGH5) et ceci dans le cadre d'une application en adéquation avec les principes de la chimie verte. La synthèse de tensioactifs à base de sucres issus de la biomasse renouvelable par biocatalyse présente un intérêt majeur dans ce contexte. De plus, les APG font partie des tensioactifs non ioniques présentant de nombreux avantages qui seront décrits dans le prochain paragraphe.

3. Propriétés, applications et production des polyglycosides d'alkyles (APG)

Les APG sont composés d'une chaîne alkyle hydrophobe et d'un ou plusieurs sucres qui composent la partie hydrophile. Ce sont des tensioactifs non ioniques qui présentent la propriété de diminuer l'énergie de surface entre deux phases (par exemple, air/eau ou eau/huile). Le caractère plus ou moins

hydrophile d'un tensioactif peut être apprécié à l'aide de l'échelle HLB (« hydrophile lipophile balance »). Cette échelle va de 0 pour les molécules les plus lipophiles à 20 pour les plus hydrophiles. Cette valeur permet d'estimer rapidement l'hydrophilicité d'un tensioactif, ainsi que ses propriétés (Kunieda and Shinoda, 1985).

Le nombre d'applications faisant appel aux APG est considérable et en pleine expansion. Ceci est dû aux avantages qu'ils possèdent sur les autres classes de tensioactifs en termes de biodégradabilité, d'innocuité dermatologique et d'impact sur l'environnement (Pantelic and Cuckovic, 2014). Les principales applications des APG sont classiquement en tant que composants de détergents et produits de nettoyage, produits cosmétiques, ou d'ingrédients alimentaires (von Rybinski and Hill, 1998; Greffe, 2004; Heyd et al., 2008; Reis et al., 2013a; Gudiña et al., 2013). Un domaine émergent pour l'utilisation de biosurfactants -et donc des APG- est le secteur de la santé. C'est alors pour leurs activités antibactérienne, antifongique et antivirale, en plus de leur caractère antiadhésif contre les agents pathogènes, que ces molécules deviennent pertinentes (Pantelic and Cuckovic, 2014). De plus, il a récemment été démontré que certains tensioactifs pourraient avoir un effet bénéfique dans le traitement de certains cancers. Le lipopeptide surfactine, par exemple, induit l'apoptose dans les cellules cancéreuses du sein (Cao et al., 2010). Ces applications démontrent que certains APG, bien caractérisés, peuvent devenir des molécules à forte valeur ajoutée et incitent donc à la découverte et au développement de molécules avec de nouvelles structures et fonctions pour des applications innovantes (Zhao et al., 2000; Kitamoto et al., 2009; Gudiña et al., 2013).

Les APG sont généralement produits par voie chimique, soit à base de résidus de la pétrochimie, soit par la combinaison de sucres, comme le D-glucose issu de l'amidon, et d'acides gras, par catalyse acide à haute température. Il en résulte un mélange d'APG avec des teneurs en sucres et des longueurs de chaînes alkyles variées (Iglauer et al., 2009). Outre la voie chimique, certaines souches bactériennes ou de levures sont capables de réaliser naturellement la synthèse de biosurfactants. Les sophorolipides par exemple, composés d'acides gras à 16 ou 18 carbones et d'un sophorose (disaccharide d'unités D-glucosyles liés en β -1,2), sont produits principalement par les souches de levures *Candida bombicola* et *Candida picola* (Van Bogaert et al., 2007). L'optimisation des procédés a permis l'obtention de rendements élevés de plus de 100 g/L de culture et a donc ouvert la voie à une exploitation industriellement viable de ces souches pour l'utilisation des sophorolipides dans la formulation de produits domestiques, en particulier au Japon et en Corée (Marchant and Banat, 2012). Ces méthodes de production de tensioactifs, plus respectueuses de l'environnement que les méthodes chimiques, sont aujourd'hui plébiscitées par les industriels comme par la société civile.

4. Conversion d'une endoglycocéramidase de type II (EGC II) en transglycosylase

L'enzyme que nous avons ciblée pour notre étude appartient à la famille des endoglycocéramidases (EGC) qui hydrolysent des glycosphingolipides (GSL), plus spécifiquement des gangliosides. La

particularité de ces substrats est de posséder une partie hydrophobe constituée d'un céramide (i.e. une sphingosine liée à un acide gras) et une partie hydrophile constituée d'un oligosaccharide à base de lactose (Figure 2). Cette dualité, propre aux molécules amphiphiles, est une opportunité pour la synthèse de potentiels agents tensioactifs tels que les APG.

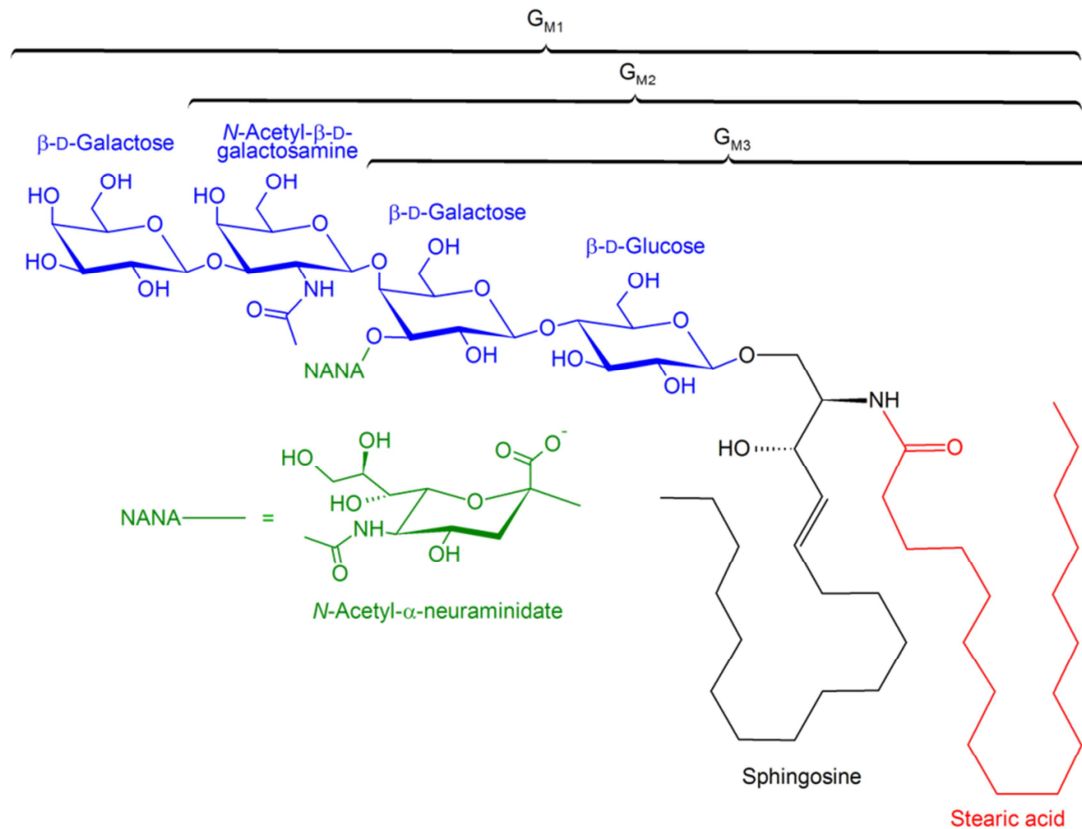


Figure 2 : Structures des gangliosides G_{M1}, G_{M2} et G_{M3}. Nomenclature des gangliosides en deux lettres et un chiffre : G = ganglioside; M, D, T ou Q en indice pour 1, 2, 3 ou 4 résidus d'acide sialique dans l'oligosaccharide et 3, 2 ou 1 en indice pour 2, 3 ou 4 résidus glycosyles dans la chaîne à l'exception des acides sialiques. Exemple : un ganglioside G_{M1} contient un chaînon oligosaccharidique de 4 sucres plus un acide sialique.

a. Trois familles d'EGC et leurs spécificités de substrats

Les EGC sont classées en trois catégories suivant leur spécificité de substrat parmi les GSL. Les GSL sont des constituants des membranes cellulaires d'organismes allant des bactéries à l'homme. 90 % des GSL de mammifère sont composés d'une base de type « lactosyl céramide » (β -D-Galp-(1-4)- β -D-Glcp-cer) et sont les substrats des EGC I et II ; 10 % possèdent un résidu β -D-galactosyle comme unité osidique "réductrice" au motif céramide et sont hydrolysés par la famille des EGC III, autrement appelées EGALC pour endogalactosylcéramidase (Figure 3).

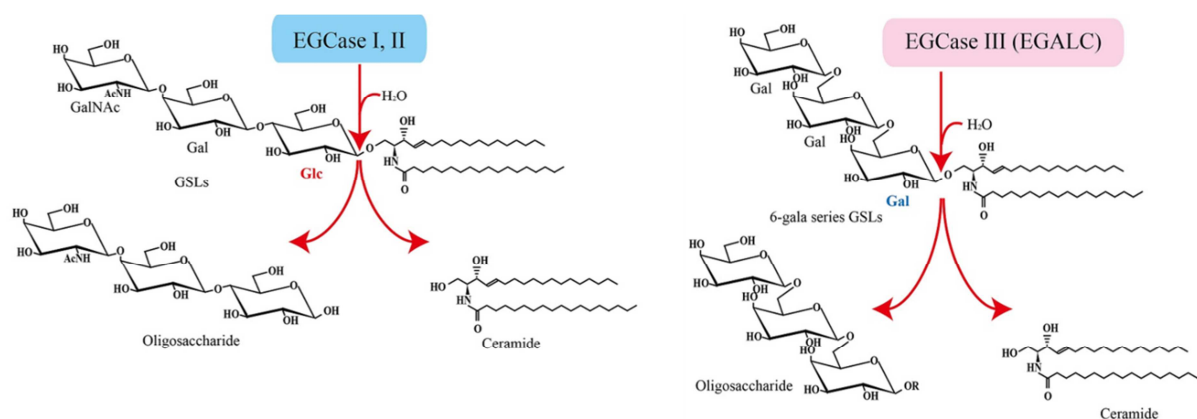


Figure 3 : Spécificité et site d'action des 3 familles d'EGC

<http://jcgddb.jp/GlycoPOD/protocolShow.action?nodeId=t175> Web.10,10,2017

Ces 3 familles d'enzymes appartiennent à la même famille de CAZymes (GH5), et partagent donc un même repliement de type tonneau (β/α)₈. Leur différence de spécificité peut en partie s'expliquer par la présence ou l'absence d'une boucle au-dessus du site actif. Ainsi, en 2012, Ishibashi et al., ont produit l'EGC I de *Rhodococcus equi* et ont pu, en se basant sur le modèle de la structure cristalline de l'EGC II provenant de la bactérie *Rhodococcus* sp. (M. E. C. Caines et al., 2007), générer un modèle prédictif tridimensionnel de l'enzyme. Cette comparaison montre qu'une boucle très labile (Pro145 à Gly154) se trouvant au-dessus du site actif de l'enzyme EGC II est absente sur l'EGC I. Le mutant EGC II délété de cette boucle voit son activité hydrolytique augmentée d'un facteur 2 sur les substrats β -D-Galp-(1-3)- β -D-GalpNAc-(1-4)- α -D-NeuAc-(2-3)- β -D-Galp-(1-4)- β -D-Glcp-Cer (GM1a) et α -D-Galp-(1-4)- β -D-Galp-(1-4)- β -D-Glcp-Cer (Gb3Cer) et démontre ainsi que cette boucle est impliquée dans la spécificité de substrat des enzymes (Ishibashi et al., 2012).

b. Choix de l'enzyme modèle

L'enzyme choisie est l'EGC II (EC 3.2.1.123) provenant de la bactérie *Rhodococcus* sp. M-777. Elle est la seule enzyme figurant parmi les 18 séquences référencées à ce jour dans la famille des GH5, sous-famille 28 dont la structure 3D est disponible (Aspeborg et al., 2012a). L'EGC II présente un repliement de type tonneau (β/α)₈ en N-ter et un domaine β sandwich en C-ter (Figure 4) et appartient au clan A dans la classification CAZy (<http://www.cazy.org/>) (Henrissat, 1991). Elle se compose de 483 acides aminés pour une masse moléculaire de 52 218 Da. Les structures de cette enzyme sous sa forme *apo* ou en complexe avec différents substrats ont été résolues en 2007 par Caines et al. et se trouvent sous les codes 2OYK, 2OYL, 2OYM, (Caines et al., 2007a) 2OSW, 2OSX, 2OSY (Caines et al., 2007b) de la Protein Data Bank (PDB). Sa structure dévoile un site actif en cohérence avec la dualité de son substrat, à savoir une grande cavité polaire qui accueille la moitié saccharidique et un couloir hydrophobe qui interagit avec le céramide. L'EGC II est une *endo*-enzyme, qui hydrolyse les

GSLs entre le céramide et l'oligosaccharide contrairement aux *exo*-enzymes qui hydrolysent leurs substrats par unité monosaccharidique.

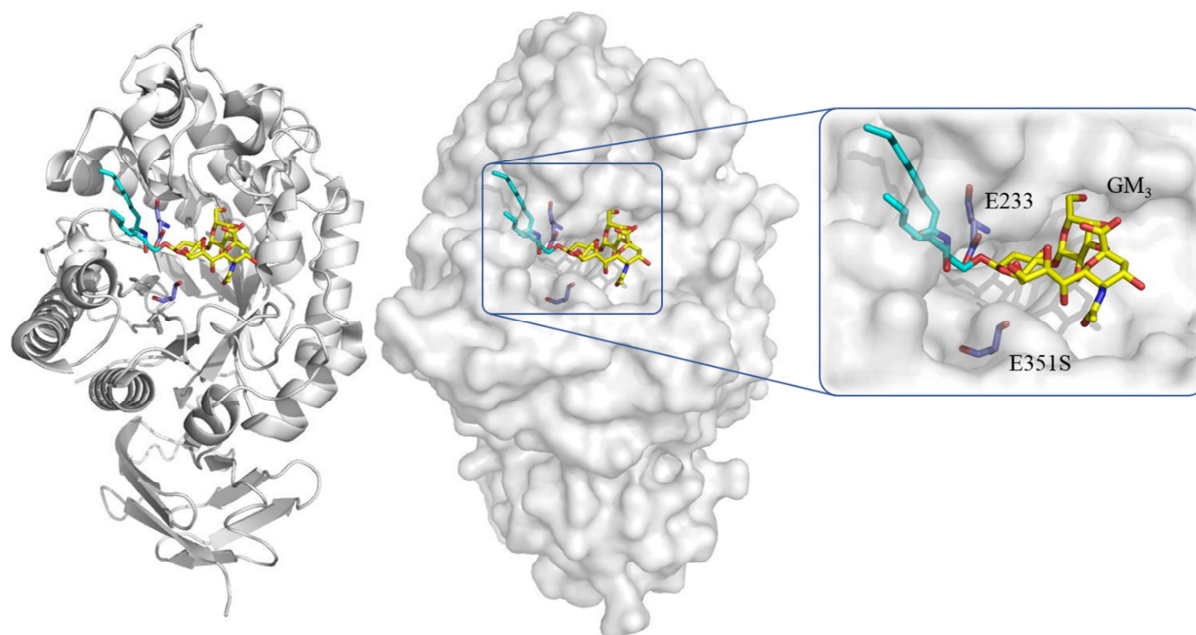


Figure 4 : **Structure tertiaire et de surface de l'EGC II de *Rhodococcus* sp. M-777 en complexe avec le substrat GM3** (code PDB 2OSX). Le céramide du GM3 est représenté en bleu et la partie oligosaccharidique en jaune. Les résidus catalytiques E233 et E351S sont représentés en violet.

La littérature nous indique que les EGC possèdent, souvent très modestement, une activité de transglycosylation sur certains alcools, céramides ou sur le détergent Triton X-100 (Li et al., 1991; Ashida et al., 1993; Horibata et al., 2001; Ishibashi et al., 2007a, 2009, 2012). Il est donc potentiellement envisageable d'augmenter cette activité par ingénierie de ces enzymes.

Des travaux s'appuyant sur le concept des glycosynthases (Malet and Planas, 1998; Moracci et al., 1998; Mackenzie et al., 1998) appliqués à l'EGC II de *Rhodococcus* sp. ont démontré que cette enzyme peut en effet être un outil de choix pour la synthèse de glycoconjugués d'intérêts et particulièrement pour la synthèse de GSLs (Vaughan et al., 2006a; Rich et al., 2011a; Rich and Withers, 2012).

5. Stratégie utilisée pour la conversion de l'EGC II en TG

Dans un premier temps, nous avons déterminé les paramètres cinétiques de l'EGC II pour l'hydrolyse du substrat donneur β -cellobioside de 2-chloro-4-nitrophenyle (CNP- β -Celb) qui possède, comme les gangliosides ou lactosides (substrats naturels), une unité D-glucosyle en première position après la liaison éther, se positionnant donc au niveau du sous-site -1 (Tableau 1). Nous constatons que l'efficacité catalytique de l'EGC II est faible pour ce substrat, l'enzyme affichant une valeur de k_{cat}/K_M

de $0,165 \text{ s}^{-1}.\text{mM}^{-1}$. Les seuls paramètres cinétiques concernant l'EGC II sur substrat naturel ont été déterminés pour l'hydrolyse du GM_{1a} ($k_{\text{cat}}/K_M = 13,26 \text{ s}^{-1}.\text{mM}^{-1}$), un substrat pour lequel l'EGC II a une activité faible, comparé à d'autres GSL. Malgré cela, il est constaté que l'efficacité catalytique sur CNP- β -Celb est 80 fois inférieure à celle sur GM_{1a} .

En outre, indépendamment de toute comparaison à d'autres EGC sur différents substrats, si l'on regarde l'efficacité catalytique de 1882 enzymes (toutes classes confondues) (Bar-Even et al., 2011), on constate que la valeur moyenne du ratio k_{cat}/K_M est de $125 \text{ s}^{-1}.\text{mM}^{-1}$, ce qui situe l'efficacité catalytique de l'EGC II pour le CNP- β -Celb parmi les plus faibles.

Tableau 1 : Paramètres cinétiques de l'EGC II pour l'hydrolyse du CNP- β -Celb et GM_{1a}

Substrat	$k_{\text{cat}} (\text{s}^{-1})$	$K_M (\text{mM})$	$k_{\text{cat}}/K_M (\text{s}^{-1}.\text{mM}^{-1})$
CNP- β -Celb	3,7	22, 38	0,16
¹ GM_{1a}	5,7	0,43	13,26

¹Valeurs pour GM_{1a} d'après (Ishibashi et al., 2012)

Au vu de notre postulat de départ qui requiert une enzyme dont l'efficacité catalytique (k_{cat}/K_M) est faible (Figure 1), nous pouvons considérer que, pour cette enzyme et ce substrat, ce premier prérequis est obtenu. Le défi que nous avons eu à relever a donc été de trouver le meilleur accepteur possible pour ce substrat donneur.

Notre choix initial s'est porté sur des alcanols simples, présentant une seule fonction hydroxyle sur une chaîne alkyle linéaire. Des réactions de transfert de cellobiosyle (à partir de CNP- β -Celb) sur des accepteurs tels que le pentanol ou l'octanol n'ont pas permis de détecter des produits de synthèse par CCM. Dans le but d'augmenter les chances de reconnaissance des molécules acceptrices par notre enzyme, nous avons alors ciblé des molécules structurellement plus proches des substrats naturels de l'enzyme. Ainsi, nous avons porté notre choix sur des accepteurs portant soit une fonction hydroxyle supplémentaire (1,3-diols), soit une fonction carbonyle supplémentaire (δ -hydroxycétones, Figure 5 A). Ces groupements présents sur la partie aglycone (céramide) des GSL permettent un plus grand mimétisme des accepteurs avec les substrats naturels des EGC.

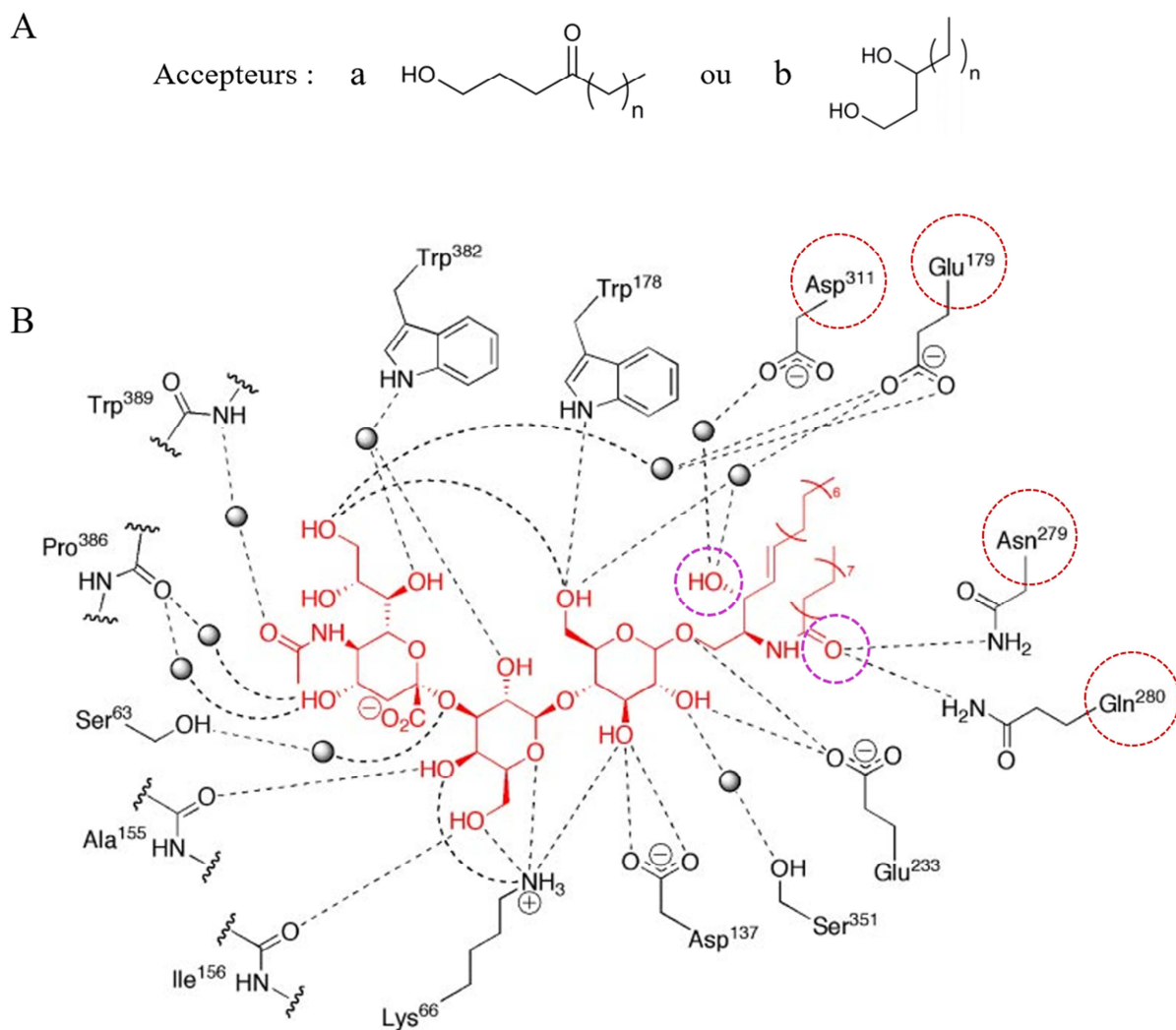


Figure 5 : **Molécules acceptrices et résidus ciblés pour la mutagenèse à saturation.** A. Dérivés aliphatiques fonctionnalisés utilisés comme accepteur pour optimiser l'affinité de l'enzyme pour ces molécules : a) δ -hydroxycétones ou b) 1,3-diols. B. Représentation schématique des interactions impliquées dans la liaison entre l'EGC II et un substrat (G_{M3}). Les 4 résidus entourés en rouges sont ciblés pour la mutagenèse afin de moduler le positionnement des molécules alkyles acceptrices sélectionnées ; les groupements hydroxyle ou carbonyle du substrat, identifiés comme critiques sont en violet sur le schéma. Les molécules d'eau sont représentées par des sphères grises. D'après (M. E. C. Caines et al., 2007).

L'utilisation de ces nouveaux accepteurs permet cette fois d'obtenir des produits de transglycosylation, mais avec une efficacité encore modérée, les produits de synthèse devenant juste détectables par CCM. Afin d'optimiser la reconnaissance de ces accepteurs, quatre résidus situés dans le sous-site accepteur, au début du tunnel hydrophobe qui reconnaît la partie aglycone des substrats de l'EGC ont été soumis à de la mutagenèse à saturation (Figure 5 B). Le criblage de ces quatre banques va permettre de mettre à jour une mutation permettant une synthèse d'APG avec des rendements très fortement améliorés par rapport à l'enzyme sauvage.

Références :

- Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., Fauré, R., O'Donohue, M.J., 2013. Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.* 30, 536–544.
- Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R., O'Donohue, M.J., 2015. Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α -L-arabinofuranosidase. *Carbohydr. Res.* 401, 64–72.
- Ashida, H., Tsuji, Y., Yamamoto, K., Kumagai, H., Tochikura, T., 1993. Transglycosylation activity of endoglycoceramidase from *corynebacterium* sp. *Arch. Biochem. Biophys.* 305, 559–562.
- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., Henrissat, B., 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12, 1.
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The Moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410.
- Bissaro, B., Durand, J., Biarnés, X., Planas, A., Monsan, P., O'Donohue, M.J., Fauré, R., 2015a. Molecular design of non-Leloir furanose-transferring enzymes from an α -L-Arabinofuranosidase: A rationale for the engineering of evolved transglycosylases. *ACS Catal.* 5, 4598–4611.
- Bissaro, B., Monsan, P., Fauré, R., O'Donohue, M.J., 2015b. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* 467, 17–35.
- Caines, M.E.C., Hancock, S.M., Tarling, C.A., Wrodnigg, T.M., Stick, R.V., Stütz, A.E., Vasella, A., Withers, S.G., Strynadka, N.C.J., 2007. The structural basis of glycosidase inhibition by five-membered iminocyclitols: The Clan A glycoside hydrolase endoglycoceramidase as a model system. *Angew. Chem. Int. Ed.* 46, 4474–4476.
- Caines, M.E.C., Vaughan, M.D., Tarling, C.A., Hancock, S.M., Warren, R.A.J., Withers, S.G., Strynadka, N.C.J., 2007. Structural and mechanistic analyses of endo-Glycoceramidase II, a membrane-associated Family 5 glycosidase in the apo and GM3 ganglioside-bound Forms. *J. Biol. Chem.* 282, 14300–14308.
- Cao, X., Wang, A., Wang, C., Mao, D., Lu, M., Cui, Y., Jiao, R., 2010. Surfactin induces apoptosis in human breast cancer MCF-7 cells through a ROS/JNK-mediated mitochondrial/caspase pathway. *Chem. Biol. Interact.* 183, 357–362.
- David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.-H., Tellier, C., 2017. Internal water dynamics control the transglycosylation/hydrolysis balance in the agarase (AgaD) of *Zobellia galactanivorans*. *ACS Catal.* 7, 3357–3367.
- Dilokpimol, A., Nakai, H., Gotfredsen, C.H., Baumann, M.J., Nakai, N., Abou Hachem, M., Svensson, B., 2011. Recombinant production and characterisation of two related GH5 endo- β -1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1814, 1720–1729.
- Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B., Mornon, J.-P., 1997. Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* 7, 277–284.
- Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C., Dion, M., 2005. Converting a β -glycosidase into a β -transglycosidase by directed evolution. *J. Biol. Chem.* 280, 37088–37097.

- Grefe, L., Bessueille, L., Bulone, V., Brumer, H., 2004. Synthesis, preliminary characterization, and application of novel surfactants from highly branched xyloglucan oligosaccharides. *Glycobiology* 15, 437–445.
- Gudiña, E.J., Rangarajan, V., Sen, R., Rodrigues, L.R., 2013. Potential therapeutic applications of biosurfactants. *Trends Pharmacol. Sci.* 34, 667–675.
- Henrissat, B., 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316.
- Heyd, M., Kohnert, A., Tan, T.-H., Nusser, M., Kirschhöfer, F., Brenner-Weiss, G., Franzreb, M., Berensmeier, S., 2008. Development and trends of biosurfactant analysis and purification using rhamnolipids as an example. *Anal. Bioanal. Chem.* 391, 1579–1590.
- Horibata, Y., Higashi, H., Ito, M., 2001. Transglycosylation and reverse hydrolysis reactions of endoglycoceramidase from the jellyfish, *Cyanea nozakii*. *J. Biochem.* 130, 263–268.
- Iglauer, S., Wu, Y., Shuler, P., Tang, Y., Goddard, W.A., 2009. Alkyl polyglycoside surfactant–alcohol cosolvent formulations for improved oil recovery. *Colloids Surf. Physicochem. Eng. Asp.* 339, 48–59.
- Ishibashi, Y., Kiyohara, M., Okino, N., Ito, M., 2007. Synthesis of fluorescent glycosphingolipids and neoglycoconjugates which contain 6-gala oligosaccharides using the transglycosylation reaction of a novel endoglycoceramidase (EGALC). *J. Biochem.* 142, 239–246.
- Ishibashi, Y., Kobayashi, U., Hijikata, A., Sakaguchi, K., Goda, H.M., Tamura, T., Okino, N., Ito, M., 2012. Preparation and characterization of EGCase I, applicable to the comprehensive analysis of GSLs, using a rhodococcal expression system. *J. Lipid. Res.* 53, 2242–2251.
- Ishibashi, Y., Nagamatsu, Y., Meyer, S., Imamura, A., Ishida, H., Kiso, M., Okino, N., Geyer, R., Ito, M., 2009. Transglycosylation-based fluorescent labeling of 6-gala series glycolipids by endogalactosylceramidase. *Glycobiology* 19, 797–807.
- Kim, M.-I., Kim, H.-S., Jung, J., Rhee, S., 2008. Crystal structures and mutagenesis of sucrose hydrolase from *Xanthomonas axonopodis* pv. *glycines*: insight into the exclusively hydrolytic amylosucrase fold. *J. Mol. Biol.* 380, 636–647.
- Kitamoto, D., Morita, T., Fukuoka, T., Konishi, M., Imura, T., 2009. Self-assembling properties of glycolipid biosurfactants and their potential applications. *Curr. Opin. Colloid Interface Sci.* 14, 315–328.
- Kone, F.M.T., Le Behec, M., Sine, J.-P., Dion, M., Tellier, C., 2008. Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.* 22, 37–44.
- Kuriki, T., Kaneko, H., Yanase, M., Takata, H., Shimada, J., Handa, S., Takada, T., Umeyama, H., Okada, S., 1996. Controlling substrate preference and transglycosylation activity of neopullulanase by manipulating steric constraint and hydrophobicity in active center. *J. Biol. Chem.* 271, 17321–17329.
- Li, Y.T., Carter, B.Z., Rao, B.N., Schweingruber, H., Li, S.-C., 1991. Synthesis of neoglycoconjugates using oligosaccharide transferring activity of ceramide glycanase. *J. Biol. Chem.* 266, 10723–10726.
- Mackenzie, L.F., Wang, Q., Warren, R.A.J., Withers, S.G., 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* 120, 5583–5584.
- Malet, C., Planas, A., 1998. From β -glucanase to β -glucansynthase: glycosyl transfer to α -glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* 440, 208–212.
- Marchant, R., Banat, I.M., 2012. Biosurfactants: a sustainable replacement for chemical surfactants? *Biotechnol. Lett.* 34, 1597–1605.

- Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., Rossi, M., 1998. Restoration of the activity of active-site mutants of the hyperthermophilic β -Glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* 37, 17262–17270.
- Ochs, M., Muzard, M., Plantier-Royon, R., Estrine, B., Rémond, C., 2011. Enzymatic synthesis of alkyl β -D-xylosides and oligoxylosides from xylans and from hydrothermally pretreated wheat bran. *Green Chem.* 13, 2380.
- Oh, Su-Won, Jang, M.-U., Jeong, C.-K., Kang, H.-J., Park, Kim, T.-J., 2008. Modulation of Hydrolysis and Transglycosylation Activity of *Thermus Maltogenic* Amylase by Combinatorial Saturation Mutagenesis. *J. Microbiol. Biotechnol.* 18, 1401–1407.
- Pantelic, I., Cuckovic, B., 2014. Alkyl Polyglucosides: An emerging class of sugar surfactants, in: *Alkyl Polyglucosides*. Woodhead Publishing Series in Biomedicine, pp. 1–19.
- Pennec, A., Daniellou, R., Loyer, P., Nugier-Chauvin, C., Ferrières, V., 2015. Araf51 with improved transglycosylation activities: one engineered biocatalyst for one specific acceptor. *Carbohydr. Res.* 402, 50–55.
- Potocki de Montalk, G., Remaud-Simeon, M., Willemot, R.M., Sarçabal, P., Planchot, V., Monsan, P., 2000. Amylosucrase from *Neisseria polysaccharea*: novel catalytic properties. *FEBS Lett.* 471, 219–223.
- Reis, R.S., Pacheco, G.J., Pereira, A.G., Freire, D.M.G., 2013. Biosurfactants: Production and Applications, in: Chamy, R. (Ed.), *Biodegradation - Life of Science*. InTech.
- Rich, J.R., Cunningham, A.-M., Gilbert, M., Withers, S.G., 2011. Glycosphingolipid synthesis employing a combination of recombinant glycosyltransferases and an endoglycoceramidase glycosynthase. *Chem. Commun.* 47, 10806.
- Rich, J.R., Withers, S.G., 2012. A chemoenzymatic total synthesis of the neurogenic starfish ganglioside LLG-3 using an engineered and evolved synthase. *Angew. Chem. Int. Ed.* 51, 8640–8643.
- Teze, D., Daligault, F., Ferrieres, V., Sanejouand, Y.-H., Tellier, C., 2015. Semi-rational approach for converting a GH36 -glycosidase into an -transglycosidase. *Glycobiology* 25, 420–427.
- Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., Tellier, C., Dion, M., 2014. Semi-rational approach for converting a GH1 β -glycosidase into a β -transglycosidase. *Protein Eng. Des. Sel.* 27, 13–19.
- Van Bogaert, I.N.A., Saerens, K., De Muynck, C., Develter, D., Soetaert, W., Vandamme, E.J., 2007. Microbial production and application of sophorolipids. *Appl. Microbiol. Biotechnol.* 76, 23–34.
- Vaughan, M.D., Johnson, K., DeFrees, S., Tang, X., Warren, R.A.J., Withers, S.G., 2006. Glycosynthase-mediated synthesis of glycosphingolipids. *J. Am. Chem. Soc.* 128, 6300–6301.
- von Rybinski, W., Hill, K., 1998. Alkyl polyglycosides—properties and applications of a new class of surfactants. *Angew. Chem. Int. Ed.* 37, 1328–1345.
- Zhao, X., Geltinger, C., Kishikawa, S., Ohshima, K., Murata, T., Nomura, N., Nakahara, T., Yokoyama, K.K., 2000. Treatment of mouse melanoma cells with phorbol 12-myristate 13-acetate counteracts mannosylerythritol lipid-induced growth arrest and apoptosis. *Cytotechnology* 33, 123–130.

Publication:

A Single Point Mutation Alters the
Transglycosylation/Hydrolysis Partition, Significantly Enhancing the
Synthetic Capability of an Endo-Glycoceramidase

ACS Catal., 2016, 6 (12), pp 8264–8275

doi: 10.1021/acscatal.6b02159

A Single Point Mutation Alters the Transglycosylation/Hydrolysis Partition, Significantly Enhancing the Synthetic Capability of an *Endo*-Glycoceramidase

Julien Durand,[†] Xevi Biarnés,[‡] Laurie Watterlot,[†] Cyrielle Bonzom,[†] Vinciane Borsenberger,[†] Antoni Planas,[‡] Sophie Bozonnet,[†] Michael J. O'Donohue,[†] Régis Faure^{*,†}

[†]LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

[‡]Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain

* Corresponding Author

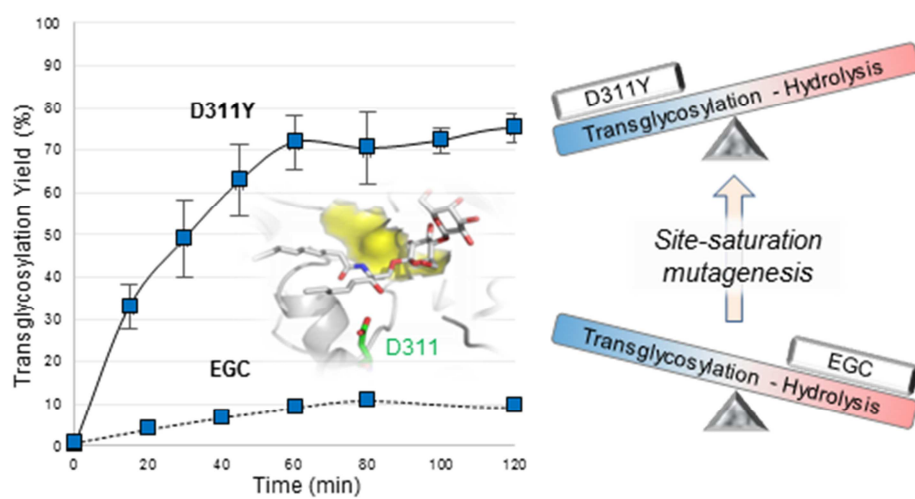
regis.faure@insa-toulouse.fr

ABSTRACT:

The mutation of D311 to tyrosine in *endo*-glycoceramidase II from *Rhodococcus* sp. and the use of a poorly recognized substrate, 2-chloro-4-nitrophenyl β -cellobioside, have provided appropriate conditions for the efficient synthesis of alkyl β -cellobioside derivatives. The mutant D311Y was characterized by a lowered K_M value for the hydrolysis of 2-chloro-4-nitrophenyl β -cellobioside and increased transglycosylation when using aliphatic 1,3-diols or alcohols bearing a δ -hydroxy ketone function as acceptors. Closer analysis revealed that the transglycosylation/hydrolysis ratio in reactions catalyzed by the mutant was completely inverted and weak secondary hydrolysis was postponed, thus providing the basis for high transglycosylation yields (between 68 and 93%). Overall, results confirm that the enhancement of transglycosylation in glycoside hydrolases can be achieved by a combination of destabilized transition states and increased recognition for acceptor molecules.

KEYWORDS: glycoside hydrolase, mutagenesis, transglycosylation, alkyl cellobiosides, molecular interactions, chemo-enzymatic glycosynthesis

Graphical abstract



INTRODUCTION

Alkyl polyglycosides (APGs) are of commercial interest, since they display non-ionic surfactant properties that are exploited for the preparation of a wide variety of products, including detergents and cleaners, personal care products, or food ingredients (von Rybinski and Hill, 1998; Greffe, 2004; Heyd et al., 2008; Reis et al., 2013b; Gudiña et al., 2013). Remarkably, they also have been shown to reduce skin irritation compared to conventional detergents (Löffler and Happle, 2003; Schwarz et al., 2012). The market for this family of molecules is expected to grow, because of the drive to reduce the use of petrochemical-based surfactants, favoring the manufacture of bio-based products using eco-friendly technologies. However, most APGs are in reality alkyl glycosides displaying an average degree of polymerization (DP) of the saccharidic moiety close to 1. This is because APGs are currently synthesized through acid-catalyzed Fischer or Koenigs-Knorr glycosidation, reductive amination or using Lewis acid catalysts (Rather and Mishra, 2013). However, the products of these reactions are mainly monosaccharide derivatives, because the anomeric centers of oligosaccharide moieties display lower reactivity and their glycosidic bonds are acid-sensitive. An alternative route to obtain a new generation of APGs displaying saccharidic moieties with higher DP is enzymatic synthesis. In this case, it is possible to benefit from the high selectivity that is a hallmark of enzymes. Several examples of enzymatic APG synthesis using glycoside hydrolases (GHs) illustrate the feasibility and efficiency of this route and demonstrate that it is possible to obtain products in good yields (Wang and Huang, 2009; Rather and Mishra, 2013). Nevertheless, naturally-occurring GHs are mostly enzymes that hydrolyze glycosidic bonds, with bond synthesis occurring either in specific conditions (high substrate concentration and low water activity) that favor so-called reverse hydrolysis, or through a mechanistically-related transglycosylation reaction that is under kinetic control (Crout and Vic, 1998; Van Rantwijk et al., 1999). Only a relatively small number of GHs, known as non-Leloir transglycosylases (TGs), naturally display strong innate ability to catalyze transglycosylation reactions. Interestingly, the key molecular determinants that differentiate TGs from hydrolytic GHs are still unclearly defined, despite considerable efforts to discover these and relevant studies on reversed transglycosylation/hydrolysis (T/H) partition (Teze et al., 2014, 2015, Bissaro et al., 2015a, 2015b; Lundemo et al., 2016; Talens-Perales et al., 2016).

The class of GHs (retaining GHs, hereafter referred to as rGHs) mostly holding the potential to perform transglycosylation reactions are those that catalyze the hydrolysis of glycosidic bonds through a double displacement mechanism that results in the retention (in the product) of the configuration of the anomeric carbon originally present in the donor substrate. The reaction trajectory of rGHs is characterized by the formation of a more or less transient glycosylated enzyme reaction intermediate that can be deglycosylated either by a water molecule (hydrolysis), or by another suitable acceptor moiety (transglycosylation), including sugars and alcohols (Figure 1) (Lundemo et al., 2013; Danby and Withers, 2016). This implies that the outcome of rGH-catalyzed reactions is influenced by

competition at the deglycosylation step between water molecules and any alternative acceptors that might be present. In most cases, because water is omnipresent at a concentration of 55 M in aqueous reaction media, hydrolysis is the principal outcome of a rGH-catalyzed reaction, although transglycosylation often occurs to some extent, with glycosylated products being obtained at low yields, except in the special case of TGs (Bissaro et al., 2015b). Moreover, in hydrolytic rGHs, yields of transglycosylation products are further diminished by secondary hydrolysis, a reaction in which the synthetic product becomes a substrate for the enzyme.

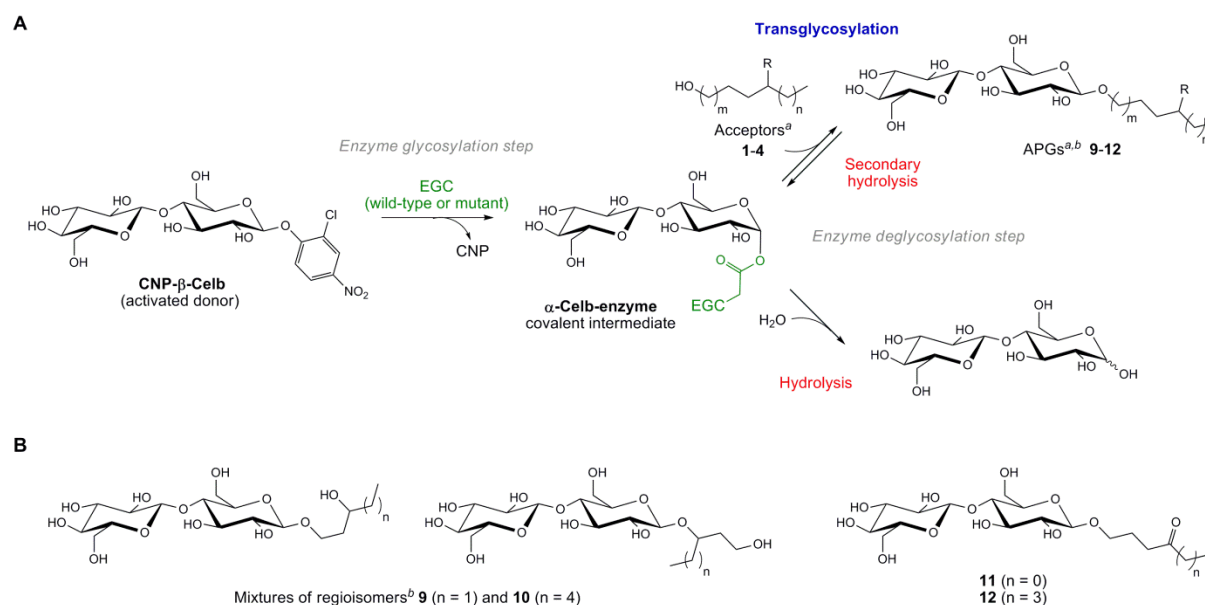


Figure 1. (A) Reactions catalyzed by the retaining GH5 *endo*-glycoceramidase II (EGC) from *Rhodococcus* sp. M-777 strain and its evolved mutant. (B) Structures of the transglycosylation products **9-12**. ^a 1,3-Diol **1** and **2** (R = OH, m = 0, n = 1 and 4 respectively) and δ -hydroxy ketones **3** and **4** (R = O, m = 1, n = 0 and 3 respectively). ^b The secondary hydroxyl group of 1,3-diol **1** and **2** was also glycosylated leading to an almost stoichiometric mixture of regioisomers of APGs **9** and **10**.

Numerous reports recount the enzymatic synthesis of APGs using rGHs displaying different sugar specificities and various alcohols (Rémond et al., 2002; Lundemo et al., 2013). Overall, previous studies have revealed that rGH-catalyzed alkylation is mainly restricted to short-chain primary alcohols, probably because the longer chains are not well accommodated in the active site of GHs and are increasingly insoluble in aqueous reaction media. To counter the latter problem and increase the yield of longer aliphatic chain APGs, some authors have deployed specific strategies, such as the use of co-solvents and vigorous shaking, or the use of immobilized enzyme systems (Ochs et al., 2011). However, an alternative strategy is to perform protein engineering, altering the enzyme's aglycone binding subsites in order to increase affinity for longer chain alcohols. In the current study, this option has been applied to type II *endo*-glycoceramidase (EGC) from *Rhodococcus* sp. M-777 strain, an

enzyme that hydrolyzes the β -glycosidic bonds of gangliosides, releasing an oligosaccharide moiety and a ceramide (Ito and Tatsuya, 1989; Ito and Yamagata, 1986; Izu et al., 1997). According to the CAZy classification (Lombard et al., 2014), this rGH belongs to family 5 (subfamily 28) (Aspeborg et al., 2012b) that is in turn assigned to clan A, the members of which bear a catalytic domain that displays $(\beta/\alpha)_8$ structural architecture. Regarding the active site topology of EGC, previous studies have shown that this enzyme presents a large polar cavity that accommodates the polyhydroxylated oligosaccharide moiety of the substrate and a narrow hydrophobic tunnel that binds the lipidic chains of the ceramide (M. E. C. Caines et al., 2007).

Previous works on *endo*-glycoceramidases have revealed their natural ability to catalyze transglycosylation reactions (Li et al., 1991; Ashida et al., 1993; Horibata et al., 2001; Ishibashi et al., 2007a, 2007b, 2009), although the reported reaction yields are not high, no doubt because the products of transglycosylation are invariably subjected to so-called secondary hydrolysis by EGC. To attenuate hydrolysis, a catalytically-impotent version of EGC has been created, using the principles of the glycosynthase methodology (Mackenzie et al., 1998; Malet and Planas, 1998; Moracci et al., 1998), which involve substitution of the catalytic nucleophile residue (E351 in EGC) by one that does not display requisite nucleophile character (e.g., Ala, Cys, Gly, or Ser). This work resulted in the creation of a hydrolytically-inactive enzyme (containing the mutation E351S) that can only catalyze transglycosylation reactions in the presence of fluorinated donor substrates displaying α -configured anomeric centers (Vaughan et al., 2006b). To improve the overall performance of this EGC-based glycosynthase, it was then submitted to random mutagenesis and subsequent screening using an ELISA-based assay. This approach ultimately procured an evolved mutant E351S-D314Y, which displayed the ability to synthesize complex glycosphingolipids in 45-86% yields starting from the acceptor D-erythro-sphingosine and tri- or penta-saccharyl donors (Hancock et al., 2009; Rich et al., 2011b; Rich and Withers, 2012).

In the current study, to develop new enzyme tools for APG synthesis and possibly to prepare glycosphingolipids (Wennekes et al., 2009; Stocker and Timmer, 2013), we have engineered EGC, targeting its ability to perform transglycosylation using a cellobiosyl moiety as the donor and simple aliphatic alcohols or derivatives (aliphatic 1,3-diols and δ -hydroxy ketones) as acceptors (Figure 1). Unlike the aforementioned study that used glycosynthase methodology, we employed a site-directed saturation mutagenesis approach, targeting amino acids that form part of the long narrow hydrophobic tunnel that constitutes the aglycone binding subsites of EGC. This strategy procured a mutant (D311Y) that displays significantly enhanced ability to catalyze transglycosylation and the means to prepare a series of alkyl cellobioside derivatives.

EXPERIMENTAL SECTION

Substrates, Materials and Chemicals. 2-Chloro-4-nitrophenyl β -cellobioside (CNP- β -Celb) was purchased from Megazyme. Pentane-1,3-diol (**1**), octane-1,3-diol (**2**), 5-hydroxypentan-2-one (**3**) and 1-hydroxyoctan-4-one (**4**) were purchased from FCH Group. Molecular biology products were purchased from New England Biolabs. The synthetic gene encoding EGC (GenScript), optimized for expression in *E. coli* (Figure S1) (Johnson et al., 2014), was cloned into the vector pET28 (Novagen) between the *NdeI* and *XhoI* sites. The resulting plasmid, pET28-EGC, was used to transform *E. coli*, selecting for its presence in cells by the inclusion of kanamycin ($50 \mu\text{g}\cdot\text{mL}^{-1}$) in the growth medium.

Creation of Site-Saturation Mutant Libraries and Site-Directed Mutagenesis. Saturation mutagenesis was carried out using the QuikChange mutagenesis kit (Agilent). The DNA template for mutagenesis was the plasmid pET28-EGC and mutations were introduced using a series of oligonucleotide primers (Eurogentec), designed using NNK degeneracy at positions E179, N279, Q280, and D311 (Table 1).

Table 1. List of primers used for the construction of the mutated sequences

Primer name	Sequence (5' to 3')
E179	GAACCCCAACCCCGCTGGNNKCTGTATTACATCCAGCCAG
N279	CCGCAGGCCATTGGCGTTNNKCAAGGTTTACCATCGGGC
Q280	CCGCAGGCCATTGGCGTTAATNNKGGTTTACCATCGGGCT
D311	GTACCCGCTGCCATTGNNKATCGGCGACGGCCACG
D314Y	GCCATTGGACATCGGCTATGGCCACGAAGGACTTG

Mismatched bases are underlined; N is A, G, C, or T; K is G or T; M is A or C; and MNN the reverse complement of NNK.

To perform mutagenesis an appropriate reaction mixture containing all of the necessary reagents (template DNA, a single primer pair, buffer, deoxyribonucleotides and thermostable DNA polymerase) was placed in a thermocycler and submitted to 16 repetitions of the following protocol: 95 °C for 30 s, 55 °C for 1 min and 68 °C for 13 min for the four NNK libraries and 7 min in the case of D314Y. After, any remaining parental DNA template was eliminated by the addition of *Dpn I* to the reaction mixture (1 h at 37 °C with 10 units of *DpnI*) before using 2 μL to transform chemically competent *E. coli* TOP10 or *E. coli* XL10-Gold cells. After growth, transformants (always > 200 cfus) were pooled and total plasmid DNA was extracted and stored at -20 °C. In this way, four mutant libraries were prepared. The successful introduction of mutations and their diversity as well as

conservation of sequence integrity were confirmed by full-length gene sequencing of five to ten clones per library (GATC Biotech, Mulhouse, France).

Mutant Library Screening. For library screening, an aliquot of the pooled plasmid DNA was used to transform *E. coli* BL21 DE3 cells. After growth, 96 cfus were selected and dispersed in 200 μ L LB medium in microplates and grown at 37 °C with shaking (700 rpm) for 16 h, thus generating pre-cultures that were used to inoculate (30 μ L) 1500 μ L of fresh TB medium in deep wells of 96-microplates. Following growth for 2.5 h at 37 °C with shaking (700 rpm) protein production was induced by the addition of IPTG (to a final concentration of 0.6 mM) and incubation resumed for 16 h at 20 °C. At the end of this period, cells were recovered using centrifugation (2,250g, 30 min) and submitted to the combined lytic effects of lysozyme (0.5 g/L) and freeze/thaw (-80 °C and 37 °C respectively) treatments. After centrifugation, the supernatants were submitted to a dual enzyme assay, measuring activity on the substrate CNP- β -Celb (5 mM) in the absence and presence of acceptors **1** and **3** (at 25 mM). The ratios between the two activities were calculated to search potential activated clones in presence of acceptor.

Expression and Purification of EGC and Mutants. *E. coli* BL21 DE3 cells containing the plasmid pET28-EGC were grown to saturation in selective LB medium (containing kanamycin, 50 μ g.mL⁻¹) for 16 h at 37 °C with shaking (140 rpm), then 0.5 mL was used to inoculate a fresh aliquot (50 mL) of culture medium and growth was pursued. Induction of protein expression was achieved by addition of IPTG (1 mM final) during the exponential growth phase (OD₆₀₀ nm of 0.5-0.6), and was allowed to continue for 18 h at 20 °C and 140 rpm. Afterwards, the culture was subjected to centrifugation (5,000g, 15 min, 10 °C) to recover the cell pellet, which was suspended in 1 mL of TALON buffer (20 mM Tris-HCl, pH 8.0, with 300 mM NaCl). Sonication (Vibracell Bioblock Scientific 72434) of the cell suspension was performed on ice during 8 min (0.5 s 'on', 1.5 s 'off', with the probe operating at 30% of maximal power), allowing the recovery of a cell free extract after centrifugation (11,000g, 30 min, 4 °C) and filtration (0.22 μ m), which was then loaded onto a Cobalt resin (Talon® Metal Affinity Resin, Clontech Laboratories, Inc.) contained in a disposable column (Bio-Spin columns, Bio-Rad). Protein elution was achieved under gravity using TALON buffer containing increasing imidazole concentrations (one volume each of 5 and 60 mM imidazole, and then two volumes of 150 mM imidazole). Fractions were analyzed using SDS-PAGE (Bio-Rad) and those containing purified enzyme were pooled and dialyzed against sodium acetate buffer (25 mM, pH 5.0) overnight at 4 °C, before being filtered (0.22 μ m) and stored at 4 °C. Average enzymes yields per liter of culture were 80 (EGC), 60 (D311Y), and 70 mg (D341Y).

Evaluating Enzymatic Activities by Thin Layer Chromatography. Reactions (20 μ L) were performed in 50 mM sodium acetate buffer (pH 4.0), containing 1 mg.mL⁻¹ BSA, Triton X100 (0.4%, v/v), 2 mM CNP- β -Celb, an acceptor (**1-4** at concentration of 2.5 to 25 mM), and enzyme (wild-type, D311Y or D314Y mutants at 150 μ g/mL), for 1 h at 37°C in PCR microplates sealed with an

aluminum membrane. Heat inactivation of the enzyme (5 min at 95 °C) was used to stop reactions and then water was removed by evaporation. The solid residue was dissolved in methanol (10 µL) and the solution was spotted onto thin layer chromatography plates (TLC - silica gel 60 F254 precoated plates, Merck). Migration was achieved using an acetonitrile/water (8:2, v/v) mobile phase and molecular species were revealed by soaking in an orcinol solution (1 g in 1 L of sulphuric acid/ethanol/water solution, 3:72.5:22.5, v/v/v), under UV illumination (254 nm) and finally by charring.

Enzyme Kinetics. Kinetic analyses were prepared in microplates (96-well format, 250 µL-well polystyrene plates) using 200 µL of 50 mM sodium acetate buffer (pH 4.0) containing Triton X100 (0.4%, v/v), CNP-β-Celb (5 mM), BSA (1 mg.mL⁻¹) and enzyme (25 µg.mL⁻¹). For the assays, microplates were incubated in an Eon spectrophotometer (BioTek) for 5 min at 37 °C prior to enzyme addition, then reactions were conducted for 15 min, monitoring at 1-min intervals the release of 2-chloro-4-nitrophenol (CNP) at 405 nm. Enzyme activity was determined by comparing data to a standard curve prepared using CNP (in the range (0 to 1 mM). Negative controls containing all of the reactants except the enzyme were always included in order to correct for spontaneous hydrolysis of the substrate and all reactions were performed at least in triplicate. The kinetic parameters K_M , k_{cat} and the efficiency coefficient k_{cat}/K_M were determined by measuring enzyme activity at various concentrations of CNP-β-Celb (0.5 to 18 mM) and then by modeling data using the Michaelis-Menten equation and the global curve fitting module in SigmaPlot 11.0 software (Systat software Inc., Ritme, Paris, France). One unit of enzyme was designated the amount of enzyme required to release 1 µmol of CNP from CNP-β-Celb per minute.

To determine the optimum pH for activity the standard assay with 2 mM CNP-β-Celb was performed in buffers (50 mM sodium acetate, pH 3.5 to 6.0, and 50 mM sodium phosphate, pH 5.0 to 7.0) covering the pH range 3.5 to 7.0, ensuring that the volume of enzyme solution did not exceed 30 µL (final volume of 200 µL) to prevent a pH change. For each pH value, a standard CNP curve was prepared. Similarly, at the optimal pH for enzyme activity, reactions were performed at eight temperatures in the range 30 to 65 °C to determine optimal temperature. Finally, the thermostability of enzymes was evaluated by incubating enzymes at 37 or 50 °C, removing samples over a 7 h-period. Afterwards, residual activity was measured at 37 °C using the assay described above containing 2 mM CNP-β-Celb.

Transglycosylation Reactions. 20 mL of 50 mM sodium acetate buffer (pH 4.0), containing Triton X100 (0.4%, v/v), BSA (1 mg.mL⁻¹), CNP-β-Celb (8 mM), an acceptor molecule (100 mM of **1-4**) was equilibrated at 37 °C for 10 min before adding D311Y (160 µg.mL⁻¹). Afterwards, the reaction was incubated for 45 min at 37 °C, under stirring and then stopped by heat inactivation (10 min at 95 °C), freezing in liquid nitrogen and lyophilisation. The APGs **5-8** were purified after acetylation (in a 1:1 v/v mixture of acetic anhydride/pyridine and in the presence of a catalytic amount of 4-dimethylaminopyridine) using automated Reveleris® flash column chromatography and a 20 to 100%

(v/v) gradient of ethyl acetate in petroleum ether. Deacetylation, performed at 0 °C in a mixture of anhydrous methanol and dichloromethane (1:1, v/v) with sodium methoxide (1 M in methanol, 0.5 eq) and acidic workup with Amberlite IR-120 (H⁺) resin afforded the expected APGs **9-12**. The compounds were purified by automated flash column chromatography using a 0 to 60% gradient of methanol in dichloromethane. NMR spectra were recorded at 298 K on a Bruker Avance II 500 spectrometer. Coupling constants (*J*) are reported in Hz, chemical shifts (δ) are given in ppm. Multiplicities are reported as follows: s = singlet, d= doublet, t= triplet, m= multiplet, br = broad, dd= doublet of doublets. Roman numerals in ascending order are given to the D-Glcp residues from the reducing end. Analysis and assignments were made using 1D (¹H, ¹³C and Jmod) and 2D (COSY, HSQC and HMBC) spectra. High-resolution mass spectrometry experiment was performed at the CRMPO (Centre Régional de Mesures Physiques de l'Ouest, University of Rennes, France) on a Thermo Fisher Q-Exactive or a Waters Q-ToF 2 spectrophotometers, equipped with an electrospray ionization source operating in positive ion detection mode (ESI⁺).

3-Acetate-pent-1-yl and 1-ethyl-(3-acetate)-prop-1-yl 2,3,4,6-tetra-O-acetyl- β -D-glucopyranosyl-(1 \rightarrow 4)-3,4,6-tri-O-acetyl- β -D-glucopyranoside (5). Mixture of regioisomers. ¹H NMR (CDCl₃, 500 MHz, 298K): δ 5.19-5.12 (m, 2H, H-3 of Glc^{I and II}), 5.06 (t, *J*_{3,4} = *J*_{4,5} = 9.7 Hz, 1H, H-4 of Glc^{II}), 4.94-4.84 (m, 3H, H-2 of Glc^{II and I} and CHO), 4.52-4.50 (m, 2H, H-6a of Glc^I and H-1 of Glc^{II}), 4.44 (d, *J*_{1,2} = 7.9 Hz, 0.4H, H-1 of Glc^I), 4.42 (d, *J*_{1,2} = 8.0 Hz, 0.6H, H-1 of Glc^I), 4.36 (br dd, *J*_{5,6a} = 4.4 and *J*_{6a,6b} = 12.5 Hz, 1H, H-6a of Glc^{II}), 4.10-4.03 (m, 2H, H-6b of Glc^{I and II}), 3.89-3.79 (m, 1H, OCH₂), 3.76 (br t, *J*_{3,4} = *J*_{4,5} = 9.4 Hz, 1H, H-4 of Glc^I), 3.67-3.63 (m, 1H, H-5 of Glc^{II}), 3.59-3.56 (m, 1H, H-5 of Glc^I), 3.54-3.49 (m, 0.4H, OCH₂), 3.45-3.40 (m, 0.6H, OCH₂), 2.13-1.98 (8s, 24H, CH₃ of acetyl), 1.86-1.72 (m, 2H, OCH₂CH₂CH), 1.60-1.53 (m, 2H, CHCH₂CH₃), 0.88-0.85 (m, 3H, CH₃ of aglycone); ¹³C NMR (CDCl₃, 125 MHz, 298K): δ 170.6-169.0 (C(=O)CH₃), 100.8, 100.8, 100.7 and 100.3 (C-1 of Glc^{I and II}), 76.5 and 76.4 (C-4 of Glc^I), 72.9 (C-3 of Glc^{II}), 72.6, 72.6, 72.6, 72.5, 72.4 and 72.3 (C-3 and C-5 of Glc^I and CHO), 72.0 (C-5 of Glc^{II}), 71.6, 71.6, 71.5 and 71.4 (C-2 of Glc^{I and II}), 67.8 (C-4 of Glc^{II}), 66.4 and 66.2 (OCH₂), 61.8 (C-6 of Glc^I), 61.5 (C-6 of Glc^{II}), 33.6 and 33.3 (OCH₂CH₂CH), 27.2 and 27.0 (CHCH₂CH₃), 21.1-20.5 (CH₃ of acetyl), 9.4 and 9.4 (CH₃ of aglycone). HR-ESI-MS *m/z* 787.2631 (0 ppm, calcd for C₃₃H₄₈O₂₀Na [M+Na]⁺ 787.26311).

3-Hydroxy-pent-1-yl and 1-ethyl-(3-hydroxy)-prop-1-yl β -D-glucopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside (9). Mixture of regioisomers. ¹H NMR (CD₃OD, 500 MHz, 298K): δ 4.41 and 4.40 (2d, *J*_{1,2} = 7.9 Hz, 1H, H-1 of Glc^{II}), 4.30 (2d, *J*_{1,2} = 7.9 Hz, 1H, H-1 of Glc^I), 4.05-3.98 (m, 1H, OCH₂), 3.92-3.83 (m, 3H, H-6a of Glc^{I and II} and H-6b of Glc^{I or II}), 3.73-3.62 (m, 4H including 1H of OCH₂, H-6b of Glc^{I or II} and CHO), 3.58-3.49 (m, 2H including H-3 of Glc^{I or II}), 3.42-3.30 (m, 3H including H-3 of Glc^{I or II}), 3.28-3.20 (m, 2H, H-2 of Glc^{I and II}), 1.84-1.77 and 1.67-1.58 (2m, 1H each, OCH₂CH₂CH), 1.54-1.40 (m, 2H, CH₂CH₃), 0.95 (t, *J* = 7.4 Hz, 3H, CH₃); ¹³C NMR (CD₃OD, 125 MHz, 298K): δ 104.6 (C-1 of Glc^{II}), 104.3 and 104.1 (C-1 of Glc^I), 80.8, 80.7 and 78.1 (C_{sugar}), 77.9

(C-3 of Glc^I or ^{II}), 76.5 and 76.5 (C_{sugar} including C-3 of Glc^I or ^{II}), 74.9, 74.9 and 74.8 (C-2 of Glc^I and ^{II}), 71.4, 71.1 and 70.9 (C_{sugar} and CHO), 68.0 (OCH₂), 62.4 and 61.9 (C-6 of Glc^I and ^{II}), 37.8 and 37.7 (OCH₂CH₂CH), 33.3 and 31.3 (CH₂CH₃), 10.3 and 10.3 (CH₃). HR-ESI-MS *m/z* 451.1790 (0 ppm, calcd for C₁₇H₃₂O₁₂Na [M+Na]⁺ 451.17915).

3-Acetate-oct-1-yl and 1-(2-acetate)-ethyl-hex-1-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1→4)-3,4,6-tri-O-acetyl-β-D-glucopyranoside (6). Mixture of regioisomers. ¹H NMR (CDCl₃, 500 MHz, 298K): δ 5.18-5.11 (m, 2H, H-3 of Glc^I and ^{II}), 5.05 (t, *J*_{3,4} = *J*_{4,5} = 9.7 Hz, 1H, H-4 of Glc^{II}), 4.96-4.85 (m, 3H, H-2 of Glc^{II} and ^I and CHO), 4.51-4.48 (m, 2H, H-1 of Glc^{II} and H-6a of Glc^I), 4.43 (d, *J*_{1,2} = 7.9 Hz, 0.6H, H-1 of Glc^I), 4.41 (d, *J*_{1,2} = 8.0 Hz, 0.4H, H-1 of Glc^I), 4.36 and 4.35 (2dd, *J*_{5,6a} = 4.5 and 4.4 and *J*_{6a,6b} = 12.5 and 12.4 Hz respectively, 1H, H-6a of Glc^{II}), 4.10-4.02 (m, 2H, H-6b of Glc^I and ^{II}), 3.87-3.73 (m, 2H, 1H of OCH₂ and H-4 of Glc^I), 3.66-3.63 (m, 1H, H-5 of Glc^{II}), 3.58-3.55 (m, 1H, H-5 of Glc^I), 3.52-3.47 (m, 0.6H of OCH₂), 3.43-3.38 (m, 0.4H of OCH₂), 2.12-1.97 (8s, 24H, CH₃ of acetyl), 1.84-1.70 (m, 2H, OCH₂CH₂CH), 1.56-1.44 (m, 2H, CHCH₂(CH₂)₃CH₃), 1.32-1.19 (m, 6H, CHCH₂(CH₂)₃CH₃), 0.88-0.85 (m, 3H, CH₃ of aglycone); ¹³C NMR (CDCl₃, 125 MHz, 298K): δ 170.6-169.0 (C(=O)CH₃), 100.8, 100.7, 100.7 and 100.3 (C-1 of Glc^I and ^{II}), 76.5 and 76.4 (C-4 of Glc^I), 72.9, 72.6, 72.6, 72.6 and 72.4 (C-3 of Glc^I and ^{II} and C-5 of Glc^I), 71.9 (C-5 of Glc^{II}), 71.6, 71.6, 71.5, 71.4, 71.4 and 71.1 (C-2 of Glc^I and ^{II} and CHO), 67.8 (C-4 of Glc^{II}), 66.3 and 66.2 (OCH₂), 61.8 (C-6 of Glc^I), 61.5 (C-6 of Glc^{II}), 34.4, 34.2, 34.1 and 33.8 (OCH₂CH₂CH and CHCH₂(CH₂)₃CH₃), 31.6, 31.5, 24.8 and 22.5 (CHCH₂(CH₂)₃CH₃), 21.1-20.5 (CH₃ of acetyl), 13.9 (CH₃ of aglycone). HR-ESI-MS *m/z* 829.3098 (0 ppm, calcd for C₃₆H₅₄O₂₀Na [M+Na]⁺ 829.31006).

3-Hydroxy-oct-1-yl and 1-(2-hydroxy)-ethyl-hex-1-yl β-D-glucopyranosyl-(1→4)-β-D-glucopyranoside (10). Mixture of regioisomers. ¹H NMR (CD₃OD, 500 MHz, 298K): δ 4.41 and 4.40 (2d, *J*_{1,2} = 7.9 Hz, 1H, H-1 of Glc^{II}), 4.30 (2d, *J*_{1,2} = 7.9 Hz, 2H, H-1 of Glc^I), 4.05-3.97 (m, 1H of OCH₂), 3.92-3.83 (m, 3H, H-6a of Glc^I and ^{II} and H-6b of Glc^I or ^{II}), 3.77-3.63 (m, 4H including 1H of OCH₂, CHO and H-6b of Glc^I or ^{II}), 3.58-3.49 (m, 2H including H-3 of Glc^I or ^{II}), 3.42-3.33 (m, 3H including H-3 of Glc^I or ^{II}), 3.28-3.21 (m, 2H, H-2 of Glc^I and ^{II}), 1.83-1.77 and 1.66-1.59 (2m, 1H each, OCH₂CH₂CH), 1.49-1.42 (m, 3H, CHCH₂(CH₂)₃CH₃ and 1H of CHCH₂(CH₂)₃CH₃), 1.39-1.25 (m, 5H of CHCH₂(CH₂)₃CH₃), 0.91 (t, *J* = 7.0 Hz, 3H, CH₃ of aglycone); ¹³C NMR (CD₃OD, 125 MHz, 298K): δ 104.6 (C-1 of Glc^{II}), 104.3 and 104.1 (C-1 of Glc^I), 80.8, 80.7, 78.1, 77.9 and 76.5 (C_{sugar}), 74.9, 74.9 and 74.8 (C-2 of Glc^I and ^{II}), 71.4 (C_{sugar}), 69.7 and 69.5 (CHO), 68.0 (OCH₂), 62.4 and 61.9 (C-6 of Glc^I and ^{II}), 38.7, 38.6, 38.3 and 38.2 (OCH₂CH₂CH and CHCH₂CH₃), 33.1, 33.1, 26.5 and 23.7 (CHCH₂(CH₂)₃CH₃), 14.4 (CH₃ of aglycone). HR-ESI-MS *m/z* 493.2263 (0 ppm, calcd for C₂₀H₃₈O₁₂Na [M+Na]⁺ 493.2261).

2-Oxo-pent-5-yl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl-(1→4)-3,4,6-tri-O-acetyl-β-D-glucopyranoside (7). ¹H NMR (CDCl₃, 500 MHz, 298K): δ 5.16 (dd, *J*_{3,4} = 9.3 Hz, 1H, H-3 of Glc^I),

5.14 (t, $J_{3,4} = 9.4$ Hz, 1H, H-3 of Glc^{II}), 5.06 (t, $J_{4,5} = 9.8$ Hz, 1H, H-4 of Glc^{II}), 4.92 (dd, $J_{2,3} = 9.4$ Hz, 1H, H-2 of Glc^{II}), 4.88 (dd, $J_{2,3} = 9.7$ Hz, 1H, H-2 of Glc^I), 4.50 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1 of Glc^{II}), 4.50 (dd, $J_{5,6a} = 2.1$ Hz, 1H, H-6a of Glc^I), 4.42 (d, $J_{1,2} = 8.0$ Hz, 1H, H-1 of Glc^I), 4.36 (dd, $J_{5,6a} = 4.4$ et $J_{6a,6b} = 12.5$ Hz, 1H, H-6a of Glc^{II}), 4.09 (dd, $J_{5,6b} = 5.1$ et $J_{6a,6b} = 12.0$ Hz, 1H, H-6b of Glc^I), 4.04 (dd, $J_{5,6b} = 2.3$ Hz, 1H, H-6b of Glc^{II}), 3.82-3.78 (m, 1H of OCH₂), 3.75 (br t, $J_{4,5} = 9.8$ Hz, 1H, H-4 of Glc^I), 3.67-3.63 (m, 1H, H-5 of Glc^{II}), 3.59-3.55 (m, 1H, H-5 of Glc^I), 3.52-3.48 (m, 1H of OCH₂), 2.48 (t, $J_{3',4'} = 7.1$ Hz, 2H, CH₂C(O)), 2.12 (s, 3H, CH₃ of aglycone), 2.12-1.97 (7s, 21H, CH₃ of acetyl), 1.87-1.78 (m, 2H, CH₂CH₂CH₂); ¹³C NMR (CDCl₃, 125 MHz, 298K): δ 208.1 (C(=O)CH₃ of aglycone), 170.5-169.0 (C(=O)CH₃ of acetyl), 100.8 (C-1 of Glc^{II}), 100.6 (C-1 of Glc^I), 76.5 (C-4 of Glc^I), 72.9 (C-3 of Glc^{II}), 72.7 (C-5 of Glc^I), 72.5 (C-3 of Glc^I), 72.0 (C-5 of Glc^{II}), 71.6 (C-2 of Glc^{II}), 71.6 (C-2 of Glc^I), 69.0 (OCH₂), 67.8 (C-4 of Glc^{II}), 61.8 (C-6 of Glc^I), 61.6 (C-6 of Glc^{II}), 39.7 (CH₂C(=O)), 29.9 (CH₃ of aglycone), 23.6 (CH₂CH₂CH₂), 20.8-20.5 (CH₃ of acetyl). HR-ESI-MS m/z 743.2368 (0 ppm, calcd for C₃₁H₄₄O₁₉Na [M+Na]⁺ 743.2369).

2-Oxo-pent-5-yl β -D-glucopyranosyl-(1 \rightarrow 4)- β -D-glucopyranoside (11). ¹H NMR (CD₃OD, 500 MHz, 298K): δ 4.41 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1 of Glc^{II}), 4.26 (d, $J_{1,2} = 7.9$ Hz, 1H, H-1 of Glc^I), 3.92-3.83 (m, 4H, 1H of OCH₂, H-6a of Glc^{I and II} and H-6b of Glc^{I or II}), 3.66 (dd, $J_{5,6a} = 5.6$ and $J_{6a,6b} = 11.9$ Hz, 1H, H-6b of Glc^{I or II}), 3.58-3.54 (m, 2H including 1H of OCH₂), 3.50 (t, $J_{2,3} = J_{3,4} = 8.9$ Hz, 1H, H-3 of Glc^{I or II}), 3.40-3.28 (m, 4H including H-3 of Glc^{I or II}), 3.24-3.20 (m, 2H, H-2 of Glc^{I and II}), 2.63 (br t, $J_{3',4'} = 7.2$ Hz, 2H, CH₂C(=O)), 2.15 (s, 3H, CH₃ of aglycone), 1.88-1.82 (m, 2H, CH₂CH₂CH₂); ¹³C NMR (CD₃OD, 125 MHz, 298K): δ 211.8 (C(=O)CH₃ of aglycone), 104.6 (C-1 of Glc^{II}), 104.2 (C-1 of Glc^I), 80.7, 78.1 and 77.9 (C_{sugar}), 76.4 and 76.4 (C-3 of Glc^{I and II}), 74.9 and 74.8 (C-2 of Glc^{I and II}), 71.4 (C_{sugar}), 69.9 (OCH₂), 62.4 and 61.9 (C-6 of Glc^{I and II}), 40.7 (CH₂C(=O)), 29.9 (CH₃ of aglycone), 24.9 (CH₂CH₂CH₂). HR-ESI-MS m/z 449.1635 (0 ppm, calcd for C₁₇H₃₀O₁₂Na [M+Na]⁺ 449.1635).

4-Oxo-oct-1-yl 2,3,4,6-tetra-O-acetyl- β -D-glucopyranosyl-(1 \rightarrow 4)-3,4,6-tri-O-acetyl- β -D-glucopyranoside (8). ¹H NMR (CDCl₃, 500 MHz, 298K): δ 5.17-5.11 (m, 2H, H-3 of Glc^{I and II}), 5.05 (br t, $J_{3,4} = J_{4,5} = 9.7$ Hz, 1H, H-4 of Glc^{II}), 4.91 (dd, $J_{2,3} = 9.3$ Hz, 1H, H-2 of Glc^{II}), 4.87 (dd, $J_{2,3} = 9.7$ Hz, 1H, H-2 of Glc^I), 4.50 (d, $J_{1,2} = 8.0$ Hz, 1H, H-1 of Glc^{II}), 4.49 (dd, $J_{5,6a} = 2.1$ and $J_{6a,6b} = 12.0$ Hz, 1H, H-6a of Glc^I), 4.42 (d, $J_{1,2} = 8.0$ Hz, 1H, H-1 of Glc^I), 4.36 (dd, $J_{5,6a} = 4.5$ et $J_{6a,6b} = 12.5$ Hz, 1H, H-6a of Glc^{II}), 4.08 (dd, $J_{5,6b} = 5.2$ et $J_{6a,6b} = 12.0$ Hz, 1H, H-6b of Glc^I), 4.04 (dd, $J_{5,6b} = 2.3$ and $J_{6a,6b} = 12.4$ Hz, 1H, H-6b of Glc^{II}), 3.82-3.77 (m, 1H of OCH₂), 3.75 (t, $J_{3,4} = J_{4,5} = 9.4$ Hz, 1H, H-4 of Glc^I), 3.66-3.63 (m, 1H, H-5 of Glc^{II}), 3.58-3.55 (m, 1H, H-5 of Glc^I), 3.51-3.47 (m, 1H of OCH₂), 2.44 (t, $J_{2',3'} = 7.2$ Hz, 2H, OCH₂CH₂CH₂C(O)), 2.37 (t, $J_{5',6'} = 7.5$ Hz, 2H, C(O)CH₂CH₂CH₂CH₃), 2.12-1.97 (7s, 21H, CH₃ of acetyl), 1.85-1.78 (m, 2H, OCH₂CH₂CH₂C(O)), 1.56-1.50 (m, 2H, C(O)CH₂CH₂CH₂CH₃), 1.32-1.27 (m, 2H, C(O)CH₂CH₂CH₂CH₃), 0.89 (t, $J_{7',8'} = 7.4$ Hz, 3H, CH₃ of aglycone); ¹³C NMR (CDCl₃, 125 MHz, 298K): δ 210.6 (C(=O) of aglycone), 170.5-169.0 (C(=O)CH₃ of acetyl), 100.8 (C-1 of Glc^{II}), 100.6 (C-1 of Glc^I), 76.5 (C-4 of Glc^I), 72.9 (C-3 of Glc^{II}), 72.7 (C-5

of Glc^I), 72.5 (C-3 of Glc^I), 72.0 (C-5 of Glc^{II}), 71.6 and 71.6 (C-2 of Glc^{I and II}), 69.1 (OCH₂), 67.8 (C-4 of Glc^{II}), 61.8 (C-6 of Glc^I), 61.6 (C-6 of Glc^{II}), 42.6 (C(O)CH₂CH₂CH₂CH₃), 38.6 (OCH₂CH₂CH₂C(O)), 25.9 (C(O)CH₂CH₂CH₂CH₃), 23.6 (OCH₂CH₂CH₂C(O)), 22.3 (C(O)CH₂CH₂CH₂CH₃), 20.8-20.5 (CH₃ of acetyl), 13.8 (CH₃ of aglycone). HR-ESI-MS *m/z* 785.2836 (0 ppm, calcd for C₃₄H₅₀O₁₉Na [M+Na]⁺ 785.28385).

4-Oxo-oct-1-yl β-D-glucopyranosyl-(1→4)-β-D-glucopyranoside (12). ¹H NMR (CD₃OD, 500 MHz, 298K): δ 4.41 (d, *J*_{1,2} = 7.9 Hz, 1H, H-1 of Glc^{II}), 4.29 (d, *J*_{1,2} = 7.9 Hz, 1H, H-1 of Glc^I), 3.90-3.83 (m, 4H, 1H of OCH₂, H-6a of Glc^{I and II} and H-6b of Glc^{I or II}), 3.68-3.62 (m, 2H including H-6b of Glc^{I or II}), 3.58-3.53 (m, 2H including 1H of OCH₂), 3.50 (t, *J*_{2,3} = *J*_{3,4} = 8.9 Hz, 1H, H-3 of Glc^{I or II}), 3.40-3.33 (m, 3H including H-3 of Glc^{I or II}), 3.24-3.20 (m, 2H, H-2 of Glc^{I and II}), 2.60 (t, *J*_{2,3} = 7.2 Hz, 2H, OCH₂CH₂CH₂C(=O)), 2.47 (t, *J*_{5,6} = 7.4 Hz, 2H, C(=O)CH₂CH₂CH₂CH₃), 1.88-1.82 (m, 2H, OCH₂CH₂CH₂C(=O)), 1.57-1.51 (m, 2H, C(=O)CH₂CH₂CH₂CH₃), 1.36-1.28 (m, 2H, C(=O)CH₂CH₂CH₂CH₃), 0.92 (s, 3H, CH₃); ¹³C NMR (CD₃OD, 125 MHz, 298K): δ 214.0 (C(O)CH₃ of aglycone), 104.6 (C-1 of Glc^{II}), 104.2 (C-1 of Glc^I), 80.7, 78.1 and 77.9 (C_{sugar}), 76.4 and 76.4 (C-3 of Glc^{I and II}), 74.9 and 74.8 (C-2 of Glc^{I and II}), 71.4 (C_{sugar}), 69.9 (OCH₂), 62.4 and 61.9 (C-6 of Glc^{I and II}), 43.3 (C(=O)CH₂CH₂CH₂CH₃), 39.9 (OCH₂CH₂CH₂C(=O)), 27.0 (C(=O)CH₂CH₂CH₂CH₃), 25.0 (OCH₂CH₂CH₂C(=O)), 23.4 (C(=O)CH₂CH₂CH₂CH₃), 14.2 (CH₃). HR-ESI-MS *m/z* 491.2100 (1 ppm, calcd for C₂₀H₃₈O₁₂Na [M+Na]⁺ 491.21045).

¹H NMR Monitoring of Enzyme Reactions and Selection of Significant Protons. Reactions (5 mL) in 20 mM sodium acetate buffer (pH 4.0), containing Triton X-100 (0.2%, v/v) BSA (0.5 mg.mL⁻¹), CNP-β-Celb or F-β-Celb as donor (5 mM) and an acceptor (62.5 mM of **1**, **2**, **3** or **4**), were incubated at 37 °C for 5 min then 200 μL aliquot was removed and designated t = 0. Afterwards, enzyme was added to start the reaction (wild-type enzyme at a final concentration of 150 μg.mL⁻¹ when using **1**, **2**, and **4**, and 120 μg.mL⁻¹ for the reaction containing **3**, or D311Y at a final concentration of 100 μg.mL⁻¹ irrespective of the acceptor) and 200 μL samples were removed at intervals over 180 min, inactivating the enzyme using a heat (95 °C for 5 min), liquid nitrogen freezing and lyophilization routine. At 120 min, more enzyme was added reaching 2x the original concentration. In the case of D311Y, all reactions were performed in triplicate, but those involving EGC were performed once only. The lyophilized samples were dissolved in 250 μL of CD₃OD and ¹H NMR spectra were recorded at 298 K on a Bruker Ascend Avance III 800 spectrometer equipped with a samplejet autosampler. Donor conversion, hydrolysis and transglycosylation activities were quantified by integrating the relevant anomeric and aromatic proton signals that are reported in Table 2.

Table 2. List of proton signals used for the tracking of the different species involved during enzymatic reactions catalyzed by EGC and mutant thereof

Compounds	Integrated proton	Signal multiplicity, δ (ppm); J (Hz)	Enzymatic reaction monitoring
CNP- β -Celb	H-6' of CNP	d, 7.43 (9.2)	Donor consumption
	H-5' of CNP	dd, 8.19 (2.8 and 9.2)	
	H-3' of CNP	d, 8.32 (2.8)	
CNP	H-6' of CNP	d, 6.68 (9.2)	CNP release (donor consumption)
	H-5' of CNP	dd, 7.94 (2.9 and 9.2)	
	H-3' of CNP	d, 8.15 (2.9)	
CNP- β -Celb	H-1 of Glc ^I	d, 5.23 (7.4)	Donor consumption
	H-1 of Glc ^{II}	d, 4.46 (7.9)	
Cellobiose	H-1 of Glc ^{Ia}	d, 5.10 (3.8)	Hydrolysis
	H-1 of Glc ^{Ib}	d, 4.50 (7.8)	
APG 9	H-1 of Glc ^I	d, 4.30 (7.9)	Transglycosylation
APG 10	H-1 of Glc ^I	d, 4.30 (7.9)	Transglycosylation
APG 11	H-1 of Glc ^I	d, 4.26 (7.9)	Transglycosylation
APG 12	H-1 of Glc ^I	d, 4.29 (7.9)	Transglycosylation

Evaluation of the Modulation from Hydrolase to Transglycosidase. The transfer rate R_T (%) of the donor substrate to the acceptor, expressed in terms of production (μmol of product/ $100 \mu\text{mol}$ of consumed substrate) was derived from the plot of product yield as a function of donor substrate conversion. As a rate, R_T is independent of the duration of the reaction and thus provides a convenient way to compare different enzymes (Bissaro et al., 2015a). Initial rates were derived from the linear regions corresponding to maximum donor consumption, which was between 16 and nearly 100% depending on the reaction, with root-mean-square deviations above 0.98.

Modeling D311Y complexed with substrates. The 3D structure of EGC (GenBank accession code: AAB67050.1, amino acids 43 to 490) bearing the mutation D311Y was modelled using MODELLER 9v8 (Sali and Blundell, 1993). Available EGC structures in the Protein Data Bank were selected for

threading with multiple templates: 2OSX (EGC II E351S from *Rhodococcus* sp. in complex with ganglioside GM3), 2OYM (EGC II from *Rhodococcus* sp. in complex with iminocyclitol) and 2OSY (EGC II E233A from *Rhodococcus* sp. lactosyl-enzyme covalent intermediate). Both the *apo*-enzyme and lactosyl-enzyme were modelled. A slow refinement protocol based on simulated annealing molecular dynamics was applied to generate an ensemble of 10 different structures for each state. CNP- β -Celb and acceptors **2** and **4** were docked separately to each modelled D311Y structure in the ensemble with AUTODOCK VINA (Trott and Olson, 2009). The docking search space was confined in a 27.8 x 31.5 x 31.5 Å³ box centered in the active site. The exhaustiveness level was set to 24 and for each docking calculation, 20 binding modes of the ligands were generated. 200 structures of D311Y in complex with each ligand were thus obtained. Only low energy binding poses were considered for analysis. Distances and surfaces were measured, and pictures of the complexes generated with VMD (Humphrey et al., 1996).

RESULTS

Acceptor Preference for Transglycosylation with the Wild-Type Enzyme. The inherent ability of EGC to transfer cellobiosyl units onto target acceptors was evaluated. Transglycosylation activity was first tested with pentan-1-ol and octan-1-ol, in the presence of 2.5% (v/v) acetonitrile, but both tests were unsuccessful. Similar tests were performed with pentan-1,3-diol (**1**) and 5-hydroxypentan-2-one (**3**), two hydroxylated molecules displaying either hydroxyl or ketone groups that can be considered to some extent mimic parts of the natural ceramide. These tests revealed that although the main reaction was hydrolysis, small amounts of transglycosylation products were obtained (Figure S2). Based on this observation, an enzyme engineering strategy was then deployed with the aim to increase the transglycosylation/hydrolysis (T/H) ratio.

Site-Saturation Mutagenesis and Detection of Transglycosylating Mutant. Resolution of the 3D structure of the EGC complexed with ganglioside revealed four amino acids that are located in the aglycone subsite of EGC, where they are thought to interact either directly with the carbonyl group of the ceramide part of glycosphingolipids (N279 and Q280), or through a water molecule (E179 and D311) with the secondary hydroxyl function of the diol from the sphingosine moiety (Figure S3) (M. E. C. Caines et al., 2007). Therefore these four residues were targeted for site-saturation mutagenesis, creating four libraries that were screened using four different potential acceptor substrates **1-4** (Figure 2) that display different carbon chain lengths and 1,3-diol and δ -hydroxyketone functions. Primary screening in 96-well microtiter plates using CNP- β -Celb as the donor substrate revealed one isolate (D311Y) that showed 2-fold higher CNP-releasing activity (i.e., a hydrolytic specific activity of 2.07 ± 0.08 IU/mg) when compared to EGC, and increased CNP-releasing activity in the presence of acceptors **1-4** (i.e., conditions favoring transglycosylation) relative to the activity in the absence of

acceptors. This result is similar to that obtained with the wild-type enzyme in the presence of 5-carbon chain acceptors **1** and **3**, but slightly more than EGC in the presence of 8-carbon acceptors **2** and **4** (Figure 2).

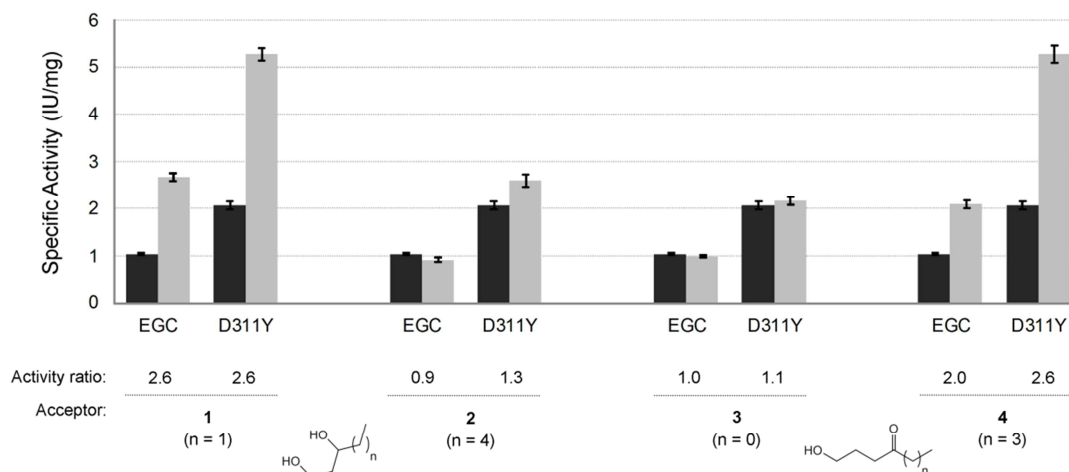


Figure 2. Measurement of the specific activities (i.e., CNP release) of EGC and D311Y on CNP- β -Celb (5 mM) alone (dark bars), or in the presence of 25 mM acceptor **1-4** (grey bars). The activity ratio is the comparison of enzyme activity in reactions performed in the presence or absence of a given acceptor.

Kinetic Analysis of the Hydrolytic Activities of EGC, D311Y and D314Y. Determination of the pH and temperature optima and thermal stability of EGC, D311Y and D314Y confirmed that compared to EGC these parameters were mostly unaltered (pH_{opt} of 4, T_{opt} of approximately 60 °C) in the mutant enzymes (Figure S4). Nevertheless, the half-life of D311Y was more than 2 h at 37 °C, compared to 45 min for EGC and 3 h for D314Y, but at 50 °C it only retained approximately 5% of its activity after 30 min (Figure S4C and Table S1), observations that are consistent with structural data suggesting that mutations at these flexible positions might reduce the B-factor (Zhang et al., 2016).

The substitution of D314 by tyrosine (i.e., D314Y) has been shown to increase the catalytic activity of the EGC glycosynthase, E351S, when this crippled mutant is used with G_{M1} -oligosaccharyl α -fluoride donor and sphingosine acceptor (Hancock et al., 2009). Therefore, D314Y was included in our study to compare with EGC and D311Y (Table 3). Apart from a slight increase of k_{cat} (1.5-fold), the catalytic parameters describing D314Y-mediated hydrolysis of CNP- β -Celb were similar to that of EGC. However, in comparison the K_{M} value measured when using D311Y (7.0 mM) was 3.2-fold lower than that observed for EGC (22.4 mM), and the performance constant $k_{\text{cat}}/K_{\text{M}}$ value was higher, even if this increase was attenuated by the slight decrease (0.8-fold) of k_{cat} .

Table 3. Steady-state kinetic parameters for the hydrolysis of CNP- β -Celb by EGC and mutants D311Y and D314Y^a

Enzyme	K_M (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (s ⁻¹ .mM ⁻¹)
EGC	22.4 \pm 3.8	3.7 \pm 0.6	0.17 \pm 0.05
D314Y	22.7 \pm 3.1	5.4 \pm 0.7	0.24 \pm 0.06
D311Y	7.0 \pm 0.8	2.9 \pm 0.3	0.42 \pm 0.08

^a All assays were carried out in triplicate at 37 °C in 50 mM sodium acetate buffer (pH 4).

Transglycosylation Activity of the Mutant D311Y. The aptitude of EGC, D311Y and D314Y to perform transglycosylation was first qualitatively assessed using various acceptors and TLC analysis (Figure S2). Although in the case of alkan-1-ol acceptors it was not possible to detect synthesis products when using either D311Y or EGC, in the case of acceptors **1-3** the transglycosylation ability of D311Y was improved compared to that of EGC. It is noteworthy that mutation D314Y, which is thought to modify the position of residue D311 (Hancock et al., 2009), did not reveal any improvement of the transglycosylation ability. Having ascertained this, the donor/acceptor ratio was optimized in favor of transglycosylation (2:25, mM/mM) (Figure S5) and large-scale transglycosylation reactions were performed, procuring APGs **9-12** at yields of 50-81%. Analysis of products **9-12** by NMR spectroscopy and mass spectrometry confirmed that these were all cellobiosides (correlation from HMBC spectra between H-1 of Glc^I and OCH_n of the aglycone and/or C-1 of Glc^I and OCH_n, with n = 1 or 2) and revealed that APGs **9** and **10** were in each case an equimolar mixture of two regioisomers, meaning that in the presence of diol acceptors D311Y was able to transfer the cellobiosyl moiety onto both the primary and secondary hydroxyl groups. Logically, in the case of APGs **11** and **12** only one product was present. Moreover, the β -stereochemistry of the newly formed *O*-glycosidic linkages in APGs **9-12** was confirmed through the determination of the experimental coupling constants ($J_{1,2} = 7.9$ Hz) and the observation of chemical shifts of H-1 of Glc^I ($\delta = 4.26-4.30$ ppm).

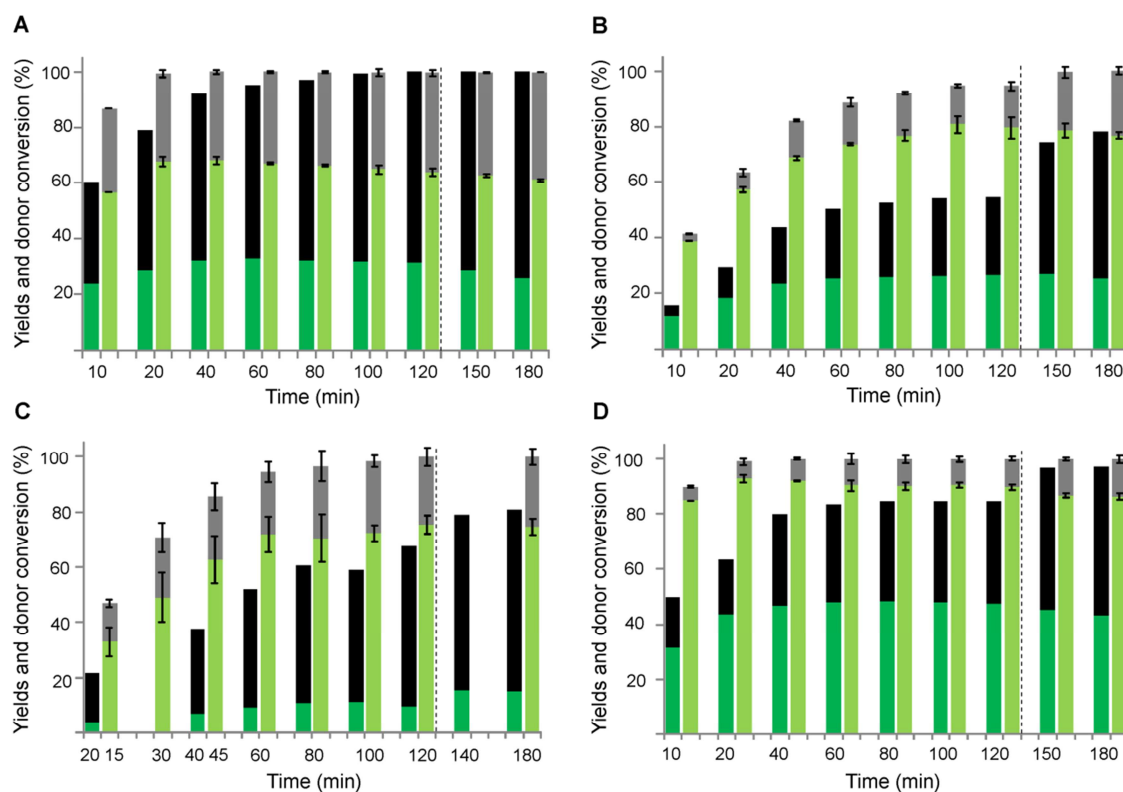


Figure 3. ^1H NMR monitoring of time-dependent CNP- β -Celb donor consumption and T/H modulation in reactions catalyzed by EGC and D311Y with acceptor (A) **1**, (B) **2**, (C) **3** and (D) **4**. Blacks and grey bars represent hydrolysis yields (% cellobiose produced) and dark and light green bars represent transglycosylation yields (% synthesis products **9-12**) for EGC or D311Y respectively. The sum (black + dark green grey + light green), corresponds to the overall progress of the reaction (i.e., donor conversion in %). Enzyme concentration is doubled after sampling at 120 min (black dotted lines), enabling to appreciate secondary hydrolysis. Reactions with D311Y and EGC were performed in triplicate and monoplicate, respectively.

^1H NMR monitoring of the time-dependent progression of both D311Y-catalyzed transglycosylation and hydrolysis revealed a marked increase of the T/H ratio when compared to EGC (Figures 3 and S6). It is noteworthy that whereas EGC fails to fully consume the donor (even when a fresh enzyme aliquot is added) D311Y quantitatively consumed all the donor in the presence of almost all the acceptors. Moreover, although D311Y-catalyzed transglycosylation was generally more efficient (i.e., faster product accumulation and higher yields) than that of EGC, irrespective of the acceptor, this was particularly evident when the shorter δ -hydroxy ketone **3** was used as the acceptor, since this reaction led to 6.8-fold increase in the transglycosylation yield (75%) compared to the equivalent EGC-catalyzed reaction (Table 4). It is noteworthy that high final yields of APGs **9-12** in reactions catalyzed by D311Y were also the result of the enzyme's inherent low secondary hydrolysis ability (< 6% of their maximal yield), this being underlined by the fact that the addition of fresh enzyme at the end of

the reaction did not affect final yields. Likewise, D311Y catalysis was characterized by increased transfer rates (R_T), in the range 1.1 to 1.8-fold, for acceptors **1**, **2** and **4**, and up to 4.2-fold for acceptor **3** (Table 4). Concerning the linearity of transglycosylation, in EGC-catalyzed reactions this was conserved up to 80% donor consumption, whereas in D311Y-catalyzed reactions this linear phase was maintained up to 95% donor consumption (Figure 4 and Table S2), underlining the fact that the replacement of D311 by a tyrosine residue confers strong transglycosylase character to EGC.

Table 4. Comparison of maximum transglycosylation yields and cellobiosyl transfer rates in EGC and D311Y catalyzed reactions containing acceptors **1-4**

Acceptor	Enzyme	Maximal yield (%) ^a	R_T ^b
Pentan-1,3-diol (1)	EGC	33	38
	D311Y	68	67
Octan-1,3-diol (2)	EGC	26	75
	D311Y	81	85
5-Hydroxypentan-2-one (3)	EGC	11	18
	D311Y	75	74
1-Hydroxyoctan-4-one (4)	EGC	48	63
	D311Y	93	93

^a The maximum transglycosylation yield was not reached at the same time for all products. ^b The transfer rate (R_T , in %) represents the number of μ moles of cellobiosyl transferred onto an acceptor for 100 μ moles of *p*NP- β -Celb consumed.

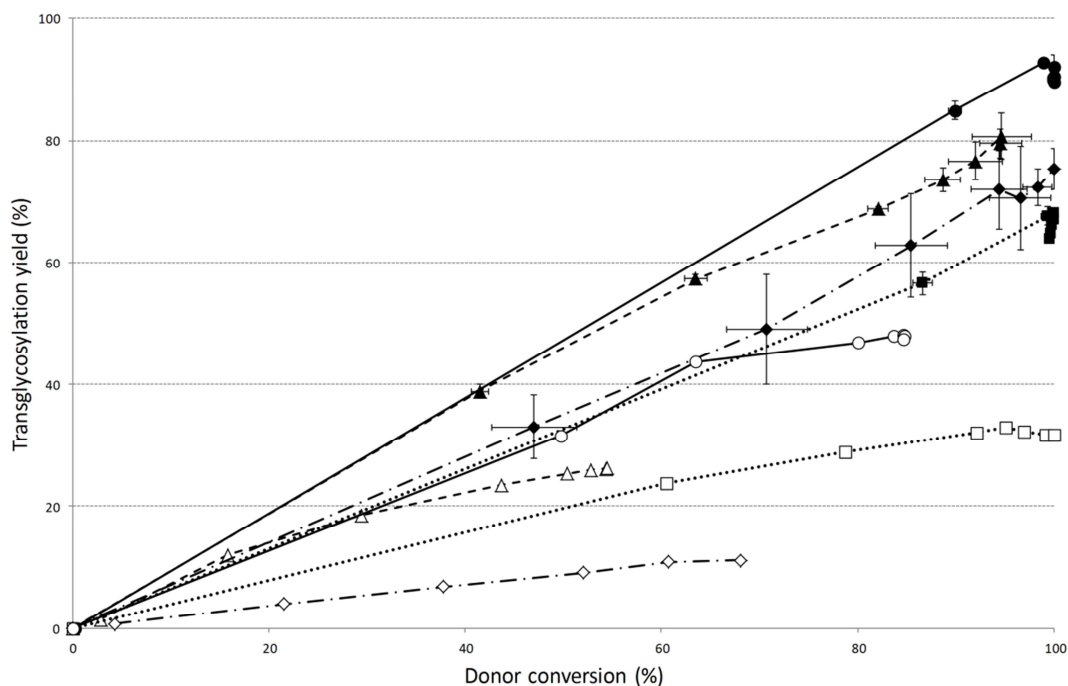


Figure 4. Transglycosylation yield evolution as a function of donor conversion with CNP- β -Celb (5 mM) as donor substrate and acceptor (62.5 mM): **1** (square), **2** (triangle), **3** (diamond) or **4** (circle), for EGC (open symbols) and D311Y (closed symbols). Transfer rates were calculated up to 120 min of reaction. Reactions with D311Y and EGC were performed in triplicate and monoplicate, respectively.

Transglycosylation using β -Cellobiosyl Fluoride as Activated Donor. To compare the impact of the leaving group on transglycosylation, β -cellobiosyl fluoride (F- β -Celb) (Konstantinidis et al., 1993) was used as the donor in reactions catalyzed by EGC and D311Y respectively and containing acceptor **3**. The transglycosylation product **11** resulting from these reactions was barely detectable (less than 2% yield) in the EGC-catalyzed reaction, whereas its yield reached 25% when using D311Y (Figure 5). Moreover, this yield could be further improved after 120 min simply by adding a fresh aliquot of D311Y. After 180 min, TLC analysis confirmed that the D311Y-catalyzed reaction was complete, whereas at the same time point in the reaction containing EGC almost all the F- β -Celb donor was either unconsumed or hydrolyzed (data not shown).

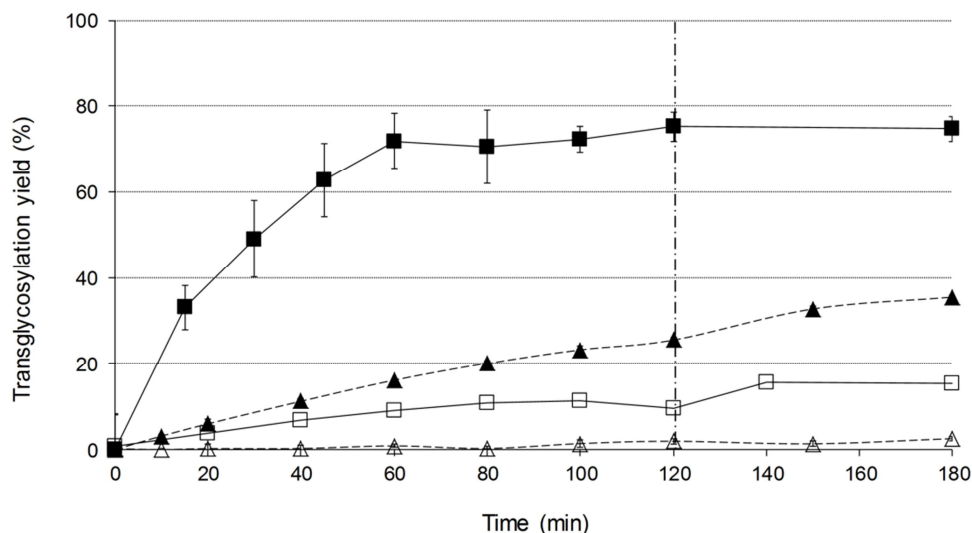


Figure 5. Transglycosylation yield evolution as a function of time with 5-hydroxypentan-2-one (**3**, at 62.5 mM) as acceptor and F-β-Celb (triangle) or CNP-β-Celb (square) as donor substrate (5 mM) for EGC (open symbols) and D311Y (closed form). Enzyme concentration is doubled after sampling at 120 min (dash-dotted lines).

Modeling D311Y in complex with donor substrate. For computational docking, the 3D structure of the D311Y:CNP-β-Celb complex was modeled using available crystallographic information (PDB ID: 2OSX), procuring a “Michaelis” complex geometry competent for catalysis (Figure 6). Analysis of this complex revealed that conformational distortion (1S_3) of the D-glucopyranosyl unit occupying subsite -1 places the anomeric carbon of the scissile bond at a distance of 3.9 Å from the nucleophile (E351) and the glycosidic oxygen at 4.4 Å from the general acid/base (E233). The leaving group (CNP) is thus located in axial orientation at subsite +1 stabilising stacking interactions with the enzyme (Y306 and the introduced Y311).

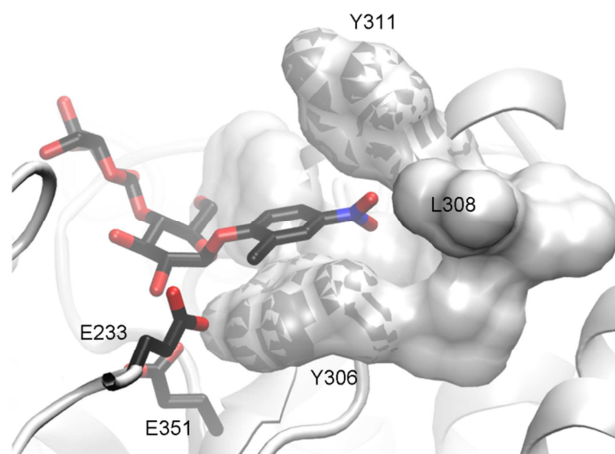


Figure 6. Model of the Michaelis complex D311Y:CNP- β -Celb. Substrate and catalytic residues are drawn using thick lines. The hydrophobic pincer formed by residues Y306, L308 and Y311 is shown in grey surface mode.

Modeling the covalent intermediate lactosyl-D311Y bound to acceptor molecules. To rationalize the outcomes of the transglycosylation reactions, the binding modes of acceptors **2** and **4** to the covalent lactosyl-D311Y intermediate were investigated. This enzyme state was chosen because it immediately precedes the transglycosylation reaction. Six different catalytically-competent geometries were obtained (Figure 7), all of which displayed correct orientations for plausible glycosyl transfer (i.e., short distances to the anomeric carbon and general acid/base) onto the alcohol moieties of the two acceptors. Regarding both acceptor **2** and **4**, in all scenarios the single aliphatic chains can adopt two different binding configurations that correspond to the two binding channels that are revealed in the crystallographic structure of EGC bound to the GM3 ganglioside (Figure S3A). Channel I is defined by hydrophobic sidechains of Y182, F235 and I276, whereas channel II is defined by the more hydrophilic sidechains R177, E179, L308 and D311. Logically, in the mutant D311Y the hydrophobicity of channel II is increased compared to that of the wild-type enzyme. This leads to the formation of a clear hydrophobic interaction between the tyrosine ring and the aliphatic chain of both acceptors **2** and **4** when these occupy channel II (right-hand side of Figure 7). When the aliphatic chain occupies channel I (left-hand side of Figure 7), hydrophobic interactions are established with Y182, F235 and I276, but not with Y311. In all scenarios, binding of the acceptor molecule is accompanied by H-bonding to either N279 or the *HO-6* of the D-glucopyranosyl unit at subsite -1, mimicking the interactions already described between EGC and the ceramide part of GM3 ganglioside.

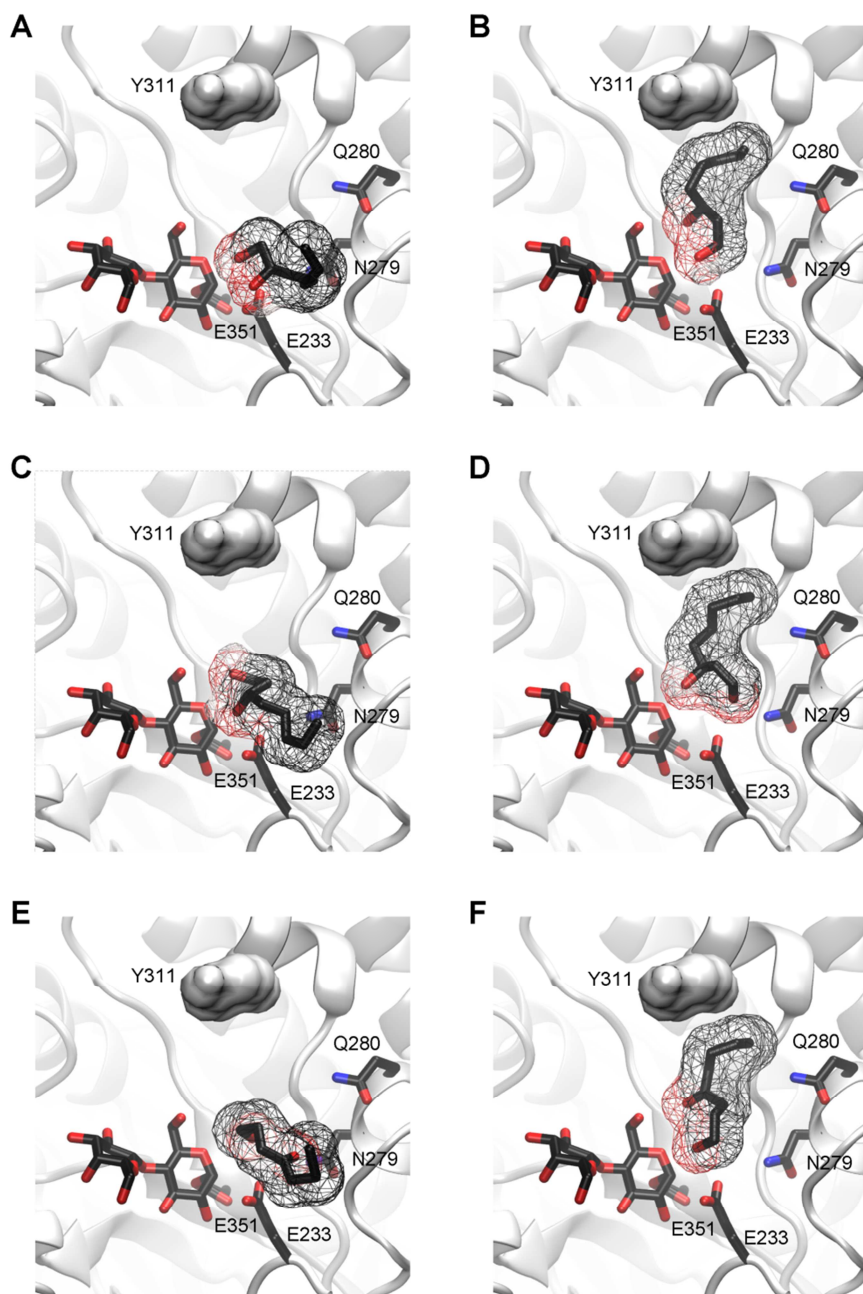


Figure 7. Model of the covalent glycosyl-D311Y intermediate:acceptors **2** and **4** complexes. Substrates and catalytic residues are drawn using thick lines. The introduced D311Y mutation is shown in a grey surface mode. Binding modes of acceptor **2** for glycosyl transfer at primary alcohol in channel I (A) and II (B). Binding modes of acceptor **2** for glycosyl transfer at secondary alcohol in channel I (C) and II (D). Binding modes of acceptor **4** in channel I (E) and II (F).

DISCUSSION

Endo-glycosceramidase II from *Rhodococcus* sp. strain M-777 is a retaining GH that hydrolyzes

complex gangliosides, which are composed of ceramide chains linked to oligosaccharides. Accordingly, this rGH has already been exploited for chemistry purposes, converting it into a glycosynthase for the synthesis of glycosphingolipids (Vaughan et al., 2006b; Hancock et al., 2009; Rich et al., 2011b; Rich and Withers, 2012; Yang et al., 2015). In the present study, the inherent properties of the active site of EGC that allow it to accommodate both sphingolipids and sugar moieties led us to consider EGC as a promising prototype for the design of enzymes for APG synthesis.

Based on previous studies aimed at improving transglycosylation in rGHs (Feng, 2005a; Osanjo et al., 2007; Kone et al., 2009; Teze et al., 2014, 2015, Arab-Jaziri et al., 2013, 2015; Bissaro et al., 2015a), it is well-established that mutations that lower the hydrolytic component of the reaction and others that increase acceptor binding can lead to overall improvements in transglycosylation. Therefore, having ascertained that EGC only poorly transfers β -cellobiosyl donor onto alcohol acceptors, it was decided to target the aglycone subsites of this enzyme using a site-saturation mutagenesis strategy to mutate four different amino acids that are thought to interact with the aglycone moiety (E179, N279, Q280 and D311) (M. E. C. Caines et al., 2007). Like in previous studies, we deployed a two-stage screening strategy that was intended to isolate mutants displaying lowered ability to catalyze the first step of the reaction (i.e., the release of CNP from the donor CNP- β -Celb in the presence of this one only), but increased ability to achieve this in the presence of additional target acceptors. The final outcome of our strategy was the isolation of D311Y. Though this clone displays greater ability to release CNP from the donor CNP- β -Celb than EGC, in contradiction to what we look for, it was the only one that exhibited an even higher activity (i.e., leaving group release) in the presence of two out of the four acceptors tested.

Previous work has revealed that rGHs that display high or even exclusive transglycosylation activity are mostly characterized by low or moderate k_{cat}/K_M values. This finding has led us to propose that so-called non-Leloir transglycosylases, which are a specific type of rGHs, are intrinsically poor enzymes that display impaired ability to generate efficient transition states (Teze et al., 2014, 2015, Bissaro et al., 2015a, 2015b). In the case of D311Y, the ability of the mutant to perform the first step of the double displacement mechanism is altered, but is apparently slightly improved when compared to EGC. However, although this appears to contradict expectations, we propose that EGC is an intrinsically very poor catalyst, because its performance constant k_{cat}/K_M is very low. Consistently, although D311Y displays enhanced k_{cat}/K_M value, the actual improvement is modest and the resulting enzyme remains an intrinsically poor catalyst for the hydrolysis of CNP- β -Celb. In fact, the modification of the T/H ratio provoked by the mutation of D311 to tyrosine is mainly due to a 3.2-fold decrease in the K_M value, which is indicative of improved donor substrate binding for transglycosylation.

Recent studies suggested that the most highly conserved residues within or around the subsite -1 are

the major determinants of donor substrate binding and thus those that have most influence on the T/H ratio (Teze et al., 2014, 2015; Bissaro et al., 2015a). In contrast, D311 is not a determinant of subsite -1, although the glutamic acid sidechain is conserved at this position in ten out of the twelve enzymes that currently form subfamily GH5_28 (Figure S7). Regarding the nature of the mutant selected in this study, it is well-established that aromatic residues are often key molecular determinants of subsites in GHs, since aromatic sidechains can stack with sugar rings. For instance, the substitution of a serine by a tryptophan in the +1 subsite of a GH5 *endo*- β -1,4-mannanase conveniently illustrates how such a mutagenesis strategy can procure higher transglycosylation capability in GHs (Dilokpimol et al., 2011b). Additionally, many studies have described positive binding sites as determinant for the T/H partition (Armand et al., 2001; van der Veen, 2001; Mark et al., 2009; Zakariassen et al., 2011; Bissaro et al., 2015a; Talens-Perales et al., 2016). Therefore, the fact that the substitution of D311 by a tyrosine procured higher transglycosylation capability in EGC is consistent with previous findings and undoubtedly results from the increased hydrophobicity and bulkiness in the aglycone subsite, notably +1 (Rivera et al., 2003; Ochs et al., 2013; Rosengren et al., 2012, 2014). These specific properties of tyrosine could alter the enzyme's behavior in several ways. Firstly, the Y311 and Y306 might stack sandwich-wise with the CNP moiety of the donor (Figure 6), an interaction that would explain the lowered K_M value of D311Y-catalyzed hydrolysis of CNP- β -Celb and, inversely, is inferred by the lowered transglycosylation level of D311Y when using F- β -Celb, a donor that lacks an aromatic aglycon moiety. Secondly, increased hydrophobicity in subsite +1 probably favors binding of longer acceptor aliphatic chains. This is clearly implied from the models of D311Y with acceptors **2** or **4** bound in channel II (right-hand side of Figure 7). Thirdly, the bulkiness of the tyrosinoyl sidechain might induce a donor/acceptor substrate interaction between the *HO-6*¹ of cellobiosyl moiety and the extra alcohol or ketone function of the acceptor. Moreover, D311 forms a salt bridge with R177. Loss of this feature in Y311 probably leads to a relative greater conformational flexibility within the active site; an important characteristic of glycosyltransferases and GHs working in transglycosylation-mode (Mark et al., 2009; Rojas-Cervellera et al., 2013; Bissaro et al., 2015b). Finally, it is noteworthy that crystallographic data reveals that D311 in EGC interacts with an ordered water molecule (c.f. structure of EGC:ganglioside complex) (M. E. C. Caines et al., 2007), thus facilitating the positioning of a nucleophilic water molecule on the β -face of the donor. Presumably, since this interaction is lost in D311Y, water is probably less well-positioned to perform hydrolysis (Hidaka et al., 2010).

It is noteworthy that in this work we used a catalytically-inefficient enzyme-substrate pair that is characterized by high transition state energy barriers (i.e., low k_{cat}/K_M value for the hydrolysis of CNP- β -Celb by EGC). In this respect, the dual effect of the mutation D311Y led to an overall improvement of catalysis that specifically favored transglycosylation. First, it procured a slightly higher - yet modest - k_{cat}/K_M value, indicative of an improved enzyme-substrate interaction and thus a lowered energy barrier associated with the glycosylation step. Second, it improved acceptor-mediated deglycosylation

of the glycosyl-enzyme intermediate, thus probably lowering the energy barrier of this step.

CONCLUSION

EGC has been successfully engineered, to create a quite efficient transglycosylase that is able to synthesize APG derivatives. Remarkably this result was achieved through a single point mutation that mostly influences interactions in the aglycone subsite, favoring first donor substrate and then acceptor substrate binding. Beyond the specific nature of our achievement, this study further underlines the fact that the switch from hydrolytic to transglycosylation activity in retaining glycoside hydrolases is brought about by fulfilling two key conditions: (i) impair the ability of the enzyme to generate efficient transition states, thus lowering its overall catalytic potency (i.e., low k_{cat}/K_M), and (ii) improve the binding of target acceptor molecules in subsite +1, thus favoring acceptor-mediated deglycosylation of the glycosyl-enzyme intermediate.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the [ACS Publications website](https://pubs.acs.org/doi/10.1021/acscatal.6b02159) at DOI:10.1021/acscatal.6b02159.

Sequence information and alignment, representation of active site, activity profiles (TLC and NMR), physico-chemical, kinetic and stability characteristics as well as transglycosylation performances.

ACKNOWLEDGMENTS

This work was performed in the framework of ERA-NETBIOSURF (grant no. 0315928A, ERA-IB10.039). NMR analyses were performed using facilities at MetaToul (Metabolomics & Fluxomics Facilities, Toulouse, France, www.metatoul.fr), which is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr), and is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, IBiSa-France, CNRS and INRA. Work requiring automated liquid handling was performed using the ICEO facility that is part of the PICT infrastructure and is financially supported by IBiSa-France, CNRS and INRA. Contribution from R.F. was partially supported by the Région Midi-Pyrénées grant DESR-Recherche 14052246 (CTP-B) and the research mobility grants from INSA Toulouse (2015). A.P acknowledges grant BIO2013-49022-C2-1-R from MINECO, Spain. The authors thank Prof. S. Cottaz and Dr S. Fort (CERMAV-CNRS) for the gift of β -cellobiosyl fluoride heptaacetate, and Prof. S. Withers for insightful discussions.

ABBREVIATIONS

APG, alkylpolyglycoside; CNP, 2-chloro-4-nitrophenol; CNP- β -Celb, 2-chloro-4-nitrophenyl β -D-Glcp-(1 \rightarrow 4)-D-Glcp (i.e. CNP- β -cellobioside); F- β -Celb, β -cellobiosyl fluoride; EGC, *endo*-glycoceramidase II from *Rhodococcus* sp. M-777 strain; GH, glycoside hydrolase; T/H, transglycosylation/hydrolysis partition; LG, leaving group; Ph, phenol; rGH, retaining GH, TG, non-Leloir transglycosylase.

REFERENCES

- Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., Fauré, R., O'Donohue, M.J., 2013. Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.* 30, 536–544.
- Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R., O'Donohue, M.J., 2015. Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α -L-arabinofuranosidase. *Carbohydr. Res.* 401, 64–72.
- Armand, S., Andrews, S.R., Charnock, S.J., Gilbert, H.J., 2001. Influence of the aglycone region of the substrate binding cleft of *Pseudomonas* xylanase 10A on catalysis. *Biochemistry* 40, 7404–7409.
- Ashida, H., Tsuji, Y., Yamamoto, K., Kumagai, H., Tochikura, T., 1993. Transglycosylation activity of endoglycoceramidase from *Corynebacterium* sp. *arch. Biochem. Biophys.* 305, 559–562.
- Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., Henrissat, B., 2012. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol. Biol.* 12, 186.
- Bissaro, B., Durand, J., Biarnés, X., Planas, A., Monsan, P., O'Donohue, M.J., Fauré, R., 2015a. Molecular design of non-Leloir furanose-transferring enzymes from an α -L-arabinofuranosidase: a rationale for the engineering of evolved transglycosylases. *ACS Catal.* 5, 4598–4611.
- Bissaro, B., Monsan, P., Fauré, R., O'Donohue, M.J., 2015b. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* 467, 17–35.
- Caines, M.E.C., Vaughan, M.D., Tarling, C.A., Hancock, S.M., Warren, R.A.J., Withers, S.G., Strynadka, N.C.J., 2007. Structural and mechanistic analyses of *endo*-Glycoceramidase II, a membrane-associated family 5 glycosidase in the apo and GM3 ganglioside-bound forms. *J. Biol. Chem.* 282, 14300–14308.
- Crout, D., Vic, G., 1998. Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis. *Current Opinion in Chemical Biology* 2, 98–111.
- Danby, P.M., Withers, S.G., 2016. Advances in enzymatic glycoside synthesis. *ACS Chem. Biol.* 11, 1784–1794.
- Dilokpimol, A., Nakai, H., Gotfredsen, C.H., Baumann, M.J., Nakai, N., Abou Hachem, M., Svensson, B., 2011. Recombinant production and characterisation of two related GH5 *endo*- β -1,4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1814, 1720–1729.
- Feng, H.-Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C., Dion, M., 2005. Converting a β -glycosidase into a β -transglycosidase by directed evolution. *J. Biol. Chem.* 280, 37088–37097.

- Grefte, L., 2004. Synthesis, preliminary characterization, and application of novel surfactants from highly branched xyloglucan oligosaccharides. *Glycobiology* 15, 437–445.
- Gudiña, E.J., Rangarajan, V., Sen, R., Rodrigues, L.R., 2013. Potential therapeutic applications of biosurfactants. *Trends Pharmacol. Sci.* 34, 667–675.
- Hancock, S.M., Rich, J.R., Caines, M.E.C., Strynadka, N.C.J., Withers, S.G., 2009. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nat. Chem. Biol.* 5, 508–514.
- Heyd, M., Kohnert, A., Tan, T.-H., Nusser, M., Kirschhöfer, F., Brenner-Weiss, G., Franzreb, M., Berensmeier, S., 2008. Development and trends of biosurfactant analysis and purification using rhamnolipids as an example. *Anal. Bioanal. Chem.* 391, 1579–1590.
- Hidaka, M., Fushinobu, S., Honda, Y., Wakagi, T., Shoun, H., Kitaoka, M., 2010. Structural explanation for the acquisition of glycosynthase activity. *J. Biochem.* 147, 237–244.
- Horibata, Y., Higashi, H., Ito, M., 2001. Transglycosylation and reverse hydrolysis reactions of *endo*-glycoceramidase from the jellyfish, *Cyanea nozakii*. *J. Biochem.* 130, 263–268.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual Molecular Dynamics. February, *Journal of Molecular Graphics* 14, 33–38.
- Ishibashi, Y., Kiyohara, M., Okino, N., Ito, M., 2007a. Synthesis of fluorescent glycosphingolipids and neoglycoconjugates which contain 6-gala oligosaccharides using the transglycosylation reaction of a novel endoglycoceramidase (EGALC). *J. Biochem.* 142, 239–246.
- Ishibashi, Y., Nagamatsu, Y., Meyer, S., Imamura, A., Ishida, H., Kiso, M., Okino, N., Geyer, R., Ito, M., 2009. Transglycosylation-based fluorescent labeling of 6-gala series glycolipids by endogalactosylceramidase. *Glycobiology* 19, 797–807.
- Ishibashi, Y., Nakasone, T., Kiyohara, M., Horibata, Y., Sakaguchi, K., Hijikata, A., Ichinose, S., Omori, A., Yasui, Y., Imamura, A., Ishida, H., Kiso, M., Okino, N., Ito, M., 2007b. A novel endoglycoceramidase hydrolyzes oligogalactosylceramides to produce galactooligosaccharides and Ceramides. *J. Biol. Chem.* 282, 11386–11396.
- Ito, M., Tatsuya, Y., 1989. Purification and characterization of glycosphingolipid-specific Endoglycosidases (Endoglycoceramidases) from a mutant strain of *Rhodococcus* sp. *J. Biol. Chem.*, The American Society for Biochemistry and Molecular Biology, Inc 264, 9510–9519.
- Ito, M., Yamagata, T., 1986. A novel glycosphingolipid-degrading enzyme cleaves the linkage between the oligosaccharide and ceramide of neutral and acidic glycosphingolipids. *J. Biol. Chem.* 261, 14278–14282.
- Izu, H., Izumi, Y., Kurome, Y., Sano, M., Kondo, A., Kato, I., Ito, M., 1997. Molecular cloning, expression, and sequence analysis of the endoglycoceramidase II gene from *Rhodococcus* species strain M-777. *J. Biol. Chem.* 272, 19846–19850.
- Johnson, K.F., Defrees, S., Withers, S., Vaughan, M., 2014. Mutant endoglycoceramidases with enhanced synthetic activity. US 2014/0147891 A1.
- Kone, F.M.T., Le Behec, M., Sine, J.-P., Dion, M., Tellier, C., 2008. Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.* 22, 37–44.
- Konstantinidis, A.K., Marsden, I., Sinnott, M.L., 1993. Hydrolyses of α - and β -cellobiosyl fluorides by cellobiohydrolases of *Trichoderma reesei*. *Biochem. J.* 291, 883–888.
- Li, Y.T., Carter, B.Z., Rao, B.N., Schweingruber, H., Li, S.-C., 1991. Synthesis of neoglycoconjugates using oligosaccharide transferring activity of ceramide glycanase. *J. Biol. Chem.* 266, 10723–10726.
- Löffler, H., Happle, R., 2003. Profile of irritant patch testing with detergents: sodium lauryl sulfate, sodium laureth sulfate and alkyl polyglucoside. *Contact Dermatitis* 48, 26–32.

- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495.
- Lundemo, P., Adlercreutz, P., Karlsson, E.N., 2013. Improved transferase/hydrolase ratio through rational design of a family 1 β -glucosidase from *Thermotoga neapolitana*. *Appl. Environ. Microbiol.* 79, 3400–3405.
- Lundemo, P., Karlsson, E.N., Adlercreutz, P., 2016. Eliminating hydrolytic activity without affecting the transglycosylation of a GH1 β -glucosidase. *Appl. Microbiol. Biotechnol.*
- Mackenzie, L.F., Wang, Q., Warren, R.A.J., Withers, S.G., 1998. Glycosynthases: mutant glycosidases for oligosaccharide synthesis. *J. Am. Chem. Soc.* 120, 5583–5584.
- Malet, C., Planas, A., 1998. From β -glucanase to β -glucansynthase: glycosyl transfer to α -glycosyl fluorides catalyzed by a mutant endoglucanase lacking its catalytic nucleophile. *FEBS Lett.* 440, 208–212.
- Mark, P., Baumann, M.J., Eklöf, J.M., Gullfot, F., Michel, G., Kallas, Å.M., Teeri, T.T., Brumer, H., Czjzek, M., 2009. Analysis of nasturtium *TmNXG1* complexes by crystallography and molecular dynamics provides detailed insight into substrate recognition by family GH16 xyloglucan *endo*-transglycosylases and *endo*-hydrolases. *Proteins Struct. Funct. Bioinforma.* 75, 820–836.
- Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., Rossi, M., 1998. Restoration of the activity of active-site mutants of the hyperthermophilic β -glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* 37, 17262–17270.
- Ochs, M., Belloy, N., Dauchez, M., Muzard, M., Plantier-Royon, R., Rémond, C., 2013. Role of hydrophobic residues in the aglycone binding subsite of a GH39 β -xylosidase in alkyl xylosides synthesis. *J. Mol. Catal. B Enzym.* 96, 21–26.
- Ochs, M., Muzard, M., Plantier-Royon, R., Estrine, B., Rémond, C., 2011. Enzymatic synthesis of alkyl β -D-xylosides and oligoxylosides from xylans and from hydrothermally pretreated wheat bran. *Green Chem.* 13, 2380.
- Osanzo, G., Dion, M., Drone, J., Solleux, C., Tran, V., Rabiller, C., Tellier, C., 2007. Directed evolution of the α -L-fucosidase from *Thermotoga maritima* into an α -L-transfucosidase. *Biochemistry* 46, 1022–1033.
- Rather, M., Mishra, S., 2013. β -Glycosidases: An alternative enzyme based method for synthesis of alkyl-glycosides. *Sustain. Chem. Process.* 1, 7.
- Reis, R.S., Pacheco, G.J., Pereira, A.G., Freire, D.M.G., 2013. Biosurfactants: production and applications, *Life of Science*. InTech.
- Rémond, C., Ferchichi, M., Aubry, N., Plantier-Royon, R., Portella, C., O'Donohue, M.J., 2002. Enzymatic synthesis of alkyl arabinofuranosides using a thermostable α -L-arabinofuranosidase. *Tetrahedron Lett.* 43, 9653–9655.
- Rich, J.R., Cunningham, A.-M., Gilbert, M., Withers, S.G., 2011. Glycosphingolipid synthesis employing a combination of recombinant glycosyltransferases and an endoglycoceramidase glycosynthase. *Chem. Commun.* 47, 10806.
- Rich, J.R., Withers, S.G., 2012. A chemoenzymatic total synthesis of the neurogenic starfish ganglioside LLG-3 using an engineered and evolved synthase. *Angew. Chem. Int. Ed.* 51, 8640–8643.
- Rivera, M.H., Lopez-Munguia, A., Soberon, X., Saab-Rincon, G., 2003. α -Amylase from *Bacillus licheniformis* mutants near to the catalytic site: effects on hydrolytic and transglycosylation activity. *Protein Eng. Des. Sel.* 16, 505–514.

- Rojas-Cervellera, V., Ardèvol, A., Boero, M., Planas, A., Rovira, C., 2013. Formation of a covalent glycosyl-enzyme species in a retaining glycosyltransferase. *Chem. - Eur. J.* 19, 14018–14023.
- Rosengren, A., Häggglund, P., Anderson, L., Pavon-Orozco, P., Peterson-Wulff, R., Nerinckx, W., Stålbrand, H., 2012. The role of subsite +2 of the *Trichoderma reesei* β -mannanase TrMan5A in hydrolysis and transglycosylation. *Biocatal. Biotransformation* 30, 338–352.
- Rosengren, A., Reddy, S.K., Sjöberg, J.S., Aurelius, O., Logan, D.T., Kolenová, K., Stålbrand, H., 2014. An *Aspergillus nidulans* β -mannanase with high transglycosylation capacity revealed through comparative studies within glycosidase family 5. *Appl. Microbiol. Biotechnol.* 98, 10091–10104.
- Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol., Journal of Molecular Biology* 243, 779–815.
- Schwarz, J.C., Klang, V., Hoppel, M., Mahrhauser, D., Valenta, C., 2012. Natural microemulsions: Formulation design and skin interaction. *Eur. J. Pharm. Biopharm.* 81, 557–562.
- Stocker, B.L., Timmer, M.S.M., 2013. Chemical tools for studying the biological function of glycolipids. *ChemBioChem* 14, 1164–1184.
- Talens-Perales, D., Polaina, J., Marín-Navarro, J., 2016. Structural dissection of the active site of *Thermotoga maritima* β -galactosidase identifies key residues for transglycosylating activity. *J. Agric. Food Chem.* 64, 2917–2924.
- Teze, D., Daligault, F., Ferrieres, V., Sanejouand, Y.-H., Tellier, C., 2015. Semi-rational approach for converting a GH36 α -glycosidase into an α -transglycosidase. *Glycobiology* 25, 420–427.
- Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., Tellier, C., Dion, M., 2014. Semi-rational approach for converting a GH1 β -glycosidase into a β -transglycosidase. *Protein Eng. Des. Sel.* 27, 13–19.
- Trott, O., Olson, A.J., 2009. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. NA-NA*.
- van der Veen, B.A., 2001. Hydrophobic amino acid residues in the acceptor binding site are main determinants for reaction mechanism and specificity of cyclodextrin-glycosyltransferase. *J. Biol. Chem.* 276, 44557–44562.
- Van Rantwijk, F., Woudenberg-van Oosterom, M., Sheldon, R.A., 1999. Glycosidase-catalysed synthesis of alkyl glycosides. *J. Mol. Catal. B Enzym.* 6, 511–532.
- Vaughan, M.D., Johnson, K., DeFrees, S., Tang, X., Warren, R.A.J., Withers, S.G., 2006. Glycosynthase-mediated synthesis of glycosphingolipids. *J. Am. Chem. Soc.* 128, 6300–6301.
- von Rybinski, W., Hill, K., 1998. Alkyl polyglycosides—properties and applications of a new class of surfactants. *Angew. Chem. Int. Ed.* 37, 1328–1345.
- Wang, L.-X., Huang, W., 2009. Enzymatic transglycosylation for glycoconjugate synthesis. *Curr. Opin. Chem. Biol.* 13, 592–600.
- Wennekes, T., van den Berg, R.J.B.H.N., Boot, R.G., van der Marel, G.A., Overkleeft, H.S., Aerts, J.M.F.G., 2009. Glycosphingolipids-nature, function, and pharmacological Modulation. *Angew. Chem. Int. Ed.* 48, 8848–8869.
- Yang, G.-Y., Li, C., Fischer, M., Cairo, C.W., Feng, Y., Withers, S.G., 2015. A FRET probe for cell-based Imaging of ganglioside-processing enzyme activity and high-throughput screening. *Angew. Chem. Int. Ed.* 54, 5389–5393.
- Zakariassen, H., Hansen, M.C., Jøranli, M., Eijsink, V.G.H., Sørli, M., 2011. Mutational effects on transglycosylating activity of family 18 chitinases and construction of a hypertransglycosylating mutant. *Biochemistry* 50, 5693–5703.

Zhang, X.-F., Yang, G.-Y., Zhang, Y., Xie, Y., Withers, S.G., Feng, Y., 2016. A general and efficient strategy for generating the stable enzymes. *Sci. Rep.* 6, 33797.

Supporting Information

A Single Point Mutation Alters the Transglycosylation/Hydrolysis Partition, Significantly Enhancing the Synthetic Capability of an *Endo*-Glycoceramidase

Julien Durand,[†] Xevi Biarnés,[‡] Laurie Watterlot,[†] Cyrielle Bonzom,[†] Vinciane Borsenberger,[†] Antoni Planas,[‡] Sophie Bozonnet,[†] Michael J. O'Donohue,[†] Régis Faure^{*,†}

[†]*LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France*

[‡]*Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Barcelona, Spain*

* Corresponding Author

regis.faure@insa-toulouse.fr

atgggcagcagccatcatcatcatcacagcagcgccctggtgccgcgagccatgggatccagcggaagcggtagcggttcgggtaccgcgctgacaccttcatatct
 gaaggatgatgacggcgagcctcattctctgctggatttaatacggcctcatctgcacaaagtgccctgacggcatgccacagttcactgaagcagatttggcgcgtgaatag
 cggacatgggtactaattttgtacgttttctgatctcttggcgctcggtggaaccggctcctggcgtatgatcaacagtaacctggatcgtgtagaagaccgttaggttggtac
 gcagagcgtggttataaagtattgctggacatgcatcaagacgtgtactcgggggcatctactccggaaggcaatagtggtaatggcgcaggtgcgattggtaattggggcaccggc
 gtgggcccacatattggatggtctgccagtggaacccccaccccgctgggaactgtattacatccagccagggcgtgatcgggcgcttgataattttggaacacgaccggcaagc
 atccggaaactggtggaacattatgcgaaagcgtggcgcgctgtagctgaccgcttcgcggaataatgatcggttggcctatgacctgatgaatgagccggttgggtggctccctg
 cagggaccggcattcgaagcgggccattagcagcaatgtaccagcgcaactactgatgccatccgtcaggtggatcaggatacttgggtttgtgtggcaccgagccattggcgt
 taatcaaggtttaccatcggttaactaaaattgatgacctcgcgcgggtcaacaacgcattgacctattgcccgcattctgtaccgctgccattggacatcggcgacggccacg
 aaggacttgcgcgcaactctgaccgatgtaaccattgatgcctggcgtgcgaacacggctcataccgcgcgctcttgggtgatgtgcctatcattctcggtgaattcggtttgat
 accacgctgcccggagcagcgcatatcattgacgtgtctatgggaccgcagcgaatgggtgcggcgcttagttattggtcgagtgatcccggcccgtggggcccgtatctgcc
 ggacggtacgcagaccttgttagtgatcttaacaagccataccctcgtgcagtgccggggaccctaccgaatggagcagcacttcggatcgctgcaattgaccattgaac
 cagatgcccctattaccgcccctacagaaatctacctgctgaggtggttccccggggatgtgcattgagaagggcggtatgcttggctgggatcgtcaatcgcgtctttta
 accgtacgcactcccggcagctggttaacctcacagtgacagttacgcccgcagcgtga

10	20	30	40	50	60	70	80	90
MGS	SHHHHHHSSG	LVPRGSHMGS	SGSGSGSGTA	LTPSYLKDDD	GRSLILRGFN	TASSAKSAPD	GMPQFTEADL	AREYADMGTN
100	110	120	130	140	150	160	170	
FVRFLISWRS	VEPAPGVYDQ	QYLDRVEDRV	GWYAERGYKV	MLDMHQDVYS	GAITPEGNSG	NGAGAIENGA	PAWATYMDGL	
180	190	200	210	220	230	240	250	
PVEPQPRWEL	YYIQPGVMRA	FDNFWNTGK	HPELVEHYAK	AWRAVADRFA	DNDAVVAYDL	MNEPFGGSLQ	GPAFEAGPLA	
				A/B				
260	270	280	290	300	310	320	330	
AMYQRTTDAI	RQVDQDTWVC	VAPQAIGVNO	GLPSGLTKID	DPRAGQQRIA	YCPHLYPLPL	DIGDHEGLA	RTLTDVTIDA	
340	350	360	370	380	390	400	410	
WRANTAHTAR	VLGDVPIILG	EFGLDITLPG	ARDYIERVYG	TAREMGAGVS	YWSSDPGPWG	PYLPDGTQTL	LVDTLNKPYP	
	Nu							
420	430	440	450	460	470	480	490	
RAVAGTPTEW	SSTSDRLQLT	IEPDAAITAP	TEIYLPFAGF	PGDVHVEGAD	VVGWDRQSRL	LIVRTPADSG	NVTVTVIPAA	

Figure S1. Sequences of the synthetic gene encoding EGC (*endo*-glycoceramidase II from *Rhodococcus* sp. strain M-777) and of its corresponding *N*-terminal (His)₆-tag protein. Numbering as by Hancock *et al.*¹ has been used. The key positions (catalytic residues, i.e. acid/base E233 and nucleophile E351 -notated A/B and Nu respectively, as well as R177 and D314) are highlighted in green, red and blue whereas the four targeted active site residues (E179, N279, Q280 and D311) are written in orange.

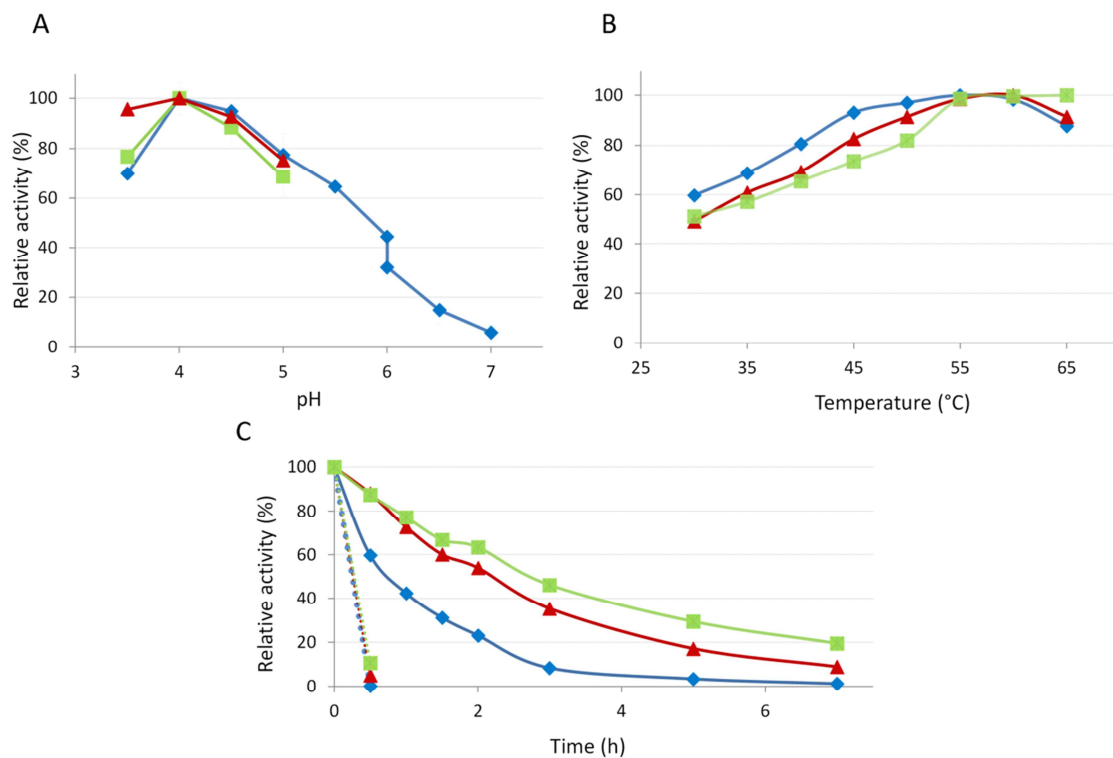


Figure S4. Activity profile at different (A) pH and (B) temperatures, and (C) thermostability at 37 °C (solid line) or 50 °C (dashed line) of EGC (◆), D311Y (▲) and D314Y (■).

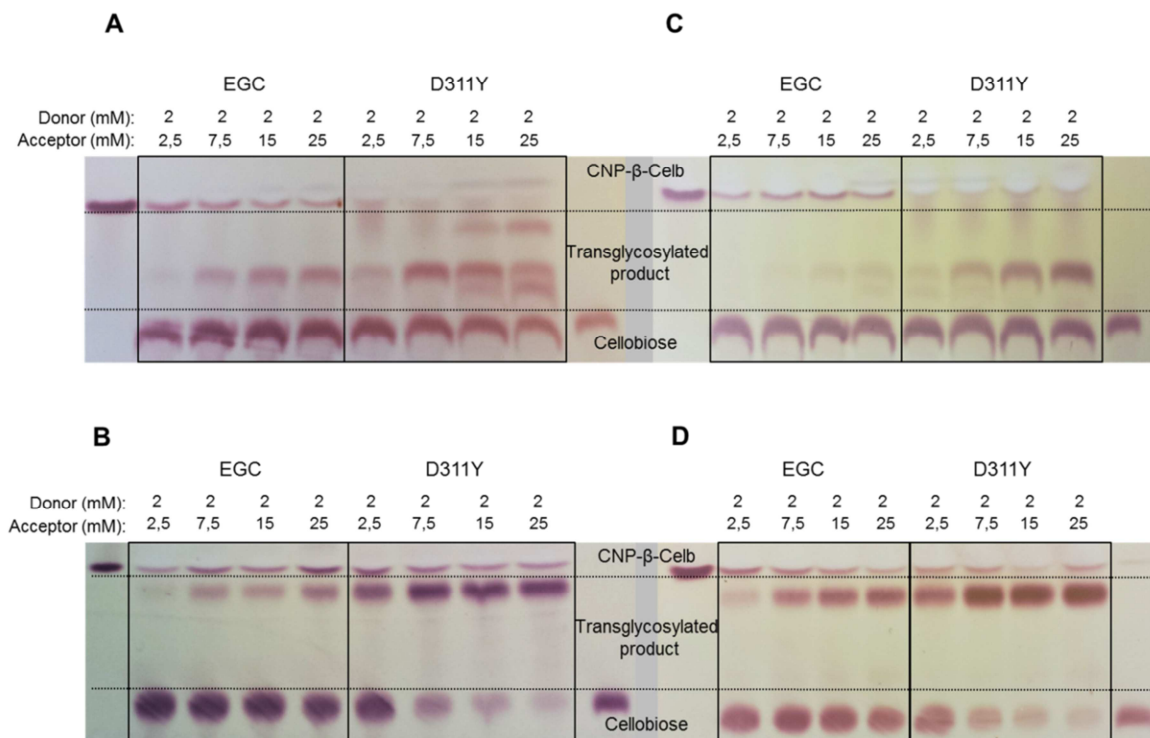


Figure S5. TLC profiles of reaction products with different ratios of donor/acceptor with acceptor 1 (A), 2 (B), 3 (C) and 4 (D). Donor substrate (CNP-β-Celb) and hydrolysis product (cellobiose) are deposited as standards.

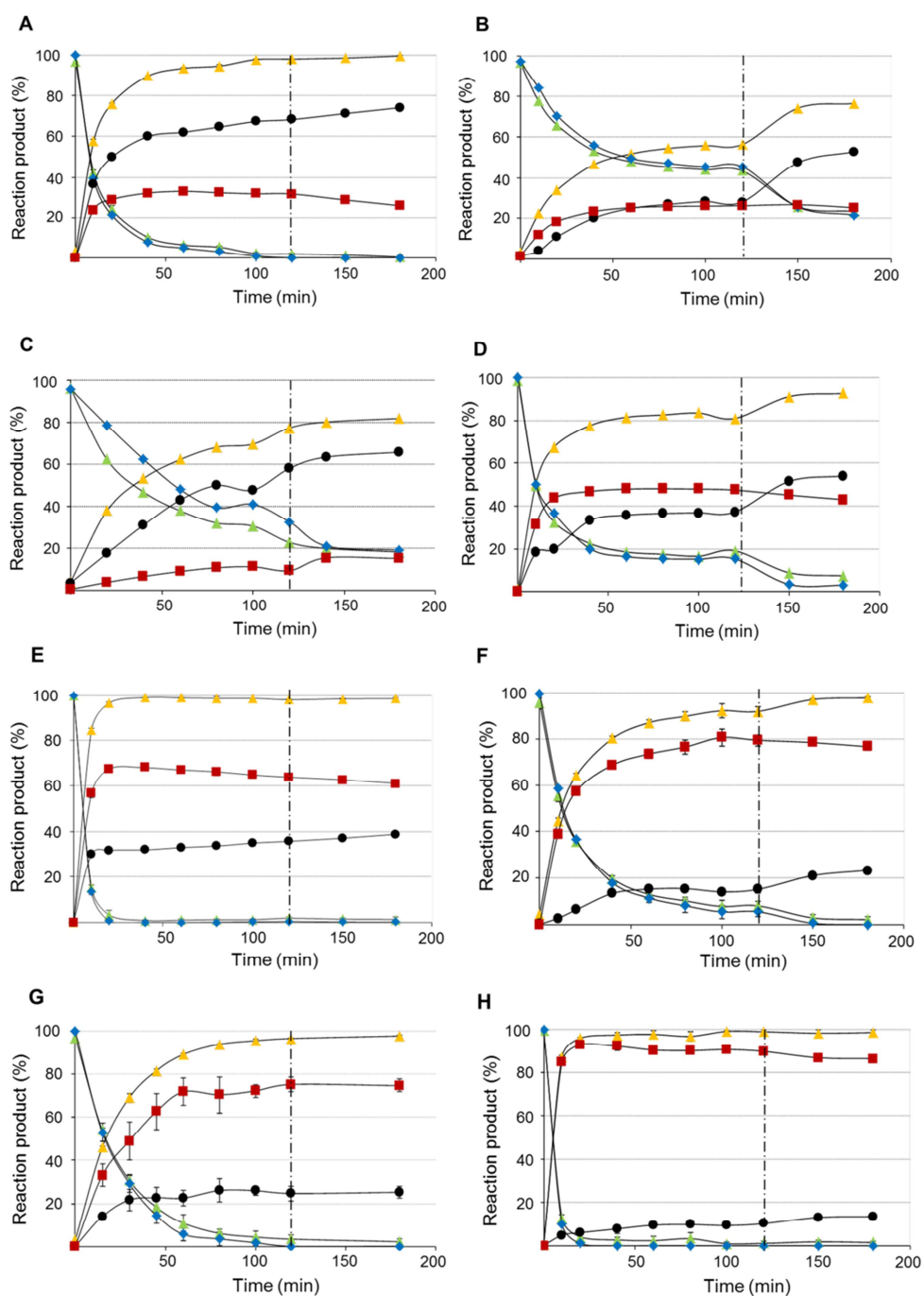


Figure S6. Kinetic of formation of reaction products for EGC (A-D) and D311Y (E-H), with CNP- β -Celb (5 mM) as donor substrate and an acceptor (62.5 mM of **1**, **2**, **3** or **4** respectively). Species involved during enzymatic reactions catalyzed by EGC and D311Y: free CNP (▲), bonded CNP (▲), CNP- β -Celb (◆), cellobiose (●) and transglycosylation product(s) (■). Enzyme concentration is doubled after sampling at 120 min (black dash-dotted lines), enabling to appreciate secondary hydrolysis.

CBL45596.1 1 MNIKNKWNFLTAQLTQKRRASGAVLGI...GILLG-----CWQKPAETQ...TDSQGRAL...ILGLN...SSSAKE
CAM03023.1 1 -----MILRRWITAFALA...LLLTTPA-----QAAQNPP...TDLDC...TLPV...GLNNGHSAKE
CBH47454.1 1 -----MLSLAAGAV...VLAGGT...AALRA-----DPAQSTA...ITDDAGRA...ILNGFN...TASAKS
BAB17317.1 1 -----MRRTRIASLAVAGS...VLAGV...TAQSSLPATGSD--SSEWSASA...ITDDAGRS...ILGNFN...TASAKS
EGC 1 -----MRRTRLVSLIVTGS...VIFGGV...AAQSSLAASSGSGSGS...TALTPS...YTKDD...GRS...ILGNFN...TASAKS
AGT95256.1 1 -----MRRRIPLSALIVAGS...VILGGI...AAQSSLPTSGS--AGGTDVPPS...ITDDAGRS...ILGNFN...TASAKS
ALU68549.1 1 -----MRRRIPLSALIVAGS...VILGGI...AAQSSLPTSGS--AGGTDVPPS...ITDDAGRS...ILGNFN...TASAKS
AKE01030.1 1 -----MRRRIPLSALIVAGS...VILGGI...AAQSSLPTSGS--AGGTSMSPS...ITDDAGRS...ILGNFN...TASAKS
BAH36401.1 1 -----MRRRIPLSALIVAGS...VILGGI...AAEGSLPTSGS--AGGTSMSPS...ITDDAGRS...ILGNFN...TASAKS
AFR06036.1 1 -----MQP...PKRGLVPAALIVLV...LLSGV...ALTRHD-----T-DAPVR...ITDDQ...GRA...FLL...GLN...TSGSAKD
ADD40377.1 1 -----MRRARWFGVLFVAV...VLLVGG...SVVRGQ-----PESRVR...ITDDE...GRA...FLL...GLN...TASAKS
ACZ84704.1 1 -----MRRPRLAAAGLALA...VAA...GGL...SAARGD-----APERVSR...ITDDR...GRV...ITL...GNFN...TASAKS

CBL45596.1 69 PETGH...EWIT...EAD...DQEVTEW...G...NFVRL...L...FDG...EP...K...GVY...DEAY...LD...VA...ER...V...R...W...Y...T...S...R...G...A...Y...V...L...D...M...H...Q...D...I...Y...S...N...K...V...G...--
CAM03023.1 54 S-ADA...AWT...G...P...G...D...AAE...ADR...G...N...S...T...R...L...L...M...F...N...S...R...E...P...R...P...C...H...Y...D...E...H...Y...L...D...A...V...A...Q...R...V...A...W...Y...E...R...A...C...M...R...V...L...D...M...H...Q...D...M...G...P...A...V...A...G...S...
CBH47454.1 53 T-PDQ...PI...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...S...O...E...Y...L...D...R...V...A...E...R...V...D...Y...A...E...R...G...Y...S...V...L...D...M...H...Q...D...L...S...S...A...T...P...S...
BAB17317.1 66 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...A...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...L...Y...S...G...A...T...P...D...
EGC 68 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...V...Y...S...G...A...T...P...E...
AGT95256.1 66 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...V...Y...S...G...A...T...P...E...
ALU68549.1 66 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...V...Y...S...G...A...T...P...E...
AKE01030.1 66 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...V...Y...S...G...A...T...P...E...
BAH36401.1 66 A-PDG...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...P...C...V...Y...D...Q...Y...L...D...R...V...E...D...R...V...G...W...Y...A...E...R...G...Y...K...V...L...D...M...H...Q...D...V...Y...S...G...A...T...P...E...
AFR06036.1 62 D-PDH...PK...F...E...S...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...E...E...G...V...Y...D...E...D...Y...L...D...V...A...B...R...V...D...W...Y...A...E...R...G...Y...H...V...L...D...M...H...Q...D...L...Y...G...R...F...T...N...T...E...
ADD40377.1 59 T-PDA...PD...I...S...E...P...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...R...P...C...V...Y...D...E...D...Y...L...D...V...A...B...R...V...D...W...Y...A...E...R...G...Y...H...V...L...D...M...H...Q...D...L...Y...G...P...G...V...T...A...K...
ACZ84704.1 59 T-PDA...PE...L...E...P...D...D...R...E...H...A...D...G...N...F...V...R...F...L...S...M...R...S...E...P...S...P...C...V...Y...D...T...K...Y...L...K...V...A...B...R...V...D...W...Y...A...K...R...G...Y...H...V...L...D...M...H...Q...D...L...Y...G...N...Q...T...P...D...
* (83%)

CBL45596.1 147 -----GNGAP...D...W...A...T...V...T...N...E...F...A...M...D...F...P...G...M...P...W...K...N...D...P...S...V...A...A...F...V...N...F...W...R...Y...Q...E...-Y...A...M...L...Q...H...Y...I...A...A...W...K...V...A...G...E...F...K...D...D...
CAM03023.1 133 SP-----GNGAP...A...W...A...T...H...F...D...G...L...P...A...T...V...-----Q...D...P...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...K...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
CBH47454.1 132 GQN-----GNGAP...A...W...A...T...Y...L...D...G...L...P...G...D...---N...D...M...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...K...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
BAB17317.1 145 GKT-----GNGAP...A...W...A...T...Y...M...D...G...L...P...E...N...E-----R...D...S...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...K...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...D...N...
EGC 147 GNSGN...G...A...G...A...I...G...N...G...A...P...A...W...A...T...Y...M...D...G...L...P...E...P...E-----Q...D...R...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...K...H...P...E...L...V...H...Y...A...K...A...W...A...D...R...F...A...D...N...
AGT95256.1 145 GNSGN...G...A...G...N...I...G...N...G...A...P...A...W...A...T...Y...M...D...G...L...P...S...P-----Q...D...R...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...Q...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
ALU68549.1 145 GNSGN...G...A...G...N...I...G...N...G...A...P...A...W...A...T...Y...M...D...G...L...P...S...P-----Q...D...R...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...Q...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
AKE01030.1 145 GNSGN...G...A...G...N...I...G...N...G...A...P...A...W...A...T...Y...M...D...G...L...P...S...P-----Q...D...R...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...Q...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
BAH36401.1 145 GNSGN...G...A...G...N...I...G...N...G...A...P...A...W...A...T...Y...M...D...G...L...P...S...P-----Q...D...R...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...N...T...T...G...Q...H...P...E...L...V...H...Y...A...N...A...W...A...D...R...F...A...G...N...
AFR06036.1 141 QHS-----GNGAP...E...W...A...T...R...I...D...L...P...R...P-----R...E...C...O...W...E...Y...Y...C...P...G...V...I...R...A...F...D...H...F...W...T...A...G...G...D...T...D...M...A...Y...A...D...A...W...R...H...V...A...D...R...F...A...D...E...
ADD40377.1 138 GDV-----GNGAP...R...W...A...T...E...T...D...G...L...P...S...A-----C...E...Q...W...E...Y...Y...C...P...G...V...I...R...A...F...D...N...F...W...G...A...S...-D...A...T...L...M...R...H...Y...A...K...A...W...G...E...V...A...D...R...F...A...D...A...
ACZ84704.1 138 GRA-----GNGAP...G...W...A...T...H...L...D...G...L...P...A...G---E...H...D...T...W...E...Y...Y...C...P...G...V...I...R...A...F...D...H...F...W...N...T...T...G...E...H...P...E...L...M...H...Y...A...G...A...W...A...D...R...F...K...N...D...
* (83%)

CBL45596.1 216 P...M...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---L...F...E...R...D...L...G...P...Y...Q...R...L...I...T...E...R...T...V...D...Q...H...W...F...E...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...T...G...
CAM03023.1 201 P...A...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---L...F...E...R...D...L...G...P...Y...Q...R...L...I...T...E...R...T...V...D...Q...H...W...F...E...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...T...G...
CBH47454.1 201 P...A...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...A...G...L...I...T...A...Y...Q...R...T...S...D...A...T...R...E...V...D...Q...D...T...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...D...G...
BAB17317.1 214 E...T...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...A...G...L...I...T...A...Y...Q...R...T...S...D...A...T...R...E...V...D...Q...D...S...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...Q...G...
EGC 223 D...A...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...A...G...L...I...T...A...Y...Q...R...T...S...D...A...T...R...E...V...D...Q...D...T...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...A...G...
AGT95256.1 221 D...T...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...G...G...P...L...A...A...Y...Q...R...T...T...N...A...T...R...E...V...D...Q...D...S...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...A...G...
ALU68549.1 221 D...T...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...G...G...P...L...A...A...Y...Q...R...T...T...N...A...T...R...E...V...D...Q...D...S...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...A...G...
AKE01030.1 221 D...T...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...G...G...P...L...A...A...Y...Q...R...T...T...N...A...T...R...E...V...D...Q...D...S...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...A...G...
BAH36401.1 221 D...T...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...G...G...P...L...A...A...Y...Q...R...T...T...N...A...T...R...E...V...D...Q...D...S...W...C...A...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...A...G...
AFR06036.1 210 P...A...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---D...F...E...A...G...L...I...T...A...Y...Q...R...T...S...D...A...T...R...E...V...D...Q...D...T...W...F...E...G...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...E...G...
ADD40377.1 206 D...A...V...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...S...G...P...L...A...E...Y...R...L...S...I...D...E...T...R...O...V...D...A...D...S...W...F...E...P...Q...A...G...V...N...G...P...S...I...P...A...H...D...P...---R...R...G...
ACZ84704.1 208 P...A...T...Y...D...L...M...N...E...P...G...G...S...P...W...P...---A...F...E...N...G...P...L...A...E...Y...R...T...I...A...Q...I...R...S...V...D...Q...D...N...W...F...E...P...Q...A...G...V...N...G...P...S...I...P...R...F...D...P...---R...A...G...
A/B (100%) ** (58%)

CBL45596.1 296 E...K...R...V...Y...A...P...H...F...Y...F...G...H...E...C...I...P...N...L...V...D...I...Q...---Q...R...D...W...N...S...N...R...V...A...E...L...N...R...--Q...K...V...P...L...Y...G...E...F...G...G...S...D...N...T...V...G...F...S...Q...V...I...N...D...V...T...A...M...Y...D...Q...
CAM03023.1 274 A...P...R...I...A...P...H...F...Y...F...T...L...G...A...G...Q...T...Y...G...A...T...A...L...V...R...N...T...R...W...F...D...A...N...A...A...T...R...R...--L...G...A...P...M...V...G...E...F...G...L...D...A...K...P...G...A...L...E...F...V...D...D...L...T...E...A...S...R...
CBH47454.1 274 E...Q...R...I...A...P...H...F...Y...F...L...P...L...D...L...G...G...Y...T...G...S...S...T...Q...V...D...A...T...D...C...W...R...G...N...V...L...R...T...A...-A...E...L...G...N...V...P...I...L...G...E...F...G...L...D...T...L...S...G...A...L...D...Y...V...N...R...Y...D...T...A...R...D...
BAB17317.1 287 A...R...I...A...P...H...F...Y...F...L...P...L...D...L...D...C...Y...S...G...F...S...T...L...D...A...T...E...T...W...R...T...S...I...E...H...V...D...T...V...L...E...G...A...V...I...L...G...E...F...G...L...D...T...L...P...G...A...D...Y...L...D...R...V...Y...T...A...R...D...
EGC 296 Q...Q...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...D...G...E...G...L...A...T...L...D...V...T...D...A...N...R...A...N...T...A...H...T...-R...V...L...D...G...V...P...I...L...G...E...F...G...L...D...T...L...P...G...A...T...D...Y...I...E...R...Y...G...T...A...R...E...
AGT95256.1 294 E...Q...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...D...Y...S...G...L...S...T...L...D...A...T...D...T...W...R...D...S...V...E...H...I...G...N...V...L...G...G...V...P...I...L...G...E...F...G...L...D...T...L...P...G...A...T...D...Y...I...D...R...Y...T...T...A...R...E...
ALU68549.1 294 E...Q...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...D...Y...S...G...L...S...T...L...D...A...T...D...T...W...R...D...S...V...E...H...I...G...N...V...L...G...G...V...P...I...L...G...E...F...G...L...D...T...L...P...G...A...T...D...Y...I...D...R...Y...T...T...A...R...E...
AKE01030.1 294 E...Q...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...D...Y...S...G...L...S...T...L...D...A...T...D...T...W...R...D...S...V...E...H...I...G...N...V...L...G...G...V...P...I...L...G...E...F...G...L...D...T...L...P...G...A...T...D...Y...I...D...R...Y...T...T...A...R...E...
BAH36401.1 294 E...Q...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...D...Y...S...G...L...S...T...L...D...A...T...D...T...W...R...D...S...V...E...H...I...G...N...V...L...G...G...V...P...I...L...G...E...F...G...L...D...T...L...P...G...A...T...D...Y...I...D...R...Y...T...T...A...R...E...
AFR06036.1 283 E...P...R...V...Y...A...P...H...F...Y...F...L...P...L...D...I...C...Q...P...E...G...D...S...G...G...I...D...A...S...D...A...N...S...E...H...V...R...L...Y...E...R...--L...E...A...P...V...I...L...G...E...F...G...L...D...M...G...A...G...A...P...E...F...V...E...R...L...E...E...T...E...T...
ADD40377.1 279 E...N...R...I...A...P...H...F...Y...F...L...P...L...D...I...C...E...D...Y...A...T...V...K...G...E...-V...D...A...A...T...T...W...S...D...N...V...L...R...T...Q...R...--L...D...A...P...V...I...L...G...E...F...G...L...D...A...R...P...G...A...L...D...Y...E...S...V...A...D...L...A...D...A...
ACZ84704.1 281 E...P...R...I...G...L...A...P...H...F...Y...F...L...P...L...D...I...C...E...D...Y...V...G...A...S...E...W...T...D...T...L...G...W...R...S...N...M...L...R...L...E...R...--Q...G...G...V...I...L...G...E...F...G...L...D...T...R...P...G...A...T...D...Y...E...R...M...E...M...T...D...E...
* (83%) Nu

CBL45596.1 370 M...A...S...W...A...T...W...A...N...D...P...C...D...W...G...L...I...D...K...D...N...---E...N...P...K...V...N...Q...V...R...A...Y...P...R...A...I...G...A...E...S...F...S...Y...H...P...E...T...R...I...F...L...Q...Y...R...P...K...V...G...V...T...G...E...F...I...F...V...E...K...R...
CAM03023.1 352 R...C...A...G...W...Y...W...S...N...D...P...C...G...W...G...P...A...P...G...G...---W...A...P...L...A...D...H...A...S...G...-----
CBH47454.1 353 M...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...N...E...E...D...G...A...---E...R...N...L...V...A...A...D...E...P...Y...P...R...A...I...A...G...E...V...R...W...A...A...D...A...D...R...L...E...F...S...Y...V...--P...N...L...S...I...T...A...P...E...I...Y...L...P...E...A...
BAB17317.1 367 M...C...A...G...V...S...Y...W...S...S...D...R...C...P...W...G...P...L...E...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...P...V...R...W...S...S...T...S...D...R...L...D...L...T...Y...R...--N...D...P...A...V...T...A...P...E...I...Y...L...P...A...A...
EGC 375 M...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...L...P...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...T...E...W...S...S...T...S...D...R...L...Q...L...T...I...E...--P...D...A...I...T...A...P...E...I...Y...L...P...E...A...
AGT95256.1 374 L...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...L...P...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...T...E...W...S...S...T...V...D...R...L...Q...L...T...F...T...--P...N...P...A...V...T...A...P...E...I...Y...L...P...E...S...
ALU68549.1 374 L...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...L...P...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...T...E...W...S...S...T...V...D...R...L...Q...L...T...F...T...--P...N...P...A...V...T...A...P...E...I...Y...L...P...E...S...
AKE01030.1 374 L...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...L...P...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...T...E...W...S...S...T...V...D...R...L...Q...L...T...F...T...--P...N...P...A...V...T...A...P...E...I...Y...L...P...E...S...
BAH36401.1 374 L...C...A...G...V...S...Y...W...S...S...D...P...C...P...W...G...P...L...P...D...G...T...---Q...T...I...L...V...D...T...N...K...P...Y...P...R...A...V...A...G...E...T...E...W...S...S...T...V...D...R...L...Q...L...T...F...T...--P...N...P...A...V...T...A...P...E...I...Y...L...P...E...S...
AFR06036.1 361 M...A...G...R...T...Y...W...S...N...D...D...G...W...G...P...N...E...E...A...D...D...G...G...L...A...P...G...L...A...H...V...N...R...A...Y...P...R...A...V...A...G...E...P...L...D...W...G...Y...D...A...E...A...G...E...L...T...V...R...Y...T...E...R...D...G...V...S...A...T...E...L...Y...L...P...S...D...
ADD40377.1 356 T...G...C...W...S...Y...W...S...N...D...P...C...G...W...G...P...N...E...E...D...G...K...---A...S...P...L...V...K...T...D...R...A...Y...P...R...A...V...A...G...E...T...A...I...D...S...D...A...S...L...T...L...F...R...--G...D...P...V...K...A...P...E...I...Y...L...P...E...R...G...
ACZ84704.1 359 L...G...M...C...R...A...Y...W...S...R...D...P...C...P...W...G...P...N...D...G...E...G...R...---P...G...L...V...P...V...Y...A...R...Y...P...R...A...I...C...G...V...E...R...Q...I...A...Y...D...R...A...A...H...S...L...T...V...R...--F...D...A...D...G...S...P...E...I...Y...L...P...A...D...
* (83%)

CBL45596.1 446 H...Y...S...Q...G...F...A...V...L...-T...Q...D...P...N...K...A...R...F...E...M...D...A...A...R...Q...L...E...Y...T...Y...--Y...P...Q...A...D...A...V...H...W...A...Y...P...L...E...D...A...P...A...G...V...T...Q...V...F...A...P...
CAM03023.1 427 T...F...P...D...G...P...D...V...---D...G...A...H...V...A...N...M...N...P...E...T...R...I...L...T...L...--R...A...P...A...T...R...Q...E...V...T...I...T...T...P...R...G...-----
CBH47454.1 441 G...F...P...G...D...I...A...V...---Q...G...A...D...V...V...G...M...D...S...Q...S...R...L...T...V...--R...S...A...P...D...A...G...E...V...T...I...T...P...T...P...A...-----
BAB17317.1 449 G...F...P...G...D...V...H...V...---E...G...A...D...V...V...G...M...D...R...Q...S...R...L...T...V...--R...T...P...A...D...S...G...N...V...T...I...T...P...T...P...A...-----
EGC 448 G...F...P...G...G...V...D...V...---A...G...A...D...I...V...M...N...P...Q...S...R...L...T...V...--E...T...A...D...D...A...G...P...V...T...I...T...P...T...P...A...S...-----
ALU68549.1 448 G...F...P...G...G...V...D...V...---A...G...A...D...I...V...M...N...P...Q...S...R...L...T...V...--E...T...A...D...D...A...G...P...V...T...I...T...P...T...P...A...S...-----
AKE01030.1 448 G...F...P...G...G...V...D...V...---A...G...A...D...I...V...M...N...P...Q...S...R...L...T...V...--E...T...A...D...D...A...G...P...V...T...I...T...P...T...P...A...S...-----
BAH36401.1 448 G...F...P...G...G...V...D...V...---A...G...A...D...I...V...M...N...P...Q...S...R...L...T...V...--E...T...A...D...D...A...G...P...V...T...I...T...P...T...P...A...G...-----
AFR06036.1 441 V...F...G...E...G...H...E...P...I...V...V...L...G...D...G...S...W...N...A...E...W...D...E...G...M...R...L...H...V...I...T...V...G...A...A...D...G...E...T...V...I...T...P...E...-----
ADD40377.1 430 F...D...G...G...E...L...D...---C...D...G...Q...S...D...M...D...A...R...R...L...V...S...V...--A...T...V...A...D...D...E...Q...H...T...I...T...T...A...K...-----
ACZ84704.1 433 E...F...P...R...G...R...V...---T...G...A...L...S...S...M...D...P...D...R...R...I...L...T...A...--T...T...A...-S...G...V...Q...I...T...I...T...P...S...-----

Figure S7. Multiple sequence alignment of the *endo*-glycosylceramidase II from *Rhodococcus* sp. strain M-777 (EGC, GenBank accession no [AAB67050.1](#)) with the eleven other GH5_28 proteins: HDN1F_20130 from *Gamma proteobacterium* HdN1 ([CBL45596.1](#)), SACE_3752 from *Saccharopolyspora erythraea* NRRL 2338 ([CAM03023.1](#)), B005_1763 from *Nocardiopsis alba* ATCC BAA-2165 ([AFR06036.1](#)), REQ_13700 from *Rhodococcus equi* 103S ([CBH47454.1](#)), *endo*-glycosylceramidase I (EGC I) from *Rhodococcus* sp. C9 ([BAB17317.1](#)), O5Y_27195 from *Rhodococcus erythropolis* CCM2595 ([AGT95256.1](#)), H351_05000 from *Rhodococcus erythropolis* R138 ([ALU68549.1](#)), XU06_27470 from *Rhodococcus erythropolis* BG43 ([AKE01030.1](#)), RER_56930 from *Rhodococcus erythropolis* PR4 ([BAH36401.1](#)), Snas_0664 from *Stackebrandtia nassauensis* DSM 44728 ([ADD40377.1](#)) and Sros_1713 from *Streptosporangium roseum* DSM 43021 ([ACZ84704.1](#)).

The four targeted active site residues (E179, N279, Q280 and D311 - numbering of EGC II from *Rhodococcus* sp. strain M-777) are highlighted (asterisk combined with their conservations - in % - and orange letters). The key positions (catalytic residues, i.e. acid/base E233 and nucleophile E351, as well as R177 and D314 - numbering of EGC II from *Rhodococcus* sp. strain M-777) are also highlighted in green, red and blue. The alignment was generated using Clustal Omega (v1.2.0)² and BoxShade (v3.21)³ was used for its presentation.

Table S1. Physico-chemical properties of EGC and mutants D311Y and D314Y for the hydrolysis of CNP- β -Celb

Enzyme	pH _{opt.}	Temp. _{opt.} (°C)	Residual act. (%)	
			30 min at 50 °C	2 h at 37 °C
EGC	4	55-60	0	23
D311Y	4	55-60	5	54
D314Y	4	55-65	11	63

Table S2. Transglycosylation performances for reactions catalyzed by EGC and D311Y

Acceptor	Enzyme	Maximal transglycosylation yield		Linearity of the transglycosylation phase		
		Time (min)	Donor consumption (%) ^a	Duration (min) ^b	Donor consumption (%) ^c	R ² value
1	EGC	60	95	20	80	> 0.99
	D311Y	40 ^d	100 ^d	40 ^d	100 ^d	> 0.99
2	EGC	120	54	10	16	> 0.98
	D311Y	100 ^d	95 ^d	100 ^d	95 ^d	> 0.99
3	EGC	120	68	80	60	> 0.99
	D311Y	120 ^d	100 ^d	120 ^d	100 ^d	> 0.99
4	EGC	80	85	40	80	> 0.98
	D311Y	20 ^d	99 ^d	40 ^d	100 ^d	> 0.99

^a This donor consumption column gives the percentage of donor consumed when the maximum yield was reached. ^b Duration of the linear transglycosylation phase starting from the addition of enzyme (at t= 0 min). ^c This donor consumption column gives the percentage of donor consumed at the end of the linear transglycosylation phase. ^d There is a clear relationship between the linearity of the transglycosylation phase and the reached maximum transglycosylation yield for the mutant D311Y.

REFERENCES:

- (1) Hancock, S.M., Rich, J.R., Caines, M.E.C., Strynadka, N.C.J., Withers, S.G., 2009. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nat. Chem. Biol.* 5, 508–514.
- (2) www.ebi.ac.uk/Tools/msa/clustalo
- (3) www.ch.embnet.org/software/BOX_form.html

Chapitre III : Evolution dirigée d'une féruloyle estérase pour améliorer sa thermostabilité et sa résistance aux solvants

Dans ce nouveau chapitre, une feruloyle estérase fongique, exprimée de façon recombinante chez *Yarrowia lipolytica*, a été soumise à un processus d'évolution dirigée en 3 étapes, afin de générer des enzymes modifiées mieux adaptées aux conditions requises pour des procédés industriels de biocatalyse. En effet, une amélioration de la thermostabilité et de la résistance aux solvants de cette enzyme offrirait une plus grande flexibilité opérationnelle pour son déploiement industriel (Doukyu and Ogino, 2010; Yeoman et al., 2010; Zhang et al., 2012b) en contribuant à i) réduire les possibilités de contamination microbienne, ii) améliorer l'accessibilité de l'enzyme aux substrats en favorisant la désorganisation de certaines matières premières, et enfin iii) créer des conditions thermodynamiquement favorables pour une activité de transfert pour la synthèse de molécules féruloylées (Vafiadi et al., 2006, 2008).

1. Température et solvants pour la biocatalyse enzymatique

Le manque de résistance intrinsèque des enzymes à la température et/ou aux solvants est souvent un frein à leurs utilisations dans des procédés industriels (Liese et al., 2006; Jemli et al., 2016). En effet, la plupart des organismes vivants se développent à des températures modérées et les enzymes sont généralement actives dans des milieux aqueux. La Nature, dans son travail d'évolution, a donc rarement optimisé les enzymes pour travailler dans d'autres conditions même s'il existe dans l'environnement des enzymes thermostables et/ou tolérantes aux solvants. Aujourd'hui, de grands groupes industriels tels que BASF, DSM ou Lonza trouvent, grâce aux enzymes actives en présence de solvants, une opportunité pour produire industriellement des composés (aspartame, acides aminés, médicaments) d'une manière à la fois viable économiquement et plus respectueuse de l'environnement, car même si la présence de solvants à retraiter n'est pas toujours évitable, cet inconvénient représente un moindre mal en comparaison aux procédés chimiques alternatifs (Schmid et al., 2001). Les enzymes thermostables sont aussi fortement appréciées, dans les nouvelles bioraffineries des produits de la biomasse carbonée par exemple, où la catalyse à haute température est avantageuse pour la bioconversion du xylane lignocellulosique en xylitol (Turner et al., 2007). Pouvoir établir des approches efficaces pour améliorer la résistance des enzymes à la température et/ou aux solvants par ingénierie enzymatique représente donc un enjeu important (Sanchez and Demain, 2011).

a. Enzymes et solvants

Les réactions en milieu organique peuvent être séparées en trois catégories :

- Systèmes [eau + solvant miscible dans l'eau] :

Ces milieux sont souvent utilisés pour augmenter la solubilité de certains substrats et pour favoriser thermodynamiquement les réactions de synthèse enzymatique au dépend de l'hydrolyse (Koudelakova et al., 2013).

- Systèmes [eau + solvant non miscible dans l'eau] :

La réaction est alors biphasique. L'enzyme, généralement soluble dans la phase aqueuse, n'est que peu affectée car les contacts avec le solvant sont limités. L'intérêt est grand si le produit formé se solubilise dans la phase organique, car cela facilite la purification du produit synthétisé. Dans ces conditions, les réactions sont souvent plus lentes, à cause de phénomènes de diffusion entre les phases (Martinek et al., 1981).

- Systèmes « quasi-anhydres » :

On parle de conditions quasi-anhydres quand la teneur en eau dans le milieu réactionnel n'excède pas 1000 ppm. Dans ce cas, les enzymes nécessitent d'être préalablement lyophilisées, immobilisées ou modifiées par ajout de molécules amphiphiles, telles que du polyéthylène glycol pour être solubles dans ces milieux (Inada et al., 1986). Les enzymes lyophilisées présentent souvent une meilleure thermostabilité mais au prix d'une activité nettement plus faible qu'en milieu aqueux. Ces systèmes conviennent pour la synthèse d'esters, de peptides ou la transformation de composés hydrophobes (Khmelnitsky et al., 1988).

L'usage de solvants pour la mise en œuvre de réactions enzymatiques présente des avantages indéniables (Tableau 1), auxquels peut s'ajouter parfois un effet thermodynamiquement favorable dans le cadre précis de réactions de transfert catalysées par des enzymes.

Tableau 1. Principaux avantages de l'utilisation de solvants dans les réactions enzymatiques.

- | |
|--|
| <ul style="list-style-type: none"> • Souvent nécessaire pour solubiliser les substrats ou les produits hydrophobes • Catalyse des réactions non réalisables en milieu aqueux • Limite les contaminations microbiennes • Facilite la purification des produits de réaction en milieu biphasique |
|--|

(Schmid et al., 2001; Serdakowski and Dordick, 2008; Doukyu and Ogino, 2010)

Par exemple, une étude comparative des rendements obtenus, avec ou sans solvant (*n*-hexane), sur la transestérification d'huile de tournesol à haute teneur en acide oléique avec du butanol par des lipases immobilisées de *Rhizomucor miehei* (Lipozyme®), a démontré que les réactions dans ce solvant permettent d'atteindre un rendement de 95 % d'ester de butyle formé contre seulement 65 % sans solvant (Dossat et al., 2002). Ce phénomène n'est toutefois pas systématique, il n'est pas observé avec toutes les enzymes et dépend de la nature du solvant choisi (Soumanou and Bornscheuer, 2003). Un solvant polaire peut également inhiber cet effet, possiblement en déstabilisant la structure tridimensionnelle de l'enzyme par des perturbations de liaisons hydrogènes et d'interactions hydrophobes.

Un aspect négatif de l'utilisation de solvants, outre celui lié à l'impact environnemental, est qu'ils causent très souvent une inactivation, plus ou moins importante des enzymes. La perte d'activité peut avoir des causes multiples qui sont listées dans le tableau 2 mais des solutions existent pour essayer de réduire cette inactivation (Klibanov, 1997).

Tableau 2. Problèmes concernant l'usage d'enzymes en milieu à faible concentration en eau (Klibanov, 1997)

	Solutions envisageables
Limitations des phénomènes de diffusion	Réaliser les réactions sous agitation
Possibilité de changement de conformation de l'enzyme	Utiliser un lyoprotecteur (sucres, PEG...) ou des molécules amphiphiles solubles dans le solvant
Perte de plasticité de l'enzyme (rigidification)	Optimiser l'activité de l'eau (a_w), en utilisant un solvant hydrophobe par exemple
Activité correspondant à un pH suboptimal	Lyophiliser l'enzyme à partir d'une solution correspondant au pH optimal de l'enzyme (mémoire du pH)

Pour augmenter le nombre encore restreint d'enzymes tolérantes aux solvants, différentes méthodes d'ingénierie ont été utilisées qui sont, soit des techniques d'évolution dirigée (Song and Rhee, 2001; Takahashi, 2005; Zumárraga et al., 2007; Kawata and Ogino, 2009), soit des approches rationnelles ou semi-rationnelles (Takagi et al., 2000; Badoei-Dalfard et al., 2010; Park et al., 2012) et qui ont permis des améliorations significatives. Ainsi, l'activité de la lipase T6 de *Geobacillus stearothermophilus* a été améliorée en présence de méthanol, solvant choisi car il est utilisé comme accepteur pour la synthèse de biodiesel, en développant une double approche, aléatoire d'abord via une stratégie d'évolution dirigée reposant sur la PCR à erreurs et rationnelle ensuite, avec des résultats complémentaires. L'approche rationnelle, par la comparaison de séquences similaires à la protéine ciblée et la création d'une structure modèle issue d'une séquence consensus, a permis d'identifier 25 mutations pouvant potentiellement améliorer la stabilité de l'enzyme en présence de solvants. Une librairie contenant ces mutations isolées ou en combinaison a alors été générée et criblée en parallèle de la banque de variants obtenus par mutagénèse aléatoire. Sur les trois mutations identifiées comme apportant les meilleurs gains de résistance au méthanol, une provient de la banque semi-rationnelle (A269T), une de la banque aléatoire (Q185L) et la troisième est détectée dans les 2 banques (H86Y). Le triple mutant n'est pas produit dans cette étude mais les doubles mutants Q185L/A269T et H86Y/A269T montrent des effets additifs qui confèrent à ces mutants des temps de demi-vie ($t_{1/2}$) en présence de 70 % de méthanol de 139 et 258 min respectivement, contre 3,9 min pour l'enzyme sauvage (Dror et al., 2014). Ces résultats sont d'autant plus intéressants que ces mutants conservent les

mêmes paramètres cinétiques que l'enzyme sauvage et que la mutation A269T confère un gain de thermostabilité de 3,4 °C sur la température de fusion (T_m).

b. Enzymes et température

Concernant la température, la Nature peut s'adapter à des conditions extrêmes et parmi les cas les plus impressionnants de protéines naturellement thermorésistantes, on trouve la protéine rubredoxine de *Pyrococcus furiosus*, archée extrémophile issue de sédiments hydrothermaux, qui possède une température de fusion (T_m) de près de 200 °C (Hiller et al., 1997). A l'opposé, certains organismes peuvent se développer ou survivre à des températures inférieures à -20 °C dans des saumures ou de la glace (Bakermans and Skidmore, 2011; Rodrigues et al., 2012) et les protéines de ces organismes psychrophiles présentent à leur tour des propriétés intéressantes car elles se caractérisent par un rendement et une stéréospécificité élevés à basse température. Ainsi, en 2014, des gènes codant pour des cytochromes P450 catalysant l'hydroxylation d'alcanes ont été identifiés dans des génomes d'organismes psychrophiles. Ces enzymes ont la capacité de convertir des *n*-alcanes en alcools plus labiles, donc possédant une meilleure biodégradabilité. Ces enzymes adaptées au froid pourraient donc participer à la bioremédiation d'hydrocarbures issus de déversements dans des environnements à basse température tel que l'Arctique (Bowman and Deming, 2014).

Depuis des années, des efforts considérables ont été faits pour analyser les déterminants moléculaires qui confèrent une stabilité thermique aux protéines hyper-thermophiles (Jaenicke, 1998). Ce qui fait principalement obstacle à l'établissement de règles simples et déterministes est que cette propriété résulte le plus souvent d'un équilibre très subtil entre différents types d'interactions stabilisatrices. De très fines modifications structurales peuvent suffire pour qu'une enzyme puisse faire face à des conditions extrêmes (Turner et al., 2007). En fin de compte, en dépit de plusieurs études soit axées sur la comparaison statistique de séquences primaires entre enzymes mésophiles et thermophiles (Chakravarty and Varadarajan, 2002; Panja et al., 2015), soit focalisées sur l'influence des interactions non-covalentes de type ponts salins par exemple dans la thermostabilité (Karshikoff and Ladenstein, 2001; Chan et al., 2011) ou encore sur la rigidification des zones de flexibilités (Yu and Huang, 2014), aucune stratégie applicable de manière rationnelle et universelle se basant sur de simples échanges d'acides aminés, n'a été établie à ce jour (Karshikoff et al., 2015). A ce titre, plusieurs stratégies d'ingénierie, rationnelle ou aléatoire, continuent d'être déployées, avec des résultats souvent équivalents (Tableau 3).

Tableau 3. Stratégies d'ingénierie pour l'augmentation de la thermostabilité enzymatique

	Exemples
Stabilisation de la structure 3D	L'ajout de ponts disulfures permet d'augmenter la température de fusion des enzymes (Matsumura et al., 1989)
Stabilisation du site actif	Mutation de résidus mobiles autour du site actif (Active Center Stabilization, ACS) (Zhang et al., 2016))
Stabilisation des régions terminales	Extrémité N-terminale (Mandrich et al., 2005; Sun et al., 2005; Zhang et al., 2010); extrémité C-terminale : (Wang et al., 2007)
Modifications post-traductionnelles	Cette étude comparative montre l'importance pour la thermostabilité de conserver la glycosylation lors d'une expression hétérologue (Fonseca-Maldonado et al., 2013)
Limiter la mobilité des résidus ayant les plus hautes valeurs de B-facteur.	L'utilisation de mutagénèse à saturation itérative (ISM) sur ces résidus a permis d'accroître la stabilité thermique de certaines enzymes (Reetz et al., 2006).

D'après Turner et al., 2007

Pour comprendre pourquoi l'amélioration de la thermostabilité concentre tant d'intérêts, il faut savoir qu'en moyenne, la vitesse de réaction des enzymes double pour une augmentation de température de 10 °C (Elias et al., 2014) et réaliser une réaction à haute température permet donc de réduire considérablement les temps de réactions et la quantité d'enzymes nécessaire, entre autres (Tableau 4).

Tableau 4. Principaux avantages à l'utilisation de températures élevées lors de réactions enzymatiques (Zamost et al., 1991; Habbeche et al., 2014)

<ul style="list-style-type: none"> • Activité enzymatique accrue • Mélange homogène plus facile • Meilleure solubilité du substrat • Taux de transfert de masse élevé • Risque de contamination réduit • Meilleure résistance aux agents dénaturants
--

Il existe donc une tendance croissante chez les industriels à vouloir utiliser des organismes microbiens et des enzymes thermophiles (Haki, 2003; Turner et al., 2007; Sanchez and Demain, 2011).

c. Lien entre thermostabilité et résistance aux solvants ?

Une des premières études qui suggèrent une corrélation entre thermostabilité et résistance aux solvants repose sur la comparaison de l'activité d'enzymes provenant de divers organismes mésophiles, thermophiles ou hyperthermophiles, testées dans des conditions identiques de température dans un milieu biphasique eau-solvant organique. En effet, la plupart des enzymes provenant d'organismes thermophiles et hyperthermophiles sont moins affectées par la présence de solvants, sans que cela ne soit confirmé pour toutes les enzymes testées (Owusu and Cowan, 1989). Depuis, il a été constaté que certains mutants ayant été sélectionnés pour une meilleure thermostabilité, obtenus soit de façon semi-rationnelle (Vazquez-Figueroa et al., 2008) soit par epPCR (Batra and Mishra, 2013), possèdent également une meilleure stabilité en présence de solvants organiques. Par exemple, les mutants thermostables de la lipase de *Bacillus subtilis* obtenus par ISM (Reetz et al., 2006; Reetz and Carballeira, 2007) présentent une stabilité et une activité résiduelle en présence d'acétonitrile, de diméthylsulfoxyde ou de diméthylformamide supérieures à l'enzyme sauvage. De plus, ces 2 gains sont proportionnels, c'est-à-dire que pour les 5 mutants testés, plus leur thermostabilité est importante, plus leur résistance aux solvants l'est aussi (Reetz et al., 2010). De manière générale, la corrélation entre ces deux propriétés n'est pas systématiquement constatée et les paramètres favorisant à la fois la thermorésistance et la résistance aux solvants ne sont pas formellement identifiés. A ce jour, aucune étude n'a visé la co-évolution simultanée d'enzymes pour cette double caractéristique de thermostabilité et de résistance aux solvants.

2. Les feruloyl estérases dans les biotechnologies

Les féruloyl estérases (FAE) (EC 3.1.1.73) sont des hydrolases à sérine, appartenant à la sous-famille des carboxylestérases qui catalysent l'hydrolyse des liaisons esters entre motif acide férulique (acide hydroxycinnamique) et polysaccharides pariétaux tels des hémicelluloses. Particulièrement présentes dans le monde microbien, où elles participent à la dégradation de la biomasse végétale (Faulds, 2010), on peut également les trouver dans les plantes où elles jouent un rôle dans le développement de la paroi des cellules végétales, ainsi que dans les cellules intestinales humaines et des ruminants pour permettre la digestion des fibres alimentaires végétales (Kern et al., 2003). Les Fae dérivent phylogénétiquement de familles très diversifiées d'enzymes. On retrouve par exemple comme enzymes ancestrales de certaines Fae, des tannases, des acétylxylane estérases, des lipases, des choline esterases ou même des xylanases (GH10 et GH11) et α -L-rhamnosidases (GH78) (Dilokpimol et al., 2016).

Cette diversité peut expliquer la grande variété de spécificité de substrat constatée entre ces différentes familles.

Les Fae utilisées dans les biotechnologies proviennent essentiellement d'organismes fongiques (Topakas et al., 2007). Leur capacité à libérer l'acide férulique et d'autres acides hydroxycinnamiques de la biomasse végétale accroît l'accessibilité des enzymes hémicellulolytiques aux substrats (Topakas et al., 2007). Il n'est donc pas étonnant de retrouver ces biocatalyseurs dans des applications telles que le prétraitement de la biomasse pour la production de biocarburants, ou l'industrie des pâtes à papiers pour leur contribution à la déconstruction de la biomasse végétale (Gopalan et al., 2015). Le deuxième intérêt des Fae réside dans leur capacité à transférer certains acides hydroxycinnamiques dont l'acide férulique sur des molécules acceptrices (Figure 1).

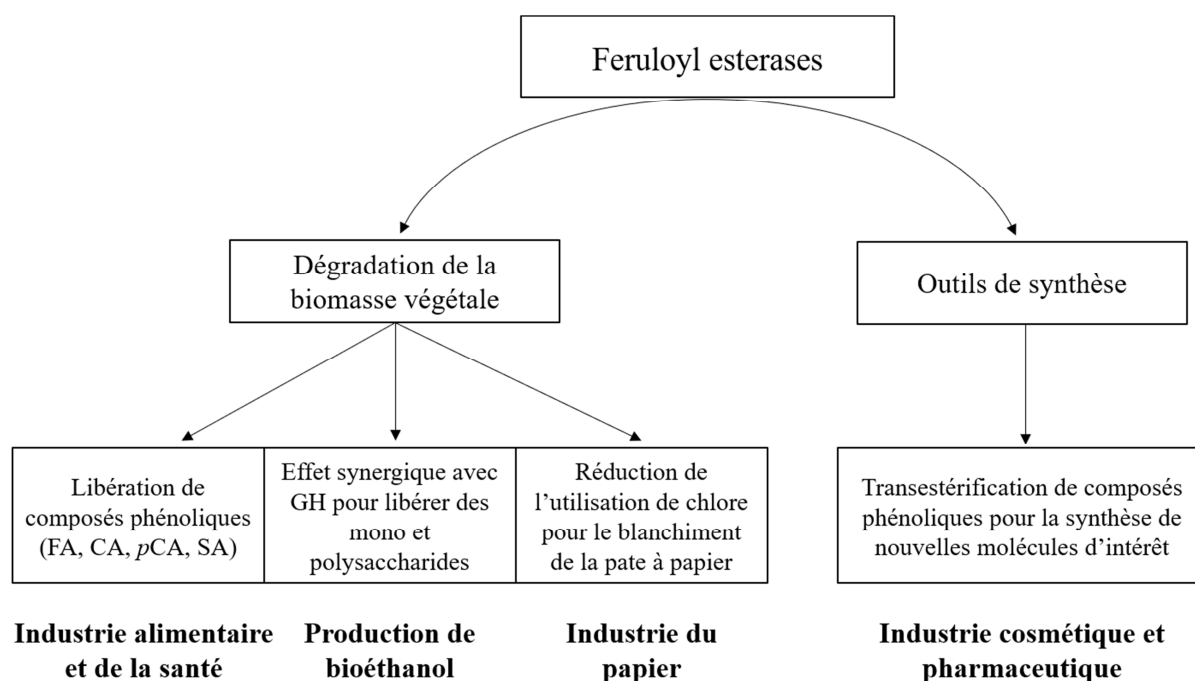


Figure 1 : **Différentes applications impliquant les feruloyl estérases.** D'après (Benoit et al., 2008). FA = acide férulique, CA = acide caféique, pCA = acide p-coumarique, SA = acide sinapique

En effet, les Fae possèdent toutes un mécanisme catalytique semblable aux protéases à serines, reposant sur une triade catalytique Ser-His-Asp/Glu conservée (Kraut, 1977) (Figure 2) qui, outre l'hydrolyse, peut dans certaines conditions mener à la transestérfication de molécules.

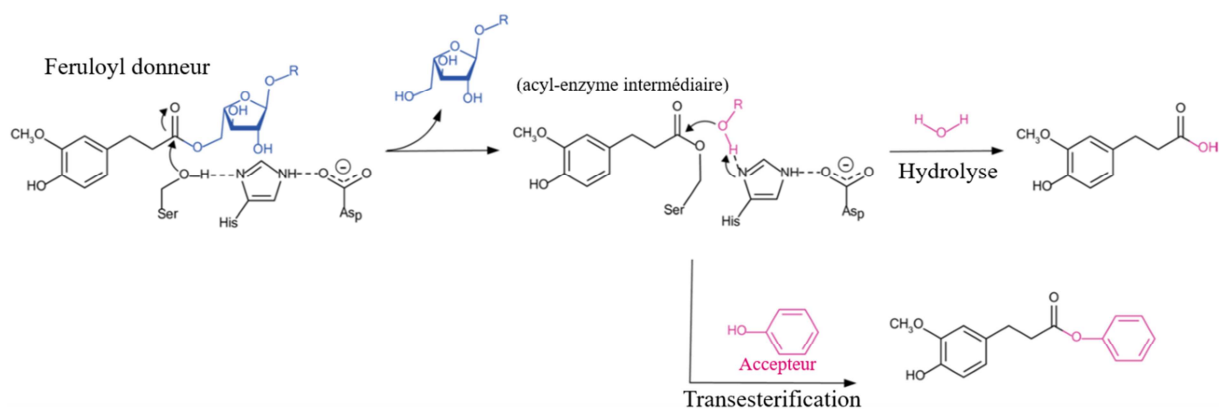


Figure 2 : **Mécanisme catalytique des féruloyle estérases.** La première étape génère un intermédiaire covalent acide férulique-enzyme et la deuxième étape finalise une réaction hydrolytique par ajout d'une molécule d'eau ou si c'est une réaction de transestérification par liaison de l'acide férulique sur une fonction hydroxyle d'une molécule acceptrice (Dilokpimol et al., 2016).

L'optimisation de l'activité de synthèse des Fae, comme pour les lipases, est généralement obtenue au travers de conditions de réactions thermodynamiquement plus favorables aux réactions de synthèses (Mastihubová et al., 2006; Vafiadi et al., 2007). Les conditions de réactions ayant potentiellement un impact sur le ratio T/H des hydrolases en général sont : l'utilisation de solvants organiques (Hansson and Adlercreutz, 2001) ou de liquides ioniques (Lang et al., 2006), l'activité de l'eau (Petersson et al., 2007), le pH (Lundemo et al., 2016), la température (Hansson and Adlercreutz, 2001), la concentration des substrats donneur et accepteur (Lang et al., 2006; Mangas-Sánchez and Adlercreutz, 2015) ou encore l'immobilisation des enzymes (Salis et al., 2008).

a. Mesure d'activité des féruloyle estérases

Jusqu'à récemment, la détection et la caractérisation des activités des Fae reposaient principalement sur l'utilisation de substrats d'origine naturelle ou synthétique, tels que des oligosaccharides estérifiés par des acides hydroxycinnamiques ou de l'acide férulique estérifié par un groupement alkyle (Donaghy et al., 1998), couplée à des méthodologies de spectrophotométrie UV (Yue et al., 2009), de chromatographie sur couche mince (CCM) (Mabinya et al., 2006) ou de chromatographie liquide à haute pression (HPLC) (Haase-Aschoff et al., 2013; Wang et al., 2004), des méthodes peu sensibles ou à faible cadence. Récemment, plusieurs nouveaux substrats, plus faciles d'utilisation, ont été reportés dans la littérature, la plupart contenant une partie 5-*O*-(*trans*-feruloyl)- α -L-arabinofuranoside liée à un chromogène (Marmuse et al., 2008; Mastihubová and Biely, 2010) ou un groupement fluorescent (Zhang et al., 2012a). Leur utilisation s'appuie sur une réaction en tandem impliquant premièrement le clivage de la liaison ester entre le résidu L-arabinosyle et l'acide férulique par une Fae suivi de la libération quasi-concomitante du groupement chromogène/fluorescent lié au motif L-arabinosyle par une α -L-arabinofuranosidase. Cette deuxième réaction doit être correctement optimisée pour éviter

qu'elle ne devienne limitante et fausse le suivi de l'activité des Fae (Biely et al., 2002). Bien que ces substrats soient utiles, particulièrement pour la détection d'activité des Fae en milieu solide, ils ne sont pas adaptés aux approches de criblage à haut débit où la simplicité des protocoles expérimentaux est primordiale, à travers la mise en place d'une détection directe par exemple (Zhang et al., 2012a). En outre, l'examen de la structure d'un mutant de la féruloyle estérase A d'*Aspergillus niger* (*AnFaeA*) complexé avec un arabinoxylo-triose féruloilé révèle que cette enzyme n'interagit pas avec la partie glycosidique du substrat, ce qui suggère que les substrats artificiels ne nécessitent pas la présence du motif α -L-arabinofuranosyle pour être correctement reconnus par les Fae (Faulds et al., 2005).

b. La féruloyle estérase A d'*Aspergillus niger* (*AnFaeA*)

L'*AnFaeA* (Faulds and Williamson, 1994) appartient à la famille A dans le classement proposé en 2004 par Crepin et al., qui compte quatre familles (A à D). Les enzymes y sont classées en fonction de la nature des substrats reconnus et sur l'identité de séquence primaire des enzymes (Crepin et al., 2004). L'*AnFaeA* est une hydrolase possédant un repliement de type α/β (Ollis et al., 1992) et présente de fortes similitudes structurales avec les lipases fongiques (McAuley et al., 2004). Cette enzyme appartient à la famille des Lipases3 dans la base de données ESTHER qui est dédiée à la classification des protéines ou des domaines protéiques appartenant à la superfamille des hydrolases α/β (<http://bioweb.ensam.inra.fr/ESTHER/general?what=index>) (Lenfant et al., 2013). Sa structure cristallographique et la construction d'une série de mutants ciblés au niveau du site actif ont permis d'identifier la triade catalytique de cette enzyme (Ser133-Asp194-His247). La crevasse catalytique est surplombée d'une hélice alpha (résidus 68 à 80) et d'une boucle (résidus 226 à 244) mobiles, qui comme sur les lipases permet une conformation ouverte ou fermée du site actif. Un nombre plus élevé de résidus polaires en surface autour du site actif et un site de *N*-glycosylation au niveau du couvercle confèrent une nature plus hydrophile au site actif de l'*AnFaeA* que pour les lipases de la famille des Lipases3 (Hermoso et al., 2004; McAuley et al., 2004).

3. Stratégie utilisée pour l'évolution de *AnFaeA*

Les résultats décrits dans ce chapitre s'inscrivent dans le plan de travail du projet OPTIBIOCAT, un programme européen focalisé sur l'étude des biocatalyseurs de type estérases (glucuronoyle et féruloyle estérases), visant à mieux connaître ces enzymes et à développer leurs usages en mettant en œuvre des procédés innovants pour la production d'esters gras phénoliques et de sucres esters dont les propriétés antioxydantes sont recherchées en formulation cosmétique mais aussi dans l'industrie alimentaire. Un lot de travail y est dédié à l'ingénierie de plusieurs estérases, soit rationnelle soit aléatoire. C'est sur ce dernier point que notre équipe a été sollicitée pour améliorer les capacités de résistance à la température et aux solvants de l'*AnFaeA*.

Un premier travail a concerné la mise au point des conditions de production recombinante de notre protéine-cible. En effet, la production d'une protéine fonctionnelle et en quantité suffisante pour sa détection est un pré-requis indispensable à la mise en œuvre de stratégies haut-débit d'ingénierie aléatoire (Pourmir and Johannes, 2012). Idéalement, *E. coli* est l'hôte de choix pour de tels travaux, des décennies d'étude avec cet organisme modèle ayant permis d'aboutir à un catalogue de souches et de plasmides très développé. En 2006, Benoit et al., ont démontré qu'il était possible d'obtenir l'*AnFAEA* sous forme active à partir de culture d'*E. coli*, mais le protocole développé nécessitait une laborieuse étape de dénaturation et renaturation des corps d'inclusion où l'enzyme était concentrée (Benoit et al., 2007). Afin d'essayer de bénéficier de la facilité de mise en œuvre d'*E. coli*, différents systèmes d'expression, sensés favoriser la production d'enzyme soluble (étiquette thioredoxine, production périplasmique voire même extracellulaire) ont été testés mais la protéine n'a jamais été obtenue sous forme active. Naturellement, la levure *Yarrowia lipolytica* a alors été envisagée comme hôte alternatif, principalement parce que, tout comme *Pichia pastoris*, ce micro-organisme permet les glycosylations post-traductionnelles qui semblent indispensable à la structuration 3D *in vivo* correcte de l'enzyme (Benoit et al., 2006) mais bénéficie en plus d'une optimisation de la souche spécifiquement pour la réalisation de projet d'évolution dirigée (Bordes et al., 2007; Duquesne et al., 2012).

Ensuite, pour faciliter la mesure des activités Fae, nous avons synthétisé et caractérisé plusieurs substrats comportant des groupements partants chromogènes tels que des dérivés indolyliques (5-bromo-4-chloro indol-3-yle ou le 5-bromo indol-3-yle) ou des dérivés à base de 4-nitrocatechol (4NTC), couplés à des acides féruliques ou *p*-coumariques. Nous avons démontré que ces substrats chromogènes peuvent servir efficacement soit en milieu solide (dosage qualitatif) soit en solution (dosage quantitatif), pour la détection et la mesure d'activité des Fae, et sont adaptés même à des applications à haut-débit. Ces travaux sont explicités dans deux publications (Gherbovet et al., 2016 et Gherbovet et al., 2017 soumis) présentées en annexe.

C'est sur la base de ces deux étapes préalables qu'un travail d'évolution dirigée de la protéine *AnFaeA* a pu être entamé pour générer des variants plus résistants à la température et/ou à la présence de solvants.

Références :

- Badoei-Dalfard, A., Khajeh, K., Asghari, S.M., Ranjbar, B., Karbalaeei-Heidari, H.R., 2010. Enhanced activity and stability in the presence of organic solvents by increased active site polarity and stabilization of a surface loop in a metalloprotease. *J. Biochem.* 148, 231–238.
- Bakermans, C., Skidmore, M.L., 2011. Microbial metabolism in ice and brine at -5°C . *Environ. Microbiol.* 13, 2269–2278.
- Batra, J., Mishra, S., 2013. Organic solvent tolerance and thermostability of a β -glucosidase co-engineered by random mutagenesis. *J. Mol. Catal. B Enzym.* 96, 61–66.
- Benoit, I., Asther, M., Sulzenbacher, G., Record, E., Marmuse, L., Parsiegla, G., Gimbert, I., Asther, M., Bignon, C., 2006. Respective importance of protein folding and glycosylation in the thermal stability of recombinant feruloyl esterase A. *FEBS Lett.* 580, 5815–5821.
- Benoit, I., Coutard, B., Oubelaid, R., Asther, M., Bignon, C., 2007. Expression in *Escherichia coli*, refolding and crystallization of *Aspergillus niger* feruloyl esterase A using a serial factorial approach. *Protein Expr. Purif.* 55, 166–174.
- Benoit, I., Danchin, E.G.J., Bleichrodt, R.-J., de Vries, R.P., 2008. Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnol. Lett.* 30, 387–396.
- Biely, P., Mastihubová, M., van Zyl, W.H., Prior, B.A., 2002. Differentiation of feruloyl esterases on synthetic substrates in α -arabinofuranosidase-coupled and ultraviolet-spectrophotometric assays. *Anal. Biochem.* 311, 68–75.
- Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* 70, 493–502.
- Bowman, J.S., Deming, J.W., 2014. Alkane hydroxylase genes in psychrophile genomes and the potential for cold active catalysis. *BMC Genomics* 15, 1120.
- Chakravarty, S., Varadarajan, R., 2002. Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study. *Biochemistry* 41, 8152–8161.
- Chan, C.-H., Yu, T.-H., Wong, K.-B., 2011. Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS ONE* 6, e21624.
- Crepin, V.F., Faulds, C.B., Connerton, I.F., 2004. Functional classification of the microbial feruloyl esterases. *Appl. Microbiol. Biotechnol.* 63, 647–652.
- Dilokpimol, A., Mäkelä, M.R., Aguilar-Pontes, M.V., Benoit-Gelber, I., Hildén, K.S., de Vries, R.P., 2016. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol. Biofuels* 9.
- Donaghy, J., Kelly, P.F., McKay, A.M., 1998. Detection of ferulic acid esterase production by *Bacillus* spp. and *Lactobacilli*. *Appl. Microbiol. Biotechnol.* 50, 257–260.
- Dossat, V., Combes, D., Marty, A., 2002. Lipase-catalysed transesterification of high oleic sunflower oil. *Enzyme Microb. Technol.* 30, 90–94.
- Doukyu, N., Ogino, H., 2010. Organic solvent-tolerant enzymes. *Biochem. Eng. J.* 48, 270–282.
- Dror, A., Shemesh, E., Dayan, N., Fishman, A., 2014. Protein engineering by random mutagenesis and structure-guided consensus of *Geobacillus stearothermophilus* lipase T6 for enhanced stability in methanol. *Appl. Environ. Microbiol.* 80, 1515–1527.
- Duquesne, S., Bordes, F., Fudalej, F., Nicaud, J.-M., Marty, A., 2012. The yeast *Yarrowia lipolytica* as a generic tool for molecular evolution of enzymes, in: Sandoval, G. (Ed.), *Lipases and Phospholipases*. Humana Press, Totowa, NJ, pp. 301–312.

- Elias, M., Wieczorek, G., Rosenne, S., Tawfik, D.S., 2014. The universality of enzymatic rate-temperature dependency. *Trends Biochem. Sci.* 39, 1–7.
- Faulds, C.B., 2010. What can feruloyl esterases do for us? *Phytochem. Rev.* 9, 121–132.
- Faulds, C.B., Molina, R., Gonzalez, R., Husband, F., Juge, N., Sanz-Aparicio, J., Hermoso, J.A., 2005. Probing the determinants of substrate specificity of a feruloyl esterase, *AnFaeA*, from *Aspergillus niger*. *FEBS J.* 272, 4362–4371.
- Faulds, C.B., Williamson, G., 1994. Purification and characterization of a ferulic acid esterase (FAE-III) from *Aspergillus niger*: specificity for the phenolic moiety and binding to microcrystalline cellulose. *Microbiology* 140, 779–787.
- Fonseca-Maldonado, R., Vieira, D.S., Alpointi, J.S., Bonneil, E., Thibault, P., Ward, R.J., 2013. Engineering the pattern of protein glycosylation modulates the thermostability of a GH11 Xylanase. *J. Biol. Chem.* 288, 25522–25534.
- Gherbovet, O., Fauré, R., Ferreira, F., Durand, J., Ragon, M., Hostyn, G., Record, E., Bozonnet, S., O’Donohue, M.J., 2016. Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities. *J. Mol. Catal. B Enzym.* 126, 24–31.
- Gopalan, N., Rodríguez-Duran, L.V., Saucedo-Castaneda, G., Nampoothiri, K.M., 2015. Review on technological and scientific aspects of feruloyl esterases: A versatile enzyme for biorefining of biomass. *Bioresour. Technol.* 193, 534–544.
- Haase-Aschoff, P., Linke, D., Berger, R.G., 2013. Detection of feruloyl- and cinnamoyl esterases from *basidiomycetes* in the presence of interfering laccase. *Bioresour. Technol.* 130, 231–238.
- Habbeche, A., Saoudi, B., Jaouadi, B., Haberra, S., Kerouaz, B., Boudelaa, M., Badis, A., Ladjama, A., 2014. Purification and biochemical characterization of a detergent-stable keratinase from a newly thermophilic actinomycete *Actinomadura keratinilytica* strain Cpt29 isolated from poultry compost. *J. Biosci. Bioeng.* 117, 413–421.
- Haki, G., Rakshit, S.K., 2003. Developments in industrially important thermostable enzymes: a review. *Bioresour. Technol.* 89, 17–34.
- Hansson, T., Adlercreutz, P., 2001. Enhanced transglucosylation/hydrolysis ratio of mutants of *Pyrococcus furiosus* β -glucosidase: Effects of donor concentration, water content, and temperature on activity and selectivity in hexanol. *Biotechnol. Bioeng.* 75, 656–665.
- Hermoso, J.A., Sanz-Aparicio, J., Molina, R., Juge, N., González, R., Faulds, C.B., 2004. The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *J. Mol. Biol.* 338, 495–506.
- Hiller, R., Zhou, Z.H., Adams, M.W., Englander, S.W., 1997. Stability and dynamics in a hyperthermophilic protein with melting temperature close to 200 degrees C. *Proc. Natl. Acad. Sci. U. S. A.* 94, 11329–11332.
- Inada, Y., Takahashi, K., Yoshimoto, T., Ajima, A., Matsushima, A., Saito, Y., 1986. Application of polyethylene glycol-modified enzymes in biotechnological processes: organic solvent-soluble enzymes. *Trends Biotechnol.* 4, 190–194.
- Jaenicke, R., 1998. What ultrastable globular proteins teach us about protein stabilization. *Biochem.* 63, 312–321.
- Jemli, S., Ayadi-Zouari, D., Hlima, H.B., Bejar, S., 2016. Biocatalysts: application and engineering for industrial purposes. *Crit. Rev. Biotechnol.* 36, 246–258.
- Karshikoff, A., Ladenstein, R., 2001. Ion pairs and the thermotolerance of proteins from hyperthermophiles: a “traffic rule” for hot roads. *Trends Biochem. Sci.* 26, 550–556.
- Karshikoff, A., Nilsson, L., Ladenstein, R., 2015. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J.* 282, 3899–3917.

- Kawata, T., Ogino, H., 2009. Enhancement of the organic solvent-stability of the LST-03 lipase by directed evolution. *Biotechnol. Prog.* NA-NA. 25, 1605–1611
- Kern, S.M., Bennett, R.N., Needs, P.W., Mellon, F.A., Kroon, P.A., Garcia-Conesa, M.-T., 2003. Characterization of metabolites of hydroxycinnamates in the in vitro model of human small intestinal epithelium Caco-2 Cells. *J. Agric. Food Chem.* 51, 7884–7891.
- Khmelnitsky, Y.L., Levashov, A.V., Klyachko, N.L., Martinek, K., 1988. Engineering biocatalytic systems in organic media with low water content. *Enzyme Microb. Technol.* 10, 710–724.
- Klibanov, A.M., 1997. Why are enzymes less active in organic solvents than in water? *Trends Biotechnol.* 15, 97–101.
- Koudelakova, T., Chaloupkova, R., Brezovsky, J., Prokop, Z., Sebestova, E., Hesseler, M., Khabiri, M., Plevaka, M., Kulik, D., Kuta Smatanova, I., Rezacova, P., Ettrich, R., Bornscheuer, U.T., Damborsky, J., 2013. Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew. Chem. Int. Ed.* 52, 1959–1963.
- Kraut, J., 1977. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* 46, 331–358.
- Lang, M., Kamrat, T., Nidetzky, B., 2006. Influence of ionic liquid cosolvent on transgalactosylation reactions catalyzed by thermostable β -glycosylhydrolase CelB from *Pyrococcus Furiosus*. *Biotechnol. Bioeng.* 95, 1093–1100.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2013. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* 41, D423–D429.
- Liese, A., Seelbach, K., Wandrey, C. (Eds.), 2006. *Industrial Biotransformations*; in: *Industrial biotransformations*. Wiley-VCH.
- Lundemo, P., Karlsson, E.N., Adlercreutz, P., 2016. Eliminating hydrolytic activity without affecting the transglycosylation of a GH1 β -glucosidase. *Appl. Microbiol. Biotechnol.*
- Mabinya, L.V., Mafunga, T., Brand, J.M., 2006. Determination of ferulic acid and related compounds by thin layer chromatography. *Afr. J. Biotechnol.* 5, 1271–1273.
- Mandrich, L., Merone, L., Pezzullo, M., Cipolla, L., Nicotra, F., Rossi, M., Manco, G., 2005. Role of the N terminus in enzyme activity, stability and specificity in thermophilic esterases belonging to the HSL family. *J. Mol. Biol.* 345, 501–512.
- Mangas-Sánchez, J., Adlercreutz, P., 2015. Enzymatic preparation of oligosaccharides by transglycosylation: A comparative study of glucosidases. *J. Mol. Catal. B Enzym.* 122, 51–55.
- Marmuse, L., Asther, M., Fabre, E., Navarro, D., Lesage-Meessen, L., Asther, M., O'Donohue, M., Fort, S., Driguez, H., 2008. New chromogenic substrates for feruloyl esterases. *Org. Biomol. Chem.* 6, 1208–1214.
- Martinek, K., Semenov, A.N., Berezin, I.V., 1981. Enzymatic synthesis in biphasic aqueous-organic systems. I. Chemical equilibrium shift. *Biochim. Biophys. Acta BBA - Enzymol.* 658, 76–89.
- Mastihubová, M., Biely, P., 2010. Preparation of regioselectively feruloylated *p*-nitrophenyl α -L-arabinofuranosides and β -D-xylopyranosides—convenient substrates for study of feruloyl esterase specificity. *Carbohydr. Res.* 345, 1094–1098.
- Mastihubová, M., Mastihuba, V., Bilaničová, D., Boreková, M., 2006. Commercial enzyme preparations catalyse feruloylation of glycosides. *J. Mol. Catal. B Enzym.* 38, 54–57.
- Matsumura, M., Becktel, W.J., Levitt, M., Matthews, B.W., 1989. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6562–6566.
- McAuley, K.E., Svendsen, A., Patkar, S.A., Wilson, K.S., 2004. Structure of a feruloyl esterase from *Aspergillus niger*. *Acta Crystallogr. D Biol. Crystallogr.* 60, 878–887.

- Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., others, 1992. The α/β hydrolase fold. *Protein Eng.* 5, 197–211.
- Owusu, R.K., Cowan, D.A., 1989. Correlation between microbial protein thermostability and resistance to denaturation in aqueous: organic solvent two-phase systems. *Enzyme Microb. Technol.* 11, 568–574.
- Panja, A.S., Bandopadhyay, B., Maiti, S., 2015. Protein thermostability is owing to their preferences to non-polar smaller volume amino acids, variations in residual physico-chemical properties and more salt-bridges. *PLOS ONE* 10, e0131495.
- Park, H.J., Joo, J.C., Park, K., Yoo, Y.J., 2012. Stabilization of *Candida antarctica* lipase B in hydrophilic organic solvent by rational design of hydrogen bond. *Biotechnol. Bioprocess Eng.* 17, 722–728.
- Petersson, A.E.V., Adlercreutz, P., Mattiasson, B., 2007. A water activity control system for enzymatic reactions in organic media. *Biotechnol. Bioeng.* 97, 235–241.
- Pourmir, A., Johannes, T.W., 2012. Directed evolution: Selection of the host organism. *Comput. Struct. Biotechnol. J.* 2, 1–7.
- Reetz, M.T., Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* 2, 891–903.
- Reetz, M.T., Carballeira, J.D., Vogel, A., 2006. Iterative Saturation Mutagenesis on the basis of B-factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.* 45, 7745–7751.
- Reetz, M.T., Soni, P., Fernández, L., Gumulya, Y., Carballeira, J.D., 2010. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem. Commun.* 46, 8657.
- Rodrigues, D.F., Bergholz, P.W., Tiedje, J.M., Ayala-del-Río, H.L., Bakermans, C., Vishnivetskaya, T.A., 2012. Genomic and expression analyses of cold-adapted microorganisms, in: Whyte, L.G., Miller, R.V. (Eds.), *Polar microbiology: Life in a deep freeze*. American Society of Microbiology, pp. 126–155.
- Salis, A., Pinna, M., Monduzzi, M., Solinas, V., 2008. Comparison among immobilised lipases on macroporous polypropylene toward biodiesel synthesis. *J. Mol. Catal. B Enzym.* 54, 19–26.
- Sanchez, S., Demain, A.L., 2011. Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Org. Process Res. Dev.* 15, 224–230.
- Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M., Witholt, B., 2001. Industrial biocatalysis today and tomorrow. *Nature* 409, 258–268.
- Serdakowski, A.L., Dordick, J.S., 2008. Enzyme activation for organic solvents made easy. *Trends Biotechnol.* 26, 48–54.
- Song, J.K., Rhee, J.S., 2001. Enhancement of stability and activity of phospholipase A1 in organic solvents by directed evolution. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* 1547, 370–378.
- Soumanou, M.M., Bornscheuer, U.T., 2003. Improvement in lipase-catalyzed synthesis of fatty acid methyl esters from sunflower oil. *Enzyme Microb. Technol.* 33, 97–103.
- Sun, J.-Y., Liu, M.-Q., Xu, Y.-L., Xu, Z.-R., Pan, L., Gao, H., 2005. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. *Protein Expr. Purif.* 42, 122–130.
- Takagi, H., Hirai, K., Maeda, Y., Matsuzawa, H., Nakamori, S., 2000. Engineering subtilisin E for enhanced stability and activity in polar organic solvents. *J. Biochem.* 127, 617–625.

- Takahashi, T., 2005. Molecular cloning of the gene encoding vibrio metalloproteinase vimelysin and isolation of a mutant with high stability in organic solvents. *J. Biochem.* 138, 701–710.
- Topakas, E., Vafiadi, C., Christakopoulos, P., 2007. Microbial production, characterization and applications of feruloyl esterases. *Process Biochem.* 42, 497–509.
- Turner, P., Mamo, G., Karlsson, E., 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. *Microb. Cell Factories* 6, 9.
- Vafiadi, C., Topakas, E., Alderwick, L.J., Besra, G.S., Christakopoulos, P., 2007. Chemoenzymatic synthesis of feruloyl D-arabinose as a potential anti-mycobacterial agent. *Biotechnol. Lett.* 29, 1771–1774.
- Vafiadi, C., Topakas, E., Alissandratos, A., Faulds, C.B., Christakopoulos, P., 2008. Enzymatic synthesis of butyl hydroxycinnamates and their inhibitory effects on LDL-oxidation. *J. Biotechnol.* 133, 497–504.
- Vafiadi, C., Topakas, E., Christakopoulos, P., Faulds, C.B., 2006. The feruloyl esterase system of *Talaromyces stipitatus*: Determining the hydrolytic and synthetic specificity of *TsFaeC*. *J. Biotechnol.* 125, 210–221.
- Vazquez-Figueroa, E., Yeh, V., Broering, J.M., Chaparro-Riggers, J.F., Bommaris, A.S., 2008. Thermostable variants constructed via the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media. *Protein Eng. Des. Sel.* 21, 673–680.
- Wang, J.-H., Tsai, M.-Y., Chen, J.-J., Lee, G.-C., Shaw, J.-F., 2007. Role of the C-terminal domain of *Thermus thermophilus* trehalose synthase in the thermophilicity, thermostability, and efficient Production of Trehalose. *J. Agric. Food Chem.* 55, 3435–3443.
- Wang, X., Geng, X., Egashira, Y., Sanada, H., 2004. Purification and characterization of a feruloyl esterase from the intestinal bacterium *Lactobacillus acidophilus*. *Appl. Environ. Microbiol.* 70, 2367–2372.
- Yeoman, C.J., Han, Y., Dodd, D., Schroeder, C.M., Mackie, R.I., Cann, I.K.O., 2010. Thermostable enzymes as biocatalysts in the biofuel industry, in: *Advances in Applied Microbiology*. Elsevier, pp. 1–55.
- Yu, H., Huang, H., 2014. Engineering proteins for thermostability through rigidifying flexible sites. *Biotechnol. Adv.* 32, 308–315.
- Yue, Q., Yang, H.J., Li, D.H., Wang, J.Q., 2009. A comparison of HPLC and spectrophotometrical methods to determine the activity of ferulic acid esterase in commercial enzyme products and rumen contents of steers. *Anim. Feed Sci. Technol.* 153, 169–177.
- Zamost, B.L., Nielsen, H.K., Starnes, R.L., 1991. Thermostable enzymes for industrial applications. *J. Ind. Microbiol.* 8, 71–81.
- Zhang, S., Zhang, K., Chen, X., Chu, X., Sun, F., Dong, Z., 2010. Five mutations in N-terminus confer thermostability on mesophilic xylanase. *Biochem. Biophys. Res. Commun.* 395, 200–206.
- Zhang, S.-B., Ma, X.-F., Pei, X.-Q., Liu, J.-Y., Shao, H.-W., Wu, Z.-L., 2012a. A practical high-throughput screening system for feruloyl esterases: Substrate design and evaluation. *J. Mol. Catal. B Enzym.* 74, 36–40.
- Zhang, S.-B., Pei, X.-Q., Wu, Z.-L., 2012b. Multiple amino acid substitutions significantly improve the thermostability of feruloyl esterase A from *Aspergillus niger*. *Bioresour. Technol.* 117, 140–147.
- Zhang, X.-F., Yang, G.-Y., Zhang, Y., Xie, Y., Withers, S.G., Feng, Y., 2016. A general and efficient strategy for generating the stable enzymes. *Sci. Rep.* 6, 33797.

Zumárraga, M., Bulter, T., Shleev, S., Polaina, J., Martínez-Arias, A., Plou, F.J., Ballesteros, A., Alcalde, M., 2007. *In vitro* evolution of a fungal laccase in high concentrations of organic cosolvents. Chem. Biol. 14, 1052–1064.

Publication (*in preparation*):

Directed evolution of a type A feruloyl esterase for increased
thermostability and organic solvent tolerance

Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance

Julien Durand¹, Victor Chaumeton¹, Olga Gherbovet¹, Mélanie Ragon¹, Peter Jütten², Joana L.A. Bras³, Catarina I.P.D. Guerreiro³, Carlos M.G.A. Fontes³, Alexander Piechot², Jean-Marc Nicaud⁴, Régis Fauré¹, Michael J. O'Donohue¹, Sophie Bozonnet^{1,*}

¹ LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

² Taros Chemicals GmbH & Co. KG, Dortmund, Germany

³ NZYTech genes & enzymes, Lisboa, Portugal

⁴ Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

* Corresponding Author

sophie.bozonnet@insa-toulouse.fr

ABSTRACT:

A methodology to express, generate and screen diversity for *Aspergillus niger* feruloyl esterase A (*AnFaeA*) was developed, tested and implemented in a high-throughput fashion, using an ad hoc chromogenic probe. In the first round of evolution, a total of ~10,000 random mutant clones were generated and screened to isolate twelve mutants with improved thermal and/or solvent stability. Then, two protein engineering approaches were applied in parallel: i) a random recombination was performed through Staggered Extension Process (StEP) while ii) 6 of the twenty-three identified mutations found in the first round hits, hypothesized to be key in the observed phenotypes, were rationally isolated or associated.

Both protein engineering approaches were complementary and successful in producing mutants with elevated thermal and/or solvent stability. The best thermo-resistant variants (N6T-A119V-S163T-V243M) possess a melting temperature (T_m) value increased by + 8.3 °C compared with the original enzyme. Concerning solvent resistance, mutant N6T-R66Q-Q117E doubles its residual activity in presence of 30% of DMSO or 22% *tert*-butanol compared to the wild-type enzyme. Substitution R66Q localized in the hinge region of the lid is mainly responsible of this new feature but this mutation causes a decrease of the turnover number.

KEYWORDS: Directed evolution, feruloyl esterases, thermostability, solvent resistance

INTRODUCTION

Ferulic acid (3-(4-hydroxy-3-methoxyphenyl)-2-propenoic acid or FA) is the most abundant hydroxycinnamic acid derivative in the cell walls of plants (Mathew and Abraham, 2004), accounting for up to 70–90% of phenolic acids found in wheat grains for example (Mattila et al., 2005; Moore et al., 2005). Because its structure comprises a phenolic moiety conjugated to a short unsaturated aliphatic chain which is functionalized by an acid group, the molecule presents important conjugation effects conferring strong antioxidant properties (Graf, 1992; Rice-Evans et al., 1996) that can be further enhanced specifically after transesterification with some alkyl chains (Kikuzaki et al., 2002; Ou and Kwok, 2004). Several *in vitro* and *in vivo* studies have shown other properties of interest for ferulic acid (Srinivasan et al., 2007), including notably, inhibition or prevention of breast, colon or lung cancers mainly due to anti-proliferative and pro-apoptotic actions (Kampa et al., 2003), or ability to absorb UV radiation (Graf, 1992), a property valuable for sunscreen composition. Nowadays, a part of commercially available ferulic acid is produced by enzymatic hydrolysis of biomass, specifically using the synergistic action of two classes of enzymes: xylanase and feruloyl acid esterases (Fae) (Saulnier and Thibault, 1999), though this remains challenging as most of the FA found in plants is covalently linked with lignin and other biopolymers, such as arabinoxylans and pectins amongst others. Specifically, Faes (EC 3.1.1.73) cleave the ester bond between ferulic acid and L-arabinofuranosyl residues of arabinoxylans. The Fae classification most commonly used today was initially proposed by Crepin *et al.* (Crepin et al., 2004), and follows a functional organization of fungal esterases in 4 classes (A-D), based on protein sequence identity and substrate specificity, namely four varieties of hydroxycinnamoyl esters: methyl 3-methoxy-4-hydroxycinnamate (MFA), methyl 3,4-dihydroxycinnamate (MCA), methyl 4-hydroxycinnamate (MpCA), and methyl 3,5-dimethoxy-4-hydroxycinnamate (MSA). More recently, a new classification of twelve Fae families was proposed, taking into account a larger diversity of enzymes of bacterial, fungal and plant origin and based on computational analysis of carefully selected sequence descriptors (Udatha et al., 2011). Some Faes can also be found in the Carbohydrate Active enZymes (CAZy) database (Lombard et al., 2014) (<http://www.cazy.org/>), in the Carbohydrate Esterase family 1 (CE1), except for one characterized bifunctional enzyme combining xylanase and feruloyl-esterase activities, which belongs to family 10 of the glycoside hydrolase (GH10) (Dodd et al., 2009). Finally, the ESTHER database, which focuses on the structural super-family of α/β -hydrolase folding enzymes, can provide interesting information about Fae and their relationships to distantly related lipases (Lenfant et al., 2013). As a matter of fact, Faes share the same two-step reaction mechanism as lipases, serine proteases and other esterases, involving the formation of a covalent acyl-enzyme intermediate, with the help of a conserved Ser-His-Asp catalytic triad, with Ser acting as a nucleophile, His as general base for both the acylation and deacylation steps, and Asp as an helper in the stabilization of transition state (Wong, 2006).

By taking advantage of the reactive carboxylic group of the ferulic acid, new molecules can be synthesized through esterification, either to add potential antioxidant properties to a hydroxylated molecule of interest (Vafiadi et al., 2008; Wen et al., 2016), or to alter the solubility of the resulting molecule (Kikugawa et al., 2016). Thus, Faes, in addition to helping the deconstruction of lignocellulosic biomass, also have a natural innate capability to synthesize, under specific operating conditions, a wide range of bioactive molecules by transesterification of ferulic ester onto different acceptor molecules (Topakas et al., 2005; Vafiadi et al., 2006). However, for these reactions, organic solvents or ionic liquids (Vafiadi et al., 2009; Zeuner et al., 2012; Antonopoulou et al., 2017) are often used to facilitate solubility of hydrophobic substrates, to allow catalysis of a wide variety of chemical reactions that are not feasible in aqueous media or to shift the thermodynamic equilibria to favor synthesis over hydrolysis (Serdakowski and Dordick, 2008; Doukyu and Ogino, 2010) even though enzymes in general exhibit very low performances in non-aqueous media. Similarly, high temperatures can favour transferuloylation reaction (Sun et al., 2014), but many natural enzymes are temperature sensitive, thus increasing their thermostability is of major interest for their use in industrial processes. In the field of Faes, several examples of such works have been reported, using rational approaches such as the engineering of disulfide bridges (Yin et al., 2015a) or random mutagenesis, either focused on specific residues selected based on their high B-factor values through iterative saturation mutagenesis (Yin et al., 2015b) or on the whole esterase gene via error-prone PCR (Cao et al., 2015) but studies amelioration of solvent resistance of Fae have not been reported yet to our knowledge.

AnFaeA is one of the most characterized esterase, cloned and studied since 1997 (De Vries et al., 1997), recombinantly produced in a variety of microorganisms, either after homologous (Record et al., 2003) or heterologous (Juge et al., 2001; Benoit et al., 2007) expression, and even used in a number of chimeric protein construction studies aiming at improving lignocellulose degradation (Levasseur et al., 2004, 2005, 2006). The 3D structure of *AnFaeA* (Hermoso et al., 2004) confirmed a classic α/β hydrolase scaffold, stabilised by 3 disulfide bridges, and the presence of a glycan chain attached to Asn79. This post-translational modification, which was found to participate against thermic denaturation in the native protein (Benoit et al., 2006), is located at the C-term of the lid (residue 68-80, *AnFaeA* numbering), a flexible region known to play a key role in interfacial activation of lipases. Moreover, as the amino acid sequence of this flap region contains also more hydrophilic residues than fungal lipases, it was postulated that this lid remains always open in contrast to what happens with lipases (Hermoso et al., 2004). This enzyme possesses a great interest for many industrial applications (Tabka et al., 2006; Tapin et al., 2006) and since it was shown to display moderate thermostability (Benoit et al., 2006), attempts were made to improve its temperature resistance (Benoit et al., 2006; Zhang and Wu, 2011; Zhang et al., 2012; J. Li et al., 2015; J.-J. Li et al., 2015).

In this work, we tackled the challenge of improving both the thermostability and solvent resistance of *Aspergillus niger* feruloyl-esterase A (*AnFaeA*) by a directed evolution approach. Using two specially

synthesized chromogenic ferulate substrates (Gherbovet et al., 2016), we have implemented an original and robust high-throughput screening strategy to identify temperature and solvent resistant mutants from a first error-prone PCR *AnFaeA* library. Then, through rational and random reassembly of interesting mutations, we have brought further the amelioration of our target esterase, and succeeded in combining both temperature and solvent resistance in the best hits. Finally, the profile of mutations is analysed in light of the available structures and reported knowledge on this enzyme.

EXPERIMENTAL SECTION

Substrates and chemicals

The chromogenic substrates 4-nitrocatechol-1-yl ferulate (4NTC-Fe) and 5-bromo-4-chloroindol-3-yl ferulate (X-Fe) were either synthesized *in house* (Gherbovet et al., 2016) or provided (4NTC-Fe only) by Taros Chemicals (Dortmund, Germany). Tryptone and yeast nitrogen bases (without amino acids and without ammonium sulphate, YNBww) were purchased from Difco (Paris, France). Kanamycin was purchased from Euromedex (Souffelweyersheim, France). Other chemicals, unless stated otherwise, were purchased from Sigma-Aldrich (Illkirch, France). Restriction enzymes were obtained from Invitrogen (Cergy Pontoise, France) or New England Biolabs (Evry, France).

Strains and culture conditions

XL1-Blue chemically competent *E. coli* strain (Invitrogen, Carlsbad, U.S.A.), used for transformation and for amplification of recombinant plasmid DNA, was grown at 37 °C in Luria-Bertani medium supplemented with kanamycin (40 µg.L⁻¹) for the selection of transformants. The growth media used for *Yarrowia lipolytica* strain JMY1212 (*MATA ura3-302 leu2-270-LEU2-zeta xpr2-322 Δlip2 Δlip7 Δlip8, Leu⁺, Ura⁻*) has been described elsewhere (Barth and Gaillardin, 1997). The yeasts were grown either on YPD (10 g.L⁻¹ yeast extract, 10 g.L⁻¹ bactopectone, 10 g.L⁻¹ of glucose) or for Ura⁺ transformant selection on YNB (1.7 g.L⁻¹ YNBww, 2 g.L⁻¹ of casamino acids, 10 g.L⁻¹ glucose, 5 g.L⁻¹ NH₄Cl in 50 mM potassium phosphate buffer pH 6) (Bordes et al., 2007). For solid media 15 g.L⁻¹ of agar from Biokar Diagnostics (Beauvais, France) was added. The plates were kept at 28 °C for growth.

Creation of random mutant libraries and site directed mutagenesis

- Plasmid

The *Y. lipolytica* codon optimised (NZYTech, Lisbon, Portugal) *AnFaeA* coding sequence (Genbank: Y09330) was cloned into the JMP62-TEF-ppLIP2-LIP2 plasmid (Duquesne et al., 2014) between the *BsrGI* and *AvrII* sites, in place of the *Lip2* gene, under the control of the constitutive promoter TEF, and in fusion with the signal peptide from Lip2, allowing protein secretion in the culture medium. The resulting plasmid JMP62TEF-*AnFaeA* plasmid contains a kanamycin resistance gene and an URA3 marker gene that allow selection in *E. coli* on kanamycin (40 µg.mL⁻¹) containing media and in *Y.*

lipolytica on URA⁻ media, respectively. Homologous recombination into *Y. lipolytica* genome is facilitated by two zeta platforms at both ends of the cloning cassette, bordered by two *NotI* restriction sites enabling the isolation of the transformation insert (Bordes et al., 2007).

- **Site directed mutagenesis**

Plasmid JMP62TEF-*AnFaeA*, with or without mutations, was used as DNA template. Mutations were inserted with the QuikChange mutagenesis kit (Agilent) following the manufacturer's instructions. The introduction of mutations was confirmed by DNA sequencing (GATC Biotech, Mulhouse, France).

- **Random mutant libraries**

The protocol followed to generate the random mutant libraries using JMP62TEF-*AnFaeA-D93G* as DNA template, which was constructed by site directed mutagenesis as described above, relied on 3 successive PCR, described here (Duquesne et al., 2012). All primers employed in this study are referenced in Table 1.

Table 1. List of primers used in this study. Only sequences (5'-3') of forward primers used in directed mutagenesis are given, the reverse primer being a reverse complement

Directed mutagenesis	
D93G-fw	CGTGGACCTCACAACCGTTGCACTGAGGG
N6T-fw	CAGAAGCGAGTGTACACCTCCGAGGACCTCTAC
R66Q-fw	AGATCATTACTGTTTTCCAAGGCACCGGTTCCGA
Q117E-fw	GGAGTCTCTGGTTAAGGAGCAGGCCTCG
A119V-fw	GGTTAAGCAGCAGGTCTCGCAGTACCCCG
S163T-fw	CCTTCGGAGAGCCTCGAACCGGAAACCAG
V243M-fw	GGGTGGACAGGGCATGAACGATGCCCA
Sequences used for creating the random mutant libraries	
PCR1_fw	GATCCCCACCGGAATTGC
PCR1_rv	GGAGAACTGCGGCCTCAGAAGGAGTGATGG
PCR2_fw	CCATCACTCCTTCTGAGGCCGCAGTTCTCC
PCR2_rv	GGAGTTCTTCGCCACCCC
PCR3_fw	CCGCCTGTCCGGAAACCGCGTTCAGGTGGAACAGGACCACC
PCR3_rv	CCGCACTGAAGGGCTTTGTGAGAGAGGTAACGCCG
Coding sequences recovery of the selected mutants in the <i>Y. lipolytica</i> genome	
gDNA_fw	AGCCAAGGGTATAAAAAGAC
gDNA_rv	CCGCACTGAAGGGCTTTGTGAGAGAGGTAACGCCG
StEP recombination	
StEP_fw	CCTTGTCAACTCCACACC
StEP_rv	CGGATGACTAACTCTCCAGAGCG
Sequencing of the selected mutants	
Seq_mut_fw	CCTTGTCAACTCCACACC
Seq_mut_rv	GTAGATAGTTGAGGTAG

Briefly, a first faithful PCR (PCR1), to amplify the first section of the expression cassette, composed of the zeta1 region, URA3 marker (*ura3d1*) and TEF promoter (pTEF), was made in 50 μ L with the plasmid JMP62TEF-*AnFaeA-D93G* as DNA template (20 ng), dNTP mix (800 μ M), primer mix (0.5 μ M each), and Phusion DNA Polymerase (2.5 IU, New England Biolabs) using the following temperature cycle: one initial denaturing step at 98 °C for 30 sec, followed by 30 cycles of (98 °C for 10 sec, 60 °C for 20 sec, 72 °C for 1 min), and a final elongation step at 72 °C for 5 min. A second PCR (PCR2) amplified the other part of the expression cassette, composed of the *AnFaeA-D93G* coding sequence and zeta2 region by error prone PCR (epPCR) using the GeneMorph II Random Mutagenesis kit from Agilent. Reactions were performed (50 μ L each) with the JMP62TEF-*AnFaeA-D93G* plasmid as DNA template (20 ng), dNTP mix (800 μ M), primers mix (3.5 ng. μ L⁻¹ each), and Mutazyme II DNA polymerase (2.5 IU, Agilent), and the following temperature cycle: one initial denaturing step at 95 °C for 2 min, followed by 35 cycles (of 95 °C for 1 min, 57 °C for 1 min, 72 °C for 1.5 min), and a final elongation step at 72 °C for 10 min. The reassembly PCR (50 μ L, PCR3) was made in two steps. First, to merge the two amplicons, 10 ng of each PCR products 1 and 2 which were purified with the QIAquick PCR Purification Kit (Qiagen) were mixed with dNTP mix (2,5 mM) and Phusion DNA Polymerase (2.5 IU) and submitted to one initial denaturing step at 98 °C for 30 sec, followed by 10 cycles of (98 °C for 10 sec, 57 °C for 30 sec, 72 °C for 2.5 min), and a final elongation step at 72 °C for 10 min. Then, 2.5 μ L of primer PCR3_fw and PCR3_rv (10 μ M each) were added in the reaction, and the reaction was allowed to continue with one initial denaturing step at 95 °C for 30 sec, followed by 20 cycles of (98 °C for 10 sec, 57 °C for 30 sec, 72 °C for 2,5 min), and a final elongation step at 72 °C for 10 min. Rates of mutations were evaluated as follow: DNA products obtained after the epPCR (PCR 2) were cloned into PCR-Blunt II-TOPO plasmids (Invitrogen), the product of ligation was transformed into *E. coli* TOP10 strains and plated on LB kanamycin plates. Purified mutant plasmid DNA extracted from randomly selected transformants were then sequenced (GATC Biotech, Mulhouse, France).

***Y. lipolytica* transformation and screening method**

Yeast cells were transformed, using the lithium acetate method (Le Dall et al., 1994), with the PCR3 product for the construction of random mutant libraries or after a *NotI* digestion for mutants obtained by site directed mutagenesis.

- First screen on solid medium

Transformed yeasts were spread on QTrays bioassay plates, containing 250 mL of solid YNB medium (URA⁻) covered with a membrane of polyvinylidene difluoride (PVDF, 0.22 μ m, 20 x 20 cm) (Merck Millipore) at a density of between 500 and 3,000 transformants per membrane. After 48 h of growth at 28 °C, membranes were delicately transferred on agarose containing X-Fe chromogenic substrate. X-Fe was first prepared at 70 mg.mL⁻¹ in DMSO and diluted at a final concentration of 150 μ g.mL⁻¹ in a solution of 100 mM sodium phosphate buffer pH 6 and agarose 1.2% (w/w), maintained at 55 °C after boiling. Colonies containing active mutant enzymes appear blue after 1 h incubation at 37 °C.

- **Liquid screening and Hit confirmation**

The blue colonies were picked and organized in 96-well microtiter plates containing 200 μL of YPD medium and then cultured overnight. 10 μL of this preculture was used to inoculate 1 mL YPD in deep well plates. D93G mutant is inoculated as control in each MP (3 wells). The deep well plates were incubated in a Multitron plate shaker (INFORS) under 700 rpm at 28 °C, for 36 h. and then centrifuged (2,250 g, 50 min). Supernatants were recovered and buffered with sodium phosphate buffer (5X) for a 100 mM pH 6 final concentration. Screening in microtiter plate format was automatized using a Genesis RSP-200 liquid handling automat (Tecan). Four different kinetics were measured at 21 °C in conditions resumed in Table 2. The heat shock was applied on aliquots of supernatants using a Peltier block. In all assay conditions, 20 μL of buffered supernatant is added to 180 μL of substrate to start the reaction. At regular intervals (6, 12 and 18 min) 40 μL of reaction were automatically removed and mixed with 160 μL of 2 M Na_2CO_3 to stop the reaction and allow absorbance reading on a spectrophotometer (Tecan Infinite M200 PRO) at 530 nm. Residual activity is calculated as the ratio of the activity after heat shock or in presence of solvent over the activity in standard condition (no heat shock and no solvent). Screening of two microtiter plates of the starting mutant D93G allowed calculating mean and standard deviation of hydrolysis in each condition.

Table 2. Concentration of substrate and solvents in the different screening assays.

	4NTC-Fe (mM)	DMSO (% , v/v)	<i>tert</i> -Butanol (% , v/v)
Standard activity	1.8	4	0
DMSO	1.8	25 or 30	0
<i>tert</i> -Butanol	1.8	4	20
Heat shock (15 min at 50 °C)	1.8	4	0

All assays were carried out in triplicate at 37 °C in 100 mM sodium phosphate buffer pH 6.

Genomic DNA extraction of *Y. lipolytica*

Genomic DNA (gDNA) of the 12 mutants selected after the first round of evolution was extracted as follow: 1.5 mL of an overnight culture in YPD inoculated from an isolated clone was centrifuged 5 min at 15,800 g and the pellet was suspended in 500 μL of sorbitol buffer (1 M sorbitol, 0.1 M Tris-HCl buffer pH 8 and 0.1 M EDTA). 50 μL of zymolyase (100T, Seikagaku Corp Coger) at 3 $\text{mg}\cdot\text{mL}^{-1}$ and 50 μL of 0.28 M β -mercaptoethanol were added before incubating 1 h at 37 °C. Supernatants were removed after centrifugation (15,800 g, 1 min), the pellet was suspended in 500 μL of TE buffer and 50 μL of 20% SDS and then incubated for 20 min at 65 °C. 200 μL of 5 M potassium acetate buffer was added, mixed and incubated 30 min on ice. The supernatant was collected after centrifugation (15,800 g, 5 min) mixed with 700 μL of pure *iso*-propanol and incubated during 10 min

at room temperature. After centrifugation (15,800 g, 5 min), the pellet was resuspended in 500 μL of 70% ethanol and centrifuged again. 400 μL of TE buffer containing 100 $\mu\text{g}\cdot\text{mL}^{-1}$ of RNase A was added on the pellet, and let to incubate 15 min at 37 °C. 40 μL of 2.5 M sodium acetate pH 5.2 and 1 mL of 100% ethanol were added for DNA precipitation. After centrifugation (15,800 g, 5 min), the pellet was washed with 700 μL of 70% ethanol, centrifuged again for 2 min; the supernatant was removed and once the pellet was well dried, resuspended in 100 μL of sterile water. The gDNA was stored at 4 °C before quantification by Nanodrop (Thermo Fisher Scientific) and use.

Staggered Extension Process (StEP) recombination

To recover the 12 coding sequences of the selected mutants from each corresponding purified gDNA, PCR were made using the primers referenced in Table 1. Each PCR, in a total volume of 50 μL , contained 10 μL of Phusion (5X) reaction buffer, 15 ng of gDNA, dNTP mix (600 μM), primers (0.8 μM each), and Phusion DNA Polymerase (2.5 IU), using the following temperature cycle: one initial denaturing step at 98 °C for 30 sec, followed by 32 cycles of (98 °C for 20 sec, 60 °C for 20 sec, 72 °C for 1.5 min), and a final elongation step at 72 °C for 5 min.

The StEP (Zhao, 2004; Zhao and Zha, 2006) recombination reaction, in 50 μL , contained: 5 μL 10X Taq DNA Polymerase Buffer, 15 pg of each of the twelve purified PCR products from the selected mutants, 600 μM of dNTP mix, 0.4 μM each of primers StEP_fw and StEP_rv, and Taq DNA Polymerase (2.5 IU, Thermo Fisher Scientific), and was submitted to one initial denaturing step at 95 °C for 30 sec, followed by 100 cycles of (95 °C for 20 sec, 60 °C for 1 sec), and a final elongation step at 72 °C for 3 min.

Thereafter, plasmid JMP62TEF-*AnFaeA* and StEP PCR products were double digested by *AvrII* and *BsrGI* and the fragments of interest were purified. The linearized plasmids and the digested inserts were then ligated (16 °C, overnight) in 20 μL reactions, containing T4 DNA ligase (2.5 IU, Thermo Fisher Scientific) and an inserts/plasmid molar ratio of 3-1, for 100 ng of total DNA. Several aliquots of 5 μL of the ligation product were used to transform 50 μL of commercial XL1-Blue competent cells (Agilent). The bacteria were then spread onto LB-agar Petri dishes containing kanamycin (40 $\mu\text{g}\cdot\text{mL}^{-1}$) and allowed to grow overnight at 37 °C. From the libraries of mutants obtained, fourteen isolated colonies were cultured in 3 mL of liquid LB medium containing kanamycin to purify plasmid DNA which was sequenced to check the effective recombination. All remaining colonies were scraped and collected in liquid LB medium containing kanamycin (40 $\mu\text{g}\cdot\text{mL}^{-1}$), added to the surface of each plate. All the cell suspensions were pooled and the total volume was centrifuged (6,800 g, 3 min, 12 °C), providing a single cell pellet which was used to purify plasmid DNA, stored at -20 °C until further use.

Y. lipolytica competent cells were transformed with small aliquots of this StEP purified plasmid DNA, linearized by *NotI*. Transformed yeasts were obtained on selective URA⁻ YNB medium and mutants

were then grown in microtiter plates and screened as described previously, except for the heat shock which was 30 min long, at 55 °C. Mutants D93G, N6T and R66Q-Q117E were added in each 96-well microtiter plate to indicate levels of activity for the template protein, the best thermostable and the most solvent resistant mutants from the first round of evolution respectively.

Sequencing of the selected mutants

PCR were made on gDNA, in 50 µL reaction containing the adequate buffer, primers gDNA_fw and StEP_rv (0.8 µM each, Table 1), 200 ng of gDNA, dNTP mix (600 µM), and Phusion DNA Polymerase (2.5 IU), and following this protocol: one initial denaturing step at 98 °C for 30 sec, followed by 30 cycles of (98 °C for 10 sec, 55 °C for 30 sec, 72 °C for 2.5 min), and a final elongation step at 72 °C for 10 min. PCR products were sent for sequencing with the following pair of oligonucleotide primers: Seq_mut_fw and Seq_mut_rv (Table 1).

Purification and activity assays using 4NTC-Fe

- Purification

For quantitative assay, cultures (100 or 200 mL) were performed in YPD medium. After growth for 32 h at 28 °C, cultures were centrifuged (10,000 g, 12 min), the supernatants filtered (0.22 µm) and total proteins were precipitated with 85% (m/v) ammonium sulphate. After centrifugation (8,000 g, 25 min), the protein pellet was resuspended in 4 mL of 100 mM sodium phosphate buffer pH 6 and 2 mL of this solution was directly injected (1 mL.min⁻¹) onto a Fast Flow Phenyl sepharose column, 7 x 25 mm (GE Healthcare), equilibrated with the binding buffer (50 mM sodium phosphate buffer pH 7 containing 1M ammonium sulphate), by using a liquid chromatography system AKTA-Pure (GE Healthcare). Elution was performed in two steps: first, most contaminant proteins were eliminated using 6 column volumes of 20% binding buffer and 80% elution buffer (50 mM sodium phosphate buffer pH 7), then *AnFaeA* or the mutant proteins were recovered thanks to an elution with 6 column volumes of 100% elution buffer. Enzyme purity was checked with SDS-PAGE (BioRad) and fractions containing pure enzyme were pooled and quantified by Nanodrop.

- Activity assays using 4NTC-Fe

4NTC-Fe was first dissolved in 100% DMSO at 50 mM and was then diluted in 100 mM sodium phosphate buffer pH 6 to achieve a final concentration of 2 mM 4NTC-Fe in 4% (v/v) DMSO. For Fae activity assay, 450 µL of this solution was preincubated at 24 or 40 °C for 4 min and then mixed with 50 µL of purified enzyme solution. The release of 4NTC was monitored in a discontinuous assay, removing 50 µL reaction aliquots every 3 min over a 21 min period. Upon removal, the aliquots were immediately mixed in a 96-well microtiter plates filled with 200 µL/well of an aqueous solution of 2 M Na₂CO₃ and kept on ice; absorbance at 530 nm was recorded using a 96-well microtiter plate reader (Infinite M200 PRO, TECAN). Activity was determined using a standard absorbance curve made of

50 μL of different concentrations of 4NTC (from 0.01 to 1 mM) in 100 mM sodium phosphate buffer pH 6, 4% DMSO added to 200 μL of 2 M Na_2CO_3 . One unit of Fae activity is defined as the amount of enzyme releasing 1 μmol of 4NTC per min.

For mutant characterisation, different conditions were tested: measurement of the specific activity in the above-mentioned conditions at 24 and 40 $^\circ\text{C}$, of residual activity after a heat shock of 5 min at 50 $^\circ\text{C}$ or 10 min at 55 $^\circ\text{C}$, or in presence of solvent, either DMSO (30%, v/v) or *tert*-butanol (22%, v/v). After heat shock, kinetics were made at 40 $^\circ\text{C}$, whereas in presence of solvent, kinetics were performed at 24 $^\circ\text{C}$. Influence of solvent on absorbance measurements was also taken into account, and corresponding standard curves were made with the same solvent concentrations. Negative controls containing all of the reactants except the enzyme were included in order to monitor and correct for spontaneous hydrolysis of the substrate (Gherbovet et al., 2016). All measurements were made in triplicate, unless otherwise stated.

- Differential Scanning Fluorimetry (DSF)

To determine the melting temperature of purified esterases, 15 μM of enzymes were mixed with Sypro 10X in 20 μL final volume of 100 mM sodium phosphate buffer pH 6. Reactions occurred in a CFSX 96 TOUCH Real-Time PCR detection system (BioRad) with a fluorescence measure each 0.3 $^\circ\text{C}$ from 20 to 80 $^\circ\text{C}$. Each run were made in triplicate except for parental D93G made 9 times (two triplicate with two different productions and one after six month storage at 4 $^\circ\text{C}$).

- N-Access and B-Factor determination

NAccess values were obtained using NAccess V2.1.1 (Atomic Solvent Accessible Area Calculations) and B-factor values using B-FITTER with the *AnFaeA* sequence (PDB file: 1USW).

RESULTS

Heterologous expression and validation of *AnFaeA*-D93G mutant as starting point.

AnFaeA and mutants were expressed and secreted in the culture medium of *Y. lipolytica*, a yeast strain that was proven very reliable for the recombinant extracellular production of enzymes (Cambon et al., 2010). Production yields were in the range 7-40 $\text{mg}\cdot\text{L}^{-1}$. The molecular mass of purified enzyme estimated on SDS PAGE was 36 kDa in contrast to the 28 kDa expected, an observation already noted by De Vries et al., 1997, who assumed that acidity of this enzyme ($\text{pI}_{\text{exp}} = 3.3$) disturbs interactions with SDS and reduces mobility in the gel.

As a previous work showed that mutation D93G slightly improves the thermostability of *AnFaeA* (Zhang and Wu, 2011), we decided to construct this mutant by site-directed mutagenesis. A very moderate rise in residual activity (+1%) after a heat shock of 5 min at 50 $^\circ\text{C}$ and in melting temperature (T_m , + 0.6 $^\circ\text{C}$) compared with wt *AnFaeA* confirmed the data by Zhang *et al.* (Table 3).

Though modestly improved, *AnFaeA-D93G* was then used as a starting template for directed evolution to further improve the thermal stability and increase the solvent resistance of the enzyme.

A two-step screening strategy was designed to select optimized mutants, based first on isolation of active mutants on solid medium containing a chromogenic ferulate substrate (X-Fe), followed by a series of kinetics with 4NTC-Fe substrate (Gherbovet et al., 2016) using culture supernatant. Solid screening was validated using the template mutant *AnFaeA-D93G*. After growth, the membrane over which transformed yeasts were plated, was transferred on an X-Fe containing buffered agarose and colonies were shown to generate a homogeneous blue colour. Liquid screening was calibrated with culture supernatant of recombinant yeasts grown in microtiter plates, and was based on the parallel kinetic measurement of four 4NTC-Fe hydrolysis reactions either in standard conditions, or after heat shock was applied to culture supernatant, or in the presence of either of the two targeted solvents, DMSO and *tert*-butanol. These data showed a mean value of residual activity of the template esterase of 28 ± 5 ; 53 ± 5 and $29 \pm 3\%$ after either a heat shock of 15 min at 50 °C, or in presence of 20% DMSO (v/v) or 20% *tert*-butanol (v/v) respectively. Mutants with residual activity above the mean value plus three times the standard deviation were considered best suited for the tested condition and selected for confirmation (Figure S1).

Creation of diversity and selection of 1st generation hits

By varying the quantity of DNA template during the epPCR and assaying the resulting rate of active (blue) colonies on solid medium containing X-Fe, we determined the amplification conditions allowing to obtain around 20% active colonies (Figure S2). Sequencing of 8 mutants, revealed a mutation charge of 5 mutations per gene. Thereafter, approximately 10,000 transformants were obtained in *Y. lipolytica* after three sets of epPCR with the DNA coding sequence of *AnFaeA-D93G* as template. A total of 2,046 active clones were then selected and assayed using our validated two-step automatized screening procedure. The most resistant clones (thirty-one, sixteen towards temperature, fifteen towards solvent presence) were individually isolated and underwent a second identical screen. Thereby, we could confirm the robustness of our developed screen, as the mean value of the residual activity obtained on every series of six isolated clones of each selected mutant was in perfect agreement with the initial value obtained in the first experiment, with a very low standard error (Table S1). The residual activity ratios for the twelve best confirmed mutants after heat shock or in the presence of solvent are represented in Figure 1A, as well as the amino acid mutations found in each corresponding clone (Figure 1B). A total of twenty-two different mutations was identified in this set of improved esterases. Mutants 1 to 3 present a better residual activity in presence of one or both solvents. In particular, mutant 2 - bearing mutations R66Q and Q117E - was found +18 and +10% more active in presence of 25% DMSO or 20% *tert*-butanol respectively, whereas template *AnFaeA-D93G* exhibits a drastic loss of almost 60 and 70% activity in the same conditions. Mutants 4 to 10 present a better residual activity after a thermal shock of 15 min at 50 °C (from +20 to +70%). Three

thermostable mutants (6, 8 and 10) possess a mutation on the N6 position (mutated in serine, threonine or isoleucine respectively), and of the three, the mono-mutant N6T was the most resistant to heat shock, with no loss of activity compared to a 70% decrease for the parental enzyme. Interestingly, despite the fact that these two features were screened separately, mutants 11 and 12 presented both a better thermostability (+45 and +60% respectively) and a better solvent resistance (at least for one solvent) than *AnFaeA*-D93G.

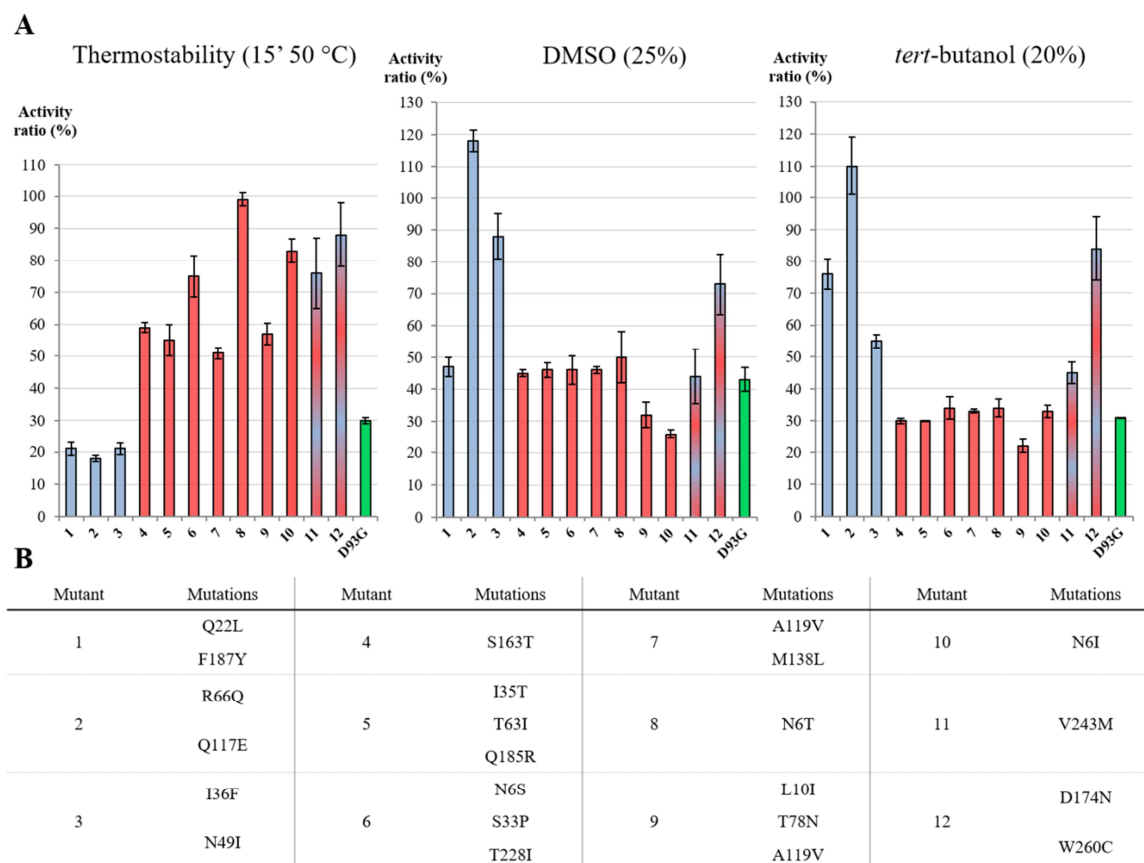


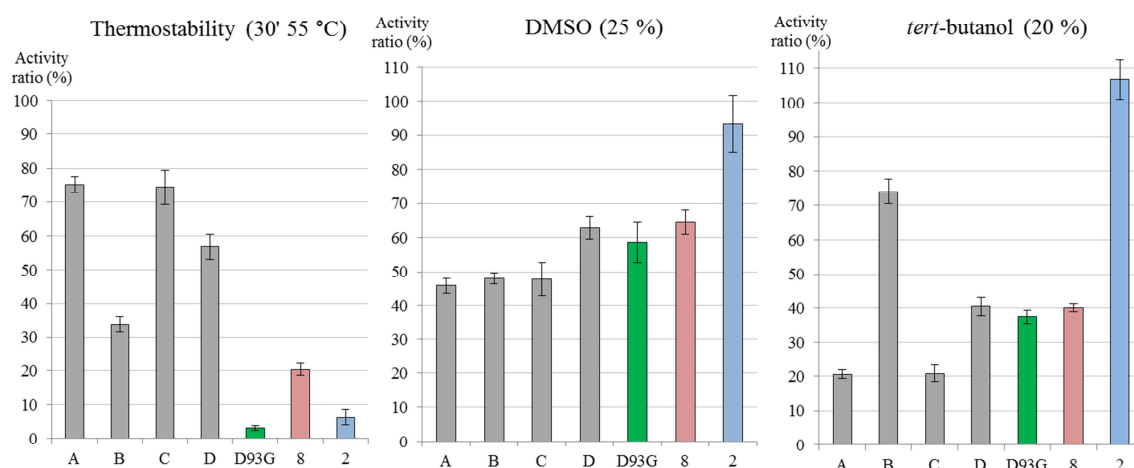
Figure 1. **Characteristics of selected improved mutants after completion of first round of evolution.** (A) Residual activities after heat shock or in presence of solvents (% in v/v), expressed in % and compared to the starting template mutant *AnFaeA*-D93G (green). In blue, mutants with improved resistance to at least one solvent; in red, mutants with a better thermostability; mutants improved for both properties are shown in blue and red. (B) Mutations found in each of the twelve selected clones (the original D93G mutation from the template, present in all mutants, is not repeated).

2nd generation library

Genomic DNA of the twelve selected hits from the first round of directed evolution was used to amplify and purify *AnFaeA* mutant genes, which were used as matrices to generate a new mutant library, where ideally, all twenty-two identified mutations would be randomly recombined. The sequences of fourteen arbitrarily isolated *E. coli* clones bearing shuffled DNA, showed an excellent reassembly of the matrix sequences as twenty (out of twenty-two) mutations belonging to all twelve

sequences were found (Table S2). 550 mutants were screened under the same conditions as for the first generation library, except for the thermal shock that was elongated (30 min, 55 °C) to increase the stringency of the selection. Four mutants (A to D) were selected and sequenced after comparison with reference mutants *AnFaeA*-D93G (template), N6T and R66Q-Q117E ((best thermostable and most solvent resistant from first round, respectively) (Figure 2). All four mutants presented a better thermostability than N6T, but unfortunately, no clone with a better solvent resistance than clone R66Q-Q117E was found. After sequencing, mutants A and C revealed the same combination of mutations (N6T-A119V-S163T-V243M), which is consistent with a very similar activity profile and thus confirms the robustness of the screening. Indeed, both clones exhibit a significantly improved thermostability, as around 75% residual activity remains after heat shock while less than 5% for D93G or 20% for N6T (mutant 8) does under similar treatment. Mutant B combines a better resistance in *tert*-butanol and a better thermostability. Finally mutant D, with mutations N6T and Q185R, was shown to exhibit a significantly improved thermostability compared to N6T alone (+40%).

A



B

Mutant	Mutations	Mutant	Mutations	Mutant	Mutations	Mutant	Mutations
A	N6T A119V S163T V243M	B	N6I Q22L F187Y	C	N6T A119V S163T V243M	D	N6T Q185R

Figure 2. Characteristics of selected mutants after StEP recombination. (A) Residual activities after heat shock (30 min at 55 °C) or in presence of solvents (% in v/v), expressed in % and compared to the starting template mutant *AnFaeA*-D93G (green), the best thermostable first round mutant (N6T) (red) and the best first round solvent-resistant mutant (R66Q-Q117E) (blue). (B) Mutation combination found in mutants (A-D) (the original D93G mutation from the template, present in all mutants, is not noted).

AnFaeA mutants A, B and D were produced, purified and characterised (Table 3), and were all confirmed with superior residual activity after heat shock and with higher T_m values than D93G (+8.3,

+7.4 and +8.3 °C for A, B and D respectively). Mutant B (N6I-Q22L-F187Y), in addition to its improved thermostability, also presented a good residual activity (60%) in the presence of *tert*-butanol.

Rational recombination by site directed mutagenesis

In the light of the mutations observed in the optimized mutants, six mono-mutants were produced and purified. As mutations S163T, N6T and V243M were found isolated in first round mutants 4, 8 and 11 respectively, they were obviously considered as positively affecting *AnFaeA*-D93G thermoresistance. Mutation A119V, found in two different thermostable multiple mutants was also picked for site-directed mutagenesis, whereas mutations R66Q and Q117E were chosen so as to separate their respective importance in the enhancement of the solvent resistance of mutant 2, all the more as no mutant more active in presence of solvent were isolated after StEP recombination and screening.

Mutation N6T confirmed a significant effect on the thermostability of the enzyme, as indeed it led to a residual activity after heat shock of $47 \pm 7\%$ and with a T_m value of 55.5 ± 0.1 °C, to compare with $17 \pm 1\%$ and 48.7 ± 0.5 °C for D93G respectively. Similarly, but to a lesser extent, mutations A119V, S163T and V243M caused a gain in terms of thermostability but at the expense of a loss of solvent resistance for V243M.

Although double-mutant 2 (R66Q-Q117E) has been highlighted as more active in the presence of both solvents, the isolated mutation Q117E did not show any improvement toward DMSO or *tert*-butanol resistance. On the contrary, R66Q is clearly less affected by the presence of solvents (54 ± 6 and $65 \pm 4\%$ residual activity against 27 ± 6 and $33 \pm 5\%$ for *AnFaeA*-D93G with 30% DMSO or 22% *tert*-butanol respectively) but, shows a lower specific activity (approximately 4 fold) and a poorer thermostability (loss of 3.4 °C on the T_m value compared to D93G) (Table 3). These results were confirmed as the double mutant R66Q-Q117E (equivalent to mutant 2) exhibited a better resistance in presence of solvent but less than R66Q alone and showed a very low specific activity and thermostability.

On the other hand, combining N6T and R66Q mutations allowed a combined effect of both mutations, which resulted in an improvement of both thermostability and solvent resistance, but yet with a lowered specific activity at 40 °C than *AnFaeA*-D93G. Combining N6T with S163T improved the T_m value (57 °C) which is higher than the one of both mono-mutant taken separately. Concerning solvent resistance, N6T-R66Q-Q117E seems to be the less affected mutant by the presence of DMSO or *tert*-butanol ($60 \pm 13\%$ and $70 \pm 15\%$ residual activity respectively) (Table 3).

Table 3. Characterization of selected mutants after recombination by the StEP method as well as rationally obtained

		SA at 40 °C ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	Residual activity (%)		T_m	
			Heat shock (5 min 50 °C)	^b DMSO (30%) ^a	^b <i>tert</i> -Butanol (22%) ^a	
	<i>AnFaeA</i>	52.8 ± 1.7	16 ± 1	18 ± 6	44 ± 9	48.1 ± 0.2
	D93G	39.6 ± 1.4	17 ± 1	27 ± 6	33 ± 5	48.7 ± 0.5
^a Mutants from StEP			Heat shock (10 min 55 °C)	DMSO (30%) ^a	<i>tert</i> -Butanol (22%) ^a	
(A)	N6T-A119V- S163T-V243M	54.8 ± 5.0	54 ± 8	9 ± 1	10 ± 2	57.0 ± 0.0
(B)	N6I-Q22L-F187Y	18.3 ± 2.8	44 ± 7	26 ± 4	60 ± 9	56.1 ± 0.0
(D)	N6T-Q185R	34.9 ± 2.9	55 ± 5	21 ± 4	25 ± 3	57.0 ± 0.0
^a Rational mono mutants			Heat shock (5 min 50 °C)	DMSO (30 %)	<i>t</i> -but (22 %)	
	N6T	45.9 ± 7.1	47 ± 7	31 ± 6	44 ± 8	55.5 ± 0.1
	R66Q	10.7 ± 0.4	12 ± 0	54 ± 6	65 ± 4	45.3 ± 0.0
	Q117E	42.6 ± 3.6	18 ± 2	25 ± 3	30 ± 4	48.7 ± 0.2
	A119V	39.3 ± 4.1	21 ± 2	24 ± 3	26 ± 2	49.7 ± 0.2
	S163T	42.8 ± 5.2	36 ± 4	30 ± 3	30 ± 2	51.9 ± 0.1
	V243M	33.4 ± 0.9	22 ± 1	10 ± 1	10 ± 1	48.9 ± 0.1
^a Rational combination			Heat shock (10 min 55 °C)	DMSO (30 %)	<i>t</i> -but (22 %)	
	R66Q-Q117E	4.5 ± 0.1	6 ± 1	41 ± 4	52 ± 9	46.2 ± 0.2
	N6T-R66Q	11.8 ± 0.2	27 ± 2	53 ± 6	58 ± 6	52.0 ± 0.2
	N6T-S163T	52.5 ± 8.0	28 ± 5	41 ± 6	37 ± 8	57.0 ± 0.0
	N6T-V243M	49.4 ± 8.2	17 ± 4	16 ± 4	15 ± 3	54.9 ± 0.0
	N6T-R66Q-Q117E	7.8 ± 0.1	29 ± 1	60 ± 13	70 ± 15	53.9 ± 0.5
	N6T-S163T-V243M	56.1 ± 8.3	44 ± 8	18 ± 6	15 ± 5	56.3 ± 0.2
	N6T-R66Q-S163T	8.3 ± 1.2	16 ± 6	43 ± 3	47 ± 4	53.1 ± 0.0

^a All the mutants contain the mutation D93G included in the parental sequence; ^b % of solvent in 100 mM sodium phosphate buffer pH 6 (v/v)

Kinetic parameters of *AnFaeA* and mutants

Kinetic parameters for a total of nine mutants were determined against 4NTC-Fe substrate (Table 4). A slight increase in catalytic efficiency was confirmed for the template D93G (Zhang and Wu, 2011), but no other additional mutation further improved the hydrolytic activity of the esterase. Though most of the substitutions only moderately affect the kinetic parameters of the mutant enzymes, k_{cat} was significantly decreased for proteins containing the R66Q substitutions, and a ca. 3x higher K_M was observed for the V243M mutant.

Table 4. Kinetic parameters of *AnFaeA* and mutants on 4NTC-Fe substrate

Enzyme	K_M (mM)	k_{cat} (s^{-1})	k_{cat} / K_M ($s^{-1}.mM^{-1}$)
<i>AnFaeA</i>	0.14 ± 0.01	30.8 ± 0.8	213.2
D93G	0.12 ± 0.01	39.0 ± 1.5	314.2
^a N6T	0.08 ± 0.01	25.1 ± 0.6	301.1
^a R66Q	0.21 ± 0.02	4.7 ± 0.1	22.7
^a V243M	0.34 ± 0.02	31.5 ± 0.6	92.9
^a N6T-Q185R	0.19 ± 0.02	31.8 ± 1.2	172.0
^a N6I-Q22L-F187Y	0.08 ± 0.01	21.0 ± 0.5	275.9
^a N6T-R66Q-Q117E	0.16 ± 0.02	9.9 ± 0.3	61.3
^a N6T-A119V-S163T-V243M	0.23 ± 0.01	34.7 ± 0.6	152.5

^a Note that all these mutants contain the D93G mutation included in the parental sequence.

All assays were carried out in triplicate at 37 °C in 100 mM sodium phosphate buffer pH 6.

DISCUSSION

Faes have consistently shown great interest for their potential application towards ferulic acid production from agro-industrial by-products, such as wheat bran or coffee pulp for example (Topakas et al., 2007; Benoit et al., 2008; Dilokpimol et al., 2016). Recently, Faes have benefitted from a gain of interest for their ability to synthesize rather than hydrolyse feruloylated molecules (Antonopoulou et al., 2017). Lipases such as Cal-B can be used for the synthesis of ferulic acid derivatives with good yields (Sun et al., 2007), but are often inhibited by conjugation phenomena in the hydroxycinnamic donor (Vafiadi et al., 2008). Therefore, as Faes show natural higher substrate specificity, they represent valuable tools for the production of regioselective products such as sugar ferulates (Topakas et al., 2005; Vafiadi et al., 2006). However, these reactions are mostly performed in the presence of high concentration of solvents (Vafiadi et al., 2008) and sometimes performed at high-temperature (Sun et al., 2013, 2014), conditions where most enzymes suffer from fast denaturation. Therefore, increasing solvent resistance and thermostability of Faes represents valuable goals for green chemistry purposes, and directed evolution an adequate methodology to generate improved enzymes (Farinas et al., 2001; Molina-Espeja et al., 2016). Hence, the objective of this work was to apply such strategies to

enhance both the thermostability and stability in DMSO or *tert*-butanol of feruloyl esterase A from *Aspergillus niger*, which has naturally relatively poor performance in both conditions.

Here, the yeast *Y. lipolytica* was chosen as host to perform molecular engineering of *AnFaeA* as it allows high transformation efficiency while limiting multiple gene integration (Bordes et al., 2007) and was later characterized by its very high reproducibility in protein expression level (Duquesne et al., 2012). *AnFaeA* production levels were in the range of what was obtained in *E. coli* (30 mg.L⁻¹) but the protein was soluble, active and easily recovered from the supernatant whereas it is found in inclusion bodies in *E. coli*. On the contrary, this level of production is ca. 10x lower than production reported in *P. pastoris* (Juge et al., 2001), but our specific activity (52 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$) is 100x higher than the one of *AnFaeA* produced also in *P. pastoris* (0.52 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$, (Zhang et al., 2012) and tested on 4-nitrophenyl ferulate, a chromogenic substrate closely related to 4NTC-Fe. This significantly increased specific activity might be due to a more adequate post-translational glycosylation.

Thanks to the recently developed chromogenic substrates X-Fe and 4NTC-Fe (Gherbovet et al., 2016), it was possible to set up a straightforward two-stage selection strategy, combining solid and then liquid screening at high-throughput. Interestingly, the second stage of the process relies on a simple assay in buffer, without detergent such as Triton X-100, necessary for the complete solubilisation of alternative colored substrates 4-nitrophenyl ferulate and 2-chloro-4-nitrophenyl ferulate (Mastihuba et al., 2011; Zhang et al., 2012), but which results very often in bubbles formation that can interfere with final absorbance measurement. This resulted in the set-up of a very efficient and robust screening, as illustrated by i) a remarkable reproducibility of activity assays in microtiter plates, even when performed from independently cultured clones, and ii) our ability to highlight identical mutations in different mutants (A119V or the repeated appearance of Asn6 mutation during the first round screening) or identical mutants (mutants A and C during the second round screening), out of the immense theoretical sequence space generated by our epPCR or recombination processes.

The location of mutations found in mutants showing the most interesting properties are shown in Figure 4. Looking globally at the mutations identified in the first generation hits with increased thermostability, we can first observe that out of the 14 residues that were found mutated eight are located on loops, five on helices and only one (T63I) is found on a beta sheet. This is clearly in accordance with previous works of Mandrich et al., who observed that -when comparing the structures of two carboxylesterases from either a thermophilic or a mesophilic *Bacillus*- most changes were localized mostly in loops regions or α -helices (Mandrich et al., 2004).

Among the mutations that we carefully investigated as isolated mutations, it is interesting to note that those improving the most the thermostability were already found as single mutations in hit clones (N6T and S163T). Concerning residue N6, its importance could already be demonstrated by the fact

that three of the nine thermostable mutants identified during the first cycle of evolution present a substitution at this position, and that it was systemically found in all second generation StEP mutants. It was mutated into three different residues, *i.e.* T, S (two polar and uncharged amino acids with an alcohol functional group) or I, and substitutions N6T and N6I were found in mono-mutants (8 and 10, Figure 1), reinforcing the idea that they are involved in the thermostability of *AnFaeA*. This asparagine is not present in the parental sequence, it replaces an Ile in the native *A. niger* sequence, and its alteration is a result of the cloning process, to fuse the mature *AnFaeA* sequence to the prepro region of *Y. lipolytica* Lip2 protein to allow extracellular production of the recombinant esterase (Figure S3). The thermostability of the resulting fused construction is close to that of the wild-type protein (Zhang et al., 2012). It has been shown that some surface asparagine residues can undergo natural deamination reactions which are unfavorable for thermostability (Yang and Zubarev, 2010), and that substitution of these asparagine residues in *Bacillus licheniformis* (Declerck et al., 2000) or *Bacillus* KR-8104 (Rahimzadeh et al., 2012) α -amylases improves their thermostability. Here we believe that these 3 mutations contribute to the overall stabilization of the N-terminal end of the protein through optimization of non-covalent interactions. This result is comforted by the observation of another mutation in the N-term of *AnFaeA*, mutation L10I, in thermostable mutant 9. The important role played by the N-terminal region towards stabilization against temperature denaturation has already been demonstrated in many types of enzymes like glycoside hydrolases (Sun et al., 2005; Zhang et al., 2010) or esterases (Fuciños et al., 2011; Mandrich et al., 2005). It was shown for example that the truncation of disordered residues, *i.e.* not involved in a secondary structure (α -helix or β -strand) in the N-terminal of an *Aspergillus niger* GH10 xylanase increased significantly its thermostability (Liu et al., 2011). Similarly, though it was not investigated independently, mutation W260C, observed in 1st generation hit 12 combined with mutation D174N, could play an equivalent role, as it concerns the ultimate amino-acid of the protein. Interestingly, the newly introduced cysteine residue would be located at a bonding distance to Cys29, therefore creating an alternative and possibly more stabilizing disulfide bridge than the native one, between Cys29 and Cys258.

The single substitution, S163T, found in mutant 4, improves its thermostability (+20% residual activity after heat-shock and +3 °C in T_m) though it only contains an extra methyl group on the C $_{\beta}$ of the amino acid. This residue is located in the core of the enzyme at a distance of ~8 Å from the catalytic serine (S133) and close to a small internal pocket without electronic density. We could suppose that addition of this methyl group fills this pocket and makes the structure more compact at the vicinity of the active site. When these two substitutions (N6T and S163T) were associated, the highest melting temperature was measured (+8.3 °C compared to template esterase), in conjunction with an increase of the specific activity at 40 °C, indicating a synergistic effect of these mutations. However, the residual activity after a 10 min heat shock at 55 °C was not ameliorated in proportion.

From all the rationally constructed mutants, the triple mutant N6T-S163T-V243M was the best thermostable mutant, with an increase in specific activity of 16%, a residual activity after heat shock of 27% and a T_m superior by 7.6 °C to the template's. Effect of mutation V243M remains unclear, as when isolated, it did not significantly improve the enzyme thermostability. It is located in the loop 226-244 that confines and shapes the active site, and was supposed to play an important role in substrate recognition as it is aligned with both catalytic His and Ser residues, as seen on *AnFaeA* crystal structure (Hermoso et al., 2004). These authors also suggest that Val243 could play a role in substrate binding, as, along with Ile196 and Leu199, it forms a hydrophobic pocket that could host a diferulic motif. The single mutation V243M caused indeed a slight decrease of the specific activity, but more importantly led to K_m value 3 times higher than the template's, confirming its involvement in the affinity towards the substrate. The effect of the substitution of this valine residue could also be the result of a local rigidification of the structure as this amino-acid displays the highest B-value (ranked 18/260 with the software B-fitter) among all identified mutations. It has been shown that modifications of local degrees of rigidity in a protein can modulate its temperature resistance profile. Indeed, Reetz and co-workers could either increase or decrease the thermostability of a protein, by carefully selecting for site-saturation mutagenesis target residues with high or low flexibility, respectively, evaluated by their B-values (Reetz et al., 2006; Reetz and Carballeira, 2007).

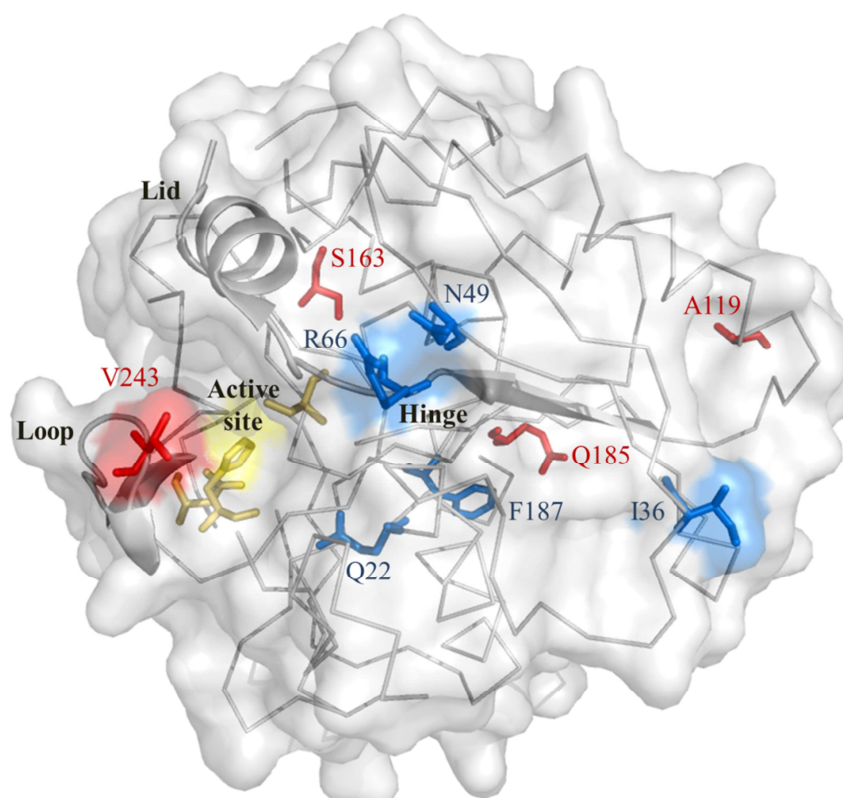


Figure 4. **3D structure of *AnFaeA* and localization of identified mutations**, using the crystallographic structure from (Benoit et al., 2006) (PDB code 2IX9).

Finally, this optimized combination (N6T-S163T-V243M) was re-created and further improved with the addition of mutation A119V when this quadruple mutant was successfully isolated twice after screening randomly shuffled esterase in the StEP procedure. It is quite impressive to remark that all these single mutations originate from distinct parents, thus emphasizing the precision and neatness of the recombination procedure developed in this work. Because substitution A119V, when studied independently, exhibited only very moderate improvement towards temperature stability (+3% residual activity after heat shock and +1 °C in T_m compared to template protein), and no change in terms of specific activity, it was not then associated with others in our rational recombination study. Nevertheless, associated with mutations N6T, S163T and V243M, though its catalytic efficiency was half that of the template enzyme, it resulted in the best *AnFaeA* variant obtained in this study in terms of thermostability.

Concerning solvent resistance, most of the improvements observed in selected mutants are related to mutations R66Q. Initially isolated in mutant 2, in association with substitution Q117E, the single mutant was constructed and characterized, and was proven to be responsible for doubling the residual activity of the enzyme in presence of either 30% DMSO or 22% *tert*-butanol. Unfortunately this was coupled to a drastic decrease of the enzyme specific activity (-75%), confirmed by a 10x reduced catalytic efficiency, and lower resistance to temperature compared to the template enzyme. The rational association of R66Q with N6T logically combined the benefits of the two substitutions, as the double mutant simultaneously exhibit temperature and solvent resistance, though the former is moderate compared to improvements obtained with the mutation N6T alone. Finally, the higher solvent resistance was observed for the triple mutant N6T-R66Q-Q117E, though its rate of reaction remains 5x time lower than the template esterase.

Amino acid R66 is directly located in the N-term hinge region of *AnFaeA*, *i.e.* the short loop (65-70) located upstream from the lid, and therefore involved in the control of the motion of this domain, known to be critical for interfacial activation in related lipase enzymes. Residue N49, which was found mutated in another mutant showing a better resistance to solvents (N49I, mutant 3 from the first generation), interacts directly with R66, though only via an interaction between N49 main chain carbonyl and R66 main chain amine groups. Thus, this specific stretch of sequence seems to be important in terms of stability towards solvent denaturation, but not towards temperature resistance, as R66Q mutant T_m slightly drops. Several studies previously focused on the differences that can be observed concerning the position and the movement of the lid in 3D structures of homologous esterases, such as *A. niger* feruloyl esterase A and *A. niger* or *Thermomyces lanuginosus* lipases (Shu et al., 2009, 2011; Skjold-Jørgensen et al., 2014). Site directed mutagenesis of residues, either part of the lid or located in the hinge region, was performed but unfortunately results were only evaluated in relation to substrate preference or activation mechanism (Skjold-Jørgensen et al., 2014; Tang et al., 2015).

Although the second generation random library was also screened for increased solvent resistance, no significantly improved mutant was isolated. Only the triple mutant N6I-Q22L-F187Y, resulting from a probable single crossing-over between mutant 2 (Q22L-F187Y) and mutant 10 (N6I), presented 8 and 16% increase in residual activity in presence of 30% DMSO or 22% *tert*-butanol, respectively, compared to D93G template protein, but with a catalytic efficiency only reduced by 13% compared to the starting enzyme. In this case, the increased resistance to denaturation in presence of solvent could be related to a combined stabilization of both termini of the protein. Indeed, the amino acid at position 22 is located within 5 Å of residue W260, and within 6 Å of two other residues from the C-term loop of *AnFaeA* (T254 and S255). The restoration of the isoleucine wild-type residue at position 6 in combination with mutation Q22L could therefore improve the non-covalent interaction network that stabilizes both protein extremities.

Independently of their non-maintenance in the most resistant mutants, whether against temperature or solvent, it is interesting to remark that several mutations were concentrated on very specific stretches of sequence. For instance, mutations S33P, I35T and I36F were found in three distinct mutants, selected either for their temperature or solvent resistance. These mutations were never isolated, always combined with at least another mutation, and although they were not studied individually or recovered in second generation StEP mutant, it is tempting to postulate that this section of the esterase, corresponding to the end of the $\alpha 1\beta 2$ loop and the very beginning of sheet $\beta 2$, may play a role in the overall stabilization of the protein. A similar observation can be made with mutations Q185R and F187Y, located at the beginning of $\beta 7$ sheet, at the surface of *AnFaeA*. In a previous directed evolution project of *AnFaeA* by Zhang and coworkers, two mutations affecting these residues had already been pinpointed. In the first small-scale study, the PoPMuSiC algorithm (Dehouck et al., 2011) was applied to select four positions for mutations, among which two were found to modify positively the temperature resistance of the esterase: D93G -that we mutated in our wild-type gene to create the template DNA of our study- and S187F (Zhang and Wu, 2011). This residue at position 187 was actually already a phenylalanine in our starting protein, originating from a different genotype of *Aspergillus niger*. Then, following the construction and the screening of a random *AnFaeA* library in *P. pastoris*, the same authors obtained a mutant displaying a 2.5-fold increase of T_m at 55 °C, that was shown to possess a single mutation, Q185R (Zhang et al., 2012). As a consequence, occurrence in two independent random mutagenesis studies of mutations at identical positions strongly supports that this region plays a key role in fold-stabilisation. Moreover, both substitutions were positively screened again in our second round of evolution (F187Y in mutant B and Q185R in mutant D) confirming this hypothesis. More specifically, concerning mutation Q185R, it was postulated that because the guanidinium group of arginine allows salt-bridges or hydrogen bonds interactions in three possible directions, arginine on surface could increase stability of enzymes (Sokalingam et al., 2012, 2013), an hypothesis that was confirmed in a comparative study between thermophilic and mesophilic proteins

showing that arginine in surface are often beneficial for thermostability while glutamine are deleterious (Yokot et al., 2006).

In conclusion, we have implemented a dual random and rational evolution strategy to create *AnFaeA* mutants with improved thermostability and/or solvent resistance. First, thanks to a robust and accurate screening methodology, two successive cycles of random mutagenesis and DNA shuffling enabled to generate several improved variants, amongst which, the quadruple-mutant N6T-A119V-S163T-V243M displayed the most improved temperature resistance properties, with a T_m increased by 9 °C and a residual activity after heat shock multiplied by 4, compared to the template enzyme. In parallel, mutant N6T-R66Q-Q117E was obtained after rational recombination of some highlighted substitution, and constitutes the esterase with the highest resistance to both tested solvents, DMSO and *tert*-butanol. Interestingly, and though our methodology never aimed at selecting mutants improved for both properties, we could identify, after the StEP recombination, mutant N6T-Q22L-F187Y which shows a better resistance in presence of *tert*-butanol and a better thermostability. These results provide useful information to go further in the engineering of *AnFaeA*, and may open new perspectives towards the use of this feruloyl esterase for synthesis reactions.

REFERENCES

- Antonopoulou, I., Leonov, L., Jütten, P., Cerullo, G., Faraco, V., Papadopoulou, A., Kletsas, D., Ralli, M., Rova, U., Christakopoulos, P., 2017. Optimized synthesis of novel prenyl ferulate performed by feruloyl esterases from *Myceliophthora thermophila* in microemulsions. *Appl. Microbiol. Biotechnol.* 101, 3213–3226.
- Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol. Rev.* 19, 219–237.
- Benoit, I., Asther, M., Sulzenbacher, G., Record, E., Marmuse, L., Parsiegla, G., Gimbert, I., Asther, M., Bignon, C., 2006. Respective importance of protein folding and glycosylation in the thermal stability of recombinant feruloyl esterase A. *FEBS Lett.* 580, 5815–5821.
- Benoit, I., Coutard, B., Oubelaid, R., Asther, M., Bignon, C., 2007. Expression in *Escherichia coli*, refolding and crystallization of *Aspergillus niger* feruloyl esterase A using a serial factorial approach. *Protein Expr. Purif.* 55, 166–174.
- Benoit, I., Danchin, E.G.J., Bleichrodt, R.-J., de Vries, R.P., 2008. Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnol. Lett.* 30, 387–396.
- Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* 70, 493–502.
- Bordes, F., Tarquis, L., Nicaud, J.-M., Marty, A., 2011. Isolation of a thermostable variant of Lip2 lipase from *Yarrowia lipolytica* by directed evolution and deeper insight into the denaturation mechanisms involved. *J. Biotechnol.* 156, 117–124.
- Cambon, E., Piamtongkam, R., Bordes, F., Duquesne, S., Laguerre, S., Nicaud, J.-M., Marty, A., 2010. A new *Yarrowia lipolytica* expression system: An efficient tool for rapid and reliable kinetic analysis of improved enzymes. *Enzyme Microb. Technol.* 47, 91–96.
- Cao, L., Chen, R., Xie, W., Liu, Y., 2015. Enhancing the thermostability of feruloyl esterase EstF27 by directed evolution and the underlying structural basis. *J. Agric. Food Chem.* 63, 8225–8233.
- Crepin, V.F., Faulds, C.B., Connerton, I.F., 2004. Functional classification of the microbial feruloyl esterases. *Appl. Microbiol. Biotechnol.* 63, 647–652.
- De Vries, R.P., Michelsen, B., Poulsen, C.H., Kroon, P.A., Van Den Heuvel, R.H., Faulds, C.B., Williamson, G., Van Den Hombergh, J.P., Visser, J., 1997. The *faeA* genes from *Aspergillus niger* and *Aspergillus tubingensis* encode ferulic acid esterases involved in degradation of complex cell wall polysaccharides. *Appl. Environ. Microbiol.* 63, 4638–4644.
- Declerck, N., Machius, M., Wiegand, G., Huber, R., Gaillardin, C., 2000. Probing structural determinants specifying high thermostability in *Bacillus licheniformis* α -amylase. *J. Mol. Biol.* 301, 1041–1057.
- Dehouck, Y., Kwasigroch, J., Gilis, D., Rooman, M., 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 12, 151.
- Dilokpimol, A., Mäkelä, M.R., Aguilar-Pontes, M.V., Benoit-Gelber, I., Hildén, K.S., de Vries, R.P., 2016. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol. Biofuels* 9.
- Dodd, D., Kocherginskaya, S.A., Spies, M.A., Beery, K.E., Abbas, C.A., Mackie, R.I., Cann, I.K.O., 2009. Biochemical analysis of a β -D-Xylosidase and a bifunctional xylanase-ferulic acid esterase from a xylanolytic gene cluster in *Prevotella ruminicola* 23. *J. Bacteriol.* 191, 3328–3338.

- Doukyu, N., Ogino, H., 2010. Organic solvent-tolerant enzymes. *Biochem. Eng. J.* 48, 270–282.
- Duquesne, S., Bozonnet, S., Bordes, F., Dumon, C., Nicaud, J.-M., Marty, A., 2014. Construction of a highly active xylanase displaying oleaginous yeast: Comparison of anchoring systems. *PLoS ONE* 9, e95128.
- Farinas, E.T., Bulter, T., Arnold, F.H., 2001. Directed enzyme evolution. *Curr. Opin. Biotechnol.* 12, 545–551.
- Fuciños, Atanes, E., López-López, O., Esperanza Cerdán, M., Isabel González-Siso, M., Pastrana, L., Luisa Rúa, M., 2011. Production and characterization of two N-terminal truncated esterases from *Thermus thermophilus* HB27 in a mesophilic yeast: Effect of N-terminus in thermal activity and stability. *Protein Expr. Purif.* 78, 120–130.
- Gherbovet, O., Fauré, R., Ferreira, F., Durand, J., Ragon, M., Hostyn, G., Record, E., Bozonnet, S., O’Donohue, M.J., 2016. Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities. *J. Mol. Catal. B Enzym.* 126, 24–31.
- Graf, E., 1992. Antioxidant potential of ferulic acid. *Free Radic. Biol. Med.* 13, 435–448.
- Hermoso, J.A., Sanz-Aparicio, J., Molina, R., Juge, N., González, R., Faulds, C.B., 2004. The crystal structure of feruloyl esterase A from *Aspergillus niger* suggests evolutive functional convergence in feruloyl esterase family. *J. Mol. Biol.* 338, 495–506.
- Juge, N., Williamson, G., Puigserver, A., Cummings, N.J., Connerton, I.F., Faulds, C.B., 2001. High-level production of recombinant *Aspergillus niger* cinnamoyl esterase (FAEA) in the methylotrophic yeast *Pichia pastoris*. *FEMS Yeast Res.* 1, 127–132.
- Kampa, M., Alexaki, V.-I., Notas, G., Nifli, A.-P., Nistikaki, A., Hatzoglou, A., Bakogeorgou, E., Kouimtoglou, E., Blekas, G., Boskou, D., others, 2003. Antiproliferative and apoptotic effects of selective phenolic acids on T47D human breast cancer cells: potential mechanisms of action. *Breast Cancer Res.* 6, 1.
- Kikugawa, M., Tsutsuki, H., Ida, T., Nakajima, H., Ihara, H., Sakamoto, T., 2016. Water-soluble ferulic acid derivatives improve amyloid- β -induced neuronal cell death and dysmnnesia through inhibition of amyloid- β aggregation. *Biosci. Biotechnol. Biochem.* 80, 547–553.
- Kikuzaki, H., Hisamoto, M., Hirose, K., Akiyama, K., Taniguchi, H., 2002. Antioxidant properties of ferulic acid and its related compounds. *J. Agric. Food Chem.* 50, 2161–2168.
- Le Dall, M.-T., Nicaud, J.-M., Gaillardin, C., 1994. Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr. Genet.* 26, 38–44.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., Chatonnet, A., 2013. ESTHER, the database of the α/β -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res.* 41, D423–D429.
- Levasseur, A., Navarro, D., Punt, P.J., Belaich, J.-P., Asther, M., Record, E., 2005. Construction of engineered bifunctional enzymes and their overproduction in *Aspergillus niger* for improved enzymatic tools to degrade agricultural by-products. *Appl. Environ. Microbiol.* 71, 8132–8140.
- Levasseur, A., Pages, S., Fierobe, H.-P., Navarro, D., Punt, P., Belaich, J.-P., Asther, M., Record, E., 2004. Design and production in *Aspergillus niger* of a chimeric protein associating a fungal feruloyl esterase and a clostridial dockerin domain. *Appl. Environ. Microbiol.* 70, 6984–6991.
- Levasseur, A., Saloheimo, M., Navarro, D., Andberg, M., Monot, F., Nakari-Setälä, T., Asther, M., Record, E., 2006. Production of a chimeric enzyme tool associating the *Trichoderma reesei* swollenin with the *Aspergillus niger* feruloyl esterase A for release of ferulic acid. *Appl. Microbiol. Biotechnol.* 73, 872–880.

- Li, J., Zhang, S., Yi, Z., Pei, X., Wu, Z., 2015. Removal of the free cysteine residue reduces irreversible thermal inactivation of feruloyl esterase: evidence from circular dichroism and fluorescence spectra. *Acta Biochim. Biophys. Sin.* 47, 612–619.
- Li, J.-J., Pei, X.-Q., Zhang, S.-B., Wu, Z.-L., 2015. Improving the thermostability of feruloyl esterase by DNA shuffling and site-directed mutagenesis. *Process Biochem.* 50, 1783–1787.
- Liu, L., Zhang, G., Zhang, Z., Wang, S., Chen, H., 2011. Terminal amino acids disturb xylanase thermostability and activity. *J. Biol. Chem.* 286, 44710–44715.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495.
- Mandrich, L., Merone, L., Pezzullo, M., Cipolla, L., Nicotra, F., Rossi, M., Manco, G., 2005. Role of the N-terminus in enzyme activity, stability and specificity in thermophilic esterases belonging to the HSL Family. *J. Mol. Biol.* 345, 501–512.
- Mathew, S., Abraham, T.E., 2004. Ferulic acid: an antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. *Crit. Rev. Biotechnol.* 24, 59–83.
- Mattila, P., Pihlava, J., Hellström, J., 2005. Contents of phenolic acids, alkyl- and alkenylresorcinols, and avenanthramides in commercial grain products. *J. Agric. Food Chem.* 53, 8290–8295.
- Molina-Espeja, P., Viña-Gonzalez, J., Gomez-Fernandez, B.J., Martin-Diaz, J., Garcia-Ruiz, E., Alcalde, M., 2016. Beyond the outer limits of nature by directed evolution. *Biotechnol. Adv.* 34, 754–767.
- Moore, J., Hao, Z., Zhou, K., Luther, M., Costa, J., Yu, L. (Lucy), 2005. Carotenoid, tocopherol, phenolic acid, and antioxidant properties of maryland-Grown soft wheat. *J. Agric. Food Chem.* 53, 6649–6657.
- Ou, S., Kwok, K.-C., 2004. Ferulic acid: pharmaceutical functions, preparation and applications in foods. *J. Sci. Food Agric.* 84, 1261–1269.
- Rahimzadeh, M., Khajeh, K., Mirshahi, M., Khayatian, M., Schwarzenbacher, R., 2012. Probing the role of asparagine mutation in thermostability of *Bacillus* KR-8104 α -amylase. *Int. J. Biol. Macromol.* 50, 1175–1182.
- Record, E., Asther, M., Sigoillot, C., Pagès, S., Punt, P.J., Delattre, M., Haon, M., van den Hondel, C.A.M.J.J., Sigoillot, J.-C., Lesage-Meessen, L., Asther, M., 2003. Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Appl. Microbiol. Biotechnol.* 62, 349–355.
- Reetz, M.T., Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* 2, 891–903.
- Reetz, M.T., Carballeira, J.D., Vogel, A., 2006. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew. Chem. Int. Ed.* 45, 7745–7751.
- Rice-Evans, C.A., Miller, N.J., Paganga, G., 1996. Structure-antioxidant activity relationships of flavonoids and phenolic acids. *Free Radic. Biol. Med.* 20, 933–956.
- Saulnier, L., Thibault, J.-F., 1999. Ferulic acid and diferulic acids as components of sugar-beet pectins and maize bran heteroxylans. *J. Sci. Food Agric. Journal of the Science of Food and Agriculture* 79, 396–402.
- Serdakowski, A.L., Dordick, J.S., 2008. Enzyme activation for organic solvents made easy. *Trends Biotechnol.* 26, 48–54.
- Shu, Z., Duan, M., Yang, J., Xu, L., Yan, Y., 2009. *Aspergillus niger* lipase: Heterologous expression in *Pichia pastoris*, molecular modeling prediction and the importance of the hinge domains at both sides of the lid domain to interfacial activation. *Biotechnol. Prog.* 25, 409–416.

- Shu, Z., Wu, J., Xue, L., Lin, R., Jiang, Y., Tang, L., Li, X., Huang, J., 2011. Construction of *Aspergillus niger* lipase mutants with oil–water interface independence. *Enzyme Microb. Technol.* 48, 129–133.
- Skjold-Jørgensen, J., Vind, J., Svendsen, A., Bjerrum, M.J., 2014. Altering the activation mechanism in *Thermomyces lanuginosus* lipase. *Biochemistry* 53, 4152–4160.
- Sokalingam, S., Madan, B., Raghunathan, G., Lee, S.-G., 2013. *In silico* study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability. *Biotechnol. Bioprocess Eng.* 18, 18–26.
- Sokalingam, S., Raghunathan, G., Soundrarajan, N., Lee, S.-G., 2012. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE* 7, e40410.
- Srinivasan, M., Sudheer, A.R., Menon, V.P., 2007. Recent advances in indian herbal drug research ferulic acid: therapeutic potential through its antioxidant property. *J. Clin. Biochem. Nutr.* 40, 92–100.
- Sun, J.-Y., Liu, M.-Q., Xu, Y.-L., Xu, Z.-R., Pan, L., Gao, H., 2005. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. *Protein Expr. Purif.* 42, 122–130.
- Sun, S., Shan, L., Liu, Y., Jin, Q., Wang, X., Wang, Z., 2007. A novel, two consecutive enzyme synthesis of feruloylated monoacyl- and diacyl-glycerols in a solvent-free system. *Biotechnol. Lett.* 29, 1947–1950.
- Sun, S., Song, F., Bi, Y., Yang, G., Liu, W., 2013. Solvent-free enzymatic transesterification of ethyl ferulate and monostearin: Optimized by response surface methodology. *J. Biotechnol.* 164, 340–345.
- Sun, S., Zhu, S., Bi, Y., 2014. Solvent-free enzymatic synthesis of feruloylated structured lipids by the transesterification of ethyl ferulate with castor oil. *Food Chem.* 158, 292–295.
- Tabka, M.G., Herpoël-Gimbert, I., Monod, F., Asther, M., Sigoillot, J.C., 2006. Enzymatic saccharification of wheat straw for bioethanol production by a combined cellulase xylanase and feruloyl esterase treatment. *Enzyme Microb. Technol.* 39, 897–902.
- Tang, L., Su, M., Yan, J., Xie, S., Zhang, W., 2015. Lid hinge region of *Penicillium expansum* lipase affects enzyme activity and interfacial activation. *Process Biochem.* 50, 1218–1223.
- Tapin, S., Sigoillot, J.-C., Asther, M., Petit-Conil, M., 2006. Feruloyl esterase utilization for simultaneous processing of nonwood plants into phenolic compounds and pulp fibers. *J. Agric. Food Chem.* 54, 3697–3703.
- Topakas, E., Vafiadi, C., Christakopoulos, P., 2007. Microbial production, characterization and applications of feruloyl esterases. *Process Biochem.* 42, 497–509.
- Topakas, E., Vafiadi, C., Stamatis, H., Christakopoulos, P., 2005. *Sporotrichum thermophile* type C feruloyl esterase (*StFaeC*): purification, characterization, and its use for phenolic acid (sugar) ester synthesis. *Enzyme Microb. Technol.* 36, 729–736.
- Udatha, D.B.R.K.G., Kouskoumvekaki, I., Olsson, L., Panagiotou, G., 2011. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnol. Adv.* 29, 94–110.
- Vafiadi, C., Topakas, E., Alissandratos, A., Faulds, C.B., Christakopoulos, P., 2008. Enzymatic synthesis of butyl hydroxycinnamates and their inhibitory effects on LDL-oxidation. *J. Biotechnol.* 133, 497–504.
- Vafiadi, C., Topakas, E., Christakopoulos, P., Faulds, C.B., 2006. The feruloyl esterase system of *Talaromyces stipitatus*: Determining the hydrolytic and synthetic specificity of *TsFaeC*. *J. Biotechnol.* 125, 210–221.

- Vafiadi, C., Topakas, E., Nahmias, V.R., Faulds, C.B., Christakopoulos, P., 2009. Feruloyl esterase-catalysed synthesis of glycerol sinapate using ionic liquids mixtures. *J. Biotechnol.* 139, 124–129.
- Wang, H.J., Le Dall, M.-T., Waché, Y., Laroche, C., Belin, J.-M., Gaillardin, C., Nicaud, J.-M., 1999. Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast *Yarrowia lipolytica*. *J. Bacteriol.* 181, 5140–5148.
- Wen, Y., Ye, F., Zhu, J., Zhao, G., 2016. Corn starch ferulates with antioxidant properties prepared by N,N'-carbonyldiimidazole-mediated grafting procedure. *Food Chem.* 208, 1–9.
- Wong, D.W.S., 2006. Feruloyl esterase: a key enzyme in biomass degradation. *Appl. Biochem. Biotechnol.* 133, 87–112.
- Yang, H., Zubarev, R.A., 2010. Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides: General. *Electrophoresis* 31, 1764–1772.
- Yin, X., Hu, D., Li, J.-F., He, Y., Zhu, T.-D., Wu, M.-C., 2015a. Contribution of disulfide bridges to the thermostability of a type A feruloyl esterase from *Aspergillus usamii*. *PLOS ONE* 10, e0126864.
- Yin, X., Li, J.-F., Wang, C.-J., Hu, D., Wu, Q., Gu, Y., Wu, M.-C., 2015b. Improvement in the thermostability of a type A feruloyl esterase, *AuFaeA*, from *Aspergillus usamii* by iterative saturation mutagenesis. *Appl. Microbiol. Biotechnol.* 99, 10047–10056.
- Yokot, K., Satou, K., Ohki, S., 2006. Comparative analysis of protein thermostability: differences in amino acid content and substitution at the surfaces and in the core regions of thermophilic and mesophilic proteins. *Sci. Technol. Adv. Mater.* 7, 255–262.
- Zeuner, B., Kontogeorgis, G.M., Riisager, A., Meyer, A.S., 2012. Thermodynamically based solvent design for enzymatic saccharide acylation with hydroxycinnamic acids in non-conventional media. *New Biotechnol.* 29, 255–270.
- Zhang, S., Zhang, K., Chen, X., Chu, X., Sun, F., Dong, Z., 2010. Five mutations in N-terminus confer thermostability on mesophilic xylanase. *Biochem. Biophys. Res. Commun.* 395, 200–206.
- Zhang, S.-B., Pei, X.-Q., Wu, Z.-L., 2012. Multiple amino acid substitutions significantly improve the thermostability of feruloyl esterase A from *Aspergillus niger*. *Bioresour. Technol.* 117, 140–147.
- Zhang, S.-B., Wu, Z.-L., 2011. Identification of amino acid residues responsible for increased thermostability of feruloyl esterase A from *Aspergillus niger* using the PoPMuSiC algorithm. *Bioresour. Technol.* 102, 2093–2096.
- Zhao, H., 2004. Staggered extension process in vitro DNA recombination. *Methods Enzymol.* 388, 42–49.
- Zhao, H., Zha, W., 2006. In vitro “sexual” evolution through the PCR-based staggered extension process (StEP). *Nat. Protoc.* 1, 1865–1871.

Supporting Information

Directed evolution of a type A feruloyl esterase for increased thermostability and organic solvent tolerance

Julien Durand¹, Victor Chaumeton¹, Olga Gherbovet¹, Mélanie Ragon¹, Peter Jütten², Joana L.A. Bras³, Catarina I.P.D. Guerreiro³, Carlos M.G.A. Fontes³, Alexander Piechot², Jean-Marc Nicaud⁴, Régis Fauré¹, Michael J. O'Donohue¹, Sophie Bozonnet^{1,*}

¹ LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

² Taros Chemicals GmbH & Co. KG, Dortmund, Germany

³ NZYTech genes & enzymes, Lisboa, Portugal

⁴ Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

* Corresponding Author

sophie.bozonnet@insa-toulouse.fr

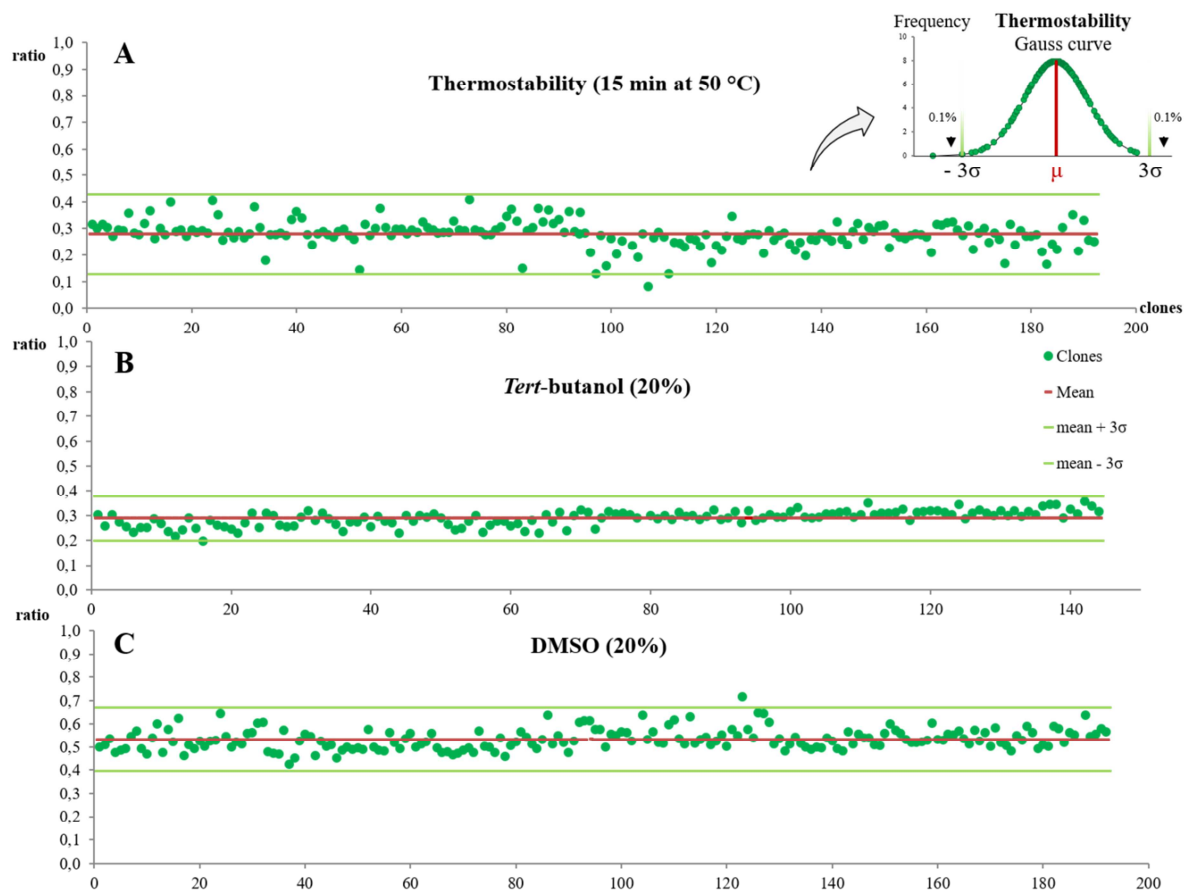


Figure S1. *AnFaeA-D93G* residual activities distribution.

Distribution of *AnFaeA-D93G* residual activities (two microtiter plates, 192 colonies): after heat shock (A), with 20% (v/v) *tert*-butanol (B), or 20% (v/v) of DMSO (C). Red line: mean value; green line: mean value \pm three times the standard deviation, used as threshold for detection of hit mutants.

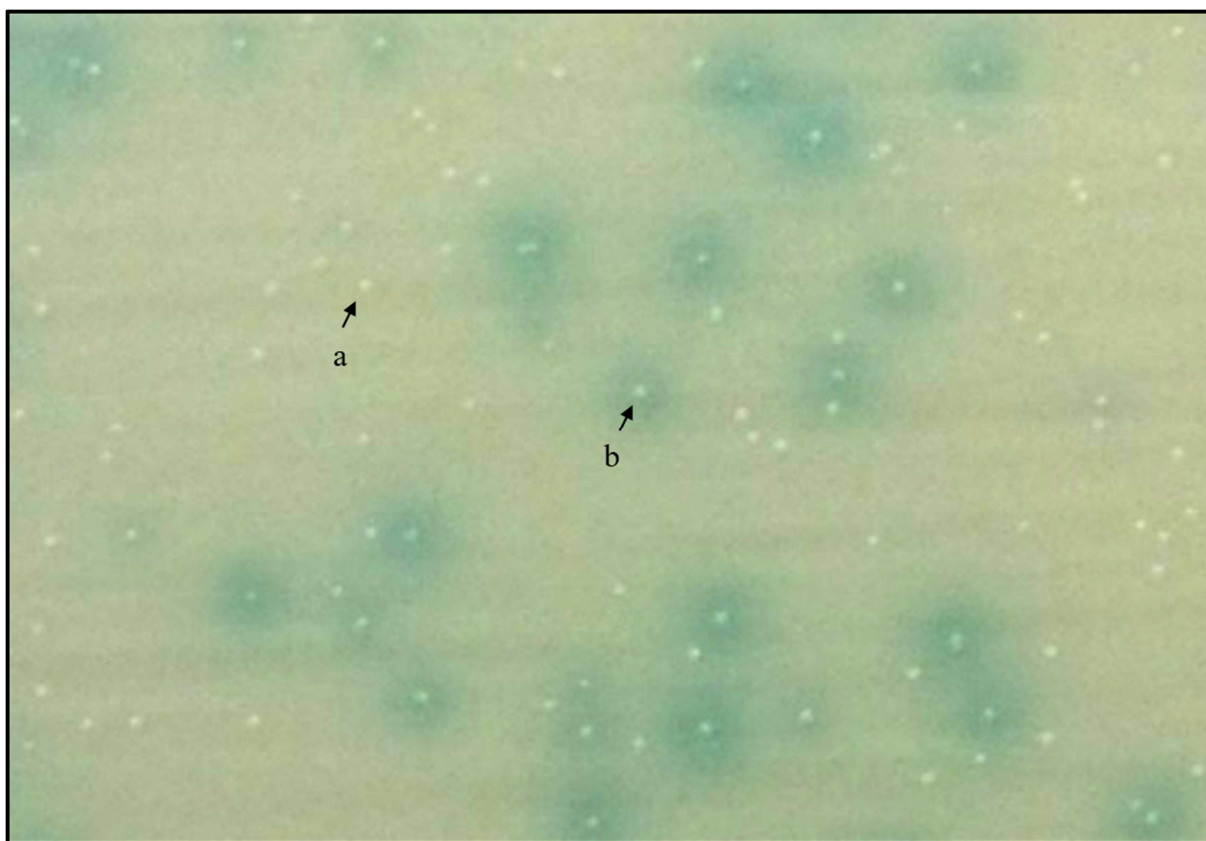


Figure S2. **Solid screening of the random libraries for selection of active clones.**

In vivo primary screening on solid medium containing X-Fe ($150 \mu\text{g}\cdot\text{mL}^{-1}$). (a) White clones, *i.e.* producing an inactive Fae and (b) slight blue colony with halo, indicating an active mutant showing parental-like activity.

Table S1. Comparaison of residual activity values between first screen and second screen, to confirm the twelve selected mutants.

Clone number	First screen (1 clone)	Confirmation (mean of 6 isolated clones)
	Residual activity	Residual activity
After heat shock (15 min at 50 °C)		
D93G	0.27	0.30 ± 0.01
4	0.63	0.59 ± 0.02
5	0.60	0.55 ± 0.05
6	0.82	0.75 ± 0.07
7	0.49	0.51 ± 0.02
8	1.03	0.98 ± 0.02
9	0.54	0.57 ± 0.03
10	0.82	0.83 ± 0.04
11	0,16 (<i>global activity ++</i>)	0.76 ± 0.11
In presence of DMSO 30 % (v/v)		
	20%	25%
D93G	0.62	0.43 ± 0.04
2	0.96	1.18 ± 0.03
3	0.94	0.88 ± 0.07
12	0.70	0.73 ± 0.10
In presence of <i>tert</i>-butanol 20% (v/v)		
D93G	0.33	0.31 ± 0.01
1	0.71	0.76 ± 0.05
2	1.24	1.10 ± 0.09

Table S2. Sequences of fourteen randomly selected mutants from the StEP recombination library

Sequenced mutant	Mutations (original clone bearing the mutation)
1	N6T (8); Q117E (2); F187Y (1); V243M (11)
2	I36F (3); N49I (3); G50S (new mutation) ^a
3	S33P (6); T78N (9); M138L (7)
4	L10I (9); T78N (9); A119V (9)
5	I35T (5); T228I (6)
6	A119V (7); M138L (7)
7	R66Q (2); F187Y (1)
8	L10I (9); D174N (12)
9	N6S (6); S33P (6)
10	A119V (7); Q185R (5)
11	N6I (10)
12	F187Y (1)
13	F187Y (1)
14	S163T (4)

^a Mutation probably generated during the StEP reaction that is not present on any of the twelve coding sequences used as a template.

Chapitre IV : Conclusions-Perspectives

Du fait de la fin programmée des ressources pétrolières, il est devenu impératif de mettre en place des solutions de remplacement plus propres pour la synthèse de molécules issues de la pétrochimie. Pour cela, une part importante des recherches actuelles en biotechnologie se concentre sur la mise en œuvre de solutions enzymatiques qui offrent des alternatives moins polluantes, économiquement viables et qui répondent aux principes de la chimie verte.

Les travaux de cette thèse ont été réalisés dans le cadre de deux projets européens distincts, totalement inscrits dans ce souci de mise au point de procédés durables, l'un sur la conversion d'une GH en transglycosylase pour la synthèse de polyglycosides d'alkyles et l'autre sur l'amélioration de la thermostabilité et de la résistance aux solvants organiques d'une féruloyle estérase. Ces deux projets ont été abordés avec un double intérêt :

- d'une part, comme évoqué précédemment, pour l'aspect applicatif, visant à fournir des enzymes mieux adaptées aux contraintes industrielles pour la synthèse de molécules d'intérêts;
- d'autre part, pour enrichir les connaissances fondamentales communes qui permettent de mieux comprendre les déterminants moléculaires qui contrôlent finement les mécanismes de résistance de la protéine à la température ou au solvant, et la partition de l'activité hydrolytique ou de synthèse des enzymes hydrolytiques.

1. Une nouvelle transglycosylase pour la synthèse de polyglucosides d'alkyle.

Dans le cadre du projet européen BIOSURF, nous avons, par une approche d'évolution semi-rationnelle, obtenu la conversion de l'activité hydrolytique de l'endoglycocéramidase de type II de la bactérie *Rhodococcus* sp. M-777 en activité de synthèse, l'objectif étant de contribuer à rationaliser tant que possible la méthodologie de conversion des rGH en transglycosylases. Cette thématique est étudiée depuis plus de 60 ans (Edelman, 1956), et a débuté peu de temps après la première description des deux mécanismes d'hydrolyse des liaisons glycosidiques par les GH (rétention ou inversion) (Koshland, 1953). Récemment, des avancées significatives ont été faites dans la compréhension des mécanismes régissant la partition de l'activité hydrolytique ou de transglycosylation des rGH par les équipes de Charles Tellier (Osanjo et al., 2007; Teze et al., 2014, 2015; Saumonneau et al., 2015; David et al., 2017), Patrick Adlercreutz (Larsson et al., 2005; Lundemo et al., 2013, 2016) ou Michael O'Donohue (Arab-Jaziri et al., 2013, 2015, Bissaro et al., 2014, 2015a) entre autres, et des revues offrent une analyse approfondie et une mise en perspective de ces avancées (Bissaro et al., 2015b; Abdul Manas et al., 2017). Plusieurs facteurs récurrents ont ainsi été mis en avant lors de ces études. Nous savons qu'il existe deux principales approches pour favoriser la transglycosylation au dépend de l'hydrolyse enzymatique :

- pour la plupart des rGH, l'hydrolyse et la transglycosylation sont des réactions concurrentielles, la proportion de chaque réaction est sous contrôle thermodynamique. La transglycosylation peut être avantagée dans des conditions spécifiques telles qu'une forte concentration en accepteurs ou une faible activité de l'eau par exemple.
- d'autre part, et c'est cette approche qui a été privilégiée lors de ce projet, la réaction de transglycosylation peut être favorisée en modifiant la configuration moléculaire du site actif pour faciliter l'accès (Ochs et al., 2013) et la reconnaissance du substrat accepteur au dépend des molécules d'eau. L'ingénierie fine de la structure protéique des rGH est un moyen efficace d'améliorer la transglycosylation et de supprimer l'hydrolyse tout en conservant des conditions de réaction éco-responsables.

Dans le cas nous concernant, l'approche semi-rationnelle visant à augmenter l'affinité de l'EGCII pour certains accepteurs, nous a permis d'identifier la mutation D311Y qui confère à cette enzyme la capacité de synthétiser efficacement des APG originaux, qui ont été purifiés et caractérisés, et dont les cinétiques de formation ont pu être suivies par RMN. Les rendements de conversion du donneur : β -cellobioside de 2-chloro-4-nitrophenyle (CNP- β -Celb) en produits de synthèse avec le mutant D311Y sont d'au moins 68 % pour les quatre accepteurs testés, atteignant même 93 % avec le 1-hydroxyoctan-4-one comme accepteur. Pour l'accepteur 5-hydroxypentan-2-one, l'efficacité de transfert est près de 7 fois meilleure que pour l'enzyme sauvage- avec un rendement maximum qui passe de 11 à 75 %. Une analyse approfondie par docking des nouvelles interactions du mutant D311Y avec le substrat donneur ou de l'intermédiaire covalent glycosyl-enzyme avec les accepteurs a permis de faire plusieurs observations. Tout d'abord, en présence du donneur CNP- β -Celb, le résidu Y311 s'associe à la Y306 pour former une poche hydrophobe qui permet de loger, par effet de stacking, le groupement partant CNP du donneur, le résidu D-glucosyle localisé dans le sous-site -1 adoptant une distorsion conformationnelle (1S_3) nécessaire à la catalyse. Ce positionnement favorable se traduit alors par la diminution du K_M constatée lors de l'hydrolyse du CNP- β -Celb. En parallèle, il a été observé que la Y311 augmente l'hydrophobicité du sous-site accepteur et permet de nouvelles interactions hydrophobes entre l'enzyme et la chaîne alkyle des accepteurs (Durand et al., 2016). Cela confirme le rôle majeur que jouent les résidus aromatiques et hydrophobes présents dans le sous-site accepteur des rGH dans les interactions enzyme-substrat (Hansson et al., 2001; Ochs et al., 2013). Rétrospectivement, il est intéressant de remarquer que cette logique se retrouve, mais inversée, dans une étude sur la xylanase 10A de *Pseudomonas cellulosa* : le remplacement en alanine de deux résidus tyrosine (220 et 255), situés au niveau des sous-sites accepteurs de l'enzyme provoque d'une part une forte diminution de l'activité de transglycosylation, et d'autre part augmente les k_{cat} et K_M d'un facteur 17 et 10 respectivement (Armand et al., 2001). La substitution ou l'incorporation de résidus aromatiques et hydrophobes provoquent de fortes modifications des paramètres cinétiques des enzymes et influent fortement sur l'activité de synthèse.

Au vu des résultats probants obtenus avec le mutant D311Y, il serait intéressant de poursuivre cette approche semi-rationnelle d'ingénierie du ratio T/H, en continuant d'explorer le rôle complémentaire que pourrait jouer les autres acides aminés initialement sélectionnés pour cette étude (E179, N279 ou Q280). Même si aucun mutant n'a été sélectionné suite au criblage des trois banques à saturations concernant ces positions, la réorganisation des interactions engendrées par l'insertion du résidu Y311 peut permettre de nouvelles options configurationnelles de cette zone de reconnaissance des accepteurs, potentiellement bénéfiques pour l'activité de synthèse. En outre, alors que la mutation D311 n'aurait dû être positive qu'avec un seul type d'accepteur -diol ou hydroxycétone-, nos résultats montrent bien que la modification du site actif est plus globale car les deux types d'accepteur ont pu être transglycosylés efficacement. Pour ce faire, selon la stratégie itérative CAST (Combinatorial Active site Saturation Test) (Reetz et al., 2006), le mutant D311Y servirait de base pour générer une nouvelle banque à saturation sur un des trois autres résidus et le mutant présentant le meilleur ratio T/H serait alors sélectionné comme mutant de départ pour générer une nouvelle banque à saturation sur la position suivante. Cette méthode permet de couvrir un espace de séquence réduit et pertinent pour identifier le mutant présentant le ratio T/H optimum pour la synthèse d'APG.

D'un point de vue plus applicatif, pour produire des APG plus « simples » et rendre cette approche de synthèse enzymatique plus compétitive, il serait judicieux :

1) de tester des substrats donneurs naturels tels que des lacto-oligosaccharides, plus disponibles et moins chers que le CNP- β -Celb, mais moins favorables car ils ne bénéficient pas du bon groupement partant que constitue le 2-chloro-4-nitrophenol,

et 2) d'utiliser des accepteurs plus abondants et donc moins onéreux comme les alcanols que nous avons testés en début de projet. Même si nous avons montré que ces accepteurs ne sont pas acceptés par l'enzyme sauvage, dans la mesure où la mutation D311Y contribue à améliorer la reconnaissance du substrat donneur grâce la poche hydrophobe formée par les résidus Y306 et Y311, on peut penser que l'activité de transfert sur ces alcanols sera en conséquence améliorée.

Dans ces deux cas, les banques à saturation générées par la méthode CAST, présentées précédemment, pourrait être criblées sur ces substrats différents, dans l'objectif d'identifier des mutants conservant un ratio T/H élevé, voire amélioré.

Outre la synthèse d'APG, cette enzyme peut également être potentiellement utile pour la synthèse de glycosphingolipides (GSL) à façon. La production de ces molécules représente un enjeu majeur pour la recherche médicale car elles sont impliquées dans de nombreux mécanismes physiologiques, et par voie de conséquence sont aussi au cœur de nombreuses pathologies, telles que les cancers, le paludisme ou les maladies auto-immunes (Wu et al., 2008). La version glycosynthase de l'EGC II a déjà démontré une capacité prometteuse pour la synthèse de ces molécules (Vaughan et al., 2006; Hancock et al., 2009; Rich et al., 2011). Il faudrait donc, dans un premier temps, tester si le mutant

D311Y permet l'emploi de céramides comme molécules acceptrices avec des rendements comparables ou meilleurs qu'avec l'EGC II glycosynthase.

Certains APG produits lors de cette étude -lorsque par exemple des dérivés 1,3-diols ont été utilisés comme accepteurs- possèdent une partie aglycone présentant une fonction hydroxyle libre et donc disponible, qui peut donc être à son tour fonctionnalisée par voie chimique (nécessite la protection/déprotection des hydroxyles présents sur la partie cellobioside) ou enzymatique : une estérification par un acide gras conduira à une molécule plus hydrophobe, et l'on peut même envisager l'usage d'un acide férulique pour générer des molécules aux propriétés antioxydantes potentielles. Ces molécules complexes pourraient, après un travail de caractérisation pointue, trouver des applications très spécifiques en tant que molécules à forte valeur ajoutée.

Ainsi, les résultats obtenus sur cette enzyme élargissent notre compréhension fondamentale sur la conversion d'enzymes hydrolytiques en outils de synthèse enzymatique et valident l'approche semi-rationnelle, qui s'appuie sur l'identification de résidus clés, associée à l'exploration exhaustive de la diversité des acides aminés que permet la mutagenèse à saturation. En outre, ce travail concret d'ingénierie enzymatique aboutit à un nouveau biocatalyseur qui pourrait offrir une alternative aux réactions chimiques pour la synthèse de nouveaux polyglucosides d'alkyles.

2. Une feruloyle estérase plus résistante à la température et aux solvants

Dans le cadre du second projet européen OPTIBIOCAT, nous avons travaillé sur une feruloyle estérase, une estérase carboxylique qui hydrolyse la liaison entre l'acide férulique et les résidus L-arabinofuranosyles des arabinoxylyanes. Certaines de ces enzymes ont une capacité naturelle à catalyser le transfert de résidus féruloyles et/ou d'acide férulique sur des molécules hydroxylées acceptrices, pouvant avoir des propriétés intéressantes, notamment antioxydantes, valorisables dans les domaines de l'alimentation, des produits pharmaceutiques ou des cosmétiques. La mise en œuvre d'une démarche d'évolution aléatoire s'appuyant sur l'établissement d'une méthode robuste de criblage à haut-débit a permis d'identifier plusieurs combinaisons de mutations conférant à la feruloyle estérase A d'*A. niger* une meilleure résistance aux solvants et/ou une meilleure thermorésistance, des conditions physico-chimiques qui favorisent l'activité de transféruoylation chez ces enzymes. Des substrats chromogènes *ad hoc* ont été synthétisés, caractérisés et validés pour le criblage à haut-débit de ces bibliothèques (Gherbovet et al., 2016). Un premier cycle d'évolution a généré près de 10 000 mutants parmi lesquels 12 ont été isolés pour leur gain de stabilité thermique et/ou leur meilleure activité résiduelle en présence de solvants. Lors d'un deuxième cycle d'évolution, une nouvelle diversité a été générée par recombinaison des 12 séquences codantes par un « processus d'extension échelonnée » (Staggered Extension Process, StEP) et a été criblée. Un total de 23 mutations a ainsi été brassé efficacement et de manière homogène, avec une qualité rarement observée dans la littérature.

Les mutations les plus intéressantes ont été étudiées individuellement et recombinaison par mutagenèse dirigée pour générer de nouvelles combinaisons encore améliorées.

C'est l'analyse approfondie de ces mutations, isolées ou associées, qui a permis de révéler l'importance de certains résidus vis-à-vis des propriétés de l'*AnFaeA*. Ainsi, la mutation N6T accroît considérablement la thermostabilité de l'enzyme de départ, cette seule mutation permettant de faire passer la température de fusion (T_m) de l'enzyme de 48,7 à 55,5 °C. Combinée à d'autres mutations, le T_m atteint 57 °C, soit un gain de 8,3 °C par rapport à l'enzyme native. L'effet est également observé sur l'activité résiduelle suite à un choc thermique, puisqu'initialement l'enzyme parent (*AnFaeA* D93G) possède une activité résiduelle de 17 % à la suite d'un choc thermique de 5 min à 50 °C alors que le mutant N6T-A119V-S163T-V243M conserve 100 % d'activité dans ces conditions. L'effet de la mutation N6T dans la partie N-terminale de l'enzyme peut être expliqué au vu des nombreux travaux qui démontrent l'importance que joue cette région dans la stabilité de nombreuses enzymes (Veprintsev et al., 1999; Sun et al., 2005; Mandrich et al., 2005; Zhang et al., 2010). L'observation de la localisation des mutations apportant un gain de thermostabilité significatif (N6T, A119V, S163T et Q185R) indique qu'elles sont éloignées du site actif; cette observation paraît logique si l'on accepte l'hypothèse que la stabilité à la température est corrélée à une rigidification globale de l'enzyme (Karshikoff et al., 2015), mais qui ne devrait pas affecter le site actif où une perte de flexibilité pourrait compromettre le positionnement des molécules substrats et risquer d'inactiver l'enzyme. D'ailleurs, la relative grande taille des enzymes par rapport à la taille très réduite du site actif peut en partie s'expliquer par le fait que des mutations sur les structures éloignées du site actif, qui jouent alors le rôle de zones tampons, permettent aux enzymes de s'adapter à de nouvelles pressions de sélection par les processus naturels d'évolution sans modification du site catalytique (Arnold et al., 2001; Bloom and Arnold, 2009). Toutefois, il a été démontré pour la première fois en 2014, lors du criblage de bibliothèques obtenues par mutagenèse à saturation sur des résidus à haut B-facteur situés dans une zone de 10 Å autour de la sérine catalytique de la lipase B de *Candida antarctica* (CAL-B), qu'augmenter la rigidité du site actif peut améliorer la stabilité de cette enzyme. Les structures cristallines de Cal-B et de ces mutants ont été obtenues et permettent de confirmer la nette diminution des valeurs de B-facteur pour plusieurs acides aminés du site actif grâce à de nouvelles liaisons hydrogènes. Les auteurs soulignent que même si cette région est critique du point de vue de l'activité enzymatique, il est nécessaire que le site actif possède une certaine rigidité pour conserver sa conformation à de hautes températures (Xie et al., 2014). C'est probablement, concernant l'*AnFaeA*, le cas du résidu V243, situé sur une boucle surplombant le site actif, à une distance de 6 Å de la sérine catalytique. En effet, il présente une valeur de B-facteur élevée, et sa mutation en méthionine, même si elle semble n'avoir qu'un effet mineur sur la thermostabilité lorsqu'elle est isolée, se retrouve dans les mutants les plus résistants à la température (N6T-S163T-V243M et N6T-A119V-S163T-V243M). A noter que la position D311 de l'EGC II est également proche du site catalytique et présente une valeur de B-facteur

élevée : sa mutation en tyrosine conduit à une augmentation de 160% de sa demi-vie à 37 °C, ce qui tend à corroborer cette théorie.

C'est la mutation R66Q, localisée au niveau de la charnière N-term du couvercle, qui, dans notre inventaire de mutations isolées, génère le plus important gain en termes d'activité résiduelle en présence de solvant. Elle permet de doubler l'activité résiduelle en présence de 30 % de DMSO ou 22 % de *tert*-butanol par rapport à l'enzyme de départ dans les mêmes conditions. Cette mutation est localisée au niveau de la charnière du couvercle surplombant le site actif (résidus 61 à 67 pour l'*AnFaeA*) ; une autre mutation, identifiée dans un mutant également résistant au solvant, concerne le résidu N49, qui est au contact direct du R66 ($< 3,5 \text{ \AA}$), ce qui semble indiquer que cette zone joue un rôle prépondérant dans l'activation de ces enzymes en présence de solvants. Dans ce deuxième cas, un acide aminé polaire non chargé (N) se trouve remplacé par un résidu possédant une chaîne latérale plus courte et hydrophobe (I), ce qui réduit la possibilité d'interactions électrostatiques au profit d'interactions hydrophobes. Le rôle clé de cette zone d'articulation pour les enzymes de cette famille a récemment été démontré sur une lipase de *Penicillium expansum* qui est structurellement très proche de la féruloyle estérase A (Tang et al., 2015). Des mutations au niveau de la région charnière, sur des résidus ciblés pour leur B-facteur élevé, ont permis d'augmenter drastiquement l'activité de la lipase, que ce soit vers un substrat natif (tributyryne, huile d'olive) ou un analogue chromogénique (*pNP*-palmitate).

Dans le domaine d'étude des lipases, c'est la nature même du "lid", le couvercle surplombant le site actif des α/β -hydrolases, qui est étudié le plus fréquemment, notamment pour son rôle dans les phénomènes d'activation en présence de solvants, à l'interface eau/huile (« activation d'interface ») (Sarda and Desnuelle, 1958; Brzozowski et al., 1991), ou pour son implication dans la résistance à la dénaturation thermique (Santarossa et al., 2005; Dror et al., 2014). Pour les lipases de la famille des Lipases3 (base de données ESTHER), le couvercle est constitué d'une l'hélice α très mobile (Peters et al., 1996; Bian et al., 2010; Barbe et al., 2011), mais peu conservée. Par contre, on peut constater que contrairement à la région charnière coté C-term ou au couvercle proprement dit, la région charnière N-term présente une séquence d'acides aminés bien conservée chez les féruloyle estérases comme chez les lipases (Figure 1), qui ne diverge nettement que sur les positions 62 et 63 (numérotation de l'*AnFaeA*). Elles sont en effet occupées exclusivement par une isoleucine et une thréonine chez les féruloyle estérases alors que l'on retrouve majoritairement une tyrosine et une valine respectivement chez les lipases.



Figure 1 : Séquences consensus correspondant à la charnière N-term du couvercle (résidus 61 à 67 pour l'*AnFaeA*) obtenues à partir de l'alignement de A) 5 séquences de féruloyle-estérase et B) 42 séquences de lipases appartenant à la famille des Lipases3 dans la base de données ESTHER. ; les séquences annotées comme putatives ou hypothétiques ont été retirées.

La création de mutants ciblés sur ces résidus, par exemple à travers une démarche de mutagenèse à saturation, constitue une perspective intéressante pour l'étude de l'impact de cette charnière sur le phénomène d'activation d'interface, qui n'est pas présent chez les Fae. La démarche inverse a souvent été explorée -mutation dans une lipase de résidus de la charnière ou du lid par un acide aminé trouvé chez une féruloyle estérase (Shu et al., 2009, 2011) - ou lors d'une étude où l'hélice α surplombant le site actif d'une lipase de *Rhizopus chinensis* a été remplacée par l'hélice α de l'*AnFaeA* et qui a permis d'orienter la spécificité de substrat de cette lipase chimère vers des substrats plus hydrophiles (Yu et al., 2014), mais elle n'a pas été appliquée dans le sens de l'amélioration des Fae. Incorporer une charnière dont la séquence serait plus proche de celle des lipases pourrait permettre de conserver une meilleure activité des Fae en présence de solvants et/ou de leur conférer une capacité "d'activation à l'interface", et présenterait une méthode d'ingénierie généralisable à d'autres Fae ou autres enzymes possédant une structure similaire.

3. Vers une féruloyle estérase améliorée pour la synthèse de molécules d'intérêt ?

Au cours de ces travaux de thèse, j'ai donc eu l'occasion de travailler sur deux enzymes hydrolytiques, une GH (EGC II) et une estérase (*AnFaeA*). Dans les deux projets, le but final, plus ou moins directement, est de pouvoir générer des variants enzymatiques, utilisables dans des réactions de synthèse, pour la production de molécules innovantes à partir d'hydrates de carbone et autres composés carbonés issus de la biomasse, dont la valorisation par des procédés doux présente un grand intérêt (Colonna, 2012).

De façon pragmatique, notre tâche sur l'estérase a consisté d'abord à créer des mutants résistants à la température et/ou aux solvants, deux paramètres physico-chimiques importants pour le déploiement

facilité de ces enzymes dans des réactions de synthèse. Toutefois, sur la base des résultats obtenus, il est d'ores et déjà intéressant de réfléchir aux moyens de développer les capacités de transféruloylation de ces mutants. Ainsi, est-ce que les connaissances acquises sur la conversion des rGH en transglycosylases sont transposables aux estérases ? Ces deux familles d'enzymes partagent un mécanisme moléculaire d'action réalisé en deux étapes, procédant via l'établissement d'un intermédiaire covalent substrat-enzyme et qui peut s'orienter soit vers l'hydrolyse, soit vers le transfert du groupement glycosyle ou féruloyle sur un groupement hydroxyle d'une molécule acceptrice (Adlercreutz, 2017). Alors que plusieurs études se sont penchées sur l'élucidation des déterminants impliqués dans la partition de l'activité T/H pour les GH, les données sur des démarches similaires pour des lipases ou des estérases sont moins nombreuses. La principale raison est que dans le cas des lipases, la mise en œuvre de réactions de transestérification avec ces catalyseurs natifs est naturellement plus favorable que pour les rGH (Debuissy et al., 2016). En effet, la plupart des lipases sont capables de travailler dans des environnements riches en acides gras libres (Canet et al., 2016), en présence de solvants (Gagnon and Vasudevan, 2011), où, par contrôle thermodynamique, les réactions d'hydrolyse ne seront pas favorisées. Ainsi, le simple contrôle des conditions de réactions, de la composition du milieu réactionnel peut suffire pour atteindre des rendements de transestérification conséquents avec des enzymes natives (Li et al., 2006). Pour autant, une démarche d'ingénierie intelligente semble une solution également adaptée pour augmenter les capacités de transfert de ce type d'enzymes.

La stratégie semi-rationnelle mise au point sur les glycoside hydrolases s'appuie sur deux piliers :

- d'une part diminuer l'activité hydrolytique de l'enzyme ciblée, soit en utilisant au départ un couple donneur/enzyme non naturel et peu productif (cas de l'EGC II et du CNP- β -Celb), soit en déstabilisant par ingénierie le site actif de l'enzyme (cas de la *TxAbf* sur *pNP- α -L-Araf*),
- et d'autre part améliorer la reconnaissance de l'accepteur par l'enzyme afin de favoriser les réactions de transfert au dépend de l'hydrolyse.

Dans un objectif de transposition de cette démarche pour faire de l'*AnFaeA* un outil efficace de transféruloylation, le mutant R66Q pourrait faire un bon candidat de départ. En effet, ce mutant, en plus de mieux tolérer la présence de solvant, présente une activité hydrolytique diminuée d'un facteur 13 par rapport à l'enzyme de départ -principalement à cause d'une diminution du k_{cat} - ce qui constitue un point de départ intéressant dans l'optique de favoriser l'activité de synthèse au dépend de l'activité hydrolytique. Des expérimentations sont en cours pour tester si l'activité de transfert de ce mutant est améliorée. A partir de là, de la même manière que ce qui a été fait pour l'EGC II, il serait intéressant, par une approche semi-rationnelle visant des résidus impliqués dans la reconnaissance de l'accepteur, de rendre l'enzyme plus affine pour une molécule acceptrice. L'identification de ces résidus pourrait se faire en dockant dans le site actif de l'*AnFaeA* les molécules acceptrices visées (glycérol, butanol,

L-arabinose par exemple), sur la base -quand cela est possible- de structures résolues de lipases homologues en complexe avec des molécules proches (Schalk-Hihi et al., 2011). L'objectif serait de pouvoir utiliser ce nouvel outil de synthèse enzymatique dans des conditions aqueuses et s'affranchir autant que possible de l'utilisation de solvants, généralement un mélange de *n*-hexane, *tert*-butanol et eau en faible proportion, qui sont aujourd'hui utilisés pour les réactions de synthèse de composés féruloylés (Couto et al., 2011; Vafiadi et al., 2006, 2007). Cette approche permettrait de générer des banques d'enzymes plus ciblées pour réduire l'espace de séquences à cribler (Dalby, 2011).

Lors du criblage effectué sur la banque de mutants de 1^{ère} génération de l'*AnFaeA*, nous avons mesuré l'activité en milieu liquide de plus de 2 000 mutants. Nous n'avons alors retenu que ceux qui affichaient des capacités de résistance à la température ou aux solvants, mais certains présentaient une activité hydrolytique clairement affectée et constituent donc des cibles intéressantes pour notre projet. Ils seront prochainement évalués pour leur capacité de transfert de l'acide férulique sur des molécules acceptrices, en mesurant simplement dans un premier temps leur potentiel à être activés en présence d'accepteur. Les meilleurs candidats seront dans un second temps soumis à un processus d'ingénierie rationnelle ou semi-rationnelle du sous-site accepteur pour optimiser l'affinité avec l'accepteur choisi.

Une approche différente, moins rationnelle et plus aléatoire, serait d'utiliser la proximité structurale de l'*AnFaeA* avec les lipases, pour générer des protéines hybrides grâce à la technique de recombinaison aléatoire de type SCRATCHY, adaptée aux séquences peu homologues (Lutz et al., 2001). Le choix pourrait se porter par exemple sur la lipase de *Penicillium expansum* qui partage 24 % d'identité de séquence avec l'*AnFaeA* (Bian et al., 2010). En effet, cette enzyme possède une grande promiscuité réactionnelle, une activation d'interface et une activité de synthèse en présence de solvants ou de liquides ioniques qui affiche de bons rendements (Zhang et al., 2011). Pour cribler ces enzymes chimères construites aléatoirement, on pourrait envisager de réaliser des réactions en milieu biphasique ou en émulsion, avec un substrat chromogénique féruloylé comme donneur, permettant de mesurer les activités avec et sans accepteur hydrophobe -par exemple des alcools à longue chaîne carbonée. Parmi les clones possédant la plus grande activation en présence d'accepteurs, pourrait se trouver un mutant possédant potentiellement le sous-site donneur de la féruloyle estérase A, spécifique à l'acide férulique, combiné au sous-site accepteur de la lipase, qui lui sera affin pour des accepteurs plus hydrophobes. Si cette enzyme chimère possède le couvercle et la charnière de la lipase lui conférant une activation d'interface, cela pourrait fournir un outil de synthèse enzymatique très intéressant pour répondre aux attentes des nombreuses applications souhaitant mettre à profit les propriétés bénéfiques de l'acide férulique (Palani Swamy and Govindaswamy, 2015).

Les travaux décrits dans cette thèse, bien que issus de projets différents, illustrent une nouvelle fois la puissance des approches d'évolution des protéines pour répondre à des questions pragmatiques -

obtenir un meilleur catalyseur pour une propriété donnée- tout en contribuant significativement à une meilleure compréhension des relations structure-activités de ces protéines. Scrupuleusement choisies en fonction des données disponibles et des questions posées, ces méthodologies d'ingénierie combinatoire permettent -si l'on dispose de cribles efficaces et robustes- quasiment toujours d'isoler des mutants améliorés et informatifs (Denard et al., 2015).

Nous avons ici déployé deux stratégies différentes, reposant chacune sur l'exploration d'un espace de séquence différent (limité par mutagenèse à saturation ou beaucoup plus vaste par mutagenèse aléatoire). Dans les deux cas, même si la diversité créée et criblée pour l'*AnFaeA* est infime au regard des possibilités existantes, nous avons été capables de fournir des variants enzymatiques répondant au cahier des charges initial. L'analyse et la compréhension des effets de ces mutations permettent enfin d'apporter une pierre au vaste édifice des connaissances concernant le domaine de l'ingénierie des enzymes de synthèse.

Références :

- Abdul Manas, N.H., Md Illias, R., Mahadi, N.M., 2017. Strategy in manipulating transglycosylation activity of glycosyl hydrolase for oligosaccharide production. *Crit. Rev. Biotechnol.* 1–22.
- Adlercreutz, P., 2017. Comparison of lipases and glycoside hydrolases as catalysts in synthesis reactions. *Appl. Microbiol. Biotechnol.* 101, 513–519.
- Arab-Jaziri, F., Bissaro, B., Dion, M., Saurel, O., Harrison, D., Ferreira, F., Milon, A., Tellier, C., Fauré, R., O'Donohue, M.J., 2013. Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.* 30, 536–544.
- Arab-Jaziri, F., Bissaro, B., Tellier, C., Dion, M., Fauré, R., O'Donohue, M.J., 2015. Enhancing the chemoenzymatic synthesis of arabinosylated xylo-oligosaccharides by GH51 α -L-arabinofuranosidase. *Carbohydr. Res.* 401, 64–72.
- Armand, S., Andrews, S.R., Charnock, S.J., Gilbert, H.J., 2001. Influence of the aglycone region of the substrate binding cleft of *Pseudomonas* Xylanase 10A on catalysis. *Biochemistry* 40, 7404–7409.
- Arnold, F.H., Wintrode, P.L., Miyazaki, K., Gershenson, A., 2001. How enzymes adapt: lessons from directed evolution. *Trends Biochem. Sci.* 26, 100–106.
- Barbe, S., Cortés, J., Siméon, T., Monsan, P., Remaud-Siméon, M., André, I., 2011. A mixed molecular modeling-robotics approach to investigate lipase large molecular motions. *Proteins Struct. Funct. Bioinforma.* 79, 2517–2529.
- Bian, C., Yuan, C., Chen, L., Meehan, E.J., Jiang, L., Huang, Z., Lin, L., Huang, M., 2010. Crystal structure of a triacylglycerol lipase from *Penicillium expansum* at 1.3 Å determined by sulfur SAD. *Proteins Struct. Funct. Bioinforma.* 78, 1601–1605.
- Bissaro, B., Durand, J., Biarnés, X., Planas, A., Monsan, P., O'Donohue, M.J., Fauré, R., 2015a. Molecular design of non-Leloir furanose-transferring enzymes from an α -L-arabinofuranosidase: a rationale for the engineering of evolved transglycosylases. *ACS Catal.* 5, 4598–4611.
- Bissaro, B., Monsan, P., Fauré, R., O'Donohue, M.J., 2015b. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* 467, 17–35.
- Bissaro, B., Saurel, O., Arab-Jaziri, F., Saulnier, L., Milon, A., Tenkanen, M., Monsan, P., O'Donohue, M.J., Fauré, R., 2014. Mutation of a pH-modulating residue in a GH51 α -L-arabinofuranosidase leads to a severe reduction of the secondary hydrolysis of transfuranosylation products. *Biochim. Biophys. Acta BBA - Gen. Subj.* 1840, 626–636.
- Bloom, J.D., Arnold, F.H., 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci.* 106, 9995–10000.
- Brzozowski, A.M., Derewenda, U., Derewenda, Z.S., Dodson, G.G., Lawson, D.M., Turkenburg, J.P., Bjorkling, F., Høge-Jensen, B., Patkar, S.A., Thim, L., 1991. A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature* 351, 491–494.
- Canet, A., Bonet-Ragel, K., Benaiges, M.D., Valero, F., 2016. Lipase-catalysed transesterification: Viewpoint of the mechanism and influence of free fatty acids. *Biomass Bioenergy* 85, 94–99.
- Colonna, P., 2012. Le carbone renouvelable dans les systèmes alimentaires, énergétiques et chimiques : Leçon inaugurale prononcée le jeudi 15 décembre 2011. Collège de France.
- Couto, J., St-Louis, R., Karboune, S., 2011. Optimization of feruloyl esterase-catalyzed synthesis of feruloylated oligosaccharides by response surface methodology. *J. Mol. Catal. B Enzym.* 73, 53–62.

- Dalby, P.A., 2011. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol.* 21, 473–480.
- David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.-H., Tellier, C., 2017. Internal water dynamics control the transglycosylation/hydrolysis balance in the agarase (AgaD) of *Zobellia galactanivorans*. *ACS Catal.* 7, 3357–3367.
- Debuissy, T., Pollet, E., Avérous, L., 2016. Enzymatic synthesis of a bio-based copolyester from poly(butylene succinate) and poly((R)-3-hydroxybutyrate): study of reaction parameters on the transesterification rate. *Biomacromolecules* 17, 4054–4063.
- Denard, C.A., Ren, H., Zhao, H., 2015. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.* 25, 55–64.
- Durand, J., Biarnés, X., Watterlot, L., Bonzom, C., Borsenberger, V., Planas, A., Bozonnet, S., O'Donohue, M.J., Fauré, R., 2016. A single point mutation alters the transglycosylation/hydrolysis partition, significantly enhancing the synthetic capability of an *endo*-glycoceramidase. *ACS Catal.* 8, 8264–8275.
- Edelman, J., 1956. The formation of oligosaccharides by enzymic transglycosylation. *Adv. Enzymol. Relat. Subj. Biochem.* 17, 189–232.
- Gagnon, M.D., Vasudevan, P.T., 2011. Effects of solvent and enzyme source on transesterification activity. *Energy Fuels* 25, 4669–4674.
- Gherbovet, O., Fauré, R., Ferreira, F., Durand, J., Ragon, M., Hostyn, G., Record, E., Bozonnet, S., O'Donohue, M.J., 2016. Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities. *J. Mol. Catal. B Enzym.* 126, 24–31.
- Hancock, S.M., Rich, J.R., Caines, M.E.C., Strynadka, N.C.J., Withers, S.G., 2009. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nat. Chem. Biol.* 5, 508–514.
- Hansson, T., Kaper, T., van der Oost, J., de Vos, W.M., Adlercreutz, P., 2001. Improved oligosaccharide synthesis by protein engineering of β -glucosidase CelB from hyperthermophilic *Pyrococcus furiosus*. *Biotechnol. Bioeng.* 73, 203–210.
- Karshikoff, A., Nilsson, L., Ladenstein, R., 2015. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J.* 282, 3899–3917.
- Koshland, D.E., 1953. Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev.* 28, 416–436.
- Larsson, J., Svensson, D., Adlercreutz, P., 2005. α -Amylase-catalysed synthesis of alkyl glycosides. *J. Mol. Catal. B Enzym.* 37, 84–87.
- Li, L., Du, W., Liu, D., Wang, L., Li, Z., 2006. Lipase-catalyzed transesterification of rapeseed oils for biodiesel production with a novel organic solvent as the reaction medium. *J. Mol. Catal. B Enzym.* 43, 58–62.
- Lundemo, P., Adlercreutz, P., Karlsson, E.N., 2013. Improved transferase/hydrolase ratio through rational design of a family 1 β -Glucosidase from *Thermotoga neapolitana*. *Appl. Environ. Microbiol.* 79, 3400–3405.
- Lundemo, P., Karlsson, E.N., Adlercreutz, P., 2016. Eliminating hydrolytic activity without affecting the transglycosylation of a GH1 β -glucosidase. *Appl. Microbiol. Biotechnol.*
- Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., Benkovic, S.J., 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proc. Natl. Acad. Sci.* 98, 11248–11253.
- Mandrich, L., Merone, L., Pezzullo, M., Cipolla, L., Nicotra, F., Rossi, M., Manco, G., 2005. Role of the N-terminus in enzyme activity, stability and specificity in thermophilic esterases belonging to the HSL family. *J. Mol. Biol.* 345, 501–512.

- Ochs, M., Belloy, N., Dauchez, M., Muzard, M., Plantier-Royon, R., Rémond, C., 2013. Role of hydrophobic residues in the aglycone binding subsite of a GH39 β -xylosidase in alkyl xylosides synthesis. *J. Mol. Catal. B Enzym.* 96, 21–26.
- Osanjo, G., Dion, M., Drone, J., Solleux, C., Tran, V., Rabiller, C., Tellier, C., 2007. Directed evolution of the α -L-Fucosidase from *Thermotoga maritima* into an α -L-Transfucosidase. *Biochemistry* 46, 1022–1033.
- Palani Swamy, S. kumaran, Govindaswamy, V., 2015. Therapeutic properties of ferulic acid and bioavailability enhancement through feruloyl esterase. *J. Funct. Foods* 17, 657–666.
- Peters, G.H., Olsen, O.H., Svendsen, A., Wade, R.C., 1996. Theoretical investigation of the dynamics of the active site lid in *Rhizomucor miehei* lipase. *Biophys. J.* 71, 119–129.
- Reetz, M.T., Wang, L.-W., Bocola, M., 2006. Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew. Chem.* 118, 1258–1263.
- Rich, J.R., Cunningham, A.-M., Gilbert, M., Withers, S.G., 2011. Glycosphingolipid synthesis employing a combination of recombinant glycosyltransferases and an endoglycoceramidase glycosynthase. *Chem. Commun.* 47, 10806.
- Sarda, L., Desnuelle, P., 1958. Action de la lipase pancréatique sur les esters en émulsion. *Biochim. Biophys. Acta* 30, 513–521.
- Saumonneau, A., Champion, E., Peltier-Pain, P., Molnar-Gabor, D., Hendrickx, J., Tran, V., Hederos, M., Dekany, G., Tellier, C., 2015. Design of an α -L-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology* cww099. 26, 261–269.
- Schalk-Hihi, C., Schubert, C., Alexander, R., Bayoumy, S., Clemente, J.C., Deckman, I., DesJarlais, R.L., Dzordzorme, K.C., Flores, C.M., Grasberger, B., Kranz, J.K., Lewandowski, F., Liu, L., Ma, H., Maguire, D., Macielag, M.J., McDonnell, M.E., Mezzasalma Haarlander, T., Miller, R., Milligan, C., Reynolds, C., Kuo, L.C., 2011. Crystal structure of a soluble form of human monoglyceride lipase in complex with an inhibitor at 1.35 Å resolution: crystal structure of monoglyceride Lipase. *Protein Sci.* 20, 670–683.
- Sun, J.-Y., Liu, M.-Q., Xu, Y.-L., Xu, Z.-R., Pan, L., Gao, H., 2005. Improvement of the thermostability and catalytic activity of a mesophilic family 11 xylanase by N-terminus replacement. *Protein Expr. Purif.* 42, 122–130.
- Tang, L., Su, M., Yan, J., Xie, S., Zhang, W., 2015. Lid hinge region of *Penicillium expansum* lipase affects enzyme activity and interfacial activation. *Process Biochem.* 50, 1218–1223.
- Teze, D., Daligault, F., Ferrieres, V., Sanejouand, Y.-H., Tellier, C., 2015. Semi-rational approach for converting a GH36 α -glycosidase into an α -transglycosidase. *Glycobiology* 25, 420–427.
- Teze, D., Hendrickx, J., Czjzek, M., Ropartz, D., Sanejouand, Y.-H., Tran, V., Tellier, C., Dion, M., 2014. Semi-rational approach for converting a GH1 β -glycosidase into a β -transglycosidase. *Protein Eng. Des. Sel.* 27, 13–19.
- Vafiadi, C., Topakas, E., Alderwick, L.J., Besra, G.S., Christakopoulos, P., 2007. Chemoenzymatic synthesis of feruloyl D-arabinose as a potential anti-mycobacterial agent. *Biotechnol. Lett.* 29, 1771–1774.
- Vafiadi, C., Topakas, E., Christakopoulos, P., Faulds, C.B., 2006. The feruloyl esterase system of *Talaromyces stipitatus*: determining the hydrolytic and synthetic specificity of TsFaeC. *J. Biotechnol.* 125, 210–221.
- Vaughan, M.D., Johnson, K., DeFrees, S., Tang, X., Warren, R.A.J., Withers, S.G., 2006. Glycosynthase-mediated synthesis of glycosphingolipids. *J. Am. Chem. Soc.* 128, 6300–6301.

- Veprintsev, D.B., Narayan, M., Permyakov, S.E., Uversky, V.N., Brooks, C.L., Cherskaya, A.M., Permyakov, E.A., Berliner, L.J., 1999. Fine tuning the N-terminus of a calcium binding protein: α -lactalbumin. *Proteins Struct. Funct. Genet.* 37, 65–72.
- Wu, D., Fujio, M., Wong, C.-H., 2008. Glycolipids as immunostimulating agents. *Bioorg. Med. Chem.* 16, 1073–1083.
- Xie, Y., An, J., Yang, G., Wu, G., Zhang, Y., Cui, L., Feng, Y., 2014. Enhanced enzyme kinetic stability by increasing rigidity within the active Site. *J. Biol. Chem.* 289, 7994–8006.
- Zhang, K.-P., Lai, J.-Q., Huang, Z.-L., Yang, Z., 2011. *Penicillium expansum* lipase-catalyzed production of biodiesel in ionic liquids. *Bioresour. Technol.* 102, 2767–2772.
- Zhang, S., Zhang, K., Chen, X., Chu, X., Sun, F., Dong, Z., 2010. Five mutations in N-terminus confer thermostability on mesophilic xylanase. *Biochem. Biophys. Res. Commun.* 395, 200–206.

ANNEXE 1 :

Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities

Journal of Molecular Catalysis B: Enzymatic 126 (2016) 24–31

doi: [10.1016/j.molcatb.2016.01.012](https://doi.org/10.1016/j.molcatb.2016.01.012)

Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities

Olga Gherbovet^{a,b,c,1}, Régis Fauré^{a,b,c,1,*}, Fernando Ferreira^{a,b,c}, Julien Durand^{a,b,c}, Mélanie Ragon^{a,b,c}, Guillaume Hostyn^{a,b,c}, Eric Record^{d,e}, Sophie Bozonnet^{a,b,c}, Michael J. O'Donohue^{a,b,c,*}

^a*Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*

^b*INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*

^c*CNRS, UMR5504, F-31400 Toulouse, France*

^d*INRA, UMR1163 Biotechnologie des Champignons Filamenteux, Aix-Marseille Université, Polytech Marseille, 163 Avenue de Luminy, CP 925, F-13288 Marseille, France*

^e*Aix-Marseille Université, UMR1163 Biotechnologie des Champignons Filamenteux, Faculté des Sciences de Luminy-Polytech, CP 925, F-13288 Marseille cedex 9, France*

* Corresponding authors. Tel: +33 5 61 55 9488.

E-mail address: regis.faire@insa-toulouse.fr (R. Fauré), michael.odonohue@insa-toulouse.fr (M.J. O'Donohue)

¹ These authors contributed equally to this work.

Abstract

Feruloyl esterases (FAEs) constitute an important sub-group of hydrolytic enzymes involved in the deconstruction of plant cell wall polysaccharides. However, in the current era of genomics and metagenomics, finding and characterizing FAEs is not straightforward, mainly due to a lack of suitable compounds for high-throughput assays. To remedy this, indolyl and 4-nitrocatechol hydroxycinnamates (i.e. *trans*-ferulate and *p*-coumarate derivatives) were synthesized in good overall yields - between 46 and 56% after 4 steps - and their usefulness as substrates for FAEs was ascertained. The hydrolysis of the ester bond of these chromogenic compounds leads to a colour change, which can be readily monitored. Overall, these compounds considerably improve upon the current situation and enable the measurement of FAE activities in both qualitative solid medium-based and quantitative liquid assays.

Keywords

Feruloyl esterase, High-throughput screening, Chromogenic substrate, Indolyl hydroxycinnamates, Nitrocatechol ferulate

Abbreviations: BI-Co, 5-bromoindol-3-yl coumarate; CinnAE, cinnamoyl esterase; CSA, camphor sulphonic acid; DIPC, *N,N'*-diisopropylcarbodiimide; DMAP, 4-(dimethylamino)pyridine; FAE, feruloyl esterase; GH, glycoside hydrolases; HTS, high-throughput screening; 4NTC, 4-nitrocatechol ferulate; 4NTC-Fe, 4-nitrocatechol-1-yl ferulate; X-Fe, 5-bromo-4-chloroindol-3-yl ferulate.

1. Introduction

Enzyme discovery is being increasingly empowered by high-throughput technologies, such as functional metagenomics and directed evolution techniques, which in turn are creating a new bottleneck linked to screening and further characterization of newly identified biocatalysts (Turner, 2009; Wang et al., 2012). Therefore, there is a pressing need to develop compounds that can provide the basis for informationally-rich high-throughput screens that will not only pinpoint target activities but also provide more detailed qualitative data on enzyme properties. So far, chromogenic substrates have proven to be very useful for high-throughput screening (Zhang et al., 2015), because these are generally relatively easy to use and can often serve in liquid assays.

The development of enzymes for biorefinery purposes is currently a major target for enzyme discovery, with techniques such as metagenomics are providing a vast reservoir of candidate enzymes. In this area, a lot of focus has been on the discovery and engineering of glycoside hydrolases (GHs) (Bastien et al., 2013; Li et al., 2009), notably cellulases. However, there is increasing recognition that other so-called auxiliary enzymes are also required for biomass hydrolysis, with these including carbohydrate esterases and particularly feruloyl esterases (FAEs) (Dumon et al., 2012; Gopalan et al., 2015; Topakas et al., 2007).

Hydroxycinnamoyl esters (e.g. *trans*-ferulic acid and *p*-coumaric acid) can constitute up to 3% (w/w, dry weight) of plant cell walls where they play vital roles in maintaining cell-wall integrity (Akin, 2008; Koseki et al., 2009). *Trans*-ferulic acid, which is the most widespread and abundant hydroxycinnamic acid in plants, is usually ester-linked to polysaccharides, such as arabinoxylans (attached to *O*-5 or *O*-2 of α -L-arabinofuranosyl side chain), xyloglucans (attached to *O*-4 to β -D-xylopyranosyl residues), arabinans (attached to *O*-2 of α -L-arabinofuranosyl units) and arabinogalactan (attached to *O*-6 of β -D-galactopyranosyl residues) (Koseki et al., 2009; Topakas et al., 2007; Wong, 2006). When linked to polysaccharides, feruloyl moieties can be engaged in di-ferulic dehydrodimers that constitute either intermolecular ester crosslinks between polysaccharides or ester-ether bridges between polysaccharides and lignin, both contribute to the overall plant cell wall architecture. To release hydroxycinnamic acids from plant cell material and thus, at least in the case of ferulic acid, disrupt the cross-linking of cell wall components many lignocellulolytic microorganisms (exemplified by some filamentous fungi) produce carbohydrate esterases (FAEs and cinnamoyl esterases, CinnAEs, E.C. 3.1.1.73) (de O Buanafina, 2009; Wong, 2006) that cleave the ester bonds linking feruloyl moieties (i.e. monomeric and/or dimeric motifs) to sugars. FAEs and CinnAEs belong to the larger category of carboxylic ester hydrolases, and, within the CAZy classification (www.cazy.org) (Lombard et al., 2014), are mostly found in family CE1. Thus far, the only mechanism attributed to FAEs and CinnAEs is a two-step reaction involving the

formation of a covalent acyl-enzyme intermediate. The reaction is catalysed by a Ser-His-Asp catalytic triad, with Ser acting as a nucleophile and His as general base for the acylation step. Presently, carbohydrate esterases are classified into four subgroups based on substrate specificity and sequence similarity (Benoit et al., 2008; Crepin et al., 2004; Udatha et al., 2011).

So far the detection and characterization of FAE activities has mostly relied on the use of UV spectrophotometry (i.e. exploiting the difference in molar absorption coefficient between esterified and free ferulic acid) or HPLC, using various natural or synthetic compounds, such as biomass-derived hydroxycinnamic-esterified oligosaccharides, methyl ferulate (**3h**, Fig. 1A) and its analogues (Hatfield et al., 1991). Recently, several substrates for FAEs have been reported in the literature (Fig. 1A), most of them containing a 5-*O*-(*trans*-feruloyl)- α -L-arabinofuranoside moiety connected to a chromogenic (e.g. 5-bromo-4-chloro-3-indolyl **1a**, 4-nitrophenyl **1d**, 2-chloro-4-nitrophenyl **1e**) (Marmuse et al., 2008; Mastihubová et al., 2003; Mastihubová and Biely, 2010) or a fluorescent (umbelliferyl **1f**) (Zhang et al., 2012) group. In addition to the multi-step synthetic pathways required to procure these compounds, their use entails a tandem reaction involving ester bond cleavage and almost concomitant release of the chromogenic group. This second reaction requires the use of a sufficient quantity of α -L-arabinofuranosidase, which implies that the overall reaction must be properly optimized to avoid a situation in which the second (auxiliary enzyme-catalyzed) reaction becomes rate-limiting (Biely et al., 2002; Marmuse et al., 2008). Consequently, although these substrates are useful they are not particularly convenient for high-throughput screening (HTS) assays, where simplicity and rapidity are paramount.

Interestingly, structure examination of a mutant of feruloyl esterase A from *Aspergillus niger* (*AnFAEA*) complexed with a feruloylated arabinoxylo-trisaccharide revealed that this enzyme does not interact with the glycosidic part of the substrate, suggesting that artificial substrates might not need to harbour sugar moieties to be properly recognized by FAEs (Faulds et al., 2005). This knowledge confirms that the preparation of α -L-arabinofuranoside-free substrates (e.g. **3e**, Fig. 1A) is a well-founded strategy and is fortunate because such compounds, combining shorter synthetic pathways and direct detection, are better adapted to HTS assays (Hegde et al., 2009; Mastihuba et al., 2002; Mastihubová et al., 2001; Zhang et al., 2012). Finally, it is noteworthy that despite the fact that previously described compounds have been shown to be potentially suitable for the detection of FAEs in solid medium (Donaghy et al., 1998; Marmuse et al., 2008), so far none of these have been successfully employed in a HTS strategy.

The ambition to design simple-to-use compounds harbouring a hydroxycinnamic acid moiety that can be deployed in HTS strategies to detect FAEs is central to the present study. Herein, we report the synthesis and the evaluation of several types of compounds including indolyl- (i.e. 5-bromo-4-

chloroindol-3-yl or 5-bromoindol-3-yl) and 4-nitrocatechol-based derivatives (4NTC) coupled to either ferulic or *p*-coumaric acids, **3a**, **3c** and **4b** (Fig. 1B). We demonstrate that these versatile chromogenic substrates can provide the basis for HTS assays for FAEs, working in either solid (qualitative assay) or liquid (quantitative assay) media.

2. Material and methods

2.1. Chemical synthesis of chromogenic hydroxycinnamates

***N*-Acetyl-5-bromo-4-chloroindol-3-yl 4-*O*-acetyl-ferulate (7a).** *Coupling method A:* *N,N*-diisopropylcarbodiimide (DIPC; 95 μ L, 0.60 mmol, 4 equiv.) and a catalytic amount of 4-(dimethylamino)pyridine (DMAP) were added to a solution of 4-*O*-acetyl-ferulic acid **5a** (Hatfield et al., 1991) (71 mg, 0.30 mmol, 2 equiv.) and *N*-acetyl-5-bromo-4-chloroindoxyl-3-ol (Witmer et al., 1991) (43 mg, 0.15 mmol, 1 equiv.) in dichloromethane (2 mL). After stirring for 1 h at room temperature, the reaction was diluted with ethyl acetate, filtered through Celite and washed with brine. The organic layers were dried and evaporated to dryness.

Coupling method B: thionyl chloride (1.27 mL, 17.3 mmol, 3 equiv.) and a drop of dimethylformamide (DMF) were added to a solution of 4-*O*-acetyl-ferulic acid (Hatfield et al., 1991) (1.37 g, 5.8 mmol, 1 equiv.) in toluene (15 mL). The resulting mixture was refluxed (95 $^{\circ}$ C) for 1 h, and then evaporated to dryness, providing the expected acyl chloride derivative (Mastihubová et al., 2001) as yellow foam in quantitative yield. *N*-acetyl-5-bromo-4-chloroindoxyl-3-ol (Witmer et al., 1991) (0.56 g, 1.9 mmol, 1 equiv.) diluted in pyridine (5 mL) and a catalytic amount of DMAP were added to a solution of the previous acyl chloride (1.48 g, 5.8 mmol, 3 equiv.) in pyridine (5 mL). After stirring for 1 h at room temperature, the reaction was diluted with ethyl acetate, washed with saturated aqueous sodium hydrogencarbonate (three times) and brine, dried and concentrated.

Flash chromatography (using gradient dichloromethane-petroleum ether, 95:5 to 100:0 v/v) followed by recrystallization from ethyl acetate afforded pure **7a** (0.58 g, 1.1 mmol, 60%) as white needles. mp 201-203 $^{\circ}$ C (from EtOAc); ^1H NMR (500 MHz; CDCl_3): δ 8.34 (1H, d, *J* 8.9 Hz, *CH* of X), 7.92 (1H, d, *J* 16.0 Hz, *CH=CHCO*₂), 7.72 (1H, s, *CH* of X), 7.60 (1H, d, *J* 8.9 Hz, *CH* of X), 7.22 (1H, dd, *J* 1.8, 8.1 Hz, *CH* of Fe), 7.18 (1H, d, *J* 1.8 Hz, *CH* of Fe), 7.11 (1H, d, *J* 8.1 Hz, *CH* of Fe), 6.61 (1H, d, *J* 16.0 Hz, *CH=CHCO*₂), 3.91 (3H, s, OMe), 2.64 (3H, s, NAc), 2.34 (3H, s, OAc); ^{13}C NMR (125 MHz; CDCl_3): δ 168.7 (C=O), 168.4 (C=O), 164.3 (C=O), 151.6 (C_q), 147.0 (CH=CHCO₂), 142.1 (C_q), 133.3 (C_q), 133.1 (C_q), 132.8 (C_q), 130.6 (CH of X), 124.6 (C_q), 123.5 (CH of Fe), 122.6 (C_q), 121.7 (CH of Fe), 118.7 (C_q), 116.5 (CH=CHCO₂), 116.4 (CH of X), 116.1 (CH of X), 111.5 (CH of Fe), 56.0 (OMe), 23.8 (NAc), 20.6 (OAc); ES-HRMS calc. for C₂₂H₁₇BrClNO₆Na ([M+Na]⁺) *m/z* 527.98255;

measured 527.9825. Elemental analysis calc. for $C_{22}H_{17}BrClNO_6$: C, 52.15; H, 3.38; N, 2.76; found: C, 51.80; H, 3.34; N, 2.72%.

5-Bromo-4-chloroindol-3-yl ferulate (3a). *Deacetylation method E:* Hydrazine acetate (45 mg, 489 μ mol, 2.5 equiv.) was added to a cooled solution of *N*-acetyl-5-bromo-4-chloroindol-3-yl 4-*O*-acetyl-ferulate **7a** (100 mg, 197 μ mol, 1 equiv.) in DMF (2 mL). After stirring for 2 h at 0 °C, the reaction was quenched by addition of ethyl acetate, washed with brine, dried, and concentrated. Flash chromatography (using gradient petroleum ether-ethyl acetate, 9:1 to 7:3 v/v) procured compound **3a** (70 mg, 166 μ mol, 85%) as green foam. 1H NMR (500 MHz; $CDCl_3$): δ 8.13 (1H, br s, *NH*), 7.88 (1H, d, *J* 15.9 Hz, *CH=CHCO*₂), 7.38 (1H, dd, *J* 0.4, 8.7 Hz, *CH* of X), 7.33 (1H, dd, *J* 0.4, 2.8 Hz, *CH* of X), 7.17 (1H, ddd, *J* 0.5, 2.0, 8.2 Hz, *CH* of Fe), 7.12 (1H, br d, *J* 2.0 Hz, *CH* of Fe), 7.10 (1H, d, *J* 8.7 Hz, *CH* of X), 6.97 (1H, d, *J* 8.2 Hz, *CH* of Fe), 6.55 (1H, d, *J* 15.9 Hz, *CH=CHCO*₂), 3.97 (3H, s, OMe); ^{13}C NMR (125 MHz; $CDCl_3$): δ 165.9 (*C=O*), 148.4 (*C*_q), 147.0 (*CH=CHCO*₂), 146.8 (*C*_q), 133.4 (*C*_q), 129.6 (*C*_q), 127.3 (*CH* of X), 126.8 (*C*_q), 124.3 (*C*_q), 123.6 (*CH* of Fe), 119.3 (*C*_q), 116.1 (*CH* of X), 114.8 (*CH* of Fe), 114.3, 114.3 (*C*_q and *CH=CHCO*₂), 111.4 (*CH* of X), 109.5 (*CH* of Fe), 56.0 (OMe); ES-HRMS calc. for $C_{18}H_{13}BrClNO_4Na$ ($[M+Na]^+$) *m/z* 443.96142; found: 443.9615. Elemental analysis calc. for $C_{18}H_{13}BrClNO_4$: C, 51.15; H, 3.10; N, 3.31; found: C, 50.19; H, 3.33; N, 3.22%.

4-Nitrocatechol-1-yl 4-*O*-acetyl-ferulate (7c). *Coupling method C:* Thionyl chloride (0.92 mL, 12.63 mmol, 3 equiv.) and a drop of DMF were added to a solution of the 4-*O*-acetyl-ferulic acid **5a** (Hatfield et al., 1991) (1.0 g, 4.21 mmol, 1 equiv.) in toluene (18 mL). The resulting mixture was refluxed (95 °C) for 1 h, and then evaporated to dryness, providing the expected acyl chloride derivative as brown foam in quantitative yield. 4-Nitrocatechol sulfate dipotassium salt (1.31 g, 4.21 mmol, 1 equiv.) diluted in DMF (6 mL) and triethylamine (0.48 mL, 3.58 mmol, 0.85 equiv.) and a catalytic amount of DMAP (0.13 mg, 1.05 mmol, 0.25 equiv.) were added to a solution of the previous acyl chloride in dichloromethane (12 mL). After stirring for 3 h at room temperature, the reaction mixture was poured into a solution of camphor sulphonic acid (CSA; 1.15 g, 5.04 mmol, 1.2 equiv.) in acetone (315 mL). Solvents were removed under reduced pressure and the residue was recovered in a saturated aqueous sodium hydrogencarbonate solution, then extracted with ethyl acetate. The combined organic extracts were washed sequentially in water then brine, dried on magnesium sulfate, filtered and evaporated. Flash chromatography (dichloromethane-acetone 1:0 to 9:1 v/v) afforded pure **7c** (779 mg, 2.08 mmol, 50%) as white powder: 1H NMR (500 MHz, $(CD_3)_2CO$): δ 8.12 (1H, s, *CH* of 4NTC), 8.11 (1H, dd, *J* 2.8, 8.8, *CH* of 4NTC), 7.89 (1H, d, *J* 16.0 Hz, *CH=CHCO*₂), 7.59 (1H, d, *J* 2.0 Hz, *CH* of Fe.), 7.38 (1H, ddd, *J* 0.5, 2.0, 8.1, *CH* of Fe), 7.21 (1H, dt, *J* 1.5, 8.8 Hz, *CH* of 4NTC), 7.17 (1H, d, *J* 8.1 Hz, *CH* of Fe), 6.84 (1H, d, *J* 16.0 Hz, *CH=CHCO*₂),

3.94 (s, 3H, OMe), 2.27 (s, 3H, OAc); ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$): δ 168.8 (C=O), 164.9 (C=O), 156.5 (C_q), 152.8 (C_q), 147.2 (CH=CHCO₂), 143.2 (C_q), 141.2 (C_q), 139.2 (C_q), 133.9 (C_q), 124.3 (CH of Fe), 123.8 (CH of 4NTC), 122.6 (CH of Fe), 120.6 (CH of 4NTC), 117.7 (CH=CHCO₂), 117.6 (CH of 4NTC), 112.7 (CH of Fe), 56.5 (OMe), 20.5 (OAc); ES-HRMS calc. for $\text{C}_{18}\text{H}_{15}\text{NO}_8\text{Na}$ ($[\text{M}+\text{Na}]^+$) m/z 396.0695; found: 396.0696.

4-Nitrocatechol-1-yl ferulate (3c). *Deacetylation method F:* Amano lipase PS (3.46 g) was added to a solution of **13** (778 mg, 2.08 mmol) in dichloromethane (38 mL) and 2-propanol (19 mL). The reaction mixture was stirred for 3 days at 37 °C. The reaction was terminated by enzyme removal using filtration. The filter cake was washed with dichloromethane and the filtrate was concentrated by evaporation of the solvents at reduced pressure to afford the product **3c** as yellow powder (635 mg, 1.91 mmol, 92%). ^1H NMR (500 MHz; $(\text{CD}_3)_2\text{CO}$): δ 8.11-8.08 (2H, m, CH of 4NTC), 7.81 (1H, d, J 15.9 Hz, CH=CHCO₂), 7.45 (1H, d, J 2.0 Hz, CH of Fe), 7.26 (1H, ddd, J 0.5, 2.0, 8.2 Hz, CH of Fe.), 7.20 (1H, dd, J 0.5, 9.9 Hz, CH of 4NTC), 6.92 (1H, d, J 8.2 Hz, CH of Fe), 6.64 (1H, d, J 15.9 Hz, CH=CHCO₂), 3.96 (3H, s, OMe); ^{13}C NMR (125 MHz; $(\text{CD}_3)_2\text{CO}$): δ 165.3 (C=O), 156.6 (C_q), 150.7 (C_q), 148.8 (C_q), 148.4 (CH=CHCO₂), 141.2 (C_q), 139.3 (C_q), 127.2 (C_q), 124.6 (CH of Fe), 123.7 (CH of 4NTC), 120.6 (CH of 4NTC), 117.6 (CH of 4NTC), 116.2 (CH of Fe), 114.0 (CH=CHCO₂), 111.6 (CH of Fe), 56.4 (OMe); ES-HRMS calc. for $\text{C}_{16}\text{H}_{13}\text{NO}_7\text{Na}$ ($[\text{M}+\text{Na}]^+$) m/z 354.0590; found: 354.0589.

N-Acetyl-5-bromoindol-3-yl 4-O-acetyl-coumarate (8b). *Coupling method A:* DIPC (0.64 mL, 4.0 mmol, 4 equiv.) and a catalytic amount of DMAP were added to a solution of 4-O-acetyl-coumaric acid **6b** (Panda et al., 2012) (412 mg, 2.0 mmol, 2 equiv.) and *N*-acetyl-5-bromo-indoxyl-3-ol (Liav et al., 1999) (254 mg, 1.0 mmol, 1 equiv.) in dichloromethane (15 mL). After stirring for 2 h at room temperature, the reaction was filtrated through Celite, washed with water, dried and concentrated under reduced pressure. Flash chromatography (dichloromethane-petroleum ether, 9:1 v/v) afforded pure **8b** (416 mg, 0.94 mmol, 94%), which crystallized in ethyl acetate as white needles. mp 156-158 °C (from EtOAc); ^1H NMR (500 MHz; CDCl_3): δ 8.38 (1H, d, J 8.9 Hz, CH of BI), 7.90 (1H, d, J 16.0 Hz, CH=CHCO₂), 7.84 (1H, s, CH of BI), 7.76 (1H, dd, J 0.5, 2.0 Hz, CH of BI), 7.67-7.64 (2H, m, CH of Co), 7.50 (1H, ddd, J 0.5, 2.0, 8.9 Hz, CH of BI), 7.21-7.19 (2H, m, CH of Co), 6.64 (1H, d, J 16.0 Hz, CH=CHCO₂), 2.63 (3H, s, NAc), 2.34 (3H, s, OAc); ^{13}C NMR (125 MHz; CDCl_3): δ 169.0 (C=O), 168.6 (C=O), 163.6 (C=O), 152.7 (C_q), 146.4 (CH=CHCO₂), 133.7 (C_q), 131.6 (C_q), 131.5 (C_q), 129.6 (2C, CH of Co), 129.1 (CH of BI), 125.4 (C_q), 122.4 (2C, CH of Co), 120.4 (CH of BI), 118.3 (CH of BI), 117.1 (C_q), 116.2 (CH=CHCO₂), 114.4 (CH of BI), 23.8 (NAc), 21.1 (OAc); ES-HRMS calc. for $\text{C}_{21}\text{H}_{16}\text{BrNO}_5\text{Na}$ ($[\text{M}+\text{Na}]^+$) m/z 464.01095; found: 464.0113.

Elemental analysis calcd for C₂₁H₁₆BrNO₅ C, 57.03; H, 3.65; N, 3.17; found: C, 56.73; H, 3.56; N, 3.10%.

5-Bromoindol-3-yl coumarate (4b). *Deacetylation method D:* *N*-acetyl-5-bromoindol-3-yl 4-*O*-acetyl-coumarate **8b** (277 mg, 0.63 mmol, 1 equiv.) was treated in methanol (15 mL) with sodium methoxide (1 M in methanol, 0.63 mL, 0.63 mmol, 1 equiv.) for 2 h at 0 °C. The solution was neutralized with Amberlite IR-120 (H⁺), filtered, and then concentrated under reduced pressure. Flash chromatography (using gradient dichloromethane then dichloromethane-methanol, 99.5:0.5 to 99:1 v/v) gave compound **4b** (146 mg, 0.41 mmol, 65%) as a slightly dark foam. ¹H NMR (500 MHz; (CD₃)₂CO): δ 7.87 (1H, d, *J* 16.0 Hz, CH=CHCO₂), 7.71 (1H, td, *J* 0.6, 2.0 Hz, CH of BI), 7.69-7.66 (2H, m, CH of Co), 7.57 (1H, dd, *J* 0.4, 2.7 Hz, CH of BI), 7.40 (1H, dd, *J* 0.6, 8.7 Hz, CH of BI), 7.27 (1H, ddd, *J* 0.4, 2.0, 8.7 Hz, CH of BI), 6.96-6.93 (2H, m, CH of Co), 6.66 (1H, d, *J* 16.0 Hz, CH=CHCO₂); ¹³C NMR (125 MHz; (CD₃)₂CO): δ 165.4 (C=O), 161.0 (C_q), 147.2 (CH=CHCO₂), 133.1 (C_q), 131.4 (2C, CH of Co), 130.5 (C_q), 127.0 (C_q), 125.7 (CH of BI), 122.9 (C_q), 120.7 (CH of BI), 116.9 (2C, CH of Co), 116.7 (CH of BI), 114.6 (CH of BI), 114.5 (CH=CHCO₂), 112.9 (C_q); ES-HRMS calc. for C₁₇H₁₂BrNO₃Na ([M+Na]⁺) *m/z* 379.98982; found: 379.9902.

2.2. Enzymes

Feruloyl esterases from *A. niger*, AnFAEA and AnFAEB, were produced and purified as previously described (Levasseur et al., 2004; Record et al., 2003).

2.3. Liquid medium-based, colorimetric assays using indolyl compounds

The semi-quantitative measurement of FAE activity was performed in a continuous assay carried out in buffered conditions in the presence of either X-Fe **3a**, or BI-Co **4b** (100 µg/mL). Solutions of the **3a** or **4b** (100 µg/mL) were prepared by diluting aliquots of the initial stock solution (20 mg/mL in DMSO) with buffer (i.e. with DMSO at a final concentration of 0.5%, v/v). Prior to the addition of enzyme solution, the reaction mixture was pre-incubated at the optimal temperature for 5 min. Reaction mixtures (200 µL) were incubated in 96-well microtiter plates under stirring at 500 rpm for 30 min. The optical densities at 650 nm were recorded on a VersaMax microplate reader (Molecular Devices). Negative controls containing all of the reactants except the enzyme were always included in order to correct for spontaneous hydrolysis of the substrates.

2.4. Liquid medium-based colorimetric assays using 4NTC-Fe

First an aliquot of the stock solution of **3c** (50 mM 4NTC-Fe) in DMSO was diluted in 100 mM sodium phosphate buffer pH 6.0 to achieve a final concentration of 2 mM 4NTC-Fe in 4% (v/v) DMSO. For the assay, 400 μ L of this solution was incubated at 40 °C for 4 min and then combined with 50 μ L of enzyme solution. After enzyme addition, 4NTC release was monitored in a discontinuous assay, removing 50 μ L aliquots every 3 min over a 21 min-period. Upon removal, the aliquots were immediately mixed in 96-well microplates with 200 μ L of an aqueous solution of ice-cold 2 M Na₂CO₃ and absorbance at 530 nm was recorded using a microplate reader (Tecan Infinite 200 PRO). Afterwards, activity was calculated using a standard absorbance curve of 4NTC (from 0.01 to 1 mM) in an aqueous solution of 2 M Na₂CO₃ containing an appropriate amount of sodium phosphate buffer (i.e. 20 mM final concentration). Negative controls containing all of the reactants except the enzyme were always included in order to monitor and correct for spontaneous hydrolysis of the substrate.

2.5. Screening of FAE(+) microorganisms in solid medium

A mixture of the strains *Yarrowia lipolytica* zeta (Ura⁻) and *Y. lipolytica* AnFaeA(Ura3⁺) were grown for 27 h at 28 °C with shaking at 650 rpm in 96-well microtiter plates containing YPD medium (composed of 10 g/L yeast extract, 20 g/L bacto-peptone and 20 g/L glucose). With the aid of a colony picker QPixII, (Genetix), the 96 cultures were then used to inoculate OmniTray plates containing solid YPD (i.e. containing 1% w/v agar) overlaid with a nitrocellulose membrane. OmniTrays were incubated at 28 °C for 36 h and then the membranes were transferred onto fresh OmniTrays containing 150 μ g/mL (0.2% final of DMSO, v/v) of either 4NTC-Fe **3c**, X-Fe **3a** or BI-Co **4b** in 50 mM sodium phosphate buffer pH 6.8, containing 1% (w/v) agarose. For 4NTC-Fe, ammonium iron citrate (0.5 mM) was also added to the medium. OmniTrays were incubated at 37 °C in order to follow the coloration of colonies producing FAE activity.

3. Results and discussion

3.1. Synthesis of substrates

The synthetic routes yielding chromogenic compounds **3a**, **3c** and **4b** are summarized in Scheme 1 and Table 1.

The 4-bromo-5-chloro-3-indolyl ferulate (X-Fe, **3a**) was synthesized starting from the 4-*O*-acetylferulic acid **5a** (Hatfield et al., 1991) (obtained from almost-quantitative acetylation of the ferulic acid). Two different pathways were used for the esterification of the carboxylic acid function. In the first strategy, coupling involved direct esterification of carboxylic acid **5a** using DIPC activation in the presence of a catalytic amount of DMAP in dichloromethane following an adaptation of the previously

reported Steglich reaction protocol (Hegde et al., 2009). The crystallized ferulate **7a** was thus obtained in 65% yield. The second strategy involved an initial treatment of **5a** with thionyl chloride catalyzed by DMF in refluxing toluene, to quantitatively yield the 4-acetoxyferuloyl chloride **5c** (Hatfield et al., 1991), followed by coupling to *N*-acetyl-5-bromo-4-chloroindoxyl-3-ol (Witmer et al., 1991) in pyridine, leading to the protected ferulate **7a** in 60% yield over two steps. Catalytic deacetylation was first attempted using sodium methoxide in a methanol/dichloromethane mixture, but this afforded the target compound **3a** in poor yield (26%) due to transesterification. When milder conditions were applied (i.e. hydrazine acetate in cold DMF), the desired chromogenic substrate **3a** was obtained in 85% yield. The same strategy was applied to access 4-bromo-indol-3-yl *p*-coumarate (BI-Co, **4b**) from *p*-coumaric acid. The latter was first acetylated in 92% yield using acetic anhydride in pyridine (Panda et al., 2012). Efficient esterification of **6b** with *N*-acetyl-5-bromo-indoxyl-3-ol (Berlin and Sauer, 1996) using DIPC/DMAP activation in dichloromethane gave *p*-coumarate **8b** in 94% yield. Catalytic deacetylation using sodium methoxide in methanol afforded the targeted chromogenic coumarate **4b** in 65% yield. The synthesis of indolyl derivatives **3a** and **4b** was achieved in 4 steps starting from commercial sources and provided overall yields of 55 and 56%, respectively.

The synthesis of a single regioisomer of 4-nitrocatechol-1-yl ferulate (4NTC-Fe, **3c**) was achieved through the selective esterification of one hydroxyl position of 4NTC using the commercially available dipotassium salt of 4-nitrocatechol sulfate. Since direct esterifications failed (see Appendix A), coupling of the feruloyl chloride derivative **5c** to 4NTC sulfate salt in DMF in the presence of catalytic amounts of trimethylamine and DMAP was performed and one-pot deprotection, using a solution of CSA in acetone (Borsenberger et al., 2013), led to **7c** in 50% overall yield. Selective deacetylation was achieved using Amano lipase PS from *Burkholderia cepacia* (Mastihubová et al., 2001) (originally classified as *Pseudomonas cepacia*) in a high 92% yield. Finally, **3c** was obtained in 46% overall yield (in 4 steps) from commercial ferulic acid and 4NTC sulfate derivative.

3.2. Evaluation and applications for hydroxycinnamoyl esterases

The stability of the three esters was evaluated over the pH range 4.0-10.0, using UV spectrophotometry to monitor the state of the compounds (Figs. A.1-A.3). Indolyl derivatives **3a** and **4b** showed very high stability over a wide pH range and up to high temperature, the first signs of degradation becoming apparent at pH 9.0 and above 70°C. These observations are consistent with the known stability of ester linkages under basic conditions. On the other hand, the ferulate **3c** was only moderately stable with hydrolysis being observed even at neutral pH and 40 °C (Fig. A.4). Nevertheless, 4NTC-Fe **3c** was stable enough to be used for the characterization of FAEs operating in the pH range 5.0-6.0 and up to 40 °C, since the appearance of free 4NTC due to spontaneous

hydrolysis can be easily corrected for by the use of proper controls without enzyme. Furthermore, it is noteworthy that although FAEs and fungal lipase share a very similar overall fold, catalytic triad and mechanism (Wong, 2006), the ester linkage of **3a**, **4b** and **3c** was stable towards various lipases (with > 1 IU/mL of commercial immobilized Novozyme 435 or Amano lipase PS, lipases Lip2 and Lip8 from *Y. lipolytica*, or lipases Lip1 and Lip4 from *Candida rugosa* - data not shown), but were cleaved by FAEs (see below).

The suitability of **3a**, **4b** and **3c** for the detection in solid medium of microorganisms expressing FAE was evaluated using wild-type *Y. lipolytica* Zeta cells (Bordes et al., 2007) as negative control and recombinant *Y. lipolytica* Zeta producing AnFAEA as the positive sample (unpublished data). To complete the evaluation, the microbial samples were also cultured on non-selective (i.e. YPD, which supports the growth of both yeast strains) and selective media (i.e. YNB, which only supports the growth of the URA3⁺, AnFAEA-producing strain), in the absence of chromogenic substrates (Fig. A.5). When grown in the presence of *Y. lipolytica* Zeta cells producing AnFAEA **3a** were characterized by the formation of turquoise haloes, while growth in the presence of **4b** and **3c** led to the formation of dark blue and brown coloration respectively (Fig. 2). Coloration in the case of **3a** and **4b** was the result of enzymatic ester bond cleavage and subsequent release of free indoxyl-3-ol groups that upon oxidation dimerized forming indigo-like insoluble dyes. (Berlin and Sauer, 1996) However, in the case of **3c** coloration was due to enzymatic ester bond cleavage followed by the chelation of free 4NTC with ferric ions (Borsenberger et al., 2012; Butterworth et al., 2004). Although all three compounds appeared to be suitable for qualitative HTS on solid medium, the compromise between sensitivity of **3c** and diffusion of 4NTC-iron(III) complex in the operating conditions limits its use to short assays (< 30 min), whereas thanks to precipitation of their related dehydroindigo **3a** and **4b** are more useful for longer assays, for example to detect weakly-expressed FAE activity (Fig. A.6). It is also noteworthy that the usefulness of **3a**, **4b** and **3c** was established in different assay conditions (i.e. with or without top agar, or when using a membrane) and using different FAE(+)-microorganisms (i.e. *A. niger* or *Y. lipolytica*) (Figs. 2 and A.7-A.8).

The indolyl-based compounds **3a** and **4b** and the 4NTC-ester derivative **3c** also proved to be suitable for use in liquid assays aimed at semi-quantitative and quantitative analyses (Fig. 3). Although the quantification of dehydroindigo in solution is quite tricky, due to their relative insolubility in aqueous solutions and the lack of suitable standards (i.e. pure dehydroindigo), **3a** and **4b** are nevertheless well-adapted for primary HTS in 96-well microtiter plates and spectrophotometric detection. The reproducibility of this type of assay was evaluated using **3a** and **4b**, and 20 mIU/mL of AnFAEs A and B in standard conditions (at 50 °C in 100 mM citrate-phosphate buffer pH 6.0). After 30 min, the mean absorbance values at 650 nm were 0.31 ± 0.03 (33 repeats) and 0.12 ± 0.01 (24 repeats) respectively, thus the coefficient of

variance (CV) of the assays was 10.8 and 11.3% respectively. Moreover, since the optimum concentration of **3a** and **4b** in this assay is 100 µg/mL, only 2 mg per microtiter plate is required, assuming a detection threshold of 2 mIU/mL.

Regarding **3c**, we found that this compound is well-suited to liquid assays and thus the quantitative characterisation of FAEs. This could be achieved using discontinuous monitoring (at 530 nm and alkaline pH) of 4NTC release, with reactions progressing from yellow to red as more free 4NTC accumulates (Fig. 3B-C).

4. Conclusions

Three compounds for the detection and characterization of FAE activity have been synthesized in good overall yields (46-56%) using quite direct (4 steps) synthetic routes. The compounds have been fully validated as FAE substrates that can be used in solid and liquid media assays, and are thus suitable for HTS. In principle, **3a**, **3c** and **4b** adequately reconstitute the essential molecular features that characterize ester linkages between hydroxycinnamoyl and sugar moieties in plant polysaccharides and accordingly will be useful for the discovery of new FAEs using approaches such as functional metagenomics or molecular gene evolution.

Acknowledgements

This work was supported by the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 613868, OPTIBIOCAT project (to O.G., J.D. and G.H.), by the French National Research Agency, grant ANR-05-PNRB-002, project SPPECABBE (to M.R.) and the Région Midi-Pyrénées grants DAER-Recherche 07009817 (to F.F.). The NMR work carried out in this work at the LISBP (Toulouse, France) was performed with the equipment of MetaToul (Metabolomics & Fluxomics Facilities, Toulouse, France, www.metatoul.fr). MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr) and is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, IBiSa-France, the Centre National de la Recherche Scientifique (CNRS) and INRA. We thank the ICEO facility dedicated to enzyme screening and discovery, and part of the Integrated Screening Platform of Toulouse (PICT, IBiSA) for providing access to its equipment. We thank Drs. Florence Bordes, Sophie Duquesne and Alain Marty for supplying cell extracts containing lipases Lip2 and Lip8 from *Y. lipolytica* and Lip1 and Lip4 from *C. rugosa*.

References

- Akin, D.E., 2008. Plant cell wall aromatics: influence on degradation of biomass. *Biofuels. Bioprod. Biorefin.* 2, 288–303.
- Bastien, G., Arnal, G., Bozonnet, S., Laguerre, S., Ferreira, F., Fauré, R., Henrissat, B., Lefèvre, F., Robe, P., Bouchez, O., Noirot, C., Dumon, C., O'Donohue, M., 2013. Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol. Biofuels.* 6, 78.
- Benoit, I., Danchin, E.G.J., Bleichrodt, R.-J., de Vries, R.P., 2008. Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. *Biotechnol. Lett.* 30, 387–396.
- Berlin, W., Sauer, B., 1996. In situ color detection of α -L-arabinofuranosidase, a “no-background” reporter gene, with 5-bromo-3-indolyl- α -L-arabinofuranoside. *Anal. Biochem.* 243, 171–175.
- Biely, P., Mastihubová, M., van Zyl, W.H., Prior, B.A., 2002. Differentiation of feruloyl esterases on synthetic substrates in a α -arabinofuranosidase-coupled and ultraviolet-spectrophotometric assays. *Anal. Biochem.* 311, 68–75.
- Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* 70, 493–502.
- Borsenberger, V., Dornez, E., Desrousseaux, M.-L., Courtin, C.M., O'Donohue, M.J., Fauré, R., 2013. A substrate for the detection of broad specificity α -L-arabinofuranosidases with indirect release of a chromogenic group. *Tetrahedron Lett.* 54, 3063–3066.
- Borsenberger, V., Ferreira, F., Pollet, A., Dornez, E., Desrousseaux, M.-L., Massou, S., Courtin, C.M., O'Donohue, M.J., Fauré, R., 2012. A versatile and colorful screening tool for the identification of arabinofuranose-acting enzymes. *ChemBioChem.* 13, 1885–1888.
- Butterworth, L.A., Perry, J.D., Davies, G., Burton, M., Reed, R.H., Gould, F.K., 2004. Evaluation of novel β -ribosidase substrates for the differentiation of Gram-negative bacteria. *J. Appl. Microbiol.* 96, 170–176.
- Crepin, V.F., Faulds, C.B., Connerton, I.F., 2004. Functional classification of the microbial feruloyl esterases. *Appl. Microbiol. Biotechnol.* 63, 647–652.
- de O Buanafina, M.M., 2009. Feruloylation in grasses: current and future perspectives. *Mol. Plant.* 2, 861–872.
- Donaghy, J., Kelly, P.F., McKay, A.M., 1998. Detection of ferulic acid esterase production by *Bacillus* spp. and *lactobacilli*. *Appl. Microbiol. Biotechnol.* 50, 257–260.
- Dumon, C., Song, L., Bozonnet, S., Fauré, R., O'Donohue, M.J., 2012. Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process. Biochem.* 47, 346–357.
- Faulds, C.B., Molina, R., Gonzalez, R., Husband, F., Juge, N., Sanz-Aparicio, J., Hermoso, J.A., 2005. Probing the determinants of substrate specificity of a feruloyl esterase, *AnFaeA*, from *Aspergillus niger*. *FEBS J* 272, 4362–4371.
- Gopalan, N., Rodríguez-Duran, L.V., Saucedo-Castaneda, G., Nampoothiri, K.M., 2015. Review on technological and scientific aspects of feruloyl esterases: A versatile enzyme for biorefining of biomass. *Bioresour. Technol.* 193, 534–544.
- Hatfield, R.D., Helm, R.F., Ralph, J., 1991. Synthesis of methyl 5-O-trans-feruloyl- α -L-arabinofuranoside and its use as a substrate to assess feruloyl esterase activity. *Anal. Biochem.* 33, 25–33.
- Hegde, S., Srinivas, P., Muralikrishna, G., 2009. Single-step synthesis of 4-nitrophenyl ferulate for spectrophotometric assay of feruloyl esterases. *Anal. Biochem.* 387, 128–129.

- Koseki, T., Fushinobu, S., Ardiansyah, Shirakawa, H., Komai, M., 2009. Occurrence, properties, and applications of feruloyl esterases. *Appl. Microbiol. Biotechnol.* 84, 803–810.
- Levasseur, A., Benoit, I., Asther, M., Asther, M., Record, E., 2004. Homologous expression of the feruloyl esterase B gene from *Aspergillus niger* and characterization of the recombinant enzyme. *Protein Express. Purif.* 37, 126–133.
- Li, L.-L., McCorkle, S.R., Monchy, S., Taghavi, S., van der Lelie, D., 2009. Bioprospecting metagenomes: glycosyl hydrolases for converting biomass. *Biotechnol. Biofuels.* 2, 10.
- Liav, A., Hansjergen, J.A., Achyuthan, K.E., Shimasaki, C.D., 1999. Synthesis of bromoindolyl 4,7-di-O-methyl-Neu5Ac: specificity toward influenza A and B viruses. *Carbohydr. Res.* 317, 198–203.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495.
- Marmuse, L., Asther, M., Fabre, E., Navarro, D., Lesage-Meessen, L., Asther, M., O'Donohue, M., Fort, S., Driguez, H., 2008. New chromogenic substrates for feruloyl esterases. *Org. Biomol. Chem.* 6, 1208–1214.
- Mastihuba, V., Kremnický, L., Mastihubová, M., Willett, J.L., Côté, G.L., 2002. A spectrophotometric assay for feruloyl esterases. *Anal. Biochem.* 309, 96–101.
- Mastihubová, M., Biely, P., 2010. Preparation of regioselectively feruloylated *p*-nitrophenyl α -L-arabinofuranosides and β -D-xylopyranosides—convenient substrates for study of feruloyl esterase specificity. *Carbohydr. Res.* 345, 1094–1098.
- Mastihubová, M., Mastihuba, V., Kremnický, L., Willett, J.L., Côté, G.L., 2001. Chemoenzymatic preparation of novel substrates for feruloyl esterases. *Syn. Lett.* 10, 1559–1560.
- Mastihubová, M., Szemesová, J., Biely, P., 2003. Two efficient ways to 2-O- and 5-O-feruloylated 4-nitrophenyl α -L-arabinofuranosides as substrates for differentiation of feruloyl esterases. *Tetrahedron Lett.* 44, 1671–1673.
- Panda, P., Appalashetti, M., Natarajan, M., Chan-Park, M.B., Venkatraman, S.S., Judeh, Z.M.A., 2012. Synthesis and antitumor activity of lapathoside D and its analogs. *Eur. J. Med. Chem.* 53, 1–12.
- Record, E., Asther, M., Sigoillot, C., Pagès, S., Punt, P.J., Delattre, M., Haon, M., van den Hondel, C.A.M.J.J., Sigoillot, J.-C., Lesage-Meessen, L., Asther, M., 2003. Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Appl. Microbiol. Biotechnol.* 62, 349–355.
- Topakas, E., Vafiadi, C., Christakopoulos, P., 2007. Microbial production, characterization and applications of feruloyl esterases. *Process Biochem.* 42, 497–509.
- Turner, N.J., 2009. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* 5, 567–573.
- Udatha, D.B.R.K.G., Kouskoumvekaki, I., Olsson, L., Panagiotou, G., 2011. The interplay of descriptor-based computational analysis with pharmacophore modeling builds the basis for a novel classification scheme for feruloyl esterases. *Biotechnol. Adv.* 29, 94–110.
- Wang, M., Si, T., Zhao, H., 2012. Biocatalyst development by directed evolution. *Bioresour. Technol.* 115, 117–125.
- Witmer, M.R., Falcomer, C.M., Weiner, M.P., Kay, M.S., Begley, T.P., Ganem, B., Scheraga, H.A., 1991. U-3'-BCIP: a chromogenic substrate for the detection of RNase A in recombinant DNA expression systems. *Nucleic Acids Res.* 19, 19–22.
- Wong, D.W.S., 2006. Feruloyl esterase: a key enzyme in biomass degradation. *Appl. Biochem. Biotechnol.* 133, 87–112.

- Zhang, Q., Peng, X., Grilley, M., Takemoto, J.Y., Chang, C.-W.T., 2015. Using fluorogenic probes for the investigation of selective biomass degradation by fungi. *Green Chem.* 17, 1918–1925.
- Zhang, S.-B., Ma, X.-F., Pei, X.-Q., Liu, J.-Y., Shao, H.-W., Wu, Z.-L., 2012. A practical high-throughput screening system for feruloyl esterases: Substrate design and evaluation. *J. Mol. Catal. B Enzym.* 74, 36–40.

Figure captions

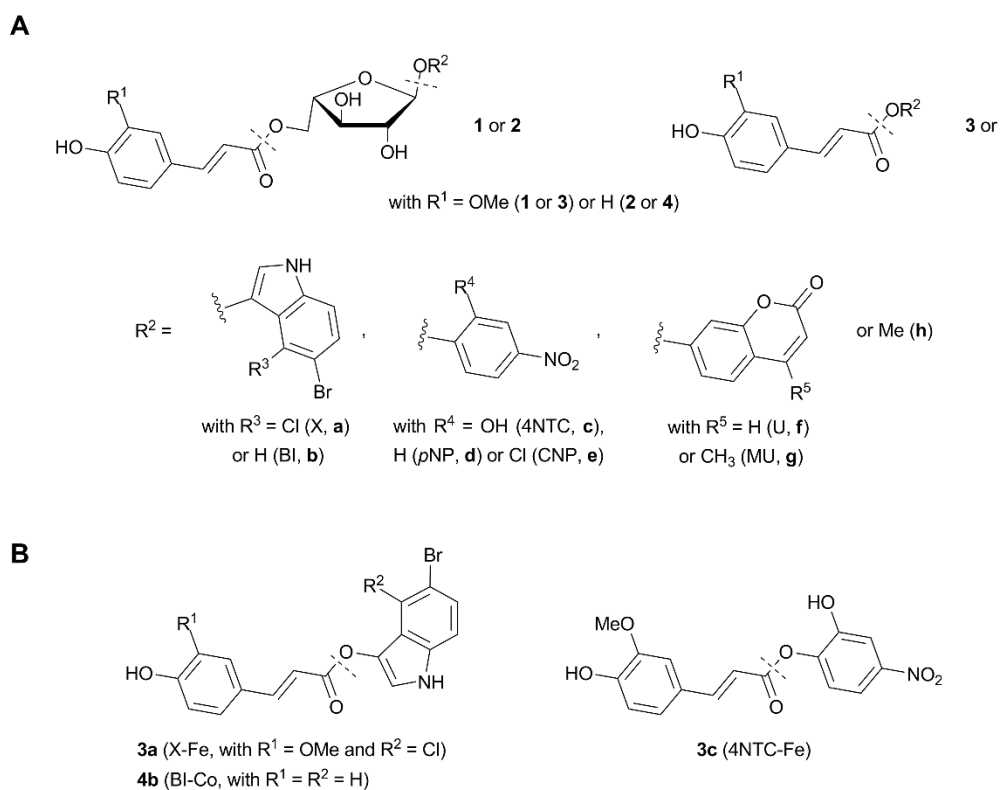
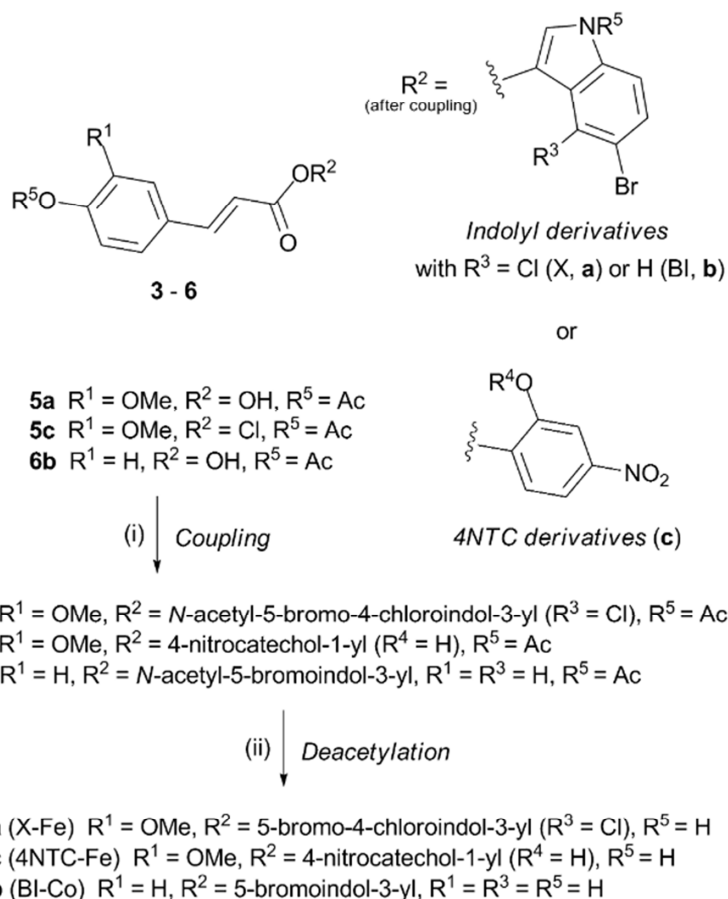


Fig. 1. Previously reported synthetic substrates (A), with (**1a,d-f,h** and **2a**) or without sugar moiety (**3d,e,g,h** and **4h**), and substrates synthesized in this work (B). All of these substrates can be used to screen for and measure FAE activities. The sites of enzymatic cleavage are marked with dashed lines.



Scheme 1. Synthesis of ferulate (**3a** and **3c**) and *p*-coumarate (**4b**) derivatives. *Reagents and conditions:* (i) **5a** or **6b**, *N*-acetyl-5-bromo-4-chloroindoxyl-3-ol or *N*-acetyl-5-bromoindoxyl-3-ol, DIPC, DMAP, CH_2Cl_2 , 2 h, **7a**: 65%, **8b**: 94%; or **5c**, *N*-acetyl-5-bromo-4-chloroindoxyl-3-ol, pyridine, DMAP, 1 h, **7a**: 60%; or **5c**, 4-nitrocatechol sulfate dipotassium salt, Et_3N , DMAP, $\text{CH}_2\text{Cl}_2/\text{DMF}$, 3 h then CSA, acetone, 30 min, **7c**: 50% (2 steps); (ii) **7a**, hydrazine acetate, DMF, 2 h, 0°C , **3a**: 85% ; or **7c**, lipase PS, $\text{CH}_2\text{Cl}_2/\text{isopropanol}$, 72 h, 37°C , **3c**: 92%; or **8b**, MeONa, MeOH, 2 h, 0°C , **4b**: 65%.

Table 1. Synthesis of ferulate (**3a** and **3c**) and *p*-coumarate (**4b**) derivatives: summary of tested pathways and yields for the coupling and deacetylation steps

Starting compound	(i) Coupling			(ii) Deacetylation		
	Method ^a	Protected compound	Yield (%)	Method ^a	Deprotected compound	Yield (%)
5a	A	7a	65	D	3a	26
				E		85
5c	B		60			
5a	A	7c	- ^b	F ^d	3c	92
5c	B		- ^c			
5c	C		50			
6b	A	8b	94	E	4b	< 10
				D		65

^a Couplig method A: *N*-acetyl-indoxyl-3-ol derivative or 4NTC sulfate dipotassium salt, DIPC, DMAP, CH₂Cl₂; Coupling method B: *N*-acetyl-indoxyl-3-ol derivative or 4NTC sulfate dipotassium salt, pyridine, DMAP; Coupling method C: 4NTC sulfate dipotassium salt, Et₃N, DMAP, CH₂Cl₂/DMF then CSA, acetone; Deacetylation method D: MeONa, MeOH; Deacetylation method E: hydrazine acetate, DMF; and Deacetylation method F: lipase PS, CH₂Cl₂/isopropanol.

^b No trace of the expected compound **7c** was detected, but formation of the corresponding *N*-(ferulic acid)acylcarbamide in 76% yield was obtained (see Appendix A).

^c No trace of the desired ferulate **7c** was detected, but di-feruloylated 4NTC was obtained in 49% yield. This was successfully deacetylated using the enzymatic method (E), procuring 4NTC-1,2-yl bis-ferulate in 87% yield (see Appendix A).

^d Other attempts to obtain **3c** using method (D) or with anhydrous K₂CO₃ in a mixture of CH₂Cl₂/MeOH, led to its decomposition.

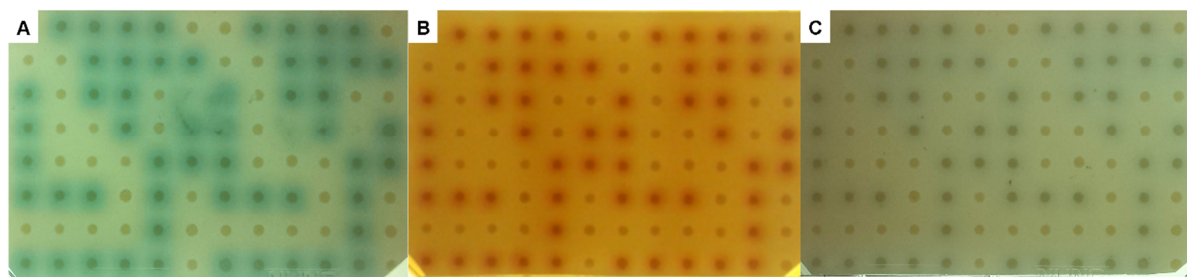


Fig. 2. Detection in solid medium of FAE activity using **3a** (A), **3c** (B) and **4b** (C). *Y. lipolytica* cells producing *AnFAEA* were characterized by the formation of turquoise (A) or dark blue (C) haloes, or the development of a brown colour (C) within and around the colony. Cells not producing *AnFAEA* failed to display haloes and thus remained uncoloured.

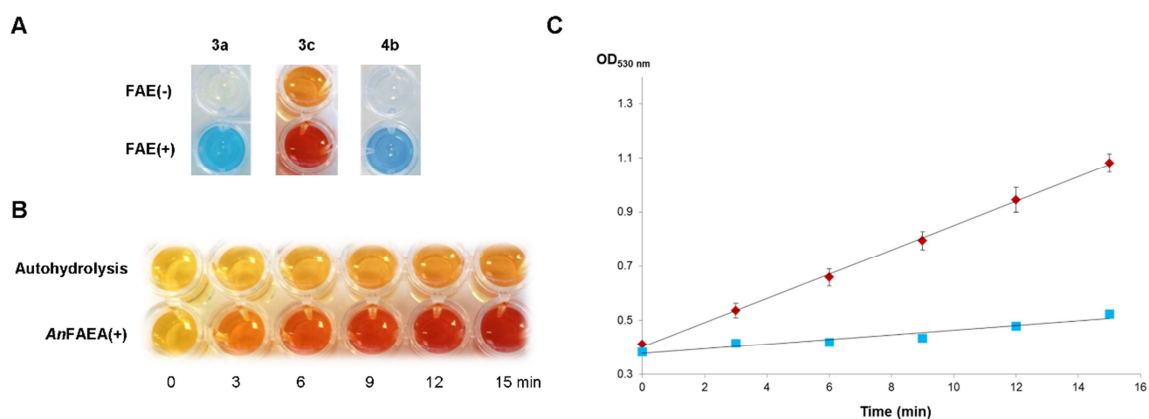


Fig. 3. Detection of FAE activity in liquid medium: hydrolysis of **3a**, **4b** and **3c** in 96-well microtiter plates. (A) Buffered, 30-min reactions in the presence or absence of FAE (+ and - respectively); (B) and (C) Discontinuous time-dependent monitoring of both autohydrolysis (■) and *AnFAEA*-mediated hydrolysis (◆) of **3c**.

Supplementary Data

Design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities

Olga Gherbovet^{a,b,c,1}, Régis Fauré^{a,b,c,1,*}, Fernando Ferreira^{a,b,c}, Julien Durand^{a,b,c}, Mélanie Ragon^{a,b,c}, Guillaume Hostyn^{a,b,c}, Eric Record^{d,e}, Sophie Bozonnet^{a,b,c}, Michael J. O'Donohue^{a,b,c,*}

^a*Université de Toulouse; INSA,UPS,INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France*

^b*INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France*

^c*CNRS, UMR5504, F-31400 Toulouse, France*

^d*INRA, UMR1163 Biotechnologie des Champignons Filamenteux, Aix-Marseille Université, Polytech Marseille, 163 Avenue de Luminy, CP 925, F-13288 Marseille, France*

^e*Aix-Marseille Université, UMR1163 Biotechnologie des Champignons Filamenteux, Faculté des Sciences de Luminy-Polytech, CP 925, F-13288 Marseille cedex 9, France*

* Corresponding authors. Tel: +33 5 61 55 9488.

E-mail address: regis.faire@insa-toulouse.fr (R. Fauré), michael.odonohue@insa-toulouse.fr (M.J. O'Donohue)

¹ These authors contributed equally to this work.

1. General methods

All reactions in organic medium were carried out under nitrogen atmosphere using commercial anhydrous or freshly distilled solvents (Armarego and Lin Chai, 2003). After work-up, organic phases were dried over anhydrous sodium sulfate and filtered. Reaction evolution was monitored by analytical thin-layer chromatography using silica gel 60 F254 precoated plates (E. Merck). Spots were visualized using UV light of 254 nm wavelength followed by soaking in a 0.1% w/v orcinol solution containing a mixture of sulfuric acid/ethanol/water (3:72.5:22.5 v/v/v) followed by charring. Purifications by column chromatography were performed using silica gel Si 60 (40-63 μm) using a bench column or using a Reveleris® flash chromatography automated system (Grace, Epernon, France) equipped with prepacked cartridges (Grace), both with the indicated eluent. Yields refer to chromatographically pure compounds. NMR spectra were recorded on a Bruker Avance II 500 spectrometer at 298 K. Chemical shifts (δ) are given in ppm with residual solvents signal as internal reference (Gottlieb et al., 1997). Coupling constants (J) are reported in Hertz (Hz) with singlet (s), doublet (d), doublet of doublet (dd), doublet of triplet (dt), doublet of doublet of doublets (ddd), broad (br). Analysis and assignments were made using 1D (^1H , ^{13}C and J-modulated spin-echo (J_{mod})) and 2D (COrelated SpectroscopY (COSY) and Heteronuclear Single Quantum Coherence (HSQC)) experiments. High-resolution mass spectra (HRMS) analyses were performed at the CRMPO (Rennes University, France) in positive ionisation mode (ES^+) on either a Waters Q-ToF 2 or a Bruker Micro-ToF-Q II spectrometer. Microanalyses were also recorded at the CRMPO (Rennes University, France).

2. Attempts at direct esterification

Initial attempts to directly esterify **5a** using the conditions described earlier failed to procure the ferulate **7c**, but led to the production of the stable *N*-(ferulic acid)acylcarbamide in 76% yield through the rearrangement of the activated *O*-(ferulic acid)acylisocarbamide intermediate, due to low nucleophilicity of the 4NTC sulfate dipotassium salt. In further experiments, the use of unprotected ferulic acid (instead of acetylated derivative **5a**) procured insoluble products. ^1H NMR of these revealed multiple alkene signals which led to the conclusion that the products were oligomerized reactant. Taking into account the fact that esterification under basic conditions with pyridine as solvent resulted mainly in the production of 4-nitrocatechol-1,2-yl bis-(*O*-acetyl-ferulate) in 49% yield, which was due to the *in situ* deprotection of the sulfate group, access to the ester **7c** was reached by the formation of 4-*O*-acetyl-feruloyl chloride intermediate.

3. Stability experiments

In order to assess the stability of the different compounds with regard to pH and temperature, dilute solutions (50 μ M) of substrate in 100 mM buffer (sodium acetate buffer at pH 3, 4 or 5; sodium phosphate buffer at pH 6, 7 or 8; and sodium carbonate buffer at pH 9 or 10), containing DMSO at final concentration (v/v) of 0.5% for **3a** (Fig. A.1) and **4b** (Fig. A.2) or 4% for **3c** (Figs. A.3-A.4), were heated at 30, 40, 50, 60, 70 and 80 °C during 1 h. After heating, the solutions were cooled to room temperature and their UV-Vis spectra were recorded.

It is noteworthy that the absorption maxima of ferulate derivatives and free ferulic acid (FA) are pH-dependant and distinct (Hatfield et al., 1991; Ralet et al., 1994). As a consequence, when both species are present, ferulates show a maximum of absorption at higher wavelengths whereas the free FA, which corresponds to spontaneous hydrolysis of the ester linkage, shows its maximum absorbance at shorter wavelength.

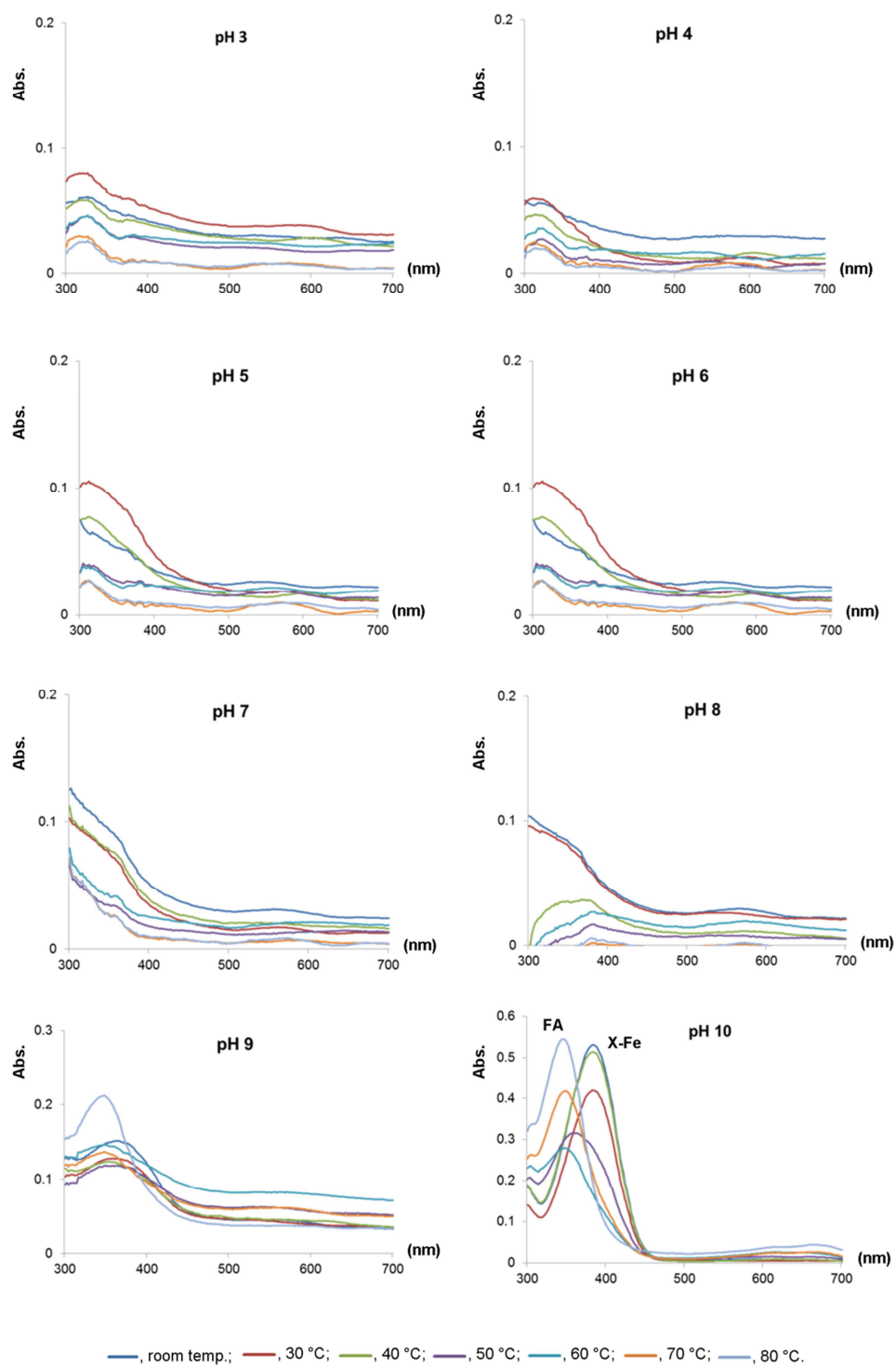


Fig. A.1. Stability of X-Fe (**3a**) monitored using UV-vis spectrophotometry.

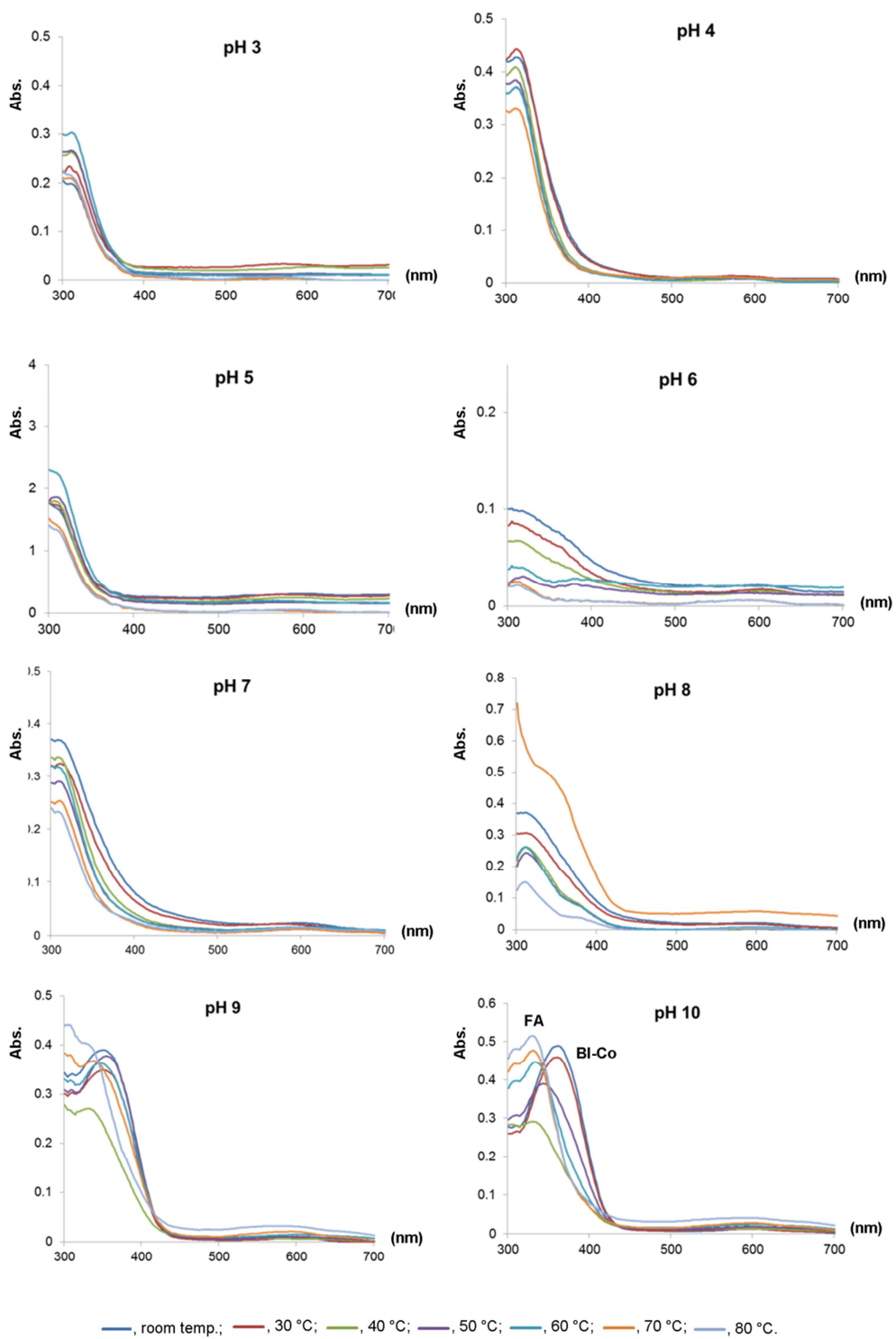


Fig. A.2. Stability of BI-Co (4b) monitored using UV-vis spectrophotometry.

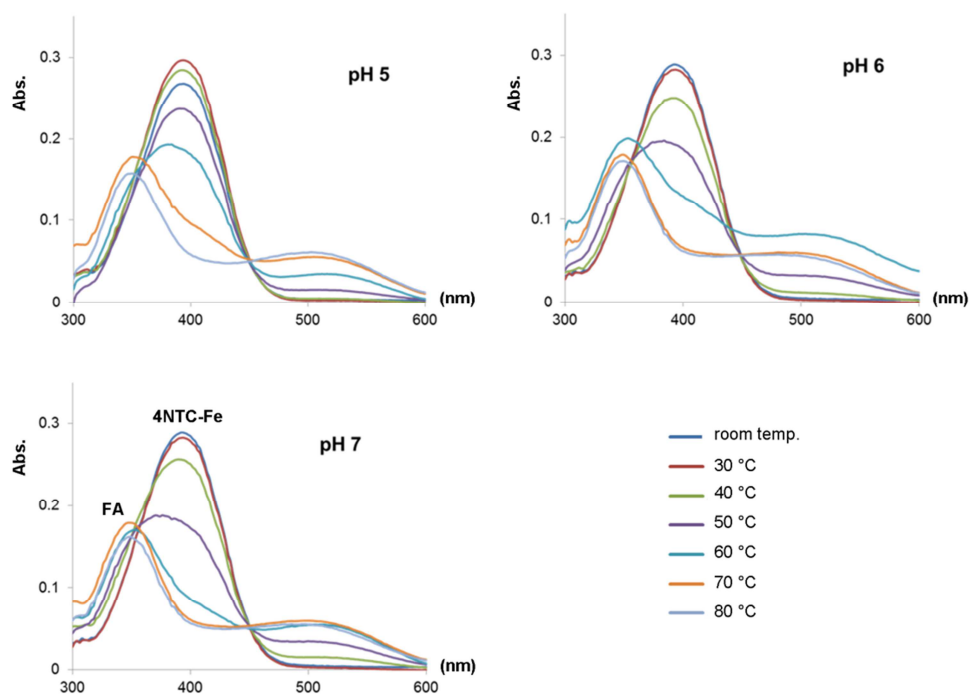


Fig. A.3. Stability of 4NTC-Fe (**3c**) monitored using UV-vis spectrophotometry.

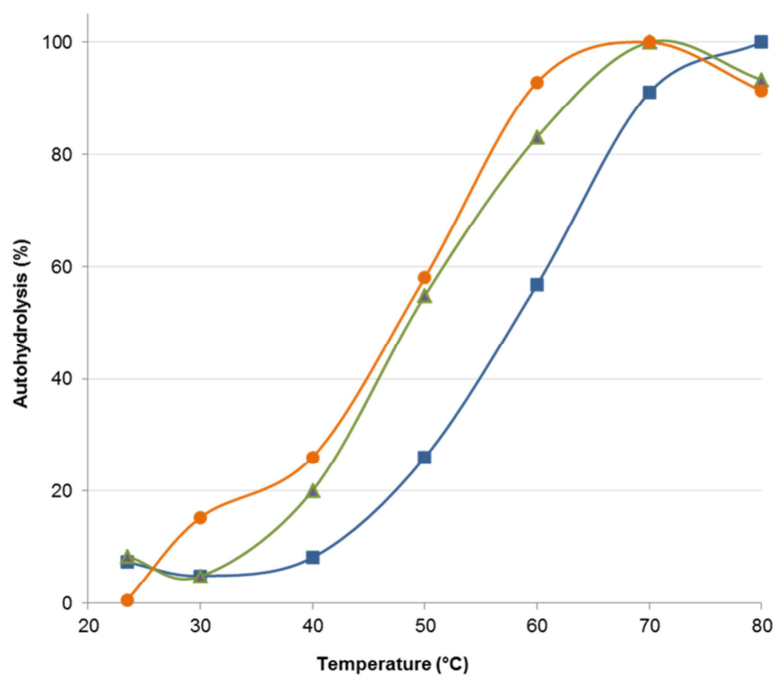


Fig. A.4. Evaluation of the autohydrolysis of 4NTC-Fe **3c** at different temperatures and pH (■, pH 5; ▲, pH 6; and ●, pH 7).

4. Solid phase FAE(+) screening in yeast *Y lipolytica* cells using membranes and automated organisation of colonies

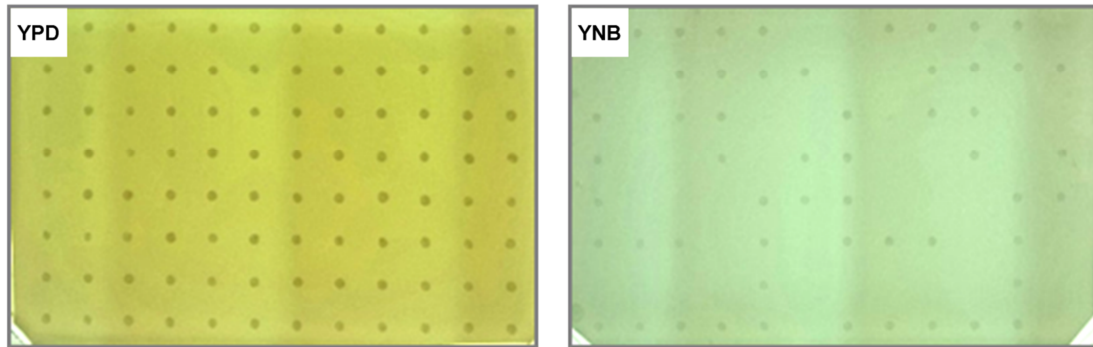


Fig. A.5. Distinguishing colonies that produce *AnFAEA* (and thus are $URA3^+$) on non-selective (YPD) and selective (YNB) media, in the absence of chromogenic substrates. Omnitray plates containing solid YPD or YNB (1.7 g/L YNB without casamino acid, 5 g/L ammonium chloride, 10 g/L glucose, 2 g/L casamino acids and 10 g/L bacto agar in 100 mM sodium phosphate buffer pH 6.0) and overlaid with a nitrocellulose membrane were inoculated using Colony picker QPixII, (Genetix) and incubated at 28 °C for 48 h.

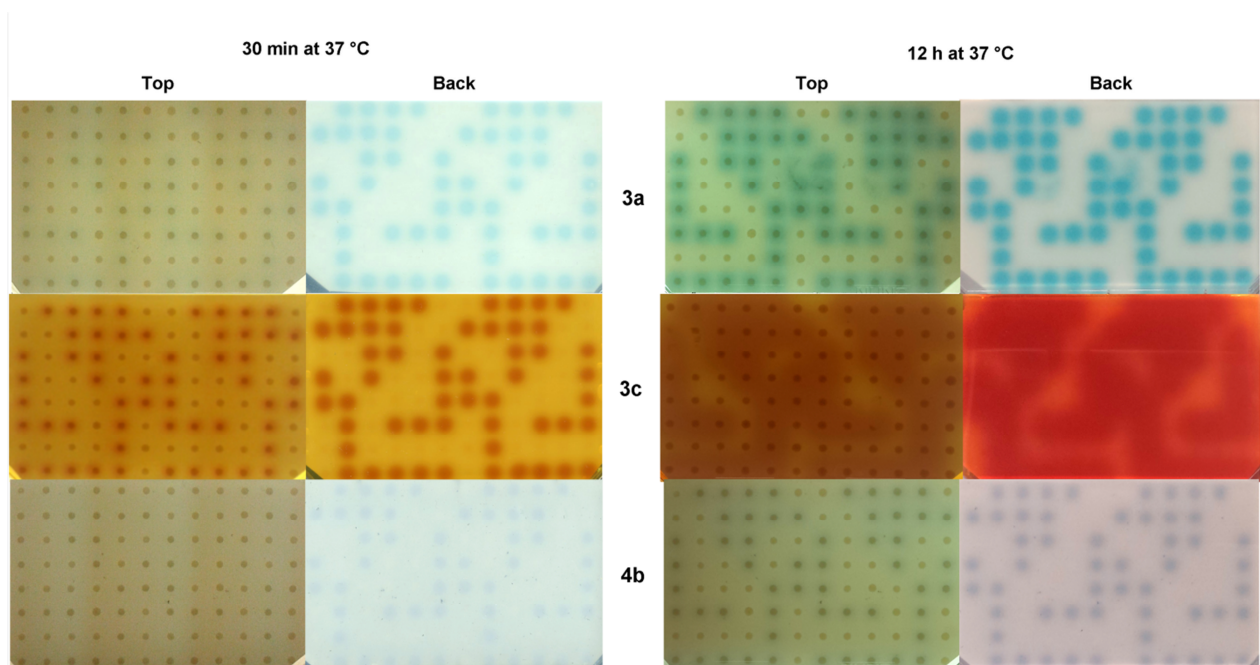


Fig. A.6. Distinguishing colonies that produce *AnFAEA* in the presence of chromogenic substrates **3a**, **3c** and **4b**. The development of coloration was observed at different times (30 min and 12 h).

5. Screening for FAE activity in yeast growing on solid medium using top agar

As an alternative to the use of nitrocellulose membrane, top agar was used to bring growing cells into contact with chromogenic substrates. To achieve this a mixture of cultures of *Y. lipolytica* AnFAEA(+) and AnFAEA(-) strains was used to inoculate solid YNB medium (1.7 g/L YNB without casamino acid, 5 g/L ammonium chloride, 20 mL/L oleic acid, 10 g/L glucose, 2 g/L casamino acids and 15 g/L bacto agar in 100 mM citrate-phosphate buffer pH 6.0) in Petri dishes. To allow colonies to form the dishes were incubated for 48 h at 30 °C before carefully overlaying the solid medium with a preparation of molten 1% (w/v) top agar containing 100 µg/mL of X-Fe **3a** diluted in DMSO (i.e. 1 mg of substrate diluted in 50 µL of DMSO and mixed with 10 mL of top agar). Once the top agar was solid, Petri dishes were incubated at 37 °C for 2 h in order to allow colour to develop.



Fig. A.7. Detection of FAE activity in *Y. lipolytica* AnFaeA(+) cells using the top agar method and monitoring the apparition on indigo-blue coloration.

6. Screening for FAE activity in fungus growing on solid medium using top agar

Spores (10 μL of 2.10^8 spores/mL) of the *Aspergillus niger* AnFaeA(+) transformant (Record et al., 2003) were plated on selective solid minimal medium (MM, without uridine, containing 6.0 g/L NaNO_3 , 0.5 g/L KCl, 1.5 g/L KH_2HPO_4 , 0.5 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g/L glucose, and trace elements (1000x stock: 4.4 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 1.0 g/L $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 1.0 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0.32 g/L $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.32 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1.5 g/L $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, and 10 g/L EDTA) in Petri dishes. To allow colonies to form, the dishes were incubated for 10 days at 30 °C, before carefully overlaying the solid medium with 8 mL molten of 1% (w/v) top agar solution containing 100 $\mu\text{g}/\text{mL}$ of X-Fe **3a** (i.e. 1 mg of substrate diluted in 50 μL of DMSO and mixed with 10 mL of top agar). To develop the coloration, the dishes were then incubated overnight at 30 °C.

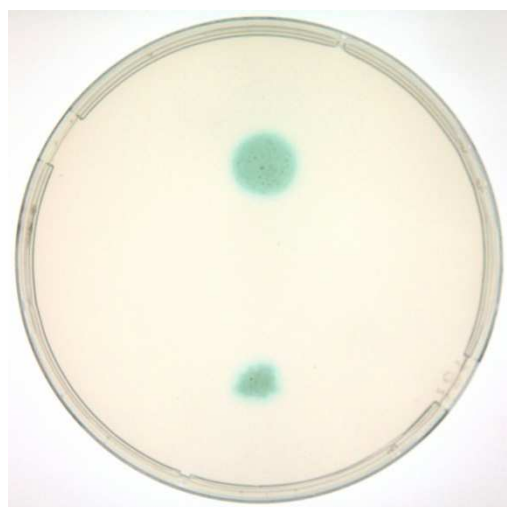


Fig. A.8. Detection of FAE activity in *A. niger* AnFaeA(+) cells by checking the indigo blue colored colonies.

Armarego, W.L.F., Lin Chai, C.L., 2003. Purification of laboratory chemicals (Fifth edition). Elsevier Inc.

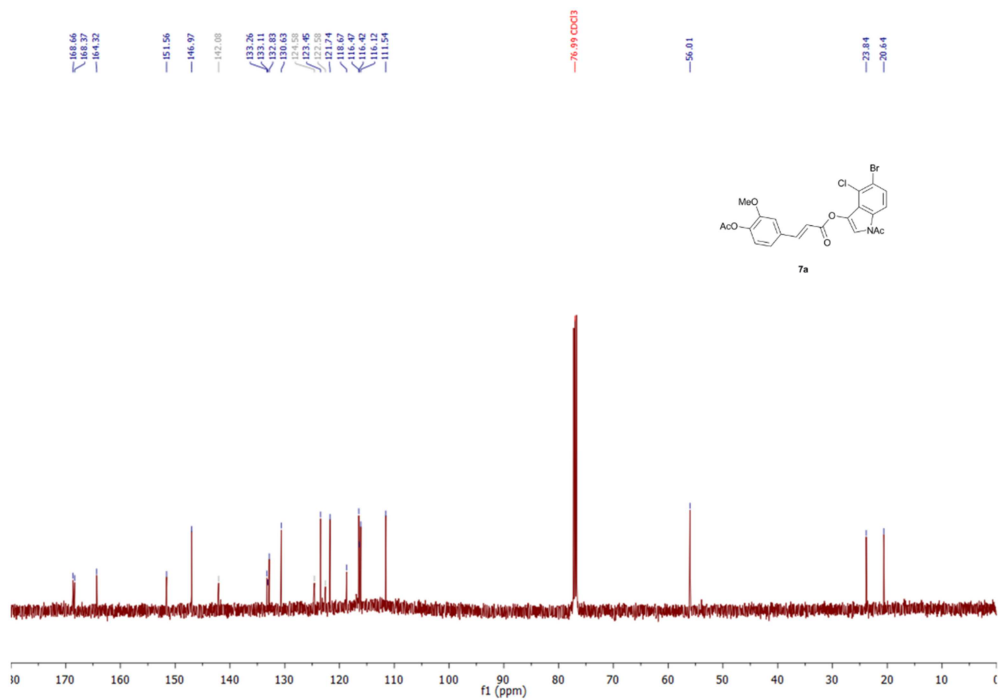
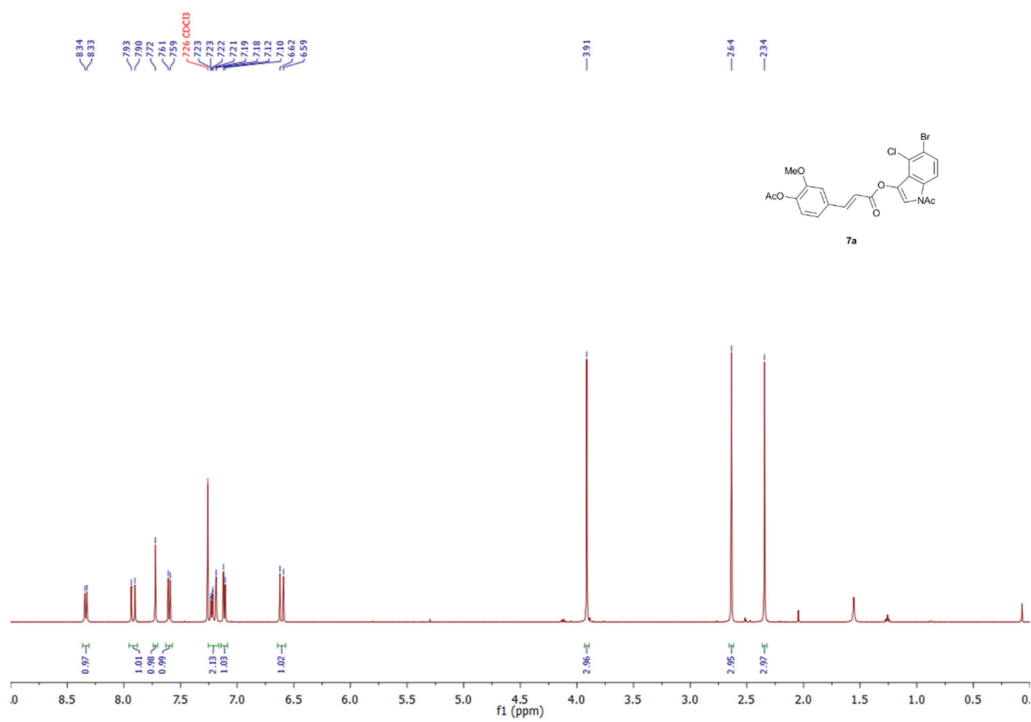
Gottlieb, H.E., Kotlyar, V., Nudelman, A., 1997. NMR chemical shifts of common laboratory solvents as trace impurities. *J Org Chem* 62, 7512–7515.

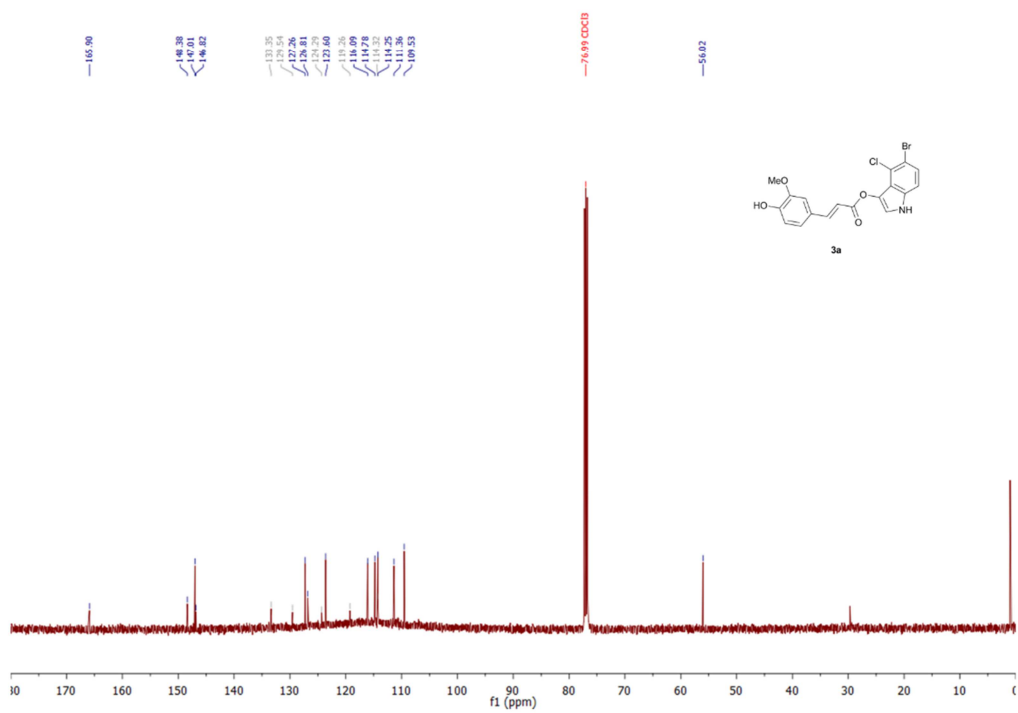
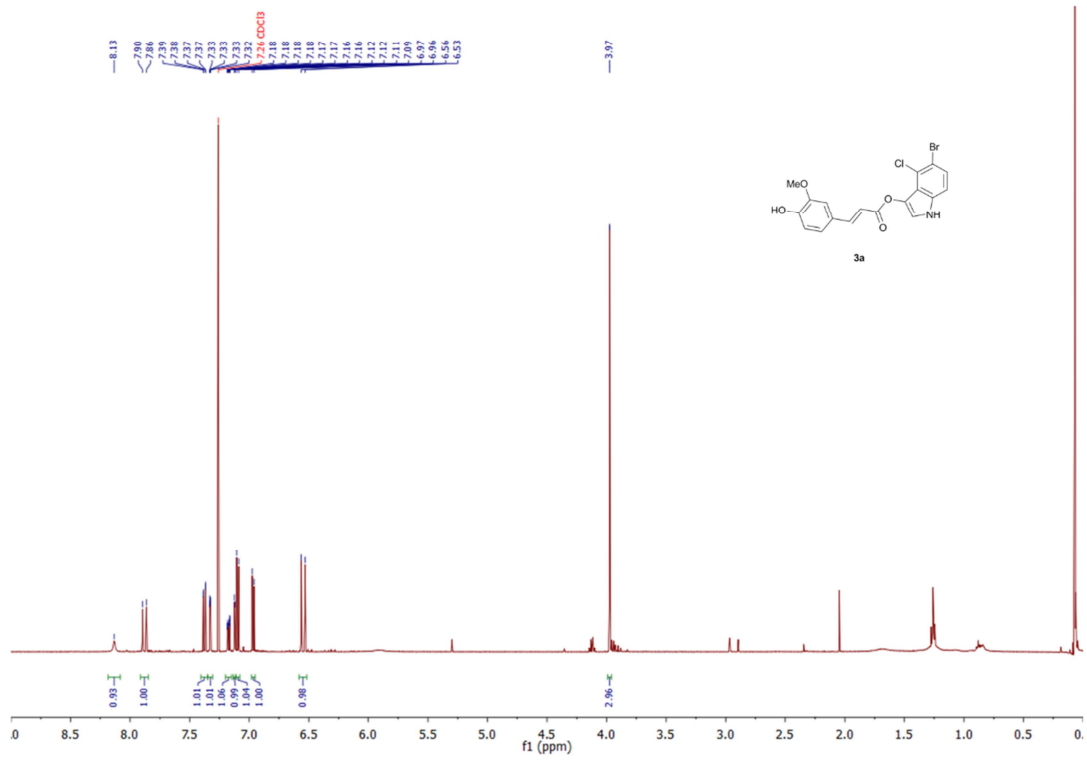
Hatfield, R.D., Helm, R.F., Ralph, J., 1991. Synthesis of methyl 5-0-trans-feruloyl- α -L-arabinofuranoside and its use as a substrate to assess feruloyl esterase activity. *Anal Biochem* 33, 25–33.

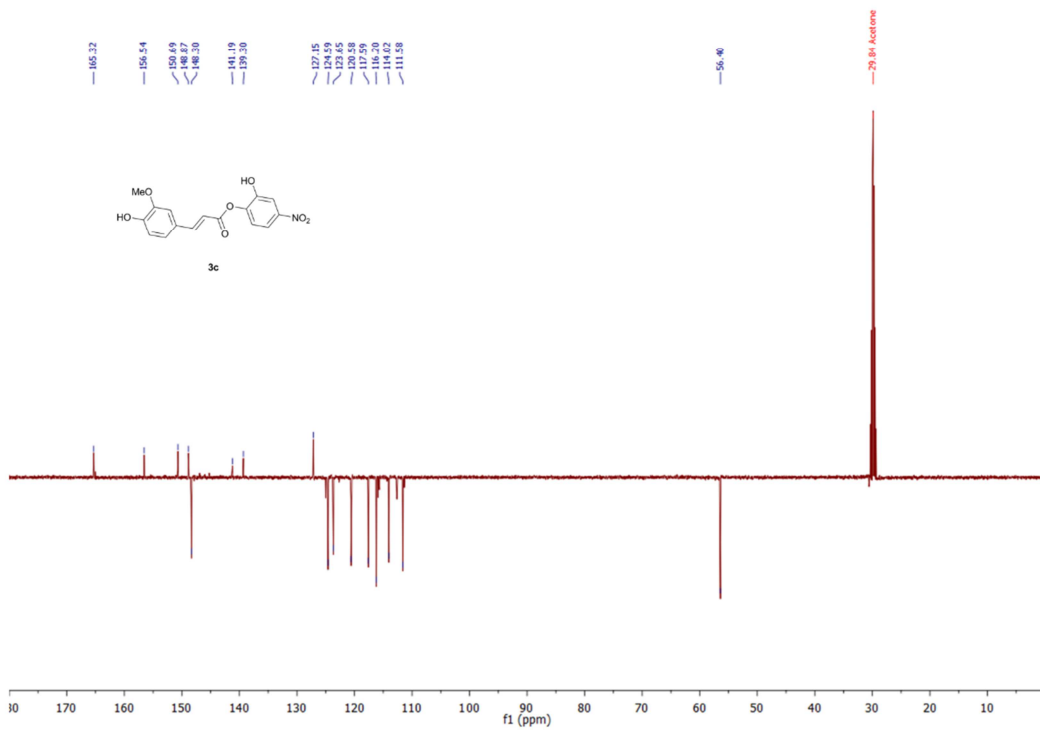
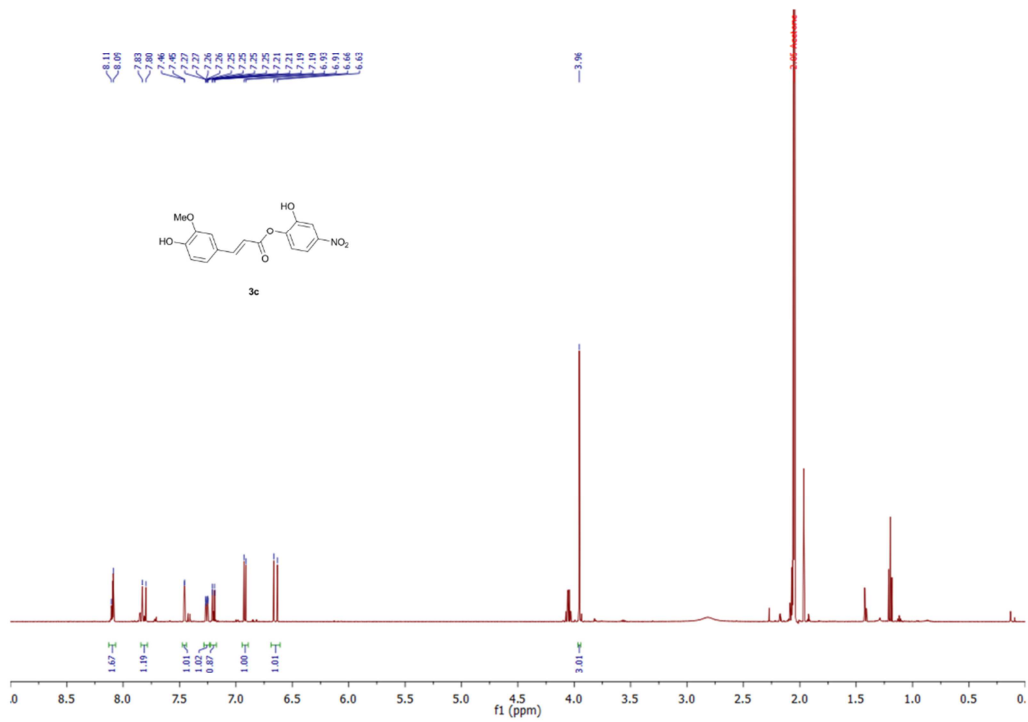
Ralet, M.-C., Faulds, C.B., Williamson, G., Thibault, J.-F., 1994. Degradation of feruloylated oligosaccharides from sugar-beet pulp and wheat bran by ferulic acid esterases from *Aspergillus niger*. *Carbohydr Res* 263, 257–269.

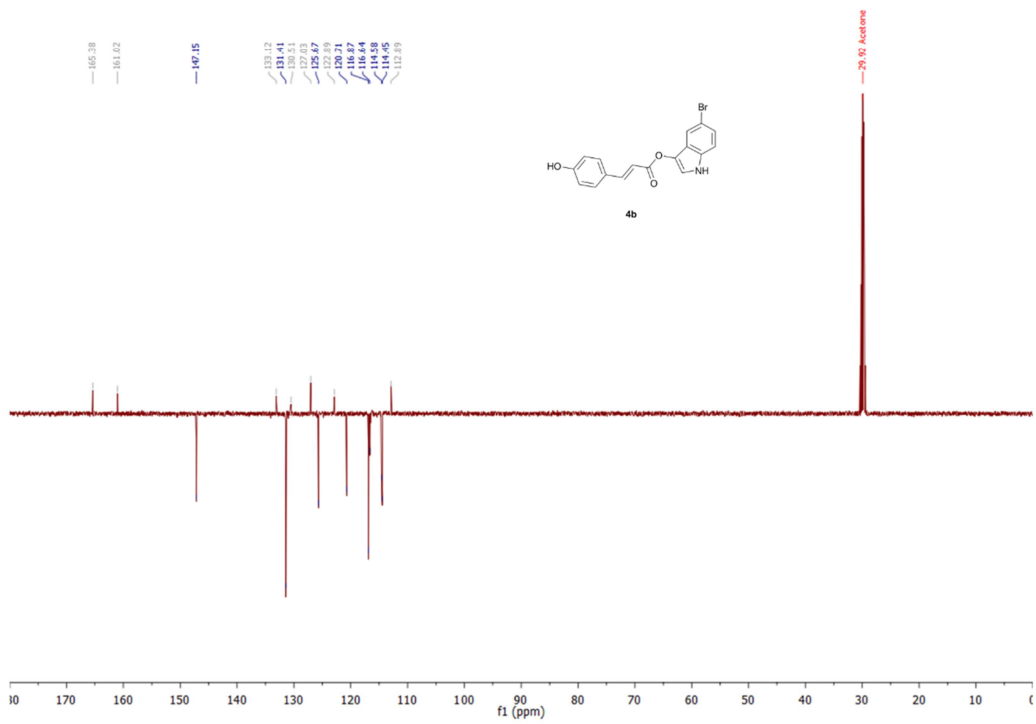
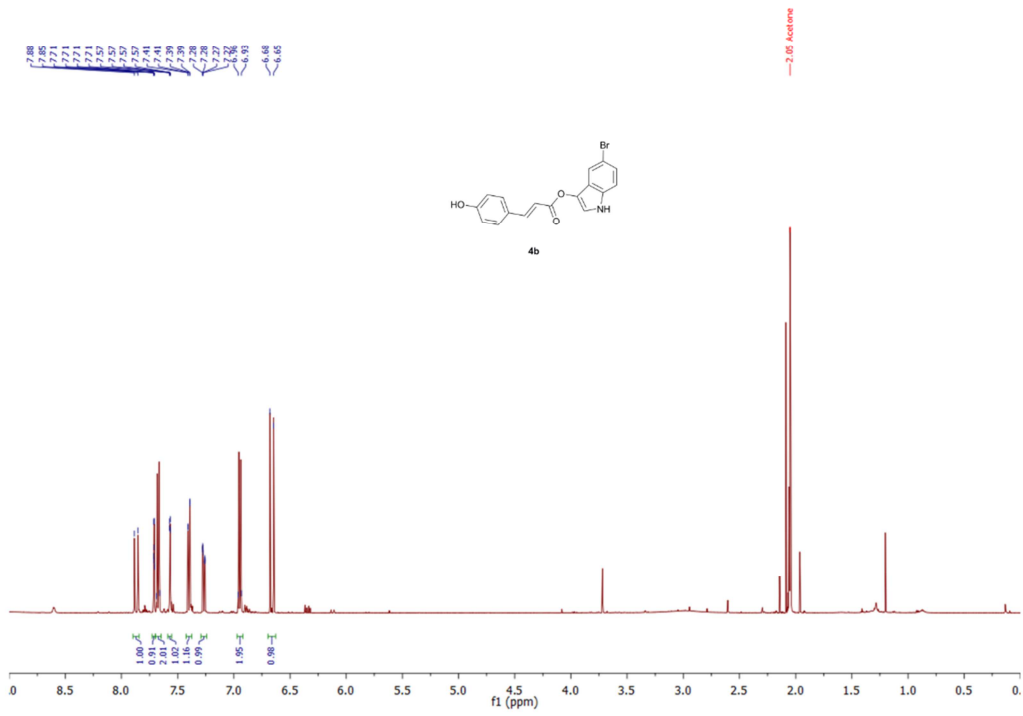
Record, E., Asther, M., Sigoillot, C., Pagès, S., Punt, P.J., Delattre, M., Haon, M., van den Hondel, C.A.M.J.J., Sigoillot, J.-C., Lesage-Meessen, L., Asther, M., 2003. Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Appl Microbiol Biotechnol* 62, 349–355. doi:10.1007/s00253-003-1325-4

7. NMR spectra









ANNEXE 2 :

Regioselective chemoenzymatic syntheses of chromogenic
substrates for feruloyl esterases

(en préparation)

Communication (in preparation)

Regioselective chemoenzymatic syntheses of chromogenic substrates for feruloyl esterases

Olga Gherbovet¹, Fernando Ferreira¹, Mélanie Ragon¹, Julien Durand¹, Sophie Bozonnet¹, Michael J. O'Donohue¹ and Régis Fauré^{1,*}

¹ LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

* Correspondence: regis.faire@insa-toulouse.fr; Tel.: +33-561-55-9488

Abstract: Although feruloyl esterases constitute an important sub-group of hydrolytic enzymes involved in the deconstruction of plant cell wall polysaccharides, their identification and detailed characterization are still challenging topics because of the availability of *ad hoc* substrates. In this respect, appropriate chromogenic substrates are valuable tools to provide quick and easy detection of enzyme-mediated hydrolysis of the ester bond, which can be readily monitored through color change, at the microbial colony level and in liquid assays. Both new and commercially available ferulate derivatives were prepared by regioselective chemoenzymatic synthetic routes. The synthetic pathways leading to 5-bromo-4-chloro-3-indolyl and 4-nitrophenyl *O*-5-feruloyl- α -L-arabinofuranosides **1** and **2** were significantly shortened, and therefore corresponding overall yields enhanced, by using enzymatic transesterification of ferulic acid from its active vinyl ester form to the primary hydroxyl group of α -L-arabinofuranosides with immobilized Lipolase 100T (from *Thermomyces lanuginosus*). Furthermore, the robust feruloylated-butanetriol 4-nitrocatechol analog **13**, containing a cleavable hydroxylated linker instead of a carbohydrate moiety, was also obtained in 29% overall yield in 3 steps, combining regioselective functionalization of 4NTC and enzymatic transesterification. Its suitability as a substrate for type A feruloyl esterase from *Aspergillus niger* was evaluated.

Keywords: biocatalysis; transesterification; screening; CAZymes; ferulic ester hydrolysis

1. Introduction

The development of white biotechnology is underpinned by advances in enzyme discovery and engineering, areas that are being driven by metagenomics and *in vitro* directed enzyme evolution. These techniques efficiently procure the means to massively discover or create new enzyme-encoding sequences, filling up databases with a considerable wealth of information. However, while resolving an early step in the biotechnology pipeline, these techniques have progressively created a new bottleneck, with the discovery rate of putative enzymes being greatly superior to that of enzyme characterization. Therefore, there is a pressing need to extend the enzymologist's toolbox, particularly with regard to informationally rich high-throughput screens that can not only attribute an activity to putative enzymes, but also procure some qualitative details on enzyme properties. In this respect, chromogenic substrates are particularly practical, because they are generally straightforward to synthesize and can provide the basis for both qualitative and quantitative assays that are compatible with automatization.

Feruloyl esterases (FAEs) are of interest, both because of their role in the deconstruction of complex plant-based materials and also as synthetic tools for the preparation of bioactive compounds.[1,2] Operating via a serine protease mechanism, FAEs (EC 3.1.1.73 and family CE1 within the CAZy classification - www.cazy.org)[3] catalyze the hydrolysis of ester bonds linking hydroxycinnamic acids to the glycosyl moieties of plant-based polysaccharides, such as arabinoxylans and arabinans. In this respect, FAEs are important components of plant cell wall-degrading enzymatic arsenal, since the hydrolysis of *trans*-ferulate-polysaccharide linkages often leads to the breakage of the intermolecular polysaccharide-polysaccharide and polysaccharide-lignin associations that characterize the lignocellulosic matrix. Moreover, since ferulic acid represents up to 3% (w/w) of plant cell wall dry weight, FAEs constitute important tools to release this commercially-valuable compound.[4]

So far, the detection and characterization of FAEs have mostly relied on either the use of HPLC or UV-visible spectrophotometry, using natural or synthetic feruloylated substrates.[5] However, the preparation of feruloylated L-arabinofuranoside-based chromogenic substrates has also been reported, with some of the resultant compounds being suitable for high-throughput screening (HTS) assays.[6,7] Nevertheless, these substrates were all obtained through multi-step synthetic pathways, which considerably limits their usefulness. To overcome this, the synthesis and use in HTS assays of L-arabinofuranoside-free chromogenic substrates (i.e., generic for all the four type of FAEs [4,8]) has also been reported. [9–12]

Herein, we report the straightforward synthesis of several feruloylated substrates, including 5-bromo-4-chloro-3-indolyl and 4-nitrophenyl *O*-5-feruloylated α -L-arabinofuranosides, **1** and **2**, and a feruloylated L-arabinofuranoside-free chromogenic substrate **13**. Substrates **1** and **2** are already

commercially available, but very costly. Therefore, this study provides the means to access these molecules more cheaply. Regarding substrate **13**, this contains a cleavable hydroxylated linker that mimics the carbohydrate moiety. The key step of these syntheses involves regioselective enzymatic transferuloylation of vinyl ferulate **3**, catalyzed by commercial Lipolase 100T (from *T. lanuginosus*).[13,14]

2. Results and Discussion

2.1. Chemoenzymatic Synthesis of 5-O-Feruloylated α -L-Arabinofuranosides

The synthesis of chromogenic 5-O-feruloylated α -L-arabinofuranosides is usually achieved using a multistep pathway that involves trapping the furanose conformation, anomeric activation, glycosidation, regioselective deprotection of the primary hydroxyl group, feruloylation and final deprotection to yield the target molecule.[5,7,11,15,16] Additionally, the temporary protection of functional groups is sometimes used during synthesis in order to facilitate certain steps. In this work, some of these steps were replaced by a one-step transesterification of the unprotected vinyl ferulate **3** using Lipolase 100T and readily available and reasonably cheap 4-nitrophenyl or 5-bromo-4-chloro-3-indolyl α -L-arabinofuranosides. This afforded the corresponding feruloylated derivatives, **1** and **2** (Figure 1A). It is noteworthy, that compounds **1** and **2** are already commercially available, but the relatively complicated way in which they are prepared makes them expensive. In this work, one-step regioselective enzymatic feruloylation of the primary hydroxyl group procured high yields (86 and 73% for the *p*NP and the indolyl derivatives respectively) that compare favorably with the overall yields (66 and 46% respectively) reported for the 3-step chemical feruloylation.[7,16] Importantly, the method described herein should be quite generic and applicable to other chromogenic α -L-arabinofuranosidic compounds, such as 4-nitrocatechol, 2-chloro-4-nitrophenyl and umbelliferyl derivatives

To demonstrate how **2** could be used to perform qualitative *in situ* screening of microbial colonies growing on solid agar medium, the compound was incorporated (300 μ g/mL) into a top agar layer along with an α -L-arabinofuranosidase from *Thermobacillus xylanilyticus* (*TxA*b_f). This provided the means to reveal microbial colonies heterologously expressing the type A FAE(+) from *Aspergillus niger* (*AnFAEA*), using an enzyme cascade wherein release of the chromogenic moiety by the *TxA*b_f was made possible by the prior release by *AnFAEA* of the ferulate moiety (Figure 1B). In an alternative demonstration, **2** was also used for colony level detection of FAE using yeast cells (*Yarrowia lipolytica*)[17] that actually coexpressed *AnFAEA* and *TxA*b_f(+) (data not shown).

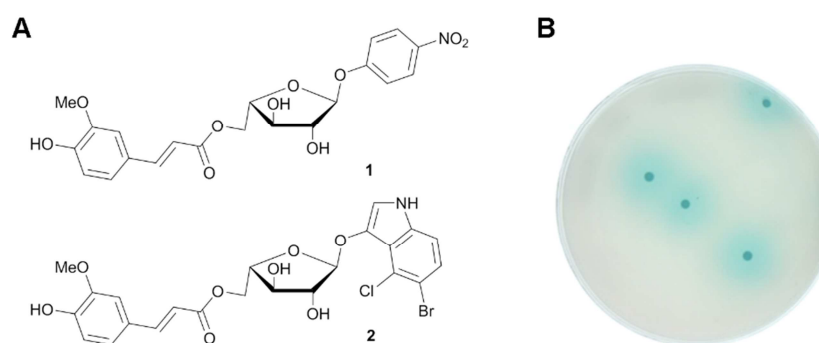
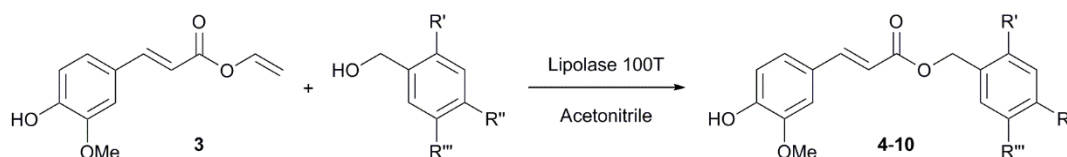


Figure 1. (A) Chemoenzymatically synthesised 5-*O*-feruloyl- α -L-arabinofuranosidic chromogenic substrates, **1** and **2**, and (B) detection of type A FAE activity on solid agar medium using **2**.

2.2. Investigating Enzymatic Transferuloylation Reaction of Non-Glycosidic Motifs

The extent to which Lipolase 100T can catalyze feruloyl transfer reactions involving substituted nitro benzylic alcohols was investigated using an excess of vinyl ferulate **3**[18] (Table 1). Only transesterification with benzylic alcohol was observed, since no phenol acylation was detected during the reaction with hydroxynitrobenzylic alcohol (Table 1, entries 2 and 7), 2-chloro-4-nitrophenol or 4-nitrocatechol (data not shown). Additionally, the exact position of the benzyl alcohol affected transfer, with *ortho*-substitutions displaying a hindered electron withdrawing group procuring the lowest target product yield, lower reactivity and greater hydrolysis of the vinyl ferulate into ferulic acid.

Table 1. Transferuloylation of substituted nitro benzylic alcohols.

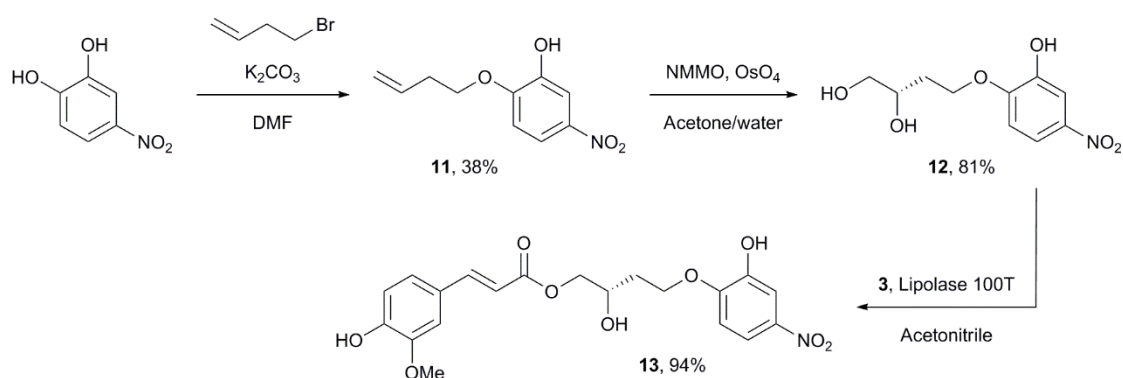


Entry	R'	R''	R'''	Compound	Yield (%) ^a	Ratio ^c
1	H	H	NO ₂	4	79	10/28/62
2	OH	H	NO ₂	5	76	19/32/49
3	Cl	H	NO ₂	6	44	31/38/31
4	F	H	NO ₂	7	46	33/28/39
5	H	NO ₂	H	8	97	10/24/66
6	NO ₂	H	H	9	- ^b	28/48/24
7	NO ₂	H	OH	10	- ^b	27/64/09

^a Yields of isolated ferulates after purification step. ^b The expected ferulates were confirmed by mass spectrometry analysis (HRMS) but the low purity of samples after purification prevented fine structural characterization by NMR. ^c Ratio (in %), determined by ¹H NMR, of the different species of feruloyl derivatives within the crude reaction mixture: remaining vinyl ferulate, ferulic acid (hydrolysis product) and ferulates **4-10**.

2.3. Synthesis of L-Arabinofuranoside Free 4NTC-linker-Fe Chromogenic Substrate **13** and its Evaluation for FAE Assays

To synthesize the chromogenic compound **13** (4NTC-linker-Fe), which contains a 4-nitrocatechol (4NTC) bound via a cleavable linker to a ferulate motif, a multistep route was devised (Scheme 1). First, a new shorter, more practical pathway towards (*S*)-4-*O*-(2-hydroxy-4-nitrophenyl)-1,2,4-butanetriol **12** was developed.[19] Compared to the previously reported 4-step synthesis, two drawbacks were circumvented, notably avoiding (*i*) the preparation of the volatile (*S*)-1-iodo-3,4-*O*-isopropylidene-3,4-butanediol intermediate and (*ii*) the use of a protected version of the chromogenic linker **12**, either for the extra hydroxyl group of the catechol or the secondary hydroxyl group of the linker.[19,9] Alkylation of 4-nitrocatechol (4NTC) with homoallylic bromide in basic conditions gave a mixture of mono- and di-alkylated 4NTC derivatives with **11** (38%) predominating because of the preferential formation of the phenolate at the *para* position (relative to the nitro group).[20] Osmium tetroxide-mediated dihydroxylation in the presence of *N*-methylmorpholine *N*-oxide (NMMO) afforded **12** in 81% yield. Finally, regioselective transferuloylation of the primary hydroxyl of triol derivative **12** with Lipolase 100T was performed and the expected chromogenic substrate **13** was isolated in high 94% yield. Thus, the synthesis of chromogenic ferulate **13** was achieved in 29% overall yield in 3 steps without the requirement to perform sensitive final deprotection.



Scheme 1. Chemoenzymatic synthesis of (*S*)-4-*O*-(2-hydroxy-4-nitrophenyl)-1-*O*-*trans*-feruloyl-1,2,4-butanetriol (**13**, 4NTC-linker-Fe).

As expected, investigation of the stability of the chromogenic substrate **13** using UV-visible spectrophotometry revealed that the presence of the alkyl-like linker gives it higher stability over a wide pH range (up to pH 9.0), irrespective of temperature. These observations were also consistent with the known stability of ester linkages under basic conditions. The usefulness of this newly synthesized 4NTC-linker-Fe was evaluated in an enzyme assay (Figure 2A-B). In order to release the 4NTC chromogenic group, compound **13** was submitted to the hydrolytic action of wild-type AnFAEA[21] at 50 °C (Figure 2C), followed by sequential addition at 0 °C throughout of sodium periodate and ethylene glycol.[19,9,22,23] It is important to mention that stoichiometric addition of

the latter compound is necessary, because an excess of periodate can react with linkerless 4NTC and procure a false absorbance reading (Figure 2B). For the quantitative characterization of FAEs assays can be performed in discontinuous mode, measuring 4NTC release at 505 nm (and alkaline pH) with the reaction progressing from colorless to red as free 4NTC accumulates.

3. Materials and Methods

3.1. Materials and General Methods

4-Nitrophenyl and 5-bromo-4-chloro-3-indolyl α -L-arabinofuranosides were purchased from CarboSynth (Compton, U.K.) and Lipolase 100T was supplied by Novozymes (Bagsvaerd, Denmark). Reaction evolution was monitored by analytical thin-layer chromatography using silica gel 60 F254 precoated plates (E. Merck). Spots were visualized using UV light of 254 nm wavelength followed by soaking in a 0.1% w/v orcinol solution containing a mixture of sulfuric acid/ethanol/water (3:72.5:22.5 v/v/v) followed by charring. Purifications by column chromatography were performed using silica gel Si 60 (40-63 μ m) using a Reveleris® flash

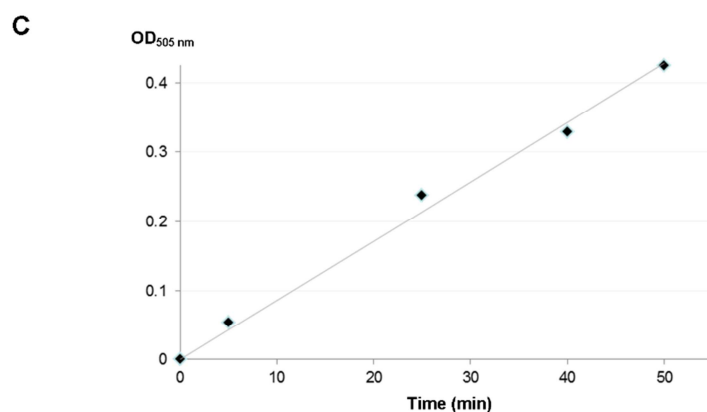
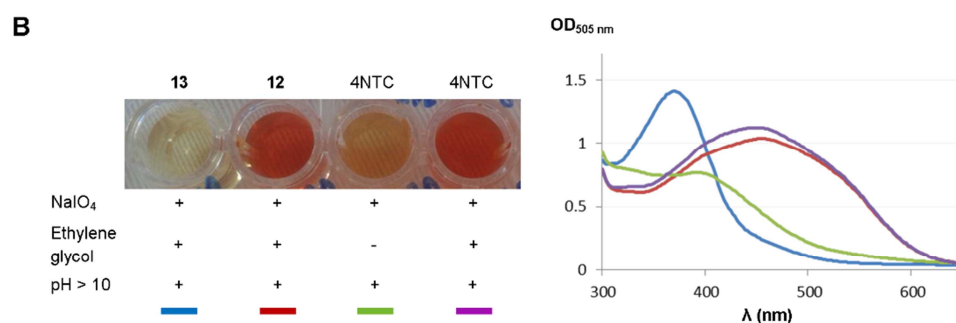
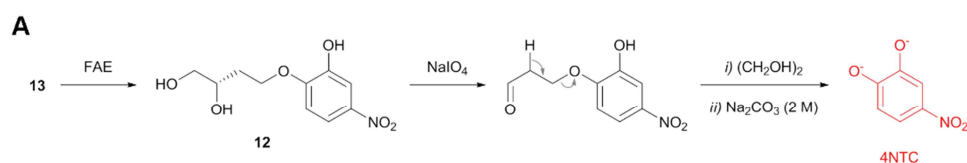


Figure 2. (A) Principle of the 4NTC release after action of FAE on **13** in the presence of sodium periodate. (B) Control reactions and (C) discontinuous time-dependent monitoring of AnFAEA-mediated hydrolysis of **13**.

chromatography automated system (Grace, Epernon, France) equipped with prepacked cartridges (Grace). NMR spectra were recorded on a Bruker Avance II 500 spectrometer at 298 K. Chemical shifts (δ) are given in ppm with residual solvents signal as internal reference.[24] Coupling constants (J) are reported in Hertz (Hz) with singlet (s), doublet (d), triplet (t), doublet of doublet (dd), doublet of doublet of doublets (ddd), broad (br) and quadruplet of triplet (qt). Analysis and assignments were made using 1D (^1H , ^{13}C and J-modulated spin-echo (Jmod)) and 2D (CORrelated SpectroscopY (COSY) and Heteronuclear Single Quantum Coherence (HSQC)) experiments. High-resolution mass spectra (HRMS) analyses were performed at PCN-ICMG (Grenoble, France).

3.2. General Procedure for Enzymatic Transesterification

The enzymatic transesterification steps were performed according to a published protocol.[13] Briefly, Lipolase 100T (1 g) was added to a solution of alcohol (0.30 mmol, 1 equiv.) and vinyl ferulate **3** (100 mg, 0.45 mmol, 1.5 equiv.) in acetonitrile (6 mL). The reaction mixture was stirred overnight at 37 °C, then filtered, filter cake washed with acetone and the filtrate was evaporated to dryness. The residue was recovered in ethyl acetate, washed with saturated aqueous sodium hydrogencarbonate (three times). Combined organic phases were dried over anhydrous sodium sulfate, filtered and concentrated under reduced pressure. Flash chromatography (gradient of ethyl acetate in petroleum ether from 0 to 50%) afforded pure ferulates **1**, **2**, **4-10** and **13**.

3-Nitrobenzyl *trans*-ferulate (4, 78 mg, 0.24 mmol, 79%), white powder. ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$): δ 8.32 (1H, t, J 1.7, CH of Bn), 8.21 (1H, ddd, J 8.3, 2.2, 1.1, CH of Bn), 7.90 (1H, ddd, J 7.6, 1.7, 1.0, CH of Bn), 7.71 (1H, t, J 7.6, CH of Bn), 7.68 (1H, d, J 15.8, CH=CHCO₂), 7.37 (1H, d, J 2.0, CH of Fe), 7.17 (1H, dd, J 8.3, 2.0, CH of Fe), 6.87 (1H, d, J 8.3, CH of Fe), 6.50 (1H, d, J 15.8, CH=CHCO₂), 5.38 (2H, s, CH₂ of Bn), 3.92 (3H, s, OMe). ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$): δ 167.2 (C=O), 150.3 (C_q), 149.4 (C_q), 148.8 (C_q) 146.7 (CH=CHCO₂), 140.2 (C_q), 135.1 (CH of Bn), 130.8 (CH of Bn), 127.3 (2C_q Fe), 124.2 (CH of Fe), 123.6 (CH of Bn), 123.4 (CH of Bn), 116.1 (CH of Bn), 115.1 (CH=CHCO₂), 111.4 (CH of Fe), 65.1 (CH₂ of Bn), 56.4 (OMe). HRMS (ESI) calc. for [M-H]⁻ C₁₇H₁₄NO₆ m/z 328.0821, found: 328.0833.

2-Hydroxy-5-nitrobenzyl *trans*-ferulate (5, 79 mg, 0.23 mmol, 76%), white powder. ^1H NMR (500 MHz, $(\text{CD}_3)_2\text{CO}$): δ 8.29 (1H, d, J 2.8, CH of Bn), 8.14 (1H, dd, J 8.9, 2.8, CH of Bn), 7.68 (1H, d, J 15.8, CH=CHCO₂), 7.38 (1H, d, J 1.8, CH of Fe), 7.17 (1H, dd, J 8.0, 1.8, CH of Fe), 7.12 (1H, d, J 8.9, CH of Bn), 6.88 (1H, d, J 8.0, CH of Fe), 6.52 (1H, d, J 15.8, CH=CHCO₂), 5.33 (2H, s, CH₂ of Bn), 3.92 (3H, s, OMe). ^{13}C NMR (125 MHz, $(\text{CD}_3)_2\text{CO}$): δ 167.5 (C=O), 162.0 (C_q) 150.3 (C_q), 148.8 (C_q), 148.7 (C_q), 146.6 (CH=CHCO₂), 141.6 (C_q),

127.4 (C_q), 126.3 (2CH of Bn), 124.3 (CH of Fe), 116.5 (CH of Bn), 116.1 (CH of Fe), 115.2 (CH=CHCO₂), 111.3 (CH of Fe), 61.1 (CH₂ of Bn), 56.4 (OMe). HRMS (ESI) calc. for [M-H]⁻ C₁₇H₁₄NO₇ *m/z* 344.0770, found: 344.0778.

2-Chloro-5-nitrobenzyl *trans*-ferulate (6, 48 mg, 0.13 mmol, 44%), white powder. ¹H NMR (500 MHz, (CD₃)₂CO): δ 8.42 (1H, br d, *J* 3.0, CH of Bn), 8.26 (1H, dd, *J* 9.0, 3.0, CH of Bn), 7.81 (1H, d, *J* 9.0, CH of Bn), 7.71 (1H, d, *J* 16.0, CH=CHCO₂), 7.40 (1H, d, *J* 1.8, CH of Fe), 7.19 (1H, dd, *J* 8.0, 1.8, CH of Fe), 6.88 (1H, d, *J* 8.0, CH of Fe), 6.56 (1H, d, *J* 16.0, CH=CHCO₂), 5.43 (2H, s, CH₂ of Bn), 3.93 (3H, s, OMe). ¹³C NMR (125 MHz, (CD₃)₂CO): δ 167.1 (C=O), 164.6 (C_q), 150.4 (C_q), 148.8 (C_q), 147.1 (CH=CHCO₂), 140.6 (C_q), 137.4 (C_q), 131.7 (CH of Bn), 127.3 (C_q), 125.2 (2CH of Bn), 124.4 (CH of Fe), 116.1 (CH of Fe), 114.7 (CH=CHCO₂), 111.4 (CH of Fe), 63.0 (CH₂ of Bn), 56.4 (OMe). HRMS (ESI) calc. for [M-H]⁻ C₁₇H₁₃ClNO₆ *m/z* 362.0431, found: 362.0431.

2-Fluoro-5-nitrobenzyl *trans*-ferulate (7, 48 mg, 0.14 mmol, 46%), white powder. ¹H NMR (500 MHz, (CD₃)₂CO): δ 8.43 (1H, dd, *J* 3.2, 2.9, CH of Bn), 8.33 (1H, ddd, *J* 9.0, 4.3, 3.0 Hz, CH of Bn), 7.67 (1H, d, *J* 16.0, CH=CHCO₂), 7.49 (1H, t, *J* 9.0, CH of Bn), 7.36 (1H, d, *J* 1.8, CH of Fe), 7.16 (1H, dd, *J* 8.0, 1.8, CH of Fe), 6.87 (1H, d, *J* 8.0, CH of Fe), 6.49 (1H, d, *J* 16.0, CH=CHCO₂), 5.39 (2H, s, CH₂ of Bn), 3.91 (3H, s, OMe). ¹³C NMR (125 MHz, (CD₃)₂CO): δ 167.1 (C=O), 165.1 (d, *J* 257.4, C_q of Bn), 150.4 (C_q of Fe), 148.8 (C_q of Fe), 146.9 (CH=CHCO₂), 127.2 (C_q of Fe), 127.1 (d, *J* 6.1, CH of Bn), 126.9 (d, *J* 10.5, CH of Bn), 126.7 (d, *J* 17.1, C_q of Bn), 124.3 (CH of Fe), 117.7 (d, *J* 24.3, CH of Bn), 116.1 (CH of Fe), 114.7 (CH=CHCO₂), 111.4 (CH of Fe), 59.6 (d, *J* 3.9, CH₂ of Bn), 56.3 (OMe). HRMS (ESI) calc. for [M-H]⁻ C₁₇H₁₃FNO₆ *m/z* 346.0727, found: 346.0731.

4-Nitrobenzyl *trans*-ferulate (8, 96 mg, 0.29 mmol, 97%), white powder. ¹H NMR (500 MHz, (CD₃)₂O): δ 8.27 (2H, m, CH of Bn), 7.72 (2H, m, CH of Bn), 7.69 (1H, d, *J* 16.0, CH=CHCO₂), 7.35 (1H, d, *J* 2.0, CH of Fe), 7.17 (1H, dd, *J* 8.2, 2.0, CH of Fe), 6.88 (1H, d, *J* 8.2, CH of Fe), 6.51 (1H, d, *J* 16.0, CH=CHCO₂), 5.38 (2H, s, CH₂ of Bn), 3.92 (3H, s, OMe). ¹³C NMR (125 MHz, (CD₃)₂CO): δ 167.1 (C=O), 150.3 (C_q), 148.8 (C_q), 148.6 (C_q) 146.7 (CH=CHCO₂), 145.4 (C_q), 129.4 (CH of Bn), 127.3 (C_q), 124.4 (CH of Fe), 124.2 (CH of Bn), 116.0 (CH of Fe), 115.0 (CH=CHCO₂), 111.4 (CH of Fe), 65.1 (CH₂ of Bn), 56.4 (OMe). HRMS (ESI) calc. for [M-H]⁻ C₁₇H₁₄NO₆ *m/z* 328.0821, found: 328.0825.

3.3. Synthesis of the 4NTC-linker-Fe 14

2-*O*-(But-3-enyloxy)-5-nitrophenol (11). To a solution of 4-nitrocatechol (1.00 g, 6.44 mmol, 1 equiv.) in dry DMF (8 mL) were added potassium carbonate (1.06 g, 7.67 mmol, 1.2 equiv.) and homoallylic bromide (670 μL, 6.51 mmol, 1 equiv.) at 40 °C. After overnight stirring at 40 °C, the reaction mixture was concentrated under reduced pressure. The residue was recovered

in ethyl acetate, washed with saturated aqueous sodium hydrogencarbonate and brine. Combined organic phases were dried over anhydrous sodium sulfate, filtered and concentrated under reduced pressure. Flash chromatography (gradient of ethyl acetate in petroleum ether from 0 to 50%) afforded **12** (516 mg, 2.47 mmol, 38%), as a solid. ¹H NMR (500 MHz, CDCl₃): δ 7.82 (1H, dd, *J* 9.0, 3.0, CH of 4NTC), 7.80 (1H, d, *J* 3.0, CH of 4NTC), 6.90 (1H, d, *J* 9.0, CH of 4NTC), 5.91-5.83 (1H, m, =CH), 5.23-5.16 (2H, m, CH₂=), 4.21 (2H, t, *J* 6.6, CH₂O), 2.63 (2H, br qt, *J* 6.6, 1.6, CH₂). ¹³C NMR (500 MHz, CDCl₃): δ 151.3 (C_q), 145.8 (C_q), 142.1 (C_q), 133.2 (=CH), 118.1 (CH₂=), 116.8 (CH of 4NTC), 110.4 (CH of 4NTC), 110.1 (CH of 4NTC), 68.5 (CH₂O), 33.2 (CH₂). HRMS (ESI) calc. for [M-H]⁻ C₁₀H₁₀NO₄ *m/z* 208.0610, found: 208.0615.

(S)-4-O-(2-Hydroxy-4-nitrophenyl)-1,2,4-butanetriol (12). [19] A solution of **12** (157 mg, 0.75 mmol, 1 equiv.) in an acetone/water mixture (2.5-1, 3.5 mL) was treated at 25 °C under stirring with *N*-methylmorpholine-*N*-oxide (106 mg, 0.90 mmol, 1.2 equiv.) and osmium tetroxide (2.5% solution in *tert*-butanol, 38 μL) and stirred at room temperature for 18 h. 10% aqueous sodium sulfite (0.5 mL) was added and stirring was prolonged for 30 min. The product was extracted with ethyl acetate (three times with 10 mL), washed with brine. Combined organic phases, dried over anhydrous sodium sulfate, filtered and concentrated under reduced pressure, afforded the expected compound **13** (165 mg, 0.68 mmol, 90%) as a solid.

(S)-4-O-(2-Hydroxy-4-nitrophenyl)-1-O-*trans*-feruloyl-1,2,4-butanetriol (13). Application of the general procedure for enzymatic transesterification with **13** (70 mg, 0.29 mmol, 1 equiv.) and vinyl ferulate **3** (94 mg, 0.43 mmol, 1.5 equiv.) to give **14** (113 mg, 0.27 mmol, 94%) as a white powder. ¹H NMR (500 MHz, (CD₃)₂CO): δ 7.79 (1H, dd, *J* 9.0, 2.8, CH of 4NTC), 7.67 (1H, d, *J* 2.8, CH of 4NTC), 7.63 (1H, d, *J* 16.0, CH=CHCO₂), 7.33 (1H, d, *J* 2.0, CH of Fe), 7.20 (1H, d, *J* 9.0, CH of 4NTC), 7.13 (1H, dd, *J* 8.2, 2.0, CH of Fe), 6.87 (1H, d, *J* 8.2, CH of Fe), 6.40 (1H, d, *J* 16.0, CH=CHCO₂), 4.47-4.42 (1H, m), 4.40-4.36 (1H, m), 4.23-4.16 (3H, m), 3.95 (3H, s, OMe), 2.18-2.11 (1H, m), 2.01-1.94 (1H, m). ¹³C NMR (125 MHz, (CD₃)₂CO): δ 167.4 (C=O), 153.6 (C_q), 150.2 (C_q), 148.8 (C_q), 147.7 (C_q), 146.0 (CH=CHCO₂), 142.4 (C_q), 127.4 (C_q), 124.0 (CH of Fe), 117.1 (CH of 4NTC), 116.1 (CH of Fe), 115.7 (CH=CHCO₂), 112.4 (CH of 4NTC), 111.3 (CH of Fe), 110.8 (CH of 4NTC), 69.0 (CH₂), 67.1 (CH₂), 66.9 (CH), 56.3 (OMe), 33.7 (CH₂). HRMS (ESI) calc. for [M-H]⁻ C₂₀H₂₀NO₉ *m/z* 418.1138, found: 418.1139.

3.4. Screening of FAE(+) microorganisms in solid medium using X- α -L-Araf-Fe **2**

Y. lipolytica AnFAEA(+) strain was used to inoculate solid YNB medium (1.7 g/L YNB without casamino acid, 5 g/L ammonium chloride, 20 mL/L oleic acid, 10 g/L glucose, 2 g/L casamino acid, and 15 g/L bacto agar in 100 mM citrate-phosphate buffer pH 5. Petri dishes were incubated for 48 h at 30 °C and then overlaid with a preparation of molten 1% (w/v) top agar containing chromogenic substrate **2** (300 μg/mL and 0.5% DMSO) and TxAbf (2 IU/mL).

Once the top agar was solid, incubation at 37 °C for 1 h allows color to develop. *Y. lipolytica* strains that contain no *AnFAEA* gene were also checked to remain colorless after addition of chromogenic substrate and auxiliary enzyme.

3.5. Liquid medium-based colorimetric assays using 4NTC-linker-Fe **14**

Discontinuous enzyme assays were performed in buffered conditions (100 mM sodium phosphate, pH 6.0) in the presence of 1.8 mM **13** and 4% DMSO. Enzymatic reactions were incubated at 50 °C for 60 min. Aliquots (25 µL) were stopped by cooling (at 0 °C) and mixed with 45 µL of cooled 10 mM NaIO₄ solution (pH 2.0). After keeping 5 min at 0 °C throughout, 45 µL of ethylene glycol were added, followed by 135 µL of 2 M Na₂CO₃ after 5 min. The optical densities at 505 nm were recorded on a VersaMax microplate reader (molecular devices). Control reactions containing **13**, **12** and 4NTC without enzyme and 4NTC without both enzyme and ethylene glycol were also prepared.

4. Conclusions

The regioselectivity of the Lipolase 100T-catalyzed transesterification of vinyl ferulate to the primary hydroxyl group of benzylic alcohols and polyhydroxylated compounds were investigated. Three compounds for the detection and/or characterization of FAE activity have thus been synthesized leading to simplified syntheses of commercially-available substrates, **1** and **2**, and the design of a new chromogenic L-arabinosyl-free substrate, 4NTC-linker-Fe **13**, obtained in 29% overall yield in 3 steps. Their suitabilities as a substrate for type A feruloyl esterase from *Aspergillus niger* were evaluated. The practical synthesis at preparative scale of this library of chromogenic probes might makes easier further investigation of feruloyl esterases.

Acknowledgments: This work was supported by the European Union's Seventh Programme for Research, Technological Development and Demonstration under Grant Agreement No 613868, OPTIBIOCAT project (to O.G. and J.D.), by the French National Research Agency, grant ANR-05-PNRB-002, project SPPECABBE (to M.R.) and the Région Midi-Pyrénées grants DAER-Recherche 07009817(to F.F.). The NMR work carried out in this work at the LISBP (Toulouse, France) was performed with the equipment of Meta-Toul (Metabolomics & Fluxomics Facilities, Toulouse, France, www.metatoul.fr). MetaToul is part of the national infrastructure MetaboHUB (The French National infrastructure for metabolomics and fluxomics, www.metabohub.fr) and is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, SICOVAL, IBiSa-France, CNRS and INRA. We thank the ICEO facility dedicated to enzyme screening and discovery, and part of the Integrated Screening Platform of Toulouse (PICT, IBiSA) for providing access to its equipment. The authors wish to acknowledge the support from the ICMG Chemistry Nanobio Platform (Grenoble, France) for HRMS analyses.

Author Contributions: O.G., S.B. and R.F. conceived and designed the experiments; O.G., F.F., M.R. and J.D. performed the experiments; O.G., F.F., M.R., S.B, M.J.O'D. and R.F. analyzed the data; O.G., S.B., M.J.O'D. and R.F wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Faulds, C. B. What can feruloyl esterases do for us? *Phytochem. Rev.* **2010**, *9*, 121–132.
2. Dumon, C.; Song, L.; Bozonnet, S.; Fauré, R.; O'Donohue, M. J. Progress and future prospects for pentose-specific biocatalysts in biorefining. *Process Biochem.* **2012**, *47*, 346–357.
3. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495.
4. Dilokpimol, A.; Mäkelä, M. R.; Aguilar-Pontes, M. V.; Benoit-Gelber, I.; Hildén, K. S.; de Vries, R. P. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol. Biofuels* **2016**, *9*, 231.
5. Ramos-de-la-Pena, A. M.; Contreras-Esquivel, J. C. Methods and substrates for feruloyl esterase activity detection , a review. *J. Mol. Catal. B Enzym.* **2016**, *130*, 74–87.
6. Biely, P.; Mastihubová, M.; van Zyl, W. H.; Prior, B. A. Differentiation of feruloyl esterases on synthetic substrates in a α -arabinofuranosidase-coupled and ultraviolet-spectrophotometric assays. *Anal. Biochem.* **2002**, *311*, 68–75.
7. Marmuse, L.; Asther, M.; Fabre, E.; Navarro, D.; Lesage-Meessen, L.; Asther, M.; O'Donohue, M.; Fort, S.; Driguez, H. New chromogenic substrates for feruloyl esterases. *Org. Biomol. Chem.* **2008**, *6*, 1208–1214.
8. Crepin, V. F.; Faulds, C. B.; Connerton, I. F. Functional classification of the microbial feruloyl esterases. *Appl. Microbiol. Biotechnol.* **2004**, *63*, 647–652.
9. Marmuse, L.; Asther, M.; Navarro, D.; Lesage-Meessen, L.; Asther, M.; Fort, S.; Driguez, H. Chromogenic substrates for feruloyl esterases. *Carbohydr. Res.* **2007**, *342*, 2316–2321.
10. Hegde, S.; Srinivas, P.; Muralikrishna, G. Single-step synthesis of 4-nitrophenyl ferulate for spectrophotometric assay of feruloyl esterases. *Anal. Biochem.* **2009**, *387*, 128–129.
11. Zhang, S.-B.; Ma, X.-F.; Pei, X.-Q.; Liu, J.-Y.; Shao, H.-W.; Wu, Z.-L. A practical high-throughput screening system for feruloyl esterases: substrate design and evaluation. *J. Mol. Catal. B Enzym.* **2012**, *74*, 36–40.
12. Gherbovet, O.; Fauré, R.; Ferreira, F.; Durand, J.; Ragon, M.; Hostyn, G.; Record, E.; Bozonnet, S.; O'Donohue, M. J. Enzymatic design of chromogenic probes for efficient screening and evaluation of feruloyl esterase-like activities. *J. Mol. Catal. B Enzym.* **2016**, *126*, 24–31.
13. Mastihubová, M.; Mastihuba, V.; Bilaničová, D.; Boreková, M. Commercial enzyme preparations catalyse feruloylation of glycosides. *J. Mol. Catal. B Enzym.* **2006**, *38*, 54–57.
14. Wang, Y.; Zhang, D.-H.; Chen, N.; Zhi, G.-Y. Synthesis of benzyl cinnamate by enzymatic esterification of cinnamic acid. *Bioresour. Technol.* **2015**, *198*, 256–261.
15. Mastihubová, M.; Szemesová, J.; Biely, P. Two efficient ways to 2-*O*- and 5-*O*-feruloylated 4-nitrophenyl α -L-arabinofuranosides as substrates for differentiation of feruloyl esterases. *Tetrahedron Lett.* **2003**, *44*, 1671–1673.
16. Mastihubová, M.; Biely, P. Preparation of regioselectively feruloylated *p*-nitrophenyl α -L-arabinofuranosides and β -D-xylopyranosides-convenient substrates for study of feruloyl esterase specificity. *Carbohydr. Res.* **2010**, *345*, 1094–1098.

17. Bordes, F.; Fudalej, F.; Dossat, V.; Nicaud, J.-M.; Marty, A. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods* **2007**, *70*, 493–502.
18. Armesto, N.; Ferrero, M.; Fernandez, S.; Gotor, V. Novel enzymatic synthesis of 4-*O*-cinnamoyl quinic and shikimic acid derivatives. *J. Org. Chem.* **2003**, *68*, 5784–5787.
19. Borsenberger, V.; Dornez, E.; Desrousseaux, M.-L.; Courtin, C. M.; O'Donohue, M. J.; Fauré, R. A substrate for the detection of broad specificity α -L-arabinofuranosidases with indirect release of a chromogenic group. *Tetrahedron Lett.* **2013**, *54*, 3063–3066.
20. Tavares, F. X.; Al-Barazanji, K. A.; Bigham, E. C.; Bishop, M. J.; Britt, C. S.; Carlton, D. L.; Feldman, P. L.; Goetz, A. S.; Grizzle, M. K.; Guo, Y. C.; Handlon, A. L.; Hertzog, D. L.; Ignar, D. M.; Lang, D. G.; Ott, R. J.; Peat, A. J.; Zhou, H.-Q. Potent, selective, and orally efficacious antagonists of melanin-concentrating hormone receptor 1. *J. Med. Chem.* **2006**, *49*, 7095–7107.
21. Record, E.; Asther, M.; Sigoillot, C.; Pagès, S.; Punt, P. J.; Delattre, M.; Haon, M.; van den Hondel, C. A. M. J. J.; Sigoillot, J.-C.; Lesage-Meessen, L.; Asther, M. Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application. *Appl. Microbiol. Biotechnol.* **2003**, *62*, 349–355.
22. Badalassi, F.; Wahler, D.; Klein, G.; Crotti, P.; Reymond, J. A versatile periodate-coupled fluorogenic assay for hydrolytic enzymes. *Angew. Chem., Int. Ed.* **2000**, *39*, 4067–4070.
23. Grognum, J.; Wahler, D.; Nyfeler, E.; Reymond, J. Universal chromogenic substrates for lipases and esterases. *Tetrahedron: Asymmetry* **2004**, *15*, 2981–2989.
24. Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR chemical shifts of common laboratory solvents as trace impurities. *J. Org. Chem.* **1997**, *62*, 7512–7515.