Composition du Jury

Pr. C. Remacle (Président du jury)
Université catholique de Louvain

Pr. P. Soumillion (Promoteur)
Université catholique de Louvain

Pr. J. Fastrez
Université catholique de Louvain

Pr. B. Elias
Université catholique de Louvain

Pr. D.B. Janssen
Rijksuniversiteit Groningen

Pr. A.D. Griffiths
Université de Strasbourg

Remerciements

Tout d'abord, j'aimerais remercier mon promoteur qui m'a confié ce projet, le Professeur Patrice Soumillion, impliqué dès le début de cette aventure. Merci de m'avoir fait confiance, d'avoir été présent au jour le jour et d'avoir relu avec tant de patience mon manuscrit. Merci pour la formation scientifique que j'ai pu acquérir au cours de ces années de thèse. J'aimerais également remercier le Professeur Jacques Fastrez pour ses conseils avisés tout au long du projet. I would like to thank the members of my jury for the time they spent reading and reviewing my thesis.

Cette thèse fut longue (1993 jours exactement…), les remerciements seront donc longs ! Mon projet m'a amenée à travailler dans plusieurs laboratoires, favorisant donc les rencontres et la naissance de belles amitiés. Merci les BIOP-LIPP-BGMB (on ne sait plus très bien) ! Tout d'abord, je tiens à remercier Vanina et Géraldine qui ont travaillé sur la pénicilline acylase avant que je commence ma thèse. Laurence, tu es la seule à avoir été là avec moi dans les labos du début à la fin de ma thèse. Merci pour ta bonne humeur, quel plaisir de t'avoir eue comme collègue, je n'écouterai plus jamais Vive La Fête ou Depeche Mode sans penser à toi ! Gene, être voisines de bureau dans un bureau exclusivement masculin était un vrai bonheur ! Tant pour les discussions scientifiques que non scientifiques… J'espère que, malgré la distance, nous continuerons à nous voir régulièrement. Merci à tous les mecs qui nous entouraient d'avoir supporté nos cancans. Bruno, désolée d'avoir fait tant souffrir tes nerfs quand il s'agissait d'ordre… Mais heureusement, tu n'es pas rancunier ! Merci de m'avoir fait une petite place sur ton bureau de rédaction pendant mes semaines de rédaction en Belgique. Bon courage pour la fin, tu n'es plus très loin ! Heykel, toi aussi, c'est pour bientôt, courage ! Philippe, merci pour la touche alsacienne en Belgique et les bons petits bredele de Noël. Val, malgré ta perpétuelle indécision, on a toujours réussi à ce que tu nous accompagnes dans nos sorties ! J-M, on a bien senti ton départ, preuve que tu ne passais pas inaperçu avec ta naïveté (oui, je n'étais pas la seule, n'oublie pas que j'ai toujours gagné tous nos paris !) et tes citations chaque fois parfaitement « adaptées » à la situation.

Merci les filles d'être arrivées pour nous soutenir, Mary, puis Anast et Steph les deux inséparables. Ce fut un plaisir de passer ces quelques années avec vous. L'ambiance « filles » du labo est entre vos mains, faites-en bon usage ! Merci Véro pour tout ce que tu as fait pour moi, sans toi ces montagnes de paperasses me seraient parues insurmontables ! Merci également à Antonio, la reine Carole, Ioana, Alex, Benjamin, Vaidas, Virginie, Gilles, Pierre, Pierre-Yves, Olivier et tous les autres qui furent un temps membres de ce labo où règne une ambiance accueillante qui m'a permis de rester motivée chaque jour.

J'aimerais particulièrement remercier le Professeur Andrew Griffiths pour ses conseils et son accueil au sein de son laboratoire à l'Université de Strasbourg, où j'ai pu apprendre les techniques microfluidiques et les appliquer à mon sujet de thèse ; ainsi qu'Isabelle sans qui le côté pratique des choses n'aurait pas été si simple. Je tiens également à remercier Michaël et son équipe, Sophie, Faith, Shigeyoshi, Annick, avec qui j'ai eu le plaisir de travailler au quotidien, qui ont eu le courage de supporter mon effet positif très marqué sur l'entropie générale de leur labo ! Merci encore pour les bons moments passés ensemble à comparer le goût des chips belges et françaises, arrosées de bières belges et alsaciennes.

I would like to thank all the LBC members for their help and advice, more specifically Abdes, Estelle, Thomas Mangeat, Oliver, Felix and Dave. I have also been very touched by your warm welcome. Special thanks to my flat mates Corinna, Lucas, Ashleigh for the wonderful moments spent together. I also want to thank Victoire, Marion and Thomas and Estelle for their kind hospitality! Estelle, j'aimerais te remercier particulièrement pour ton canapé-lit si confortable dont j'ai pu profiter pendant ces quelques mois, et l'amitié qui s'est tissée peu à peu entre nous. How nice was it to spend those evenings around Strasbourg in the company of all of you. You are all more than welcome in Paris or in Brussels; I'm waiting for you with Pêche-mel-Bush's and tuna-peaches!

La partie chimie organique n'aurait pu se faire sans l'accueil du Professeur Olivier Riant et toute son équipe dans le département de Chimie à Louvain-la-Neuve ; ainsi que les précieux conseils d'Olivier Schicke à Louvain, et d'Abdes et d'Estelle à Strasbourg. J'ai commencé sans y connaître grand-chose et grâce à vous, j'ai beaucoup appris.

Collaboration beaucoup plus brève, mais tellement efficace en termes production de résultats, merci Konstantin pour ton aide précieuse en RT-PCR ! Et de manière plus générale, merci aux Bani pour leur microscope à fluorescence. J'aimerais également remercier Fabienne Courtois pour le don du plasmide pGFP, ainsi que Michaël, Ashleigh et Geneviève qui ont pris le temps de relire mon manuscrit.

Merci à mes amis grâce à qui j'ai eu une vie sociale bien remplie durant cette thèse, je pense entre autres à Virg, Mario, Sophie, Alix et Sybille. Merci également à Valérie, Sophie, Elise, Sylvie, Auriane, Hélène, Béné et Caro pour ces pauses-midis passées ensemble ! Un merci tout particulier à mes parents qui m'ont toujours soutenue durant ces 11 années d'études sans jamais avoir bien compris ce que je faisais vraiment de ces gouttes fluo… Merci à mes frères et sœurs, Dimitri, Joséphine, Elisabeth, Augustin et Kristina pour l'ambiance chaleureuse qui a toujours régné à la maison pendant toutes ces années. Merci également à ma belle-famille, Grand-Mamy et Siméon pour leur présence, et à Gabrielle de Groër pour son accueil bienveillant lors de mon premier séjour à Strasbourg.

Merci à Antoine d'avoir supporté mes allers-retours incessants entre Bruxelles et Strasbourg[*]*, mon bureau-rédaction de 3 mètres de long installé au milieu du salon duquel j'arrivais encore à déborder. Je suis sûre que tu as cru que cette thèse ne se terminerait jamais ! Malgré tout, tu as réussi à tout supporter, preuve irréfutable que tu es donc bien le plus grand, le plus beau et le plus fort ! Je le savais !*

---

[*] *Merci la « GWU » de m'avoir chaque fois si fidèlement conduite à bon port, bravant la neige, la glace et les pannes de phares…*

Abréviations et symboles

6-APA	6-aminopenicillanic acid
7-ACA	7-aminocephalosporanic acid
7-ADCA	7-aminodeacetoxycephalosporanic acid
°C	Celsius
A	Ampere
aa	Amino acid
AC	Alternating current
AFET	Activity fed enzyme translation
amp	Ampicilline
ara	Arabinose
ARN	Acide ribonucléique
ARNm	Acide ribonucléique messager
ARNt	Acide ribonucléique de transfert
bp	Base pair
BSA	Bovine serum albumin
CA	Cephalosporin acylase
CFU	Colony forming unit
Da	Dalton
DEP	Dielectrophoresis
DsbAss	DsbA signal sequence
DTT	Dithiothréitol
ε	Molar extinction coefficient
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
em.	Emission wavelength
epPCR	Error prone PCR
ex.	Exitation wavelength
FACS	Fluorescence Activated Cell Sorter
FADS	Fluorescence Activated Droplet Sorter
g	Gram
GFP	Green Fluorescent Protein
glu	Glucose
h	Hour
HRCA	Hyperbranched rolling circle amplification
HR-EPL	High Rate error prone PCR library
IVC	In vitro compartmentalization
IVTT	In vitro transcription and translation
Kb	Kilobase
k_{cat}	Catalytic constant
kDa	Kilo Dalton
K_M	Michaelis-Menten constant

λ	Mean number of gene per droplet
L	Litre
LB	Luria Bertani (medium)
LR-EPL	Low Rate error prone PCR library
Min	Minute
NIPAB	6-nitro-3-phenylacetamino benzoic acid
OD	Optical density
ORF	Open Reading Frame
PAA	Phenylacetic acid
PAC	Penicillin G acylase
PAGE	PolyAcrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein data bank
pH	Potentiel hydrogen
PMT	Photomultiplier tube
PTFE	Poly(tetrafluoroethylene)
RC	Rolling Circle
RFU	Relative fluorescence unit
RNA	Ribonucleic acid
RNase	Endoribonuclease
rpm	Rotations par minute
RT	Room Temperature
RT-PCR	Reverse transcription PCR
s	Second
SDS	Sodium Dodecyl Sulfate
SRP	Signal Recognition Particule
ss	Signal sequence
TAE	Tris Acetate EDTA
TBS	Tris Buffered Saline
Tet	Tétracycline
TIR	translation initiation region
UV	Ultra Violet
V	Volt
WT	Wild type

Table des matières

INTRODUCTION

Chapitre 1 : La pénicilline G acylase d'*E. coli*

Introduction

Les β-lactames acylases sont des enzymes capables d'hydrolyser les chaines latérales des antibiotiques de type β-lactames, tout en laissant le noyau β-lactame intact. Plusieurs β-lactames acylases ont été découvertes et classées en fonction de leur spécificité de substrat. Elles sont séparées en deux groupes en fonction de leur spécificité pour le noyau β-lactame qui est soit l'acide 6-aminopénicillanique (6-APA) pour les pénicillines acylases soit l'acide 6-aminocéphalosporanique (7-ACA) pour les céphalosporines acylases (Figure 1.1). Les pénicillines acylases sont présentes chez les bactéries gram négatives et gram positives, les levures et les champignons filamenteux. Elles peuvent être classées en trois classes sur base de leur préférence pour la chaine latérale des pénicillines : les pénicillines G acylases, pénicilline V acylases et ampicilline acylases (Figure 1.1).[1] Les céphalosporines acylases quant à elles, sont toutes capables d'hydrolyser le glutaryl-7-ACA, mais certaines d'entre elles possèdent également une activité non négligeable pour la céphalosporine C.[2,3]

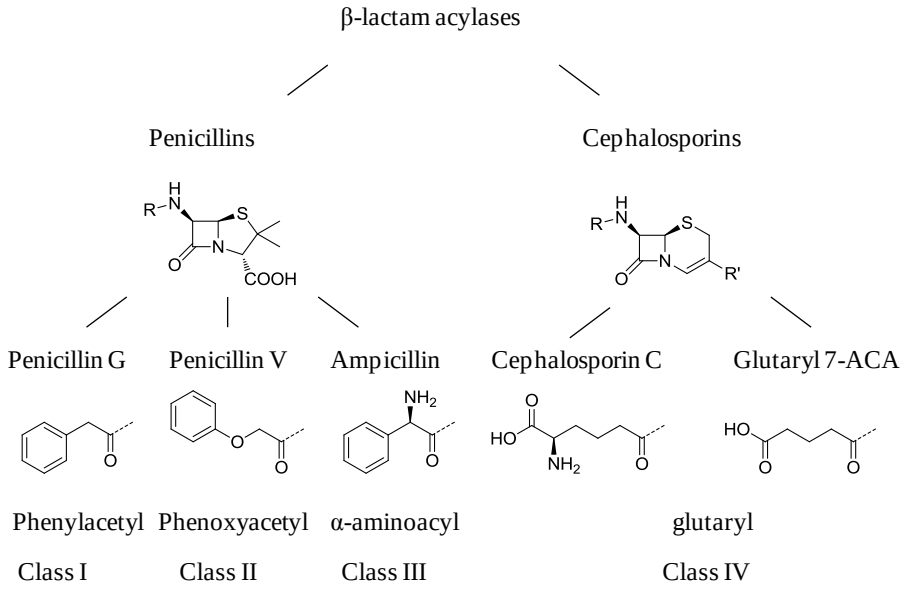


Figure 1.1 : Nomenclature des β-lactames acylases. Elles sont séparées en deux groupes : les pénicillines et céphalosporines acylases, en fonction de leurs préférences pour le noyau β-lactame (soit le 6-APA pour les pénicillines ou le 7-ACA pour les céphalosporines). Elles sont ensuite classées en quatre classes en fonction de leur préférence pour la chaine latérale « R » de l'antibiotique. (Figure adaptée de la thèse de Polderman-Tijmes, J.[4])

Nous avons choisi de travailler avec la pénicilline G acylase d'*E. coli*. Cette enzyme hétérodimérique est synthétisée sous forme d'un prépropeptide de 846 acides aminés dans le cytoplasme. Dans un premier temps, un peptide signal de 26 acides aminés permet l'exportation du peptide dans le périplasme par la voie Tat.[5] Ensuite, le propeptide interne de 54 acides aminés est excisé selon un processus autocatalytique donnant une enzyme hétérodimérique active composée d'une sous-unité α et d'une sous-unité β (Figure 1.2, ce mécanisme est expliqué en détail dans la section suivante). Son rôle physiologique reste encore à élucider, bien qu'il semblerait qu'elle soit impliquée dans le catabolisme de composés phénylacétylés. Par exemple, la pénicilline G acylase hydrolyse divers substrats esters ou amides phénylacétylés et libère de l'acide phénylacétique qui peut être utilisé comme source de carbone pour son hôte.[6,7] Cette hypothèse est soutenue par l'équipe de Maria Prieto qui a étudié le

métabolisme des cycles aromatiques chez *E. coli* et sa régulation. En effet, chez *E. coli*, le gène *pac* se trouve dans un cluster de gènes impliqués dans le catabolisme de composés aromatiques et plus particulièrement dans le métabolisme de l'acide 4-hydroxyphénylacétique, ce qui indiquerait que l'enzyme jouerait un rôle dans la dégradation et l'assimilation de composés aromatiques.[8,9]

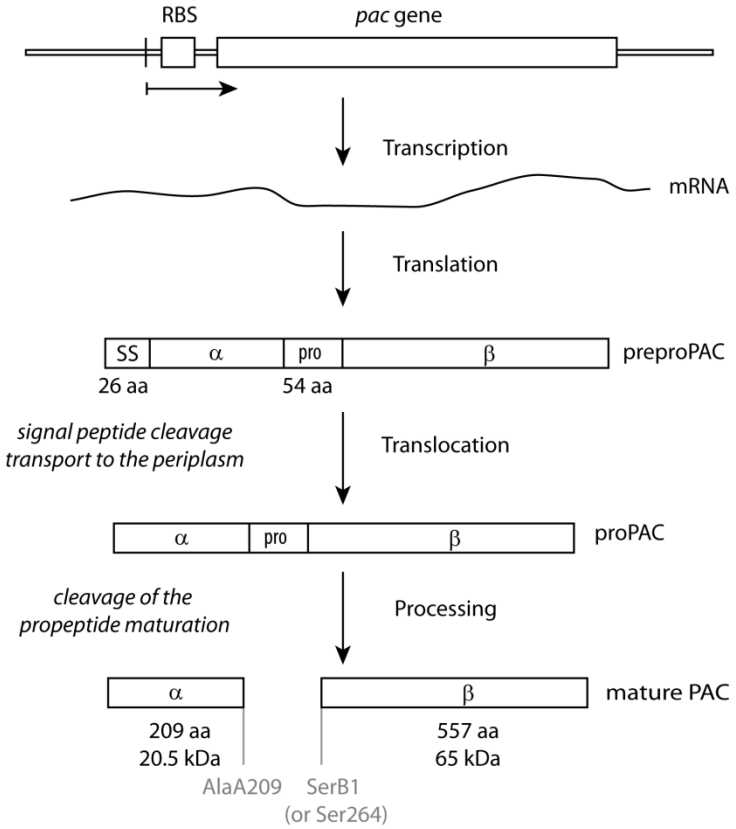


Figure 1.2 : schéma de l'expression du gène *pac* dans *E. coli*. (SS : signal sequence, aa : amino acids, pro : propetide. Figure adaptée de J. Rajendhran et P. Gunasekaran.[8])

L'expression du gène *pac* d'*E. coli* est régulée par plusieurs facteurs différents. Elle est soumise au niveau de sa transcription à la répression catabolique en présence de glucose, fructose ou glycérol,[10,11] et est également induite par l'acide phénylacétique et le cAMP.[9,12] La température influence

également l'expression de la PAC. L'enzyme est produite sous forme mature entre 24 et 30°C, mais pas à 37°C. L'inactivation de l'enzyme par la température ne peut pas en être la cause car la température optimale pour l'activité de l'enzyme est d'environ 50°C.[13,14] Il semblerait tout d'abord que la température ait un effet direct sur la conformation du précurseur, et que la protéine soit bloquée au niveau de la membrane sans être transportée vers le périplasme. De plus, il a été démontré que la régulation par la température s'opérait également au niveau de la traduction. Il est dès lors possible que la régulation au niveau traductionnel ait été sélectionnée afin d'éviter l'accumulation de protéines bloquées dans la membrane.[7,9,10,15-17]

La pénicilline G acylase d'*E. coli* peut être considérée comme étant une enzyme « modulaire », constituée de plusieurs modules indépendants, dont les activités de catalyse et de spécificité de substrat sont séparables.[18] La structure de la pénicilline acylase est en forme d'entonnoir au fond duquel se trouvent le site actif et la poche de spécificité pour la fonction phénylacétique, laissant le reste du substrat dirigé vers l'extérieur de la poche de spécificité (Figures 1.3 et 1.4).

Ces enzymes, vraisemblablement impliquées dans la mise à disposition de nutriments, ont sans doute évolué sous une pression de sélection demandant de tolérer une grande variété de substrats. C'est peut-être pour cette raison que ce type d'enzymes modulaires représente une source attractive de catalyseurs industriels, et que, grâce à leurs structures modulaires, elles présentent un intérêt en ingénierie des protéines.[18]

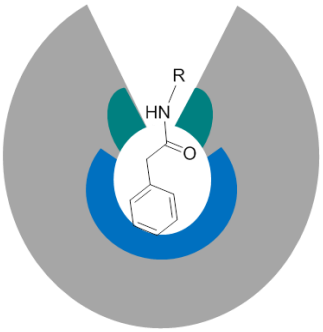


Figure 1.3 : Représentation schématique de la pénicilline acylase en présence de son substrat. L'enzyme présente une forme d'entonnoir au fond duquel se trouve une cavité hydrophobe profonde et spécifique pour la fonction phénylacétique (en bleu). Le reste du substrat pointe vers l'extérieur de la cavité, dans l'entonnoir accessible au solvant. Les résidus impliqués dans la catalyse se trouvent à l'entrée de la poche de spécificité (en vert).

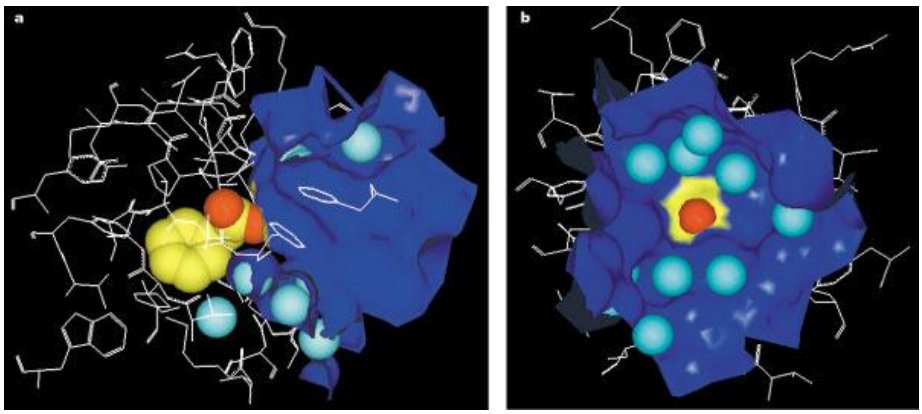


Figure 1.4 : Reconnaissance de la fonction phénylacétyle par la pénicilline G acylase. Les résidus se trouvant à une distance maximale de 10 Å du site actif sont repris en blanc. Les atomes de carbone et d'oxygène de l'acide phénylacétique sont représentés en jaune et orange respectivement. Une partie de la surface accessible au solvant est en bleu foncé. Les boules bleues claires représentent des molécules d'eau. **a**. La fonction phénylacétyle se trouve dans une poche hydrophobe, et un des atomes d'oxygène de la fonction acide pointe vers le solvant. **b**. Vue du même site actif, après une rotation de 90°. (Image provenant de Kholsa et al.,[18] faite à partir de la structure 1PNL de la PDB.)

Structure

Le site de fixation du substrat se trouve dans un creux formé par une structure α-β-β-α composée d'hélices α et de feuillets β provenant des chaines α et β de la protéine. Cette structure est caractéristique des enzymes de la superfamille des hydrolases N-terminale nucléophiles (Ntn-hydrolases),[19,20] décrite ci-après.

Famille Ntn-hydrolases

Les enzymes de cette superfamille sont toutes activées de manière autocatalytique. Elles sont produites sous forme de précurseur, qui donnera l'enzyme active après maturation et clivage d'un propetide interne. Elles ne possèdent pas la triade catalytique classiquement retrouvée dans beaucoup d'hydrolases mais disposent d'un résidu essentiel qui peut être une serine, une cystéine ou une thréonine (voir mécanisme plus loin). Malgré le fait qu'il n'y ait aucune similarité de séquence entre les différentes Ntn-hydrolases, elles possèdent toutes une structure tertiaire similaire au niveau du site catalytique : deux couches d'hélices α antiparallèles et, entre les deux, deux feuillets β antiparallèles de quatre ou cinq branches ; structure appelée noyau α-β-β-α (voir Figure 1.5).[21,22] Toutes les protéines de cette superfamille semblent catalyser l'hydrolyse de liens amides, tout en possédant une spécificité de substrat différente. Oinonen et al. ont comparé les séquences et structures tridimensionnelles de quatre Ntn-hydrolases, dont la pénicilline G acylase de *E. coli*, et ont mis en évidence que de la plupart des résidus conservés et importants pour l'activité se trouvent dans le noyau α-β-β-α. Ces résidus sont repris en quatre catégories : le nucléophile d'une part (1), et les résidus impliqués dans la stabilisation du nucléophile (2), dans la liaison du substrat dans la poche de spécificité (3) et dans la stabilisation de l'oxianion (4).[22]

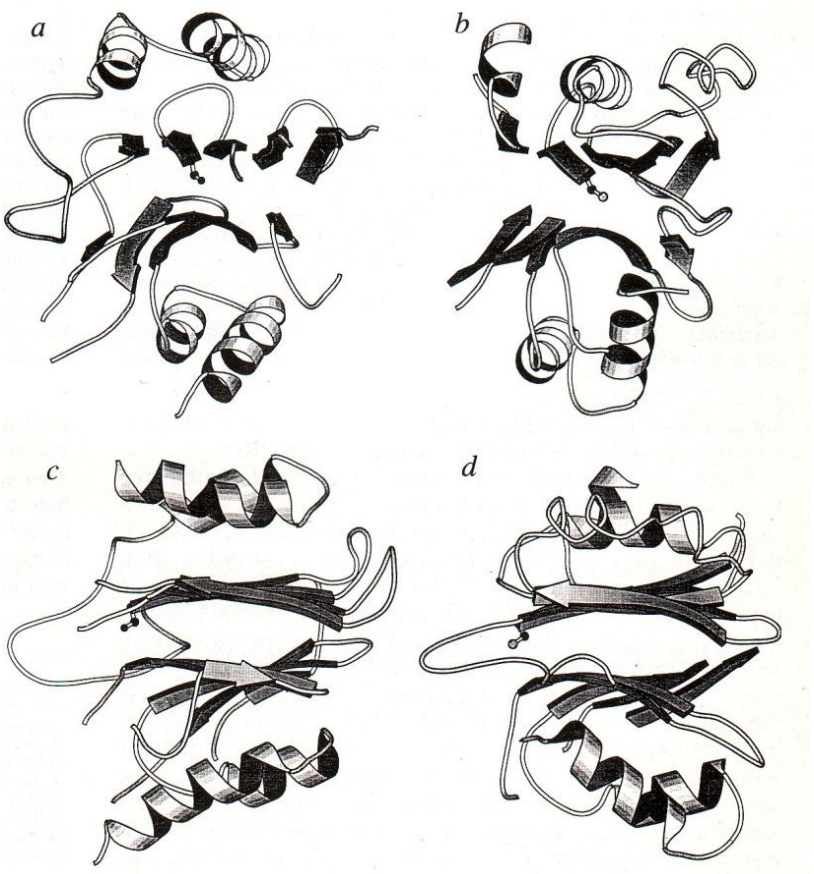


Figure 1.5 : structure commune retrouvée chez les Ntn-hydrolases. Vue de face (a, b) et de côté (c, d) de l'assemblement des structures secondaires du noyau α-β-β-α de la pénicilline G acylase (a, c) et de la glutamine PRPP amidotransférase (b, d). Figure provenant de Brannigan et al.[19]

Structure tridimensionnelle

La structure de la pénicilline G acylase d'*E. coli* a été résolue par Duggleby et al. en 1995 en présence d'acide phényle acétique (code PDB : 1pnl) et d'acide phényle méthane sulfonique (code PDB : 1pnm) qui révèlent la présence d'un site spécifique de liaison pour la chaine phénylacétyle du substrat. La protéine possède une forme tridimensionnelle qui rappelle la forme d'un rein, avec une dépression au centre où se trouve la poche de spécificité (Figures 1.6 et 1.7). De plus, un ion calcium se retrouve à l'interface entre les deux sous-unités,

coordonné par les chaines latérales des résidus AspB73, AspB76 et AspB252, ainsi que par les fonctions carbonyles des résidus ValB75, ProB205 et GluA152.[6]

L'enzyme possède une poche hydrophobe de très grande spécificité pour la fonction phénylacétyle qui se trouve derrière le site actif et est composée de résidus se trouvant sur les chaines α et β de l'enzyme.[8,22-25] La Figure 1.6 montre les résidus impliqués dans la spécificité et la catalyse dans la structure tridimensionnelle de l'enzyme.

Des études structurales approfondies de la pénicilline acylase ont montré la présence d'une conformation fermée lorsqu'aucune molécule de substrat ou d'inhibiteur n'est présente dans le site actif.[27] La poche de spécificité pour le groupement phénylacétique est alors protégée du solvant par la PheA146, et l'ArgA145 pointe également vers le site actif en faisant un pont hydrogène avec l'oxygène du carbonyl de la PheB24. Lorsque le substrat se lie à la poche de spécificité, un changement conformationnel a lieu et la poche de spécificité passe d'une conformation fermée à une conformation ouverte. Les résidus PheA146 et ArgA145 bougent pour permettre une meilleure accommodation du substrat. La PheA146 avec la GlnB23 créent alors un second site de liaison pour le noyau β-lactame. Des expériences de mutagénèse de la PheA146 ont montré qu'un cycle aromatique sur ce résidu était nécessaire pour l'interaction avec la fonction β-lactame. L'existence de cette seconde poche de spécificité a pu aussi être mise en évidence par l'inhibition de l'enzyme par le 6-APA et le 7-ADCA.[23,24,28]

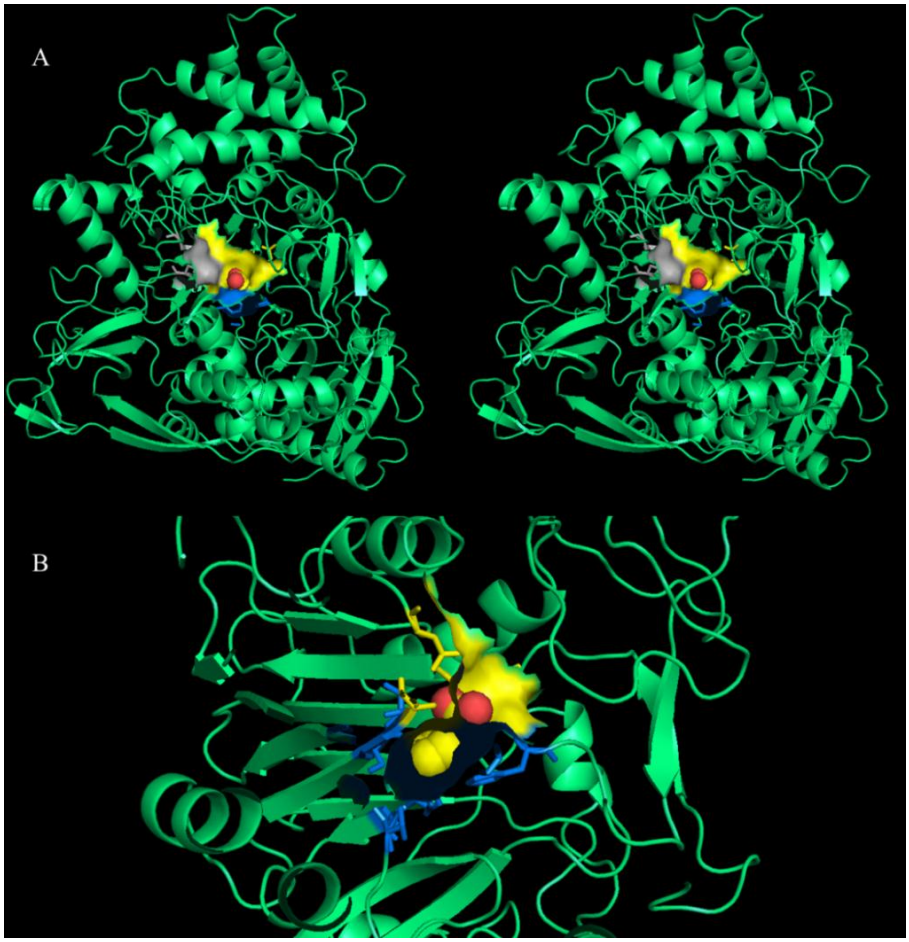


Figure 1.6 : Vue d'ensemble (A) et coupe de profil (B) de la structure de la pénicilline acylase complexée à l'acide phényl acétique. Les résidus de la poche de spécificité pour le ligand sont représentés en bâtonnets, la serine catalytique est en bâtonnets jaunes et la surface de la poche est représentée en bleu. Le ligand est en space filling jaune et rouge. (Code PDB : 1pnl.[6])

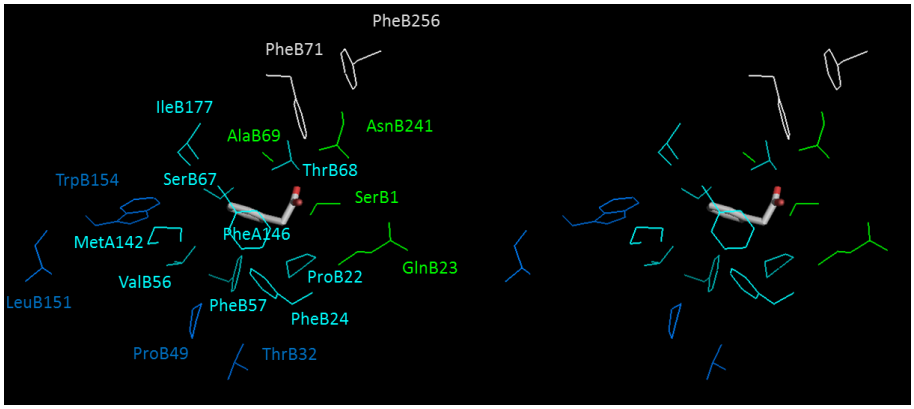


Figure 1.7 : Structure du site actif de la pénicilline acylase en présence d'acide phényle acétique (en gris et rouge), vue stéréo. Les chaînes latérales des résidus impliqués dans la catalyse sont en vert (SerB1, GlnB23, AlaB69 et AsnB241), les chaînes latérales des résidus de la poche de spécificité sont en bleu (bleu clair : résidus se trouvant à moins de ou à 5 Å de l'acide phényle acétique, bleu foncé : se trouvant à plus de 5 Å, mais étant répertoriés comme impliqués dans la spécificité pour le PAA).[6,8,22,24-26] (Code PDB : 1pnl[6])

Mécanisme

La résolution de la structure de la pénicilline acylase a permis à Duggleby et al. de mieux comprendre le rôle du résidu catalytique unique situé en position N-terminale de la chaine β.[6] Un mécanisme catalytique a été proposé puis complété par la suite (voir Figure 1.8).[29-31] La conformation du résidu SerB1 montre que la fonction hydroxyle est activée par son propre groupement α-amino. Ce dernier sert de base catalytique, de manière analogue à la diade His-Asp dans la triade catalytique des protéases à sérine classiques.[32,33] La sérine N-terminale joue donc le double rôle de nucléophile et de base catalytique. Les pKas des deux fonctions alcool (>14) et amine (~8) semblent fort éloignés pour permettre la déprotonation de la sérine par l'amine. Des études structurales suggèrent que des interactions de ces fonctions avec des groupements du substrat et des chaines latérales de résidus du site actif permettent de faciliter l'activation du nucléophile. McVey et al. ont démontré par étude cristallographique que AsnB241, en plus de son rôle dans la formation du trou de l'oxianion, aide au bon positionnement du substrat et de

sa fonction amide et interagit également avec la sérine catalytique. La GlnB23 a aussi été rapportée comme interagissant avec la serine catalytique afin sans doute d'en assurer le bon positionnement (mis en évidence à la Figure 1.8).[29,32,34]

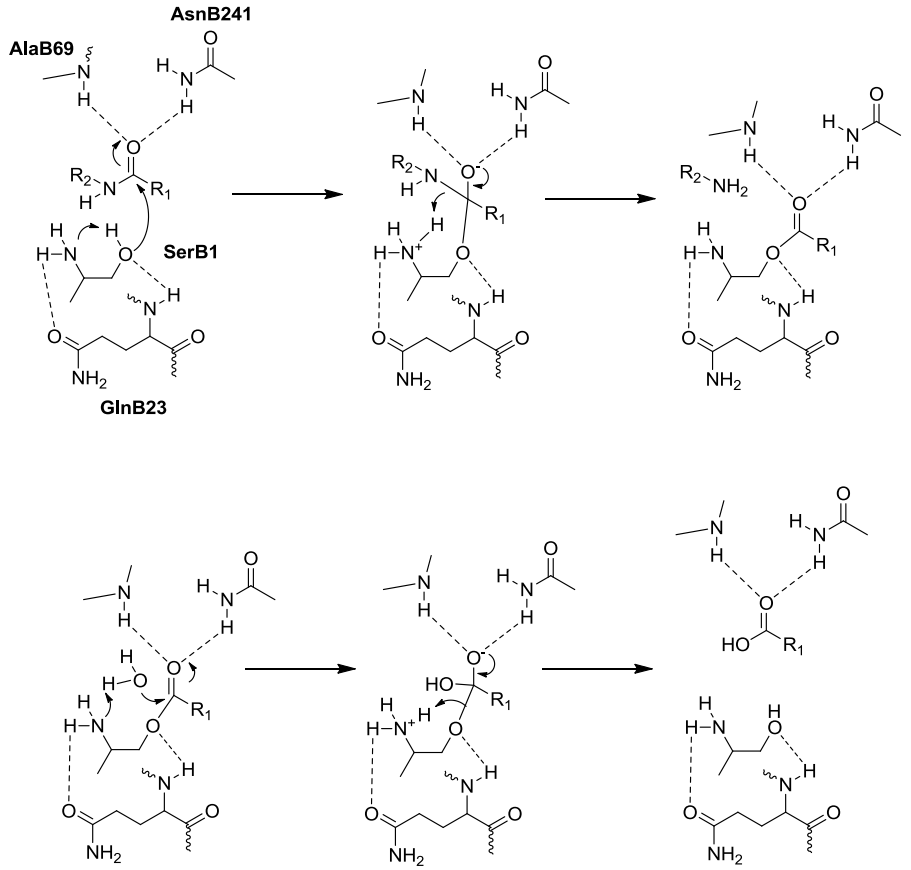


Figure 1.8 : Mécanisme catalytique de la pénicilline G acylase d'*E. coli*. Modèle proposé par Oinonen et al.[22] et Zhiryakova et al.[31] dans lequel la SerB1 est activée par sa propre chaine latérale. Un modèle plus ancien avait été proposé par Duggleby et al.[6] dans lequel l'activation de la serine catalytique se fait via une molécule d'eau. La littérature récente opte pour le mécanisme ci-dessus.[30]

Après l'attaque nucléophile de la sérine, l'intermédiaire tétraédrique chargé négativement est stabilisé dans le trou de l'oxyanion : les résidus AlaB69 et AsnB241 forment des ponts hydrogène avec l'oxygène du carbonyle via leurs azotes situées respectivement sur leur chaine principale et latérale. L'étape d'acylation se termine lorsque l'ammonium du résidu SerB1 obtenu après la catalyse basique cède un proton au groupe partant, donnant un intermédiaire covalent acyle-enzyme. Cet intermédiaire est ensuite désacylé par attaque nucléophile d'une molécule d'eau ou d'une amine (hydrolyse ou aminolyse) qui sera vraisemblablement également activée par l'amine de SerB1. Un second intermédiaire tétraédrique similaire au premier sera formé, menant au second produit et au recyclage de l'enzyme.

Maturation

La mutation Thr263Gly, dernier résidu du propeptide, entraîne le ralentissement de la maturation de l'enzyme, ce qui a permis de cristalliser et d'étudier l'enzyme sous sa forme non maturée. La structure de la proenzyme non maturée montre clairement que le propeptide se situe à l'entrée du site catalytique et en bloque l'accès (Figures 1.9 et 1.10).[21] Ignatova et al. ont démontré que le propeptide sert de chaperon lors du repliement de la pénicilline G acylase.[5] Les différences structurales entre la forme épissée et non épissée sont minimes. Dans le précurseur, Pro261 (un résidu du propeptide) se trouve dans la poche hydrophobe de haute spécificité pour le substrat de l'enzyme maturée. Pro261 induit des changements dans la poche comme par exemple le déplacement des résidus PheA146 et ArgA145, qui sont également impliqués dans l'ouverture et la fermeture du site actif de l'enzyme mature.[21]

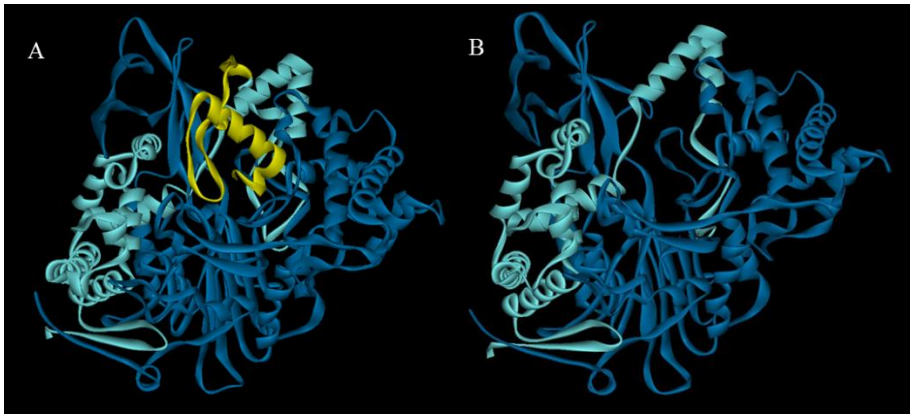


Figure 1.9 : A. Structure tridimentionnelle de la pénicilline acylase non maturée avec le propeptide (jaune), B. Structure de l'enzyme après maturation. La chaine α est représentée en bleu clair et la chaine β en bleu foncé. (A. code PDB : 1e3a,[21] B. code PBD : 1pnk[6])

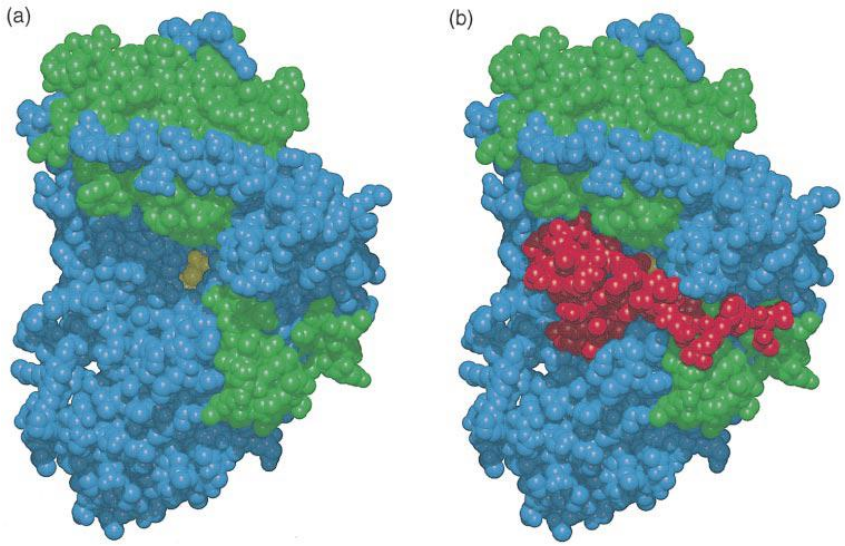


Figure 1.10 : Modèles en « space filling » de la pénicilline acylase mature (a) et du précurseur (b). Le chaine α est en vert, la chaine β en bleu et le propeptide connecteur en rouge. La sérine catalytique en jaune est visible à la base de la faille du site actif. Cette représentation montre clairement que le propeptide masque l'entrée du site actif. (Image provenant de[21])

Des molécules d'eau se trouvent à l'interface entre le propeptide et l'enzyme et forment un « canal de solvant » jusqu'à la sérine catalytique (Ser264). Dans le mutant Thr263Gly, le premier clivage qui apparait est celui de la liaison Tyr260-Pro261, probablement par attaque nucléophile du O-gamma de la Thr262. Dans la structure non épissée, la Ser264 (ou SerB1) semble avoir une orientation différente que dans l'enzyme mature : le O-gamma de la Ser264 pointe vers le carbone du carbonyl de la liaison peptidique 263-264, arrangement approprié pour un clivage autocatalytique. Après clivage de ce lien, la Ser264 devient le premier résidu de la sous-unité β, SerB1 (voir Figure 1.2).[6,21] Le premier clivage Thr260-Pro261 est peut-être également présent dans l'enzyme sauvage ce qui expliquerait l'importance de ces résidus pour la maturation de l'enzyme. La mutagénèse dirigée des résidus Thr260 et Pro261 affecte également la maturation de l'enzyme, ce qui suggère que le premier clivage entre ces résidus a également lieu dans l'enzyme sauvage.[35]

Le fait que les chaines α et β aient des structures similaires que l'enzyme soit maturée ou non et que le site actif soit protégé par le propeptide soutient l'hypothèse d'un épissage autocatalytique intramoléculaire. Cette hypothèse fut ensuite démontrée grâce à l'étude de la maturation du mutant Thr263Gly immobilisé sur une résine, où une maturation intermoléculaire était impossible.[5,21,36]

En fin de maturation, le reste du propeptide subi un clivage entre les résidus Ala209 et Ala210 qui pourrait être un clivage intermoléculaire,[36-38] cependant, le clivage propeptide-chaine α ne semble pas avoir été étudié en détail, et il n'y a pas d'informations claires dans la littérature.

Applications industrielles

L'utilisation de biocatalyseurs en industrie chimique présente de nombreux avantages grâce à leur productivité élevée ou à la très faible quantité de produits secondaires générés. Il existe un panel d'enzymes qui acceptent une grande variété de molécules, et qui peuvent être énantio- ou régio-sélectives.[39] La demande de nouveaux biocatalyseurs est en constante augmentation, principalement dans les domaines agro-alimentaire et pharmaceutique pour la synthèse de molécules complexes à haute valeur ajoutée.[40]

Les pénicillines acylases sont principalement utilisées dans la production de précurseurs des pénicillines et ampicillines semi-synthétiques, classes d'antibiotiques qui restent à ce jour les plus utilisées dans le monde. Ces précurseurs sont le 6-APA (acide 6-aminopénicilanique), le 7-ACA (acide 7-aminocéphalosporanique) et le 7-ADCA (acide 7-aminodeacetoxy-céphalosporanique). L'hydrolyse par catalyse enzymatique de la pénicilline G par les pénicillines G et V acylases permet d'isoler le 6-APA. De manière similaire, le 7-ACA peut être obtenu à partir de la céphalosporine C.[3] Celle-ci peut soit être hydrolysée par un mutant de la céphalosporine acylase (N266M),[41-44] soit être modifiée de manière à convertir sa chaine latérale α-aminoadipyl en glutaryl par une acide D-amino oxydase, qui sera ensuite hydrolysée par une céphalosporine acylase sauvage (Figure 1.11).[45] L'acide 7-aminodeacetoxy-céphalosporanique (7-ADCA) peut être obtenu soit par extension chimique ou enzymatique (par une expandase) du noyau du 6-APA, ou encore par l'hydrolyse enzymatique de la céphalosporine G par la PAC.[3,45,46] Ces trois composés serviront de point de départ pour la synthèse d'antibiotiques de type β-lactames (Figure 1.11).[47] Dans les années 1970, le 6-APA était produit par voie chimique. Cette voie de production est restée la principale pendant 15 à 20 ans avant d'être largement remplacée par la catalyse enzymatique. Le 6-APA produit via hydrolyse par la pénicilline G acylase et la

pénicilline V acylase représente actuellement une production mondiale de 9000 tonnes par an.

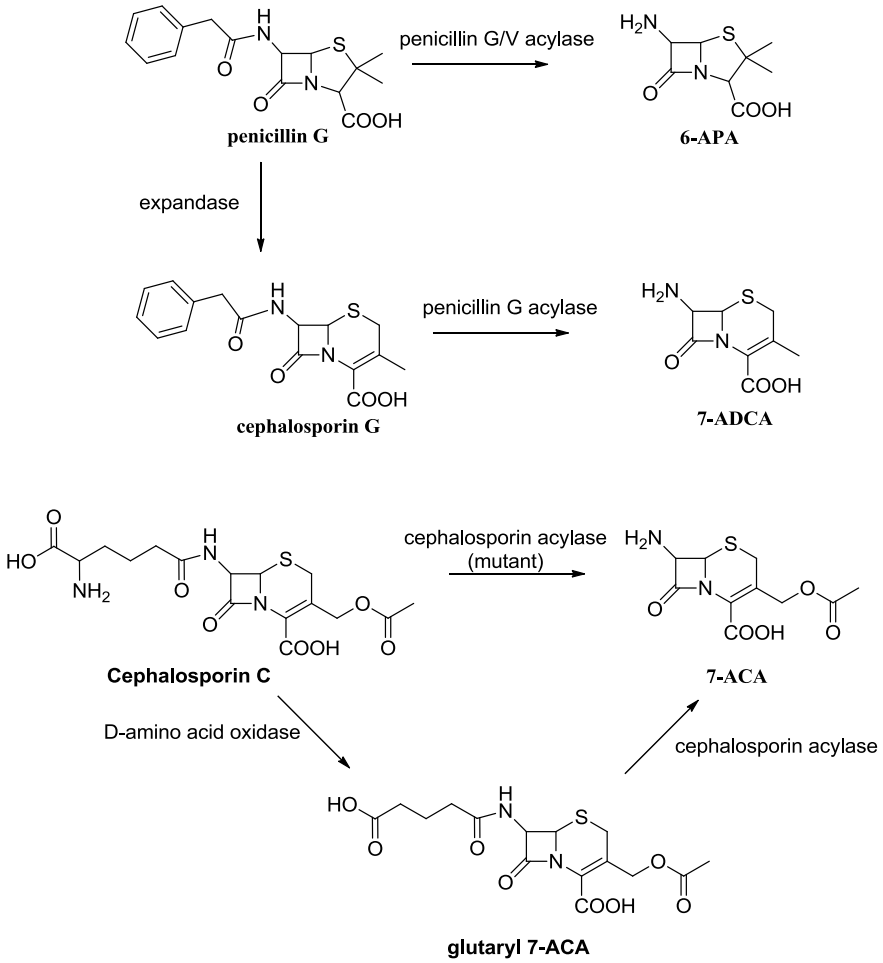


Figure 1.11 : production de noyaux β-lactames pour la production d'antibiotiques semi-synthétiques.

La pénicilline acylase peut également catalyser des réactions d'aminolyse, et donc être utilisée pour la synthèse directe de nouveaux antibiotiques semi-synthétiques en catalysant la condensation entre le 6-APA ou 7-ACA avec un acide activé ou non. La Figure 1.12 montre quelques procédés d'aminolyse utilisés industriellement par la société DSM pour la production de pénicillines et céphalosporines semi-synthétiques par des pénicillines acylases.[40]

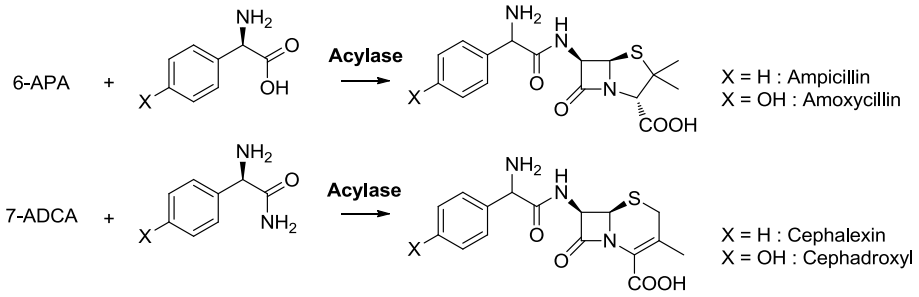


Figure 1.12 : Processus biocatalytiques utilisés par DSM pour la production de pénicillines et céphalosporines semi-synthétiques utilisant des acylases. La condensation du 6-APA ou 7-ADCA avec un dérivé acide ou amide est catalysée par une acylase afin de synthétiser les antibiotiques semi-synthétiques montrés ci-dessus. (Figure adaptée de Henke, E. et al.[48])

La synthèse de pénicillines et céphalosporines semi synthétiques réalisée par catalyse enzymatique peut se faire par contrôle thermodynamique ou cinétique.

Le contrôle thermodynamique est simple à mettre en œuvre. Dans ce cas, l'enzyme catalyse la condensation de l'acide libre avec le noyau β-lactame, nucléophile.[49,50] Le rendement maximal correspond à la concentration de produit à l'équilibre. Cette option possède certains inconvénients. L'équilibre réactionnel doit favoriser l'apparition du produit (voir Figure 1.13), or dans la gamme de pH à laquelle fonctionne l'enzyme (pH 6-8), la concentration de produit à l'équilibre est faible car l'acide et l'amine sont respectivement déprotoné et protonée.[51] L'équilibre peut être déplacé en faveur de l'apparition du produit en enlevant le produit au fur et à mesure de la réaction par cristallisation ou complexation.[52] Il est également possible de jouer sur l'équilibre thermodynamique de la réaction en diminuant l'activité de l'eau par addition de cosolvants ou en utilisant l'enzyme directement en solvant organique. L'enzyme doit alors être immobilisée afin qu'elle soit plus robuste envers les solvants organiques : par immobilisation sur un support,[53] en créant un microenvironnement hydrophile autour de l'enzyme, ou encore en

modifiant certains acides aminés hydrophiles de la surface en résidus hydrophobes.[45]

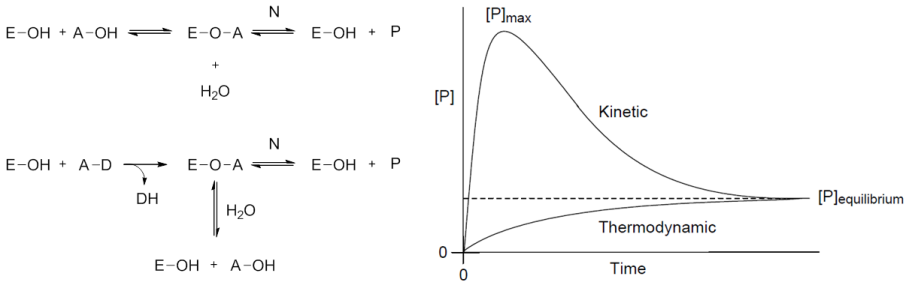


Figure 1.13 : Schéma simplifié de la réaction de synthèse d'antibiotiques catalysée par la pénicilline acylase en contrôle thermodynamique et cinétique, et graphique correspondant, représentant la concentration en produit P (antibiotique) au court du temps. E-OH : enzyme libre, A-OH : acide, E-O-A : acyle enzyme, N : nucléophile β-lactame, P : produit, A-D : acyle activé. (Figure provenant de la thèse de Michiel Janssen, 2006, adaptée)

Cependant, le contrôle cinétique de la synthèse reste la voie la plus classique, car des rendements très élevés peuvent être obtenus. Dans ce cas, une acylation de l'enzyme a lieu par un précurseur activé de la fonction acyle, suivie d'une aminolyse de l'acyle enzyme par le noyau β-lactame (voir Figure 1.13). Le produit s'accumule et atteint généralement des concentrations bien plus élevées que la concentration attendue à l'équilibre (voir Figure 1.13). Mais des réactions secondaires d'hydrolyse peuvent également avoir lieu car la pénicilline G acylase agit comme transférase et hydrolase : une compétition a lieu entre l'hydrolyse et l'aminolyse de l'acyle enzyme. L'optimisation du rendement dans le cas d'un contrôle thermodynamique ne peut se faire qu'en influençant l'équilibre de la réaction en modifiant la température, le pH ou en diminuant la concentration du produit au fur et à mesure, alors que les paramètres de l'enzyme n'influenceront pas la position de l'équilibre, mais juste le temps que prendra la réaction pour l'atteindre. Par contre, ils peuvent modifier le rendement dans le cas d'un contrôle cinétique. Son augmentation sera favorisée par l'augmentation de la vitesse d'aminolyse par rapport à celle d'hydrolyse (ratio $V_{synthèse}/V_{hydrolyse}$

ou V_s/V_h). L'augmentation de taux de conversion de l'acyle donneur par rapport au taux de conversion du produit augmentera également la concentration maximale de produit. L'évolution dirigée se présente alors comme un outil de choix pour l'optimisation du contrôle cinétique.[46]

Les pénicillines acylases sont également utilisées comme outil de déprotection sélective de groupements hydroxyl et amino, pouvant être utilisés dans des conditions douces et donc applicables aux molécules les plus fragiles,[54] ainsi que pour la résolution par hydrolyse sélective d'un énantiomère de certains mélanges racémiques, comme par exemple des mélanges d'acides aminés $\beta^{2,3}$ et β^3 avec une efficacité très élevée.[39] Immobilisée sur une résine, la Pénicilline G Acylase peut aussi être utilisée comme phase stationnaire chirale pour la séparation en HPLC d'énantiomères acidiques de structures similaires aux acides 2-aryloxy-2-arylacétiques.[55,56]

Ingénierie génétique de la pénicilline acylase

L'ingénierie de la pénicilline acylase sauvage peut améliorer ses propriétés catalytiques ou physico-chimiques et permettre une utilisation industrielle plus large de cette enzyme. Beaucoup de recherches ont été réalisées afin d'améliorer la stabilité de l'enzyme face à différentes températures[57] ou différents pHs, à la présence de solvants organiques, ou encore pour améliorer son immobilisation sur des résines.[8,45]

Les rendements de synthèse d'antibiotiques β-lactames catalysée par la pénicilline acylase en contrôle cinétique sont influencés par les propriétés catalytiques de l'enzyme. Pour l'enzyme sauvage, ces rendements sont modestes et nécessitent d'être améliorés pour un usage industriel compétitif. Plusieurs objectifs peuvent être visés : augmenter l'affinité de l'enzyme pour l'acyle activé, diminuer l'affinité pour l'acide correspondant et le produit, afin d'en diminuer l'inhibition, diminuer l'activité d'hydrolyse par rapport à l'activité d'aminolyse.

Alkema et al. ont étudié par mutagénèse dirigée l'influence de plusieurs résidus sur le ratio $V_{synthèse}/V_{hydrolyse}$ (V_s/V_h).[23,24,28,58,59]

La mutation de PheA146, impliquée dans la liaison du noyau β-lactame et du phényle acétyle, en tyrosine permet de garder la même affinité pour l'acide phényle acétique (PAA), mais la synthèse de la pénicilline G est très faible. Par contre les mutations PheA146Leu ou Ala augmentent le ratio V_s/V_h de 6 fois et diminue l'affinité pour l'acide phényle acétique, et donc son inhibition.[23,28]

Les mutants ArgA145Leu, Cys ou Lys montrent une plus faible spécificité (k_{cat}/K_M) que l'enzyme sauvage sur différents substrats, mais présentent un ratio V_s/V_h plus élevé pour la synthèse de l'ampicilline et de la pénicilline G. Dans ce cas, la plus faible liaison du noyau β-lactame améliore la réactivité d'aminolyse envers l'acyle enzyme.[59]

La mutation PheB24Ala confère à l'enzyme un ratio V_s/V_h 3 fois plus grand que pour le sauvage avec le 6-APA et le 7-ADCA, tout en gardant une affinité semblable au sauvage pour les esters activés. Son inhibition par l'acide phényle acétique est plus faible. Ce mutant est donc intéressant pour la synthèse d'ampicilline, de cefalexine, d'amoxicilline ou du cefadroxil.[8,23-25,45]

Morillas et al. ont mis en évidence que les résidus PheB71 et PheB256 font des interactions de type pi-stacking. La mutation de la PheB71 en cystéine ou leucine entraine une rotation de la PheB256, et permet alors une meilleure accommodation de la gluraryl-Leucine dans la poche de spécificité en l'enzyme, et son hydrolyse.[60] Ces variants ont été sélectionnés par sélections auxotrophes de souches d'*E. coli* en présence de gluraryl-L-Leucine comme seule source de leucine.[61] Malheureusement, ces mutants ne sont pas actifs envers le gluratyl-7-ACA. Ils sont cependant plus résistants à la dénaturation par l'urée et l'inactivation par la chaleur que l'enzyme sauvage. La substitution du résidu équivalent à PheB71 en valine est également présente dans un mutant de la pénicilline G acylase de *Kluyvera citrophila* sélectionné pour sa capacité à

hydrolyser de l'adipyl-L-Leu. Ce mutant est aussi capable d'hydrolyser le gluratyl-, valéryl-, caproyl-, heptanoyl- et phénoxyacétyl-L-Leu. Par contre il n'hydrolyse pas le gluraryl-7ACA et a perdu 90 % de son activité envers la pénicilline G.[15]

Un variant de la pénicilline acylase de *P. rettgeti* a été trouvé dans la souche Bro1 et possède la particularité d'hydrolyser le 6-bromohexanamide. Cette enzyme présente 64 % d'identité de séquence avec la pénicilline acylase d'*E. coli*. L'alignement de la séquence met en évidence la présence d'un résidu de la poche de spécificité qui n'est pas conservé : MetA142 d'*E. coli* est remplacée par une leucine dans Bro1.[32]

Oh et al. ont voulu modifier la spécificité de substrat de la pénicilline acylase (PAC) vers celle de la céphalosporine acylase (CA).[26] La comparaison des structures tridimentionnelles des deux enzymes a permis de repérer sept résidus de la poche de spécificité de PAC à muter en résidus correspondants de la CA. Le mutant obtenu possède une activité accrue sur le glutaryl-7-ACA, et son activité envers la pénicilline G est drastiquement diminuée. Ils ont choisi d'exprimer les deux chaînes de l'enzyme séparément afin d'éviter de passer par le proenzyme. Concernant la CA, il a en effet été remarqué que la maturation de l'enzyme pouvait être fortement ralentie par la présence de mutations dans le site actif ou dans la poche de spécificité.[62,63] Dans le cas de l'ingénierie de l'enzyme, il parait judicieux d'éviter les problèmes liés à la maturation de l'enzyme en expriment les deux sous-unités séparément.

L'ingénierie génétique de la pénicilline acylase est clairement intéressante tant d'un point de vue fondamental pour la compréhension du mécanisme catalytique et de la spécificité de l'enzyme que du point de vue appliqué pour améliorer les performances catalytiques de l'enzyme au niveau industriel.

Références

1. Arroyo M, de la Mata I, Acebal C, Castillon MP; *Biotechnological applications of penicillin acylases: state-of-the-art*. Appl Microbiol Biotechnol **2003**; 60: 507-14

2. Kim Y, Yoon K, Khang Y, Turley S, Hol WG; *The 2.0 A crystal structure of cephalosporin acylase*. Structure **2000**; 8: 1059-68

3. Bruggink A, Roos E, deVroom E; *Penicillin acylase in the industrial production of beta-lactam antibiotics*. Organic Process Research and Development **1998**; 2: 128-133

4. Polderman-Tijmes J; *Biochemical characterization of a-amino acid ester hydrolases*. PhD thesis **2004**

5. Ignatova Z, Hornle C, Nurk A, Kasche V; *Unusual signal peptide directs penicillin amidase from Escherichia coli to the Tat translocation machinery*. Biochem Biophys Res Commun **2002**; 291: 146-9

6. Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; *Penicillin acylase has a single-amino-acid catalytic centre*. Nature **1995**; 373: 264-8

7. Galan B, Garcia JL, Prieto MA; *The PaaX repressor, a link between penicillin G acylase and the phenylacetyl-coenzyme A catabolon of Escherichia coli W*. J Bacteriol **2004**; 186: 2215-20

8. Rajendhran J, Gunasekaran P; *Recent biotechnological interventions for developing improved penicillin G acylases*. J Biosci Bioeng **2004**; 97: 1-13

9. Prieto MA, Galan B, Torres B, Ferrandez A, Fernandez C, Minambres B, Garcia JL, Diaz E; *Aromatic metabolism versus carbon availability: the regulatory network that controls catabolism of less-preferred carbon sources in Escherichia coli*. FEMS Microbiol Rev **2004**; 28: 503-18

10. Merino E, Balbas P, Recillas F, Becerril B, Valle F, Bolivar F; *Carbon regulation and the role in nature of the Escherichia coli penicillin acylase (pac) gene*. Mol Microbiol **1992**; 6: 2175-82

11. Deutscher J; *The mechanisms of carbon catabolite repression in bacteria*. Curr Opin Microbiol **2008**; 11: 87-93

12. Gang DM, Shaikh K; *Regulation of penicillin acylase in Escherichia coli by cyclic AMP*. Biochim Biophys Acta **1976**; 425: 110-4

13. Savidge TA, Cole M; *Penicillin acylase (bacterial)*. Methods Enzymol **1975**; 43: 705-21

14. Keilmann C, Wanner G, Bock A; *Molecular basis of the exclusive low-temperature synthesis of an enzyme in E. coli: penicillin acylase*. Biol Chem Hoppe Seyler **1993**; 374: 983-92

15. Roa A, Garcia JL; *New insights into the regulation of the pac gene from Escherichia coli W ATCC 11105*. FEMS Microbiol Lett **1999**; 177: 7-14

16. Valle F, Gosset G, Tenorio B, Oliver G, Bolivar F; *Characterization of the regulatory region of the Escherichia coli penicilin acylase structural gene*. Gene **1986**; 50: 119-22

17. Dai M, Zhu Y, Yang Y, Wang E, Xie Y, Zhao G, Jiang W; *Expression of penicillin G acylase from the cloned pac gene of Escherichia coli ATCC11105. Effects of pac*R *and temperature*. Eur J Biochem **2001**; 268: 1298-303

18. Khosla C, Harbury PB; *Modular enzymes*. Nature **2001**; 409: 247-52

19. Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG; *A protein catalytic framework with an N-terminal nucleophile is capable of self-activation*. Nature **1995**; 378: 416-9

20. Schumacher G, Sizmann D, Haug H, Buckel P, Bock A; *Penicillin acylase from E. coli: unique gene-protein relation*. Nucleic Acids Res **1986**; 14: 5713-27

21. Hewitt L, Kasche V, Lummer K, Lewis RJ, Murshudov GN, Verma CS, Dodson GG, Wilson KS; *Structure of a slow processing precursor penicillin acylase from Escherichia coli reveals the linker peptide blocking the active-site cleft*. J Mol Biol **2000**; 302: 887-98

22. Oinonen C, Rouvinen J; *Structural comparison of Ntn-hydrolases*. Protein Sci **2000**; 9: 2329-37

23. Alkema WB, Dijkhuis AJ, De Vries E, Janssen DB; *The role of hydrophobic active-site residues in substrate specificity and acyl transfer activity of penicillin acylase*. Eur J Biochem **2002**; 269: 2093-100

24. Alkema WB, Hensgens CM, Snijder HJ, Keizer E, Dijkstra BW, Janssen DB; *Structural and kinetic studies on ligand binding in wild-type and active-site mutants of penicillin acylase*. Protein Eng Des Sel **2004**; 17: 473-80

25. Niersbach H, Kuhne A, Tischer W, Weber M, Wedekind F, Plapp R; *Improvement of the catalytic properties of penicillin G acylase from Escherichia coli ATCC*

11105 by selection of a new substrate specificity. Appl Microbiol Biotechnol **1995**; 43: 679-84

26. Oh B, Kim K, Park J, Yoon J, Han D, Kim Y; *Modifying the substrate specificity of penicillin G acylase to cephalosporin acylase by mutating active-site residues*. Biochem Biophys Res Commun **2004**; 319: 486-92

27. Done SH, Brannigan JA, Moody PC, Hubbard RE; *Ligand-induced conformational change in penicillin acylase*. J Mol Biol **1998**; 284: 463-75

28. Alkema WB, Hensgens CM, Kroezinga EH, de Vries E, Floris R, van der Laan JM, Dijkstra BW, Janssen DB; *Characterization of the beta-lactam binding site of penicillin acylase of Escherichia coli by structural and site-directed mutagenesis studies*. Protein Eng **2000**; 13: 857-63

29. McVey CE, Tolley SP, Done SH, Brannigan JA, Wilson KS; *A new crystal form of penicillin acylase from Escherichia coli*. Acta Crystallogr D Biol Crystallogr **1997**; 53: 777-9

30. Chilov GG, Sidorova AV, Svedas VK; *Quantum chemical studies of the catalytic mechanism of N-terminal nucleophile hydrolase*. Biochemistry (Mosc) **2007**; 72: 495-500

31. Zhiryakova D, Ivanov I, Ilieva S, Guncheva M, Galunsky B, Stambolieva N; *Do N-terminal nucleophile hydrolases indeed have a single amino acid catalytic center?* FEBS J **2009**; 276: 2589-98

32. McDonough MA, Klei HE, Kelly JA; *Crystal structure of penicillin G acylase from the Bro1 mutant strain of Providencia rettgeri*. Protein Sci **1999**; 8: 1971-81

33. Dodson G, Wlodawer A; *Catalytic triads and their relatives*. Trends Biochem Sci **1998**; 23: 347-52

34. kim MG, Lee SB; *penicillin acylase-catalyzed synthesis of beta-lactam antibiotics in water-ethanol mixture: effect of cosolvent content and chemical nature of the substrate on reaction rates and yields*. J. Mol. Cat. B: Enzymatic **1996**; 1: 201-211

35. Choi KS, Kim JA, Kang HS; *Effects of site-directed mutations on processing and activities of penicillin G acylase from Escherichia coli ATCC 11105*. J Bacteriol **1992**; 174: 6270-6

36. Kasche V, Lummer K, Nurk A, Piotraschke E, Rieks A, Stoeva S, Voelter W; *Intramolecular autoproteolysis initiates the maturation of penicillin amidase from Escherichia coli*. Biochim Biophys Acta **1999**; 1433: 76-86

37. Ignatova Z, Wischnewski F, Notbohm H, Kasche V; *Pro-sequence and Ca2+-binding: implications for folding and maturation of Ntn-hydrolase penicillin amidase from E. coli*. J Mol Biol **2005**; 348: 999-1014

38. Kasche V, Galunsky B, Ignatova Z; *Fragments of pro-peptide activate mature penicillin amidase of Alcaligenes faecalis*. Eur J Biochem **2003**; 270: 4721-8

39. Liljeblad A, Kanerva LT; *biocatalysis as a profound tool in the preparation of higly enantiopure beta-amino acids*. Tetrahedron **2006**; 62: 5831-5854

40. Schmid A, Dordick JS, Hauer B, Kiener A, Wubbolts M, Witholt B; *Industrial biocatalysis today and tomorrow*. Nature **2001**; 409: 258-68

41. Otten LG, Sio CF, van der Sloot AM, Cool RH, Quax WJ; *Mutational analysis of a key residue in the substrate specificity of a cephalosporin acylase*. Chembiochem **2004**; 5: 820-5

42. Otten LG, Sio CF, Vrielink J, Cool RH, Quax WJ; *Altering the substrate specificity of cephalosporin acylase by directed evolution of the Beta -subunit*. J Biol Chem **2002**; 277: 42121-7

43. Sio CF, Otten LG, Cool RH, Quax WJ; *Analysis of a substrate specificity switch residue of cephalosporin acylase*. Biochem Biophys Res Commun **2003**; 312: 755-60

44. Sio CF, Riemens AM, van der Laan JM, Verhaert RM, Quax WJ; *Directed evolution of a glutaryl acylase into an adipyl acylase*. Eur J Biochem **2002**; 269: 4495-504

45. Sio CF, Quax WJ; *Improved beta-lactam acylases and their use as industrial biocatalysts*. Curr Opin Biotechnol **2004**; 15: 349-55

46. Giordano RC, Ribeiro MP, Giordano RL; *Kinetics of beta-lactam antibiotics synthesis by penicillin G acylase (PGA) from the viewpoint of the industrial enzymatic reactor optimization*. Biotechnol Adv **2006**; 24: 27-41

47. Chandel AK, Rao LV, Narasu ML, Singh OV; *The realm of penicillin G acylase in beta-lactam antibiotics*. Enzyme and microbioal technology **2008**; 42: 199-207

48. Henke E, Bornscheuer UT, Schmid RD, Pleiss J; *A molecular mechanism of enantiorecognition of tertiary alcohols by carboxylesterases*. Chembiochem **2003**; 4: 485-93

49. Kasche V, Haufler U, Riechmann L; *Equilibrium and kinetically controlled synthesis with enzymes: semisynthesis of penicillins and peptides*. Methods Enzymol **1987**; 136: 280-92

50. Kasche V, Haufler U, Riechmann L; *Kinetically controlled semisynthesis of beta-lactam antibiotics and peptides*. Ann N Y Acad Sci **1984**; 434: 99-105

51. Margolin AL, Svedas VK, Berezin IV; *Substrate specificity of penicillin amidase from E. coli*. Biochim Biophys Acta **1980**; 616: 283-9

52. Schroen CG, VandeWiel S, Kroon PJ, DeVroom E, Janssen AE, Tramper J; *Equilibrium position, kinetics, and reactor concepts for the adipyl-7-ADCA-hydrolysis process*. Biotechnol Bioeng **2000**; 70: 654-61

53. Kallenberg AI, van Rantwijk F, Sheldon RA; *Immobilization of Penicillin G Acylase: The Key to Optimum Performance*. Advanced Synthesis & Catalysis **2005**; 347: 905-926

54. Rich JO, Michels PC, Khmelnitsky YL; *Combinatorial biocatalysis*. Curr Opin Chem Biol **2002**; 6: 161-7

55. Massolini G, Fracchiolla G, Calleri E, Carbonara G, Temporini C, Lavecchia A, Cosconati S, Novellino E, Loiodice F; *Elucidation of the enantioselective recognition mechanism of a penicillin G acylase-based chiral stationary phase towards a series of 2-aryloxy-2-arylacetic acids*. Chirality **2006**; 18: 633-43

56. Massolini G, Temporini C, Calleri E; *Penicillin G acylase as chiral selector in LC and CE: exploring the origins of enantioselectivity*. J Chromatogr B Analyt Technol Biomed Life Sci **2008**; 875: 20-9

57. Polizzi KM, Chaparro-Riggers JF, Vazquez-Figueroa E, Bommarius AS; *Structure-guided consensus approach to create a more thermostable penicillin G acylase*. Biotechnol J **2006**; 1: 531-6

58. Alkema WB, de Vries E, Floris R, Janssen DB; *Kinetics of enzyme acylation and deacylation in the penicillin acylase-catalyzed synthesis of beta-lactam antibiotics*. Eur J Biochem **2003**; 270: 3675-83

59. Alkema WB, Prins AK, de Vries E, Janssen DB; *Role of alphaArg145 and betaArg263 in the active site of penicillin acylase of Escherichia coli*. Biochem J **2002**; 365: 303-9

60. Morillas M, McVey CE, Brannigan JA, Ladurner AG, Forney LJ, Virden R; *Mutations of penicillin acylase residue B71 extend substrate specificity by decreasing steric constraints for substrate binding*. Biochem J **2003**; 371: 143-50

61. Forney LJ, Wong DC, Ferber DM; *Selection of amidases with novel substrate specificities from penicillin amidase of Escherichia coli*. Appl Environ Microbiol **1989**; 55: 2550-5

62. Kim S, Kim Y; *Active site residues of cephalosporin acylase are critical not only for enzymatic catalysis but also for post-translational modification*. J Biol Chem **2001**; 276: 48376-81

63. Oh B, Kim M, Yoon J, Chung K, Shin Y, Lee D, Kim Y; *Deacylation activity of cephalosporin acylase to cephalosporin C is improved by changing the side-chain conformations of active-site residues*. Biochem Biophys Res Commun **2003**; 310: 19-27

Chapitre 2 : Évolution dirigée

Introduction

« Comme il naît beaucoup plus d'individus de chaque espèce qu'il n'en peut survivre, et que, par conséquent, il se produit souvent une lutte pour la vie, il s'ensuit que tout être, s'il varie, même légèrement, d'une manière qui lui est profitable, dans les conditions complexes et quelquefois variables de la vie, aura une meilleure chance pour survivre et ainsi se retrouvera choisi d'une façon naturelle. En raison du principe dominant de l'hérédité, toute variété ainsi choisie aura tendance à se multiplier sous sa forme nouvelle et modifiée. »[1]

Dans « the origin of species »,[2] Charles Darwin émit l'hypothèse de la sélection naturelle. La théorie de l'évolution darwinienne requiert trois principes : présence de variation des individus d'une population, le caractère héréditaire de ces variations et la compétition pour la survie de chaque individu. Les individus les mieux adaptés vont pouvoir se reproduire plus facilement et prendre une place de plus en plus importante dans la population globale. Cette théorie constitue l'explication essentielle du processus d'évolution, par cycles de mutation et sélection. Au vingtième siècle, avec les découvertes de la structure de l'ADN et du code génétique, le développement de techniques de séquençage et d'amplification de l'ADN de plus en plus efficaces et les progrès des méthodes computationnelles, un regard moléculaire sur l'évolution naturelle a progressivement vu le jour. La quantité de données toujours croissante ainsi que le progrès des méthodes de traitement de données, comme les méthodes de comparaison des séquences et des structures, permettent de mettre en lumière les mécanismes évolutifs des protéines.[3]

Le monde vivant est une ressource de biocatalyseurs, dont la plupart sont de nature protéique, finement ajustés pour catalyser des réactions biochimiques dans des conditions très spécifiques. Ces biocatalyseurs possèdent de nombreux avantages tel que leur grande spécificité pour un substrat (stéréo-, chémo-, énantio-, ou encore régiospécificité), leur nature biodégradable, réutilisable, les conditions douces de pH, température et pression auxquelles ils fonctionnent. Ces facteurs en font des catalyseurs très attractifs pour l'industrie (comme expliqué dans le chapitre 1). Cependant les réactions chimiques et/ou les conditions réactionnelles d'intérêt industriel ne correspondent que rarement aux réactions et conditions naturelles de ces biocatalyseurs. Il est dès lors intéressant de pouvoir créer des biocatalyseurs qui satisfassent les besoins de l'industrie. L'homme s'est inspiré des mécanismes évolutifs depuis très longtemps, en sélectionnant des plantes ou des animaux possédant certaines caractéristiques. A l'instar de l'évolution par sélection naturelle, il est possible de mimer le processus de manière accélérée à l'échelle du laboratoire et de créer des macromolécules biologiques possédant la nouvelle fonction ou propriété recherchée. L'ingénierie permet de modifier les propriétés catalytiques d'une enzyme en en modifiant par exemple la thermostabilité, la tolérance envers des solvants organiques ou encore la spécificité. L'évolution dirigée consiste à faire subir à un gène des cycles de mutagénèse puis de sélection ou de criblage pour faire évoluer la protéine encodée vers une nouvelle fonction donnée, comme par exemple sa capacité à lier de petites molécules ou à catalyser une réaction chimique, et ainsi créer de nouveaux ligands ou biocatalyseurs. La tâche s'avère néanmoins très ardue et les réalisations, bien que très nombreuses, restent laborieuses et souvent modestes en terme de sauts évolutifs. Notamment, l'évolution d'enzymes dotées de nouveaux mécanismes catalytiques reste un défi majeur pour le vingt-et-unième siècle.

Créer de la diversité

Il est important de bien choisir la méthode de mutagénèse utilisée pour construire des collections de protéines mutantes sucsceptibles de contenir des variants d'intérêt. Les trois stratégies principales sont les suivantes : error prone PCR (ou epPCR), la mutagénèse aléatoire ciblée sur des positions choisies, ou encore le DNA shuffling.[4] Une combinaison de plusieurs de ces techniques est également envisageable. La technique la plus souvent utilisée, car elle est sans doute la plus facile à mettre en œuvre, est l'epPCR. Une PCR mutagénéisante est réalisée sur l'entièreté du gène, ce qui ne nécessite pas spécialement une bonne connaissance préalable de la structure de l'enzyme en question. Différentes techniques peuvent être utilisées. En plus du kit « GeneMorph » de Agilent Technologies utilisant un mélange de deux polymérases et permettant d'avoir entre 0 et 16 mutations par kb, l'utilisation de la taq polymérase en présence d'agents mutagènes est aussi souvent utilisée (ajout de Mn^{2+}, biais dans la concentration en dNTPs, ajout d'analogues d'acides nucléiques).[5,6]

Elle est généralement considérée comme ciblant l'entièreté de l'enzyme, mais cette technique possède de sérieux désavantages. L'epPCR est sujette à des biais sévères dus à la dégénérescence du code génétique et il a été démontré que ce biais engendre une baisse de la diversité de la banque qui peux être jusqu'à 5 fois plus faible que la diversité théorique.[7-10] Il est donc nécessaire d'avoir recours à des techniques de sélection ou de criblage à ultra haut débit pour pouvoir trouver les variants intéressants. La mutagénèse aléatoire ciblée permet de ne dégénérer que certains résidus particuliers et de créer ainsi des banques mieux contrôlées et possiblement de meilleure qualité. La qualité d'une banque tient compte de la fraction de variants « intéressants » par rapport à la diversité totale et d'un meilleur degré d'amélioration de l'activité catalytique. Elle n'est cependant qualifiable qu'à postériori en regard des résultats obtenus. Cette technique nécessite néanmoins une bonne connaissance de l'enzyme et de sa structure afin de pouvoir choisir les résidus potentiellement intéressants à

mutagénéiser.[11] Les résidus ciblés sont mutagénéisés par insertion de cassette où les codons correspondants sont généralement remplacés par des séquences dégénérées de type nnk (k représentant t ou g, soit 32 codons pour 20 acides aminés) ; permettant ainsi une représentation moins biaisée de tous les acides aminés, et en supprimant deux des trois codons stop.

La recombinaison *in vitro* de familles de gènes homologues ou de mutants issus d'une première sélection sur une banque initiale permet de créer de la diversité. Cette technique, appelée le DNA shuffling,[12-14] consiste à cliver aléatoirement les gènes parents, avant de les recombiner *in vitro* par PCR, et présente l'avantage de potentiellement combiner des mutations bénéfiques tout en éliminant les délétères.

Il existe aujourd'hui un débat dans la littérature allant soit en faveur de la mutagénèse à faible taux de mutations combinée à un grand nombre de cycles de mutations et sélections[11] soit en faveur de la mutagénèse à haut taux de mutations combinée à des approches sélectives ou de criblage à haut débit.[6] L'opinion majoritaire tend plutôt pour la première mimant l'évolution naturelle qui consiste sans doute à une incorporation de peu de mutations à la fois, notamment parce qu'une mutagénèse importante introduit statistiquement beaucoup de mutations inactivantes (des codons stop, insertions ou délétions). Une mutagénèse plus limitée réduit cependant les chances d'observer des phénomènes épistatiques où la combinaison de plusieurs mutations a un effet bénéfique nettement plus important que la somme des effets des mutations individuelles.

L'« évolvabilité » d'une protéine défini sa capacité à tolérer des changements de séquence et de fonction. Il a été considéré pendant longtemps que les protéines étaient capables de tolérer quasiment toute mutation en gardant leur structure et fonction. Aujourd'hui, l'effet délétère des mutations qui diminuent fortement leur stabilité cinétique ou thermodynamique apparait comme un frein majeur à l'évolvabilité des protéines. Seule une petite fraction

de la diversité totale des mutants présentera des mutations bénéfiques ou neutres.[15-18]

Le « neutral drift » ou « dérive neutre » est un nouvel outil utile à la fois pour l'ingénierie de nouvelles protéines mais permet également des avancées dans la compréhension fondamentale de l'évolution moléculaire (développé par D. Tawfik et F. Arnold). Il consiste en l'accumulation de mutations dans un gène tout en maintenant une pression de sélection pour la conservation de la fonction d'origine et de la structure de la protéine étudiée. Un « nuage » de séquences autour de la séquence sauvage est alors obtenu, codant pour des protéines possédant les mêmes activités et structures que la séquence de départ. Une sélection pour l'activité sauvage permet l'accumulation de ces mutations neutres et bénéfiques tout en écartant les mutations délétères. Les banques construites à partir de petites collections de séquences issues du « neutral drift » possèdent généralement une stabilité plus grande et une large variation qu'une banque créée à partir d'une séquence unique, et ont de ce fait un meilleur potentiel évolutif.[19-21]

Certaines enzymes, catalysant une réaction spécifique, peuvent aussi catalyser d'autres réactions de manière très peu efficace. Cette capacité qu'à une enzyme à catalyser plusieurs réactions différentes dans le même site catalytique s'appelle la promiscuité catalytique, et pourrait être une trace de l'évolution naturelle.[15,16,22-24] Les enzymes ancestrales étaient vraisemblablement des enzymes généralistes qui ont évolué pour donner les enzymes spécialistes d'aujourd'hui. Mais il existe toujours des activités promiscuitaires latentes : elles n'ont pas subi de pression de sélection et ne servent à rien du point de vue physiologique mais elles existent. Elles seraient plutôt là par défaut et constituent un réservoir probablement essentiel pour l'émergence de nouvelles fonctions.[25]

Evolution dirigée d'une activité enzymatique

L'évolution d'activités enzymatiques, qu'elle soit naturelle ou dirigée, nécessite un lien entre le génotype, à savoir le gène codant pour l'enzyme, et le phénotype qui représente la fonction globale de l'enzyme en présence de ses substrats. Différents types de liens ont été utilisés et développés en fonction des besoins. Dans la nature, la cellule permet la compartimentalisation des gènes avec les enzymes qu'ils codent et les substrats que ces enzymes transforment. Beaucoup de fonctions protéiques ont été évoluées en laboratoire par criblage de colonies et sélections auxotrophes. Cependant l'expression et la sélection *in vivo* présentent certains problèmes. Il est tout d'abord impossible de travailler dans des conditions différentes que celles de l'environnement cellulaire. De plus, la taille des banques est limitée par l'efficacité de la transformation des cellules avec une collection d'ADN construite *in vitro*. La toxicité de certaines protéines pose également problème, et l'organisme hôte peut aussi contourner la pression sélective par un tout autre moyen que l'évolution de la protéine d'intérêt.[26]

Des techniques *in vitro* ont été développées pour éviter l'utilisation de microorganismes. Le phage display[27] est aujourd'hui la technique la plus utilisée pour faire évoluer l'affinité de protéines pour des ligands. Il permet de lier le gène et sa protéine via la capside du phage. Les phages sont sélectionnés par capture d'affinité de la protéine pour un ligand spécifique. Il est également possible de faire évoluer des activités enzymatiques en sélectionnant des variants se liant à des inhibiteurs suicides. Cependant, les enzymes sont sélectionnées pour un turnover unique, et ne seront peut-être pas les catalyseurs les plus efficaces sur le substrat réel. De plus, la préparation des phages nécessite une étape d'expression *in vivo* de la protéine d'intérêt en fusion avec la capside du phage et son exportation vers le périplasme. Cette méthode est donc toujours sujette à la toxicité et présente des limitations qui lui sont propres. Il existe également d'autres méthodes où un lien physique est présent entre le

gène et la protéine comme par exemple le ribozome display, qui présente l'avantage d'être un processus totalement *in vitro*.[28-30] Une banque de séquences ADN d'une protéine est transcrite puis traduite *in vitro*. Les gènes d'intérêts sont fusionnés à une région « linker » sans codons stop. Lors de la traduction, le linker reste accroché à son ARNt et associé au ribosome. La protéine protubérante peut alors se replier. Le complexe formé entre le ribosome, l'ARNm et la protéine permet la sélection par affinité de protéines ligands. Ce processus totalement *in vitro* n'est pas limité par l'efficacité de transformation de microorganismes. Il est dès lors possible de travailler avec de très grandes banques.

Evolution dirigée par IVC

Il est possible de créer des cellules artificielles simplifiées qui permettent de maintenir la compartimentation entre le gène, son produit d'expression et les éventuels substrats nécessaires à l'activité de la protéine. Les protéines d'intérêt peuvent être exprimées au sein de ces cellules artificielles et un test d'activité peut y être opéré. Cette technique s'appelle la compartimentalisation *in vitro* (IVC pour *« in vitro* compartimentation ») et se base sur la génération de microgouttelettes de phase aqueuse dispersée dans une phase huileuse.[31] Un des principaux avantages de l'IVC est que cette technique permet la sélection pour activité catalytique. Il est possible d'y opérer une catalyse intermoléculaire en *trans*, où le substrat n'est pas covalemment lié au catalyseur, et donc de sélectionner pour de multiples « turnovers ».[32]

Il existe différentes manières de créer ces microcompartiments. Les premières expériences réalisées utilisaient des émulsions dites en « bulk », où la phase aqueuse est dispersée dans la phase huileuse par simple agitation, en utilisant un homogénéisateur mécanique ou encore par extrusion à travers une membrane poreuse.

Chaque goutte renfermera un variant d'une enzyme, et un test d'activité y sera opéré. Cette technique est compatible avec plusieurs systèmes d'expression différents. Il est possible d'encapsuler des bactéries ou microorganismes[33] pour une expression cytoplasmique ou périplasmique, ou présentation de protéines à la surface de la cellule.[34] L'expression cytoplasmique requiert parfois la lyse de la bactérie au sein de la goutte. Il est aussi possible d'encapsuler l'ADN en présence d'un mélange de transcription et de traduction *in vitro* et des substrats de l'enzyme ; la goutte peut être alors considérée comme une cellule artificielle simplifiée.

L'IVC a permis de faire évoluer des activités enzymatiques dont, entre autres, des DNA méthyle transférases,[31,35,36] DNA polymérases,[37,38] endonucléases de restriction,[39] phosphotriestérases,[40] β-galactosidases,[41] et thioloactonases.[42] En présence d'un système de transcription *in vitro*, il a également été possible de faire évoluer des ARN catalytiques capables de turnovers multiples.[43,44]

Sélection de catalyseurs

L'IVC fut tout d'abord utilisée pour la sélection de nouveaux catalyseurs, comme illustré par l'évolution de DNA méthyle transférases qui est schématiquement détaillée à la Figure 2.1. La protéine est traduite *in vitro* dans la goutte à partir du gène. L'ADN contient le gène de la protéine ainsi qu'une séquence de restriction-méthylation (R-M). Si l'enzyme exprimée dans la goutte est active, elle méthylera le site R-M. Ensuite l'émulsion est cassée et la réaction est stoppée. L'ADN est mis en présence de l'enzyme de restriction en question. L'ADN non méthylé sera restreint alors que le méthylé ne le sera pas, et pourra ensuite être sélectivement amplifié par PCR, puis analysé ou réintroduit dans un nouveau cycle de sélection. La sélection modèle a permis la sélection du gène de la méthyle transférase HaeIII sauvage dans un mélange contenant un excès de gènes non fonctionnels d'un facteur 10^7 en seulement deux tours de sélection.[31]

La méthode a ensuite été utilisée pour modifier la spécificité de substrat de l'enzyme.[35] Ce genre de stratégies a également été appliqué à l'évolution d'autre types d'enzymes modifiant l'ADN comme des endonucléases de restriction,[39] ou encore des ADN polymérases.[37,38] Dans ce dernier cas, la polymérase si elle est active, va pouvoir amplifier son propre gène, permettant ainsi sa sélection.

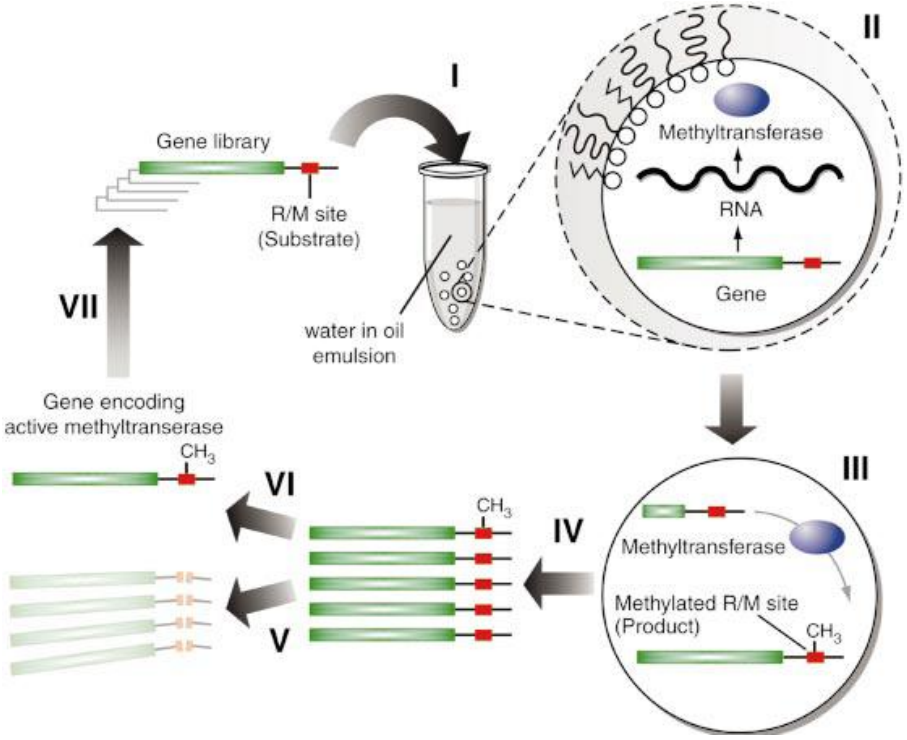


Figure 2.1 : Sélection d'ADN méthyle transférases par IVC. I. Une banque de variants de méthyl transférases ainsi qu'un mélange de transcription-traduction *in vitro* sont émulsifiés dans une émulsion eau-dans-huile. II. Les gènes sont transcrits et traduits. III. Le site de restriction-méthylation (R-M) est méthylé dans les gouttes contenant un variant actif. IV. L'émulsion est cassée, la phase aqueuse est récupérée et incubée en présence de l'enzyme de restriction correspondante. V. L'ADN non méthylé est restreint. VI. L'ADN méthylé n'est pas restreint. VII. Ce dernier est ensuite sélectivement amplifié par PCR et peut être réintroduit dans un second cycle de sélection (image provenant de[36]).

Criblage de catalyseurs par FACS

Le criblage permet l'analyse de chaque variant individuellement, induisant un meilleur contrôle de la procédure de sélection. Contrairement à la sélection, il nécessite l'analyse séquentielle de chaque variant, ce qui limite le nombre de variants analysables. Cependant, le tri par FACS (pour « fluorescence-activated cell sorting ») peut cribler des clones à une vitesse de plus de 10^7 clones par heure.

L'analyse d'activités enzymatiques par FACS requiert un substrat fluorogénique afin que le produit fluorescent puisse être détecté. Il est également nécessaire de travailler avec des doubles émulsions d'eau dans l'huile dans l'eau car une gaine de phase huileuse est incompatible avec le FACS.[45] De plus, tout phénomène de diffusion du produit dans l'huile ou vers d'autres compartiments doit être évité car il empêchera le bon déroulement de l'expérience. Il est aussi possible d'utiliser des substrats qui, après transformation par l'enzyme, se fixent à la surface des cellules contenant l'ADN d'intérêt[46,47] ou sur des microbilles sur lesquelles l'ADN d'intérêt a été immobilisé.[40] L'émulsion est alors cassée avant le tri au FACS. L'utilisation d'un substrat fluorogénique classique permet de suivre la réaction catalytique dans l'émulsion intacte. L'enzyme d'intérêt doit être exprimée au sein de la goutte de manière *in vitro*, en présence d'un mélange de transcription traduction *in vitro* (IVTT)[41] ou *in vivo* par exemple ci-dessous dans une bactérie *E. coli* (Figure 2.2).[42] Le gène et le système d'expression sont compartimentalisés avec le substrat fluorogénique dans une double émulsion eau-dans-huile-dans-eau. Après incubation, les enzymes actives transformeront le substrat en produit fluorescent. Les gouttes de la double émulsion seront ensuite triées au FACS (Figure 2.2).

Cette méthode a été utilisée entre autre pour l'évolution de l'activité thiolactonase de PON1 (mammalian serum paraoxanase), schématisée dans la

partie gauche de la Figure 2.2. La banque de mutants de PON1 a été exprimée en bactéries puis encapsulée dans une émulsion eau-dans-huile. Du substrat TBL (thiobutyrolactone) ainsi qu'un substrat fluorogénique permettant la détection de thiols ont été ajoutés par diffusion à travers la phase huileuse. Ensuite l'émulsion a été transformée en double émulsion (eau-dans-huile-dans-eau) pour pouvoir être criblée par FACS. Des mutants possédant une activité (k_{cat}/K_M) de 20 à 100 fois supérieure à l'enzyme sauvage ont été trouvés.[42]

La partie droite de la Figure 2.2 illustre la sélection *in vitro* de variants de la protéine Ebg possédant une activité β-galactosidase améliorée. Ebg est une protéine d'*E. coli* de fonction inconnue ayant une très faible activité β-galactosidase qui fut souvent utilisée comme modèle pour l'étude de l'évolution d'activités enzymatiques.[45] Dans cet exemple, la banque de variants est exprimée *in vitro* en gouttes, un substrat fluorogénique est ajouté, l'émulsion est transformée en double émulsion et criblée par FACS. Les meilleurs variants sélectionnés possèdent une activité catalytique (k_{cat}/K_M) 1700 fois supérieure à celle de l'Ebg sauvage.[41]

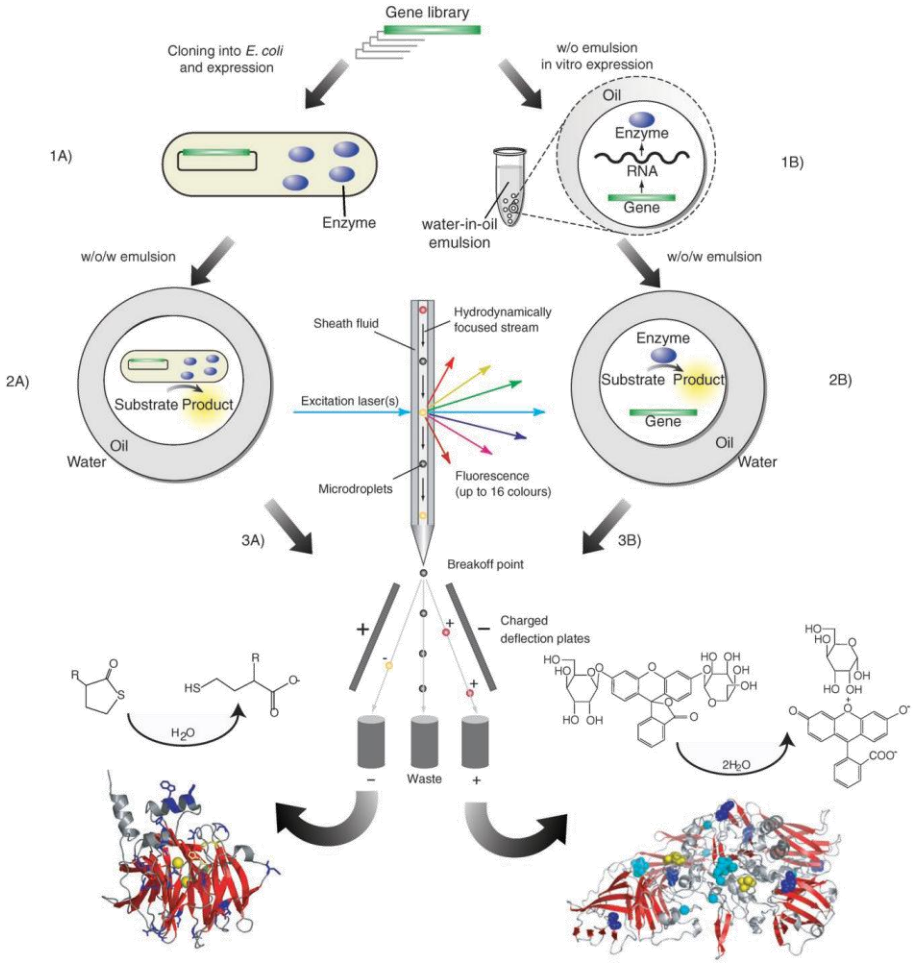


Figure 2.2 : Sélection d'émulsions par FACS. 1A. une banque de variants est clonée et exprimée dans *E. coli*. Les bactéries sont encapsulées dans une émulsion eau-dans-huile de telle sorte que chaque goutte contienne généralement maximum une bactérie. 1B. Un mélange d'IVTT contenant une banque de gènes est émulsifié de manière à avoir un gène par goutte. Les gènes sont exprimés au sein des gouttes. 2A et B. Les variants actifs convertissent le substrat en produit fluorescent. Ensuite l'émulsion eau-dans-huile est transformée en émulsion double eau-dans-huile-dans-eau. 3A et B. Les gouttes fluorescentes sont triées par FACS. Ces gouttes contiennent des variants capables d'hydrolyser le thiobutyrolactone (montré en 3.A) ou la fluorescéine di-β-galactopyranoside (3.B). Les bactéries ou les variants sélectionnés provenant des gouttes fluorescentes sont récupérés, les bactéries sont mises en culture, ou l'ADN est amplifié par PCR. Les gènes peuvent ensuite être introduits dans un second tour de sélection. (Image provenant de[48])

Manipulation d'émulsions par microfluidique

Les émulsions réalisées par les techniques classiques d'IVC (appelées émulsions en bulk) ont fait leurs preuves malgré le fait qu'elles présentent plusieurs désavantages majeurs. Tout d'abord leur grande hétérogénéité de taille de gouttes oblige la dilution importante des banques pour éviter au maximum la présence de deux variants au sein de la même goutte, ce qui augmente le nombre de gouttes à cribler pour analyser au mieux la diversité de la banque.

De plus, la composition des gouttes est difficilement modifiable après leur création. Bien qu'il soit possible d'ajouter par exemple un substrat fluorogénique hydrophobe dans la phase huileuse afin qu'il diffuse dans les gouttes,[40,43] ou de modifier le pH par addition d'acide acétique,[40] il est impossible de fusionner les gouttes avec de nouvelles gouttes contenant par exemple un nouveau tampon ou de nouveaux substrats qui ne peuvent être amenés dans la goutte par simple diffusion. De plus, l'analyse cinétique des réactions en goutte est impossible, le système d'analyse des émulsions au FACS ne permet que les mesures de type « end-point ».

L'utilisation de plateformes microfluidiques pour la création et manipulation d'émulsions est en plein essor et permet la création d'émulsions monodisperses, avec une grande maîtrise de la taille des gouttes. L'intérêt majeur de ces plateformes est leur grande modularité. Des modules individuels peuvent être combinés sur une même puce microfluidique.[49-51] Il est possible de générer des gouttes à des fréquences de 0.1 à 10 kHz (figure 2.3a).[52,53] La phase aqueuse est alors dispersée dans une phase huileuse continue qui peut être une huile organique, minérale ou fluorée. Les huiles fluorées conviennent parfaitement à la microfluidique en gouttes car elles sont compatibles avec les puces en polydimethylsiloxane (PDMS), elles sont inertes, non-miscibles avec l'eau, transparentes, hydrophobes, insolubles, et permettent ainsi la détection de fluorescence, et par leur capacité de transport de gaz elles permettent aussi la

culture de cellules en gouttes.[54-56] La génération de gouttes stables dépend principalement du choix du tensio-actif, qui doit également être biocompatible.[33]

Après leur création, les gouttes peuvent subir une série de manipulations. Leur contenu peut être mélangé (Figure 2.3b) ou alors modifié via la fusion à haut débit de chaque goutte de l'émulsion avec une autre goutte contenant un tampon différent, un substrat ou un inhibiteur. Différents systèmes de fusion ont été développés. La coalescence des gouttes peut être induite de manière active par l'application d'un champ électrique (Figure 2.3c),[57-59] par un laser[60,61] ou encore de manière passive en jouant sur la stabilité des gouttes, via des variations de la concentration en tensioactif.[62]

Ensuite, les gouttes peuvent être incubées sur puce dans des lignes de délais pour des durées allant de quelques minutes[34,63] à environ une heure (Figure 2.3d).[64] Il est également possible d'incuber les gouttes plus longtemps de manière stationnaire dans des « dropspots », des réservoirs (Figure 2.3e) ou des piège à gouttes,[65,66] mais dans ce cas, seul un nombre limité de gouttes pourront être stockées et analysées. Une alternative consiste à incuber l'émulsion en dehors de la puce dans un tube de type eppendorf (Figure 2.3j) ou un capillaire.[67] Elle peut alors être incubée pour de plus longues durées ou encore thermocyclée[68] puis être réinjectée dans un module de réinjection où les gouttes seront espacées suffisamment par de l'huile perfluorée (Figure 2.3h). La fluorescence de chaque goutte peut également être analysée quantitativement par couplage à un système de détection optique. Il est aussi possible de trier les gouttes en fonction de leur fluorescence (Figure 2.3g). La technique de tri la plus efficace actuellement est le tri par « dielectrical steering » et permet d'atteindre des fréquences de tri entre 1 et 2 kHz, avec un très faible de taux de faux positifs.[34,69]

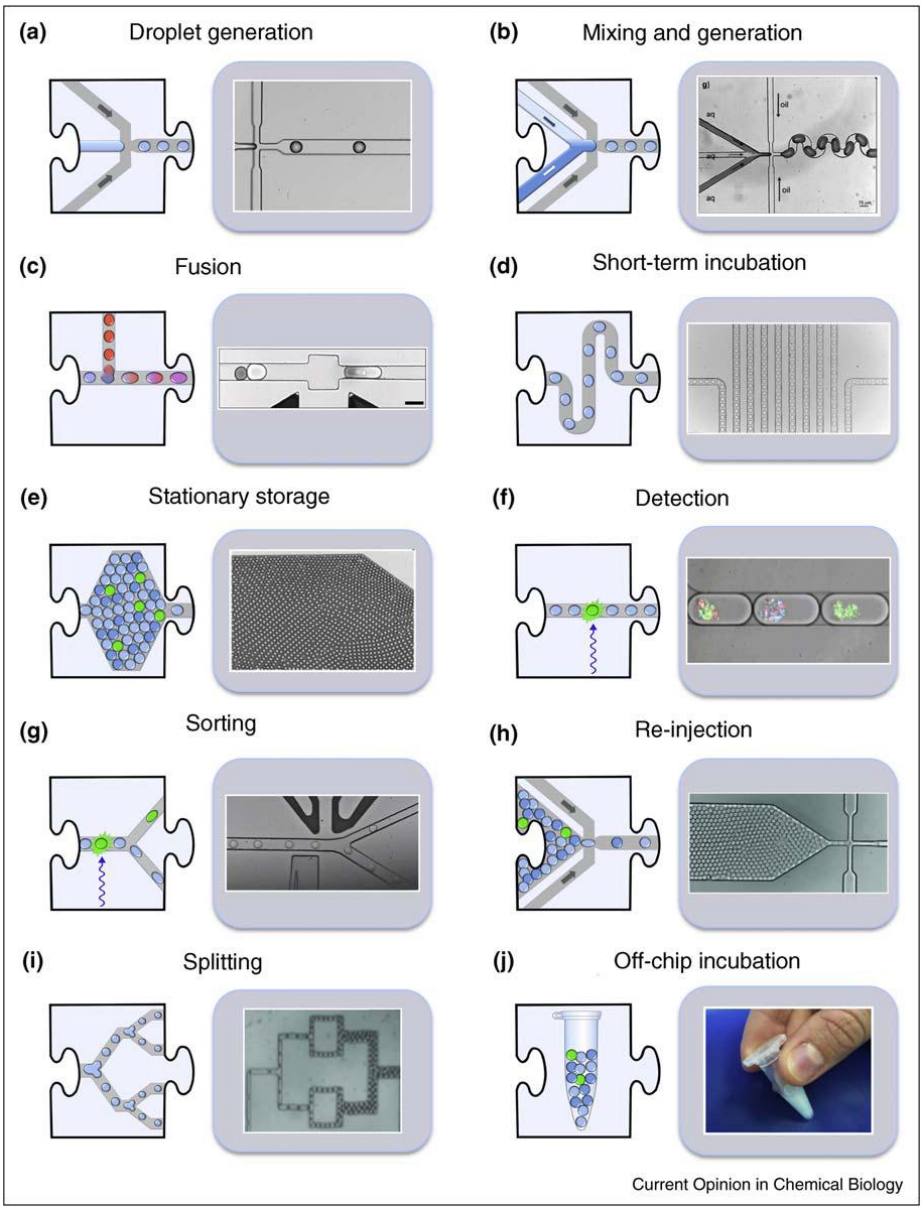


Figure 2.3 : Présentation des principaux modules utilisés pour l'utilisation de la microfluidique en gouttes. Les modules peuvent être assemblés et intégrés à une plateforme microfluidique en fonction des besoins. (a) génération de gouttes, (b) génération et mélange de gouttes, (c) fusion, (d) incubation à court terme, (e) incubation stationnaire, (f) détection de la fluorescence, (g) tri, (h) réinjection, (i) division de gouttes,[57] (j) incubation hors puce. (Figure provenant de[49])

Application de la microfluidique en gouttes pour l'évolution dirigée

La combinaison des différents modules de base vus ci-dessus permet entre autres d'amplifier des gènes en goutte, et de les traduire ensuite *in vitro*, de suivre une cinétique enzymatique en gouttes à partir d'enzymes produites *in vitro* ou *in vivo*, de trier les gouttes sur base de leur fluorescence et permet de se diriger vers une plateforme totalement intégrée pour l'évolution dirigée en microfluidique. Les quelques exemples décrits ci-dessous permettent de se faire une idée de l'avancement de la technique disponible aujourd'hui, sans pour autant faire une revue intégrale de tous les systèmes de microfluidique en gouttes opérationnels à ce jour.

Cinétique enzymatique en gouttes

Un des avantages de l'utilisation de puces microfluidiques est la possibilité de suivre la cinétique d'une réaction au sein-même des gouttes. Il est possible d'utiliser des lignes de délais permettant des incubations allant jusqu'à environ une heure. Les gouttes y sont en mouvement permanent, et leur fluorescence est mesurable en différents points correspondant à différents temps d'incubation (Figure 2.4a). Ici, les gouttes sont créées et incubées sur une même puce, et l'activité de la β-lactamase a pu être suivie.[64]

Un autre exemple de cinétique en gouttes a permis de suivre la cinétique enzymatique d'une alcaline phosphatase exprimée dans le périplasme d'*E. coli* en présence d'un substrat fluorogénique (3-*O*-methylfluorescein-phosphate) (Figure 2.5).[70] De manière similaire, Mazutis et al. ont exprimé la laccase *cotA in vitro*. Après incubation hors puce, l'émulsion a été réinjectée et fusionnée à une seconde émulsion contenant les réactifs nécessaires au test enzymatique. L'émulsion a été collectée hors puce, incubée et réinjectée pour analyse de la fluorescence à intervalles réguliers (Figure 2.6).[71]

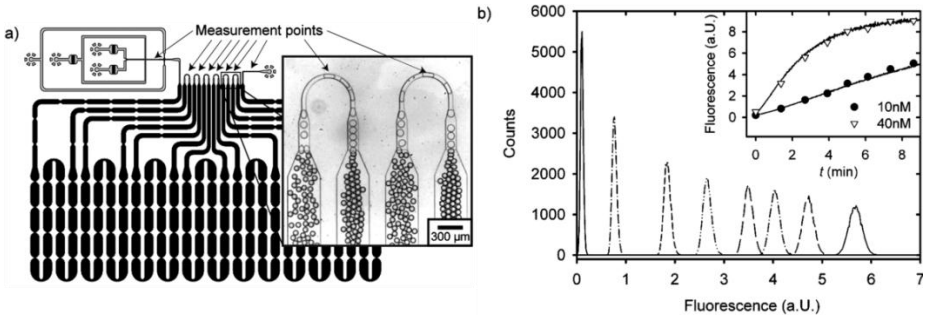


Figure 2.4 : a. Design microfluidique pour tests cinétiques en gouttes comprenant huit points de mesures au niveau des canaux peu profonds (25 µm de hauteur). Les canaux plus profonds (75 µm) présentent des restrictions qui permettent le mélange des gouttes pendant l'incubation. b. Cinétique enzymatique de la β-lactamase. Histogrammes des intensités de fluorescence de gouttes contenant l'enzyme (10 nM) et son substrat (fluorocillin, 10 µM) à chaque point de mesure. Le second graphe présente l'intensité de fluorescence moyenne des gouttes en fonction du temps d'incubation en présence de 10 nM ou 40 nM d'enzyme. Les lignes continues correspondent aux mesures faites en cuvettes. (Figure provenant de[64])

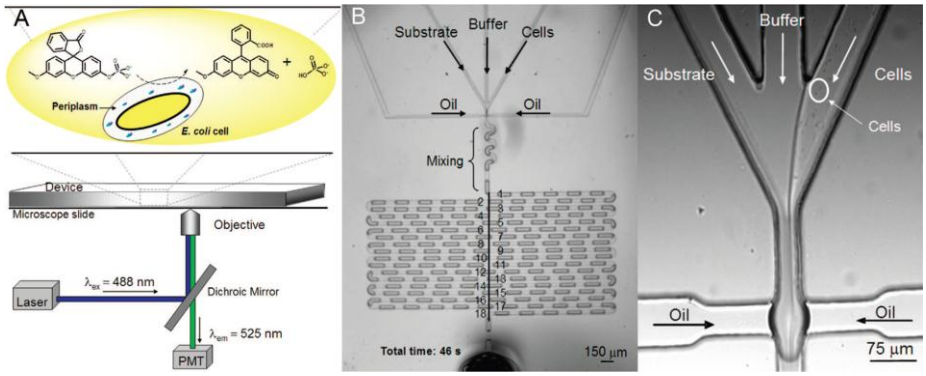


Figure 2.5 : Schéma de la cinétique enzymatique et de la puce microfluidique. A. Test fluorogénique pour l'activité de l'alkaline phosphatase exprimée dans le périplasme d'*E. coli*. La fluorescence des gouttes est mesurée à 18 positions par un PMT (photomultiplier tube). B. Les gouttes sont formées de trois solutions aqueuses injectées séparément (le substrat, le tampon et les cellules, l'image C. montre cette section en detail). Les gouttes formées sont directement mélangées et incubées dans la ligne de délais et leur fluorescence est mesurée en 18 points. (Figure provenant de[70])

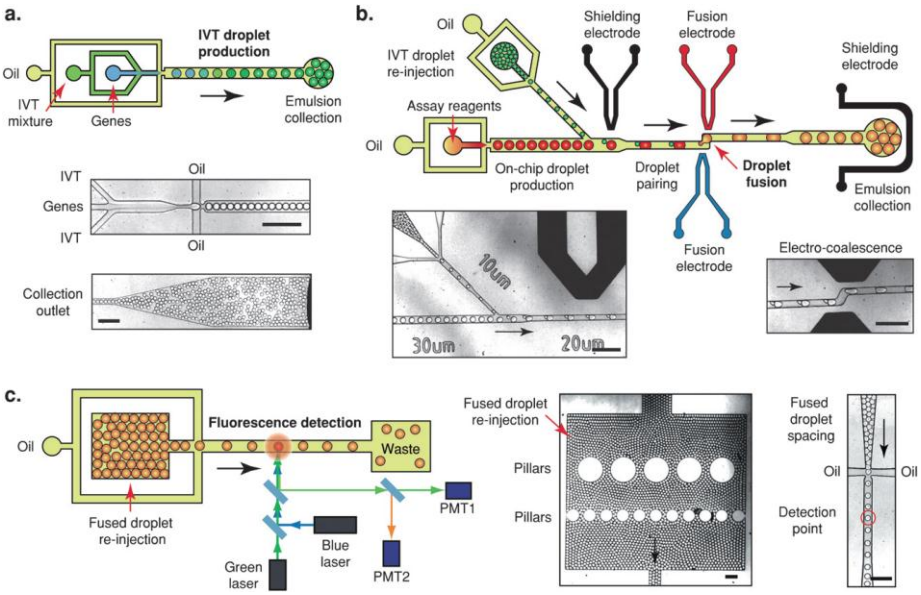


Figure 2.6 : Design d'une plateforme microfluidique permettant de suivre en gouttes la cinétique enzymatique d'une enzyme exprimée *in vitro*. a. Schéma de la puce permettant la formation de gouttes contenant l'IVTT et l'ADN codant pour la laccase cotA. b. Après incubation, l'émulsion d'IVTT (verte) est réinjectée et fusionnée avec une émulsion contenant les réactifs nécessaires au test enzymatique (rouge). c. L'émulsion fusionnée est incubée puis réinjectée après différents temps d'incubation dans une puce pour analyse de la fluorescence des gouttes. (Figure provenant de[71])

Amplification d'ADN en gouttes

Il est souvent difficile d'atteindre des concentrations de protéines suffisantes par expression *in vitro* à des concentrations d'un gène par goutte. L'évolution dirigée requiert que la majorité des gouttes ne contiennent pas plus qu'une molécule d'ADN. Il est donc nécessaire que l'ADN de la banque soit encapsulé puis amplifié en goutte par PCR ou par amplification isothermale (HRCA pour « hyperbranched rolling cirle amplification »).[67] L'amplification par PCR[68] suivie d'une fusion avec une émulsion contenant de l'IVTT est développée en détails dans le chapitre 4. L'amplification isothermale de plasmides suivie de l'expression *in vitro* du gène l*acZ* en présence d'un substrat fluorogénique afin de détecter l'activité de la β-galactosidase fut réalisée par Mazutis et al.[67] La Figure 2.7 reprend les différentes étapes de la manipulation.

Des gouttes de 2 pL contenant l'ADN et les réactifs nécessaires pour l'HRCA sont tout d'abord produites (Figure 2.6a) puis collectées et incubées hors puce à 30°C. Les gouttes sont ensuite analysées (Figure 2.6b) puis fusionnées à des gouttes de 15 pL contenant de l'IVTT et du substrat fluorogénique. Les gouttes fusionnées sont incubées puis réinjectées pour l'analyse de l'activité β-galactosidase.[67]

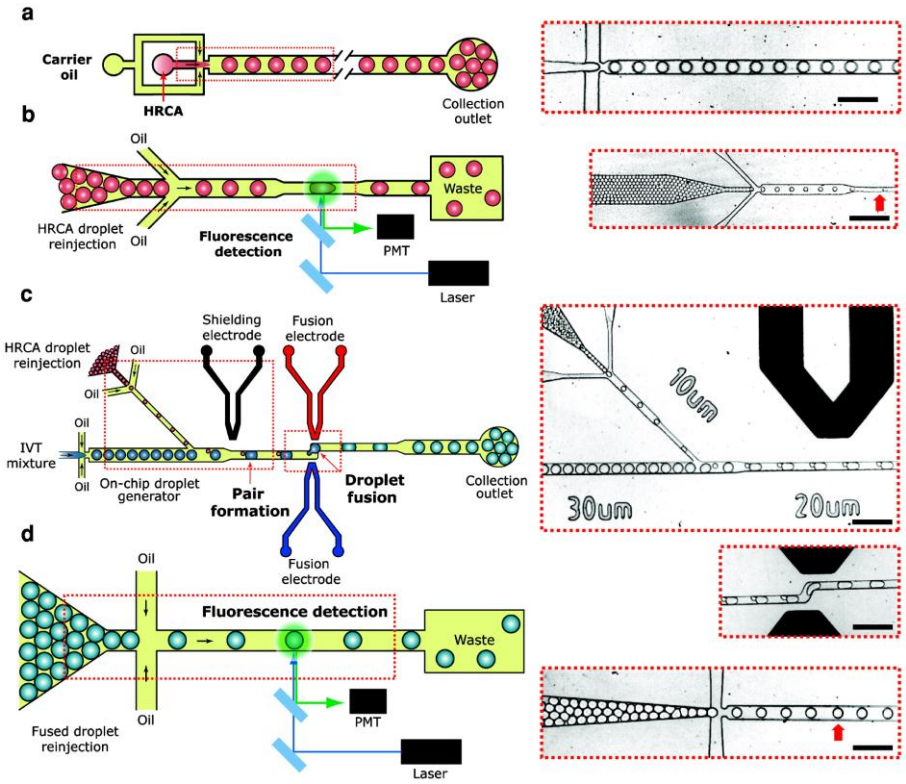


Figure 2.7 : Design d'une plateforme microfluidique permettant l'amplification isothermale de plasmides en gouttes suivie de l'expression des gènes *in vitro* et d'un test enzymatique. A. Puce pour la création d'une émulsion contenant maximum un gène lacZ par goutte, un mélange HCRA (hyperbranched rolling circle amplification) et un marqueur intercalant d'ADN double brin. B. Puce permettant l'analyse de l'émulsion HCRA grâce à la fluorescence du marqueur intercalant. C. L'émulsion (rouge) est réinjectée et fusionnée à des gouttes contenant de l'IVTT et les réactifs nécessaires au test enzymatique. (Figure provenant de[67])

Tri de banques de variants

La microfluidique en goutte a permis l'évolution dirigée de la horseradish peroxidase en deux tours de mutagénèse et criblage.[34] Le système utilisé est décrit à la Figure 2.7. La banque est exprimée à la surface de cellules de levure comme protéine de fusion à la protéine Aga2. Deux banques successives d'environs 10^7 mutants ont été créées. Au premier tour, 100 variants d'activité comparable à l'enzyme sauvage ont été récupérés et utilisés pour la création de la seconde banque. Ensuite, la seconde banque fut criblée pour une activité améliorée et des mutants 10 fois plus actifs que le sauvage ont pu être sélectionnés.

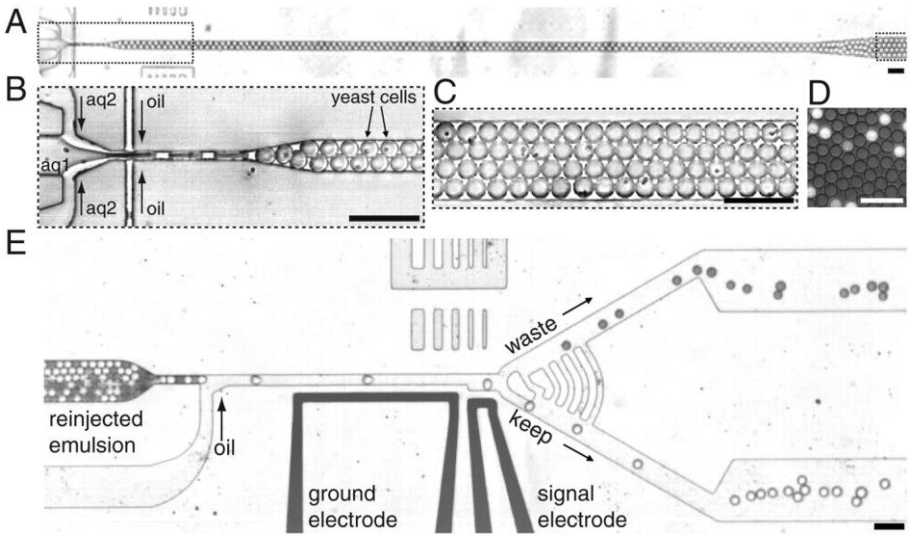


Figure 2.7 : Plateforme microfluidique utilisée par Agresti et al. pour l'évolution dirigée de la horseradish peroxidase. A. Image de la première puce utilisée, comprenant un module pour la formation des gouttes (B), et l'incubation des gouttes dans une ligne de délais d'environ 5 minutes (C et D). L'émulsion est ensuite réinjectée, analysée et triée dans la puce présentée en E. (Echelle : 80 μm, figure provenant de[34])

Références

1. Darwin C; *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle of Life*, Barbier E; *L'origine des espèces : au moyen de la sélection naturelle ou la préservation des races favorisées dans la lutte pour la vie*, Flammarion, **1992**

2. Darwin C; *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, John Murray, **1859**

3. Orengo CA, Thornton JM; *Protein families and their evolution-a structural perspective*. Annu Rev Biochem **2005**; 74: 867-900

4. Yuan L, Kurek I, English J, Keenan R; *Laboratory-directed protein evolution*. Microbiol Mol Biol Rev **2005**; 69: 373-92

5. Brakmann S, A. S; *Evolutionary Methods in Biotechnology: Clever Tricks for Directed Evolution*, Susanne Brakmann, Andreas Schwienhorst, **2005**

6. Zaccolo M, Gherardi E; *The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase*. J Mol Biol **1999**; 285: 775-83

7. Eggert T, Reetz MT, Jaeger KE; *Enzyme functionality-design, engineering and screening*, **2004**

8. Drummond DA, Iverson BL, Georgiou G, Arnold FH; *Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins*. Journal of Molecular Biology **2005**; 350: 806-816

9. Wong TS, Tee KL, Hauer B, Schwaneberg U; *Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution*. Nucleic Acids Research **2004**; 32

10. Shafikhani S, Siegel RA, Ferrari E, Schellenberger V; *Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization*. Biotechniques **1997**; 23: 304-310

11. Reetz MT, Kahakeaw D, Lohmer R; *Addressing the numbers problem in directed evolution*. Chembiochem **2008**; 9: 1797-1804

12. Coco WM, Levinson WE, Crist MJ, Hektor HJ, Darzins A, Pienkos PT, Squires CH, Monticello DJ; *DNA shuffling method for generating highly recombined genes and evolved enzymes*. Nat Biotechnol **2001**; 19: 354-9

13. Stemmer WP; *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution*. Proc Natl Acad Sci U S A **1994**; 91: 10747-51

14. Stemmer WP; *Rapid evolution of a protein in vitro by DNA shuffling*. Nature **1994**; 370: 389-91

15. Khersonsky O, Roodveldt C, Tawfik DS; *Enzyme promiscuity: evolutionary and mechanistic aspects*. Curr Opin Chem Biol **2006**; 10: 498-508

16. Tokuriki N, Tawfik DS; *Protein dynamism and evolvability*. Science **2009**; 324: 203-7

17. Bloom JD, Arnold FH; *In the light of directed evolution: Pathways of adaptive protein evolution*. Proceedings of the National Academy of Sciences of the United States of America **2009**; 106: 9995-10000

18. Bloom JD, Labthavikul ST, Otey CR, Arnold FH; *Protein stability promotes evolvability*. Proceedings of the National Academy of Sciences of the United States of America **2006**; 103: 5869-5874

19. Bershtein S, Goldin K, Tawfik DS; *Intense neutral drifts yield robust and evolvable consensus proteins*. J Mol Biol **2008**; 379: 1029-44

20. Gupta RD, Tawfik DS; *Directed enzyme evolution via small and effective neutral drift libraries*. Nat Methods **2008**; 5: 939-42

21. Peisajovich SG, Tawfik DS; *Protein engineers turned evolutionists*. Nat Methods **2007**; 4: 991-4

22. O'Brien PJ, Herschlag D; *Catalytic promiscuity and the evolution of new enzymatic activities*. Chemistry & Biology **1999**; 6: R91-R105

23. Copley SD; *Enzymes with extra talents: moonlighting functions and catalytic promiscuity*. Current Opinion in Chemical Biology **2003**; 7: 265-272

24. Babtie AC, Bandyopadhyay S, Olguin LF, Hollfelder F; *Efficient Catalytic Promiscuity for Chemically Distinct Reactions*. Angewandte Chemie-International Edition **2009**; 48: 3692-3694

25. Tawfik DS; *Messy biology and the origins of evolutionary innovations*. Nature Chemical Biology **2010**; 6: 692-696

26. Leemhuis H, Stein V, Griffiths AD, Hollfelder F; *New genotype-phenotype linkages for directed evolution of functional proteins*. Curr Opin Struct Biol **2005**; 15: 472-8

27. Smith GP, Petrenko VA; *Phage display*. Chemical Reviews **1997**; 97: 391-410

28. Hanes J, Pluckthun A; *In vitro selection and evolution of functional proteins by using ribosome display*. Proc Natl Acad Sci U S A **1997**; 94: 4937-42

29. Lipovsek D, Pluckthun A; *In-vitro protein evolution by ribosome display and mRNA display*. J Immunol Methods **2004**; 290: 51-67

30. Zahnd C, Amstutz P, Pluckthun A; *Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target*. Nat Methods **2007**; 4: 269-79

31. Tawfik DS, Griffiths AD; *Man-made cell-like compartments for molecular evolution*. Nature Biotechnology **1998**; 16: 652-656

32. Griffiths AD, Tawfik DS; *Man-made enzymes - from design to in vitro compartmentalisation*. Current Opinion in Biotechnology **2000**; 11: 338-353

33. Clausell-Tormos J, Lieber D, Baret JC, El-Harrak A, Miller OJ, Frenz L, Blouwolff J, Humphry KJ, Koster S, Duan H, Holtze C, Weitz DA, Griffiths AD, Merten CA; *Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms*. Chem Biol **2008**; 15: 427-37

34. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA; *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution*. Proc Natl Acad Sci U S A **2010**; 107: 4004-9

35. Cohen HM, Tawfik DS, Griffiths AD; *Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization*. Protein Engineering Design & Selection **2004**; 17: 3-11

36. Lee YF, Tawfik DS, Griffiths AD; *Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation*. Nucleic Acids Research **2002**; 30: 4937-4944

37. Ghadessy FJ, Ong JL, Holliger P; *Directed evolution of polymerase function by compartmentalized self-replication*. Proceedings of the National Academy of Sciences of the United States of America **2001**; 98: 4552-4557

38. Ghadessy FJ, Ramsay N, Boudsocq F, Loakes D, Brown A, Iwai S, Vaisman A, Woodgate R, Holliger P; *Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution*. Nature Biotechnology **2004**; 22: 755-759

39. Doi N, Kumadaki S, Oishi Y, Matsumura N, Yanagawa H; *In vitro selection of restriction endonucleases by in vitro compartmentalization*. Nucleic Acids Research **2004**; 32

40. Griffiths AD, Tawfik DS; *Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization*. Embo Journal **2003**; 22: 24-35

41. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD; *High-throughput screening of enzyme libraries: In vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions*. Chemistry & Biology **2005**; 12: 1291-1300

42. Aharoni A, Amitai G, Bernath K, Magdassi S, Tawfik DS; *High-throughput screening of enzyme libraries: Thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments*. Chemistry & Biology **2005**; 12: 1281-1289

43. Agresti JJ, Kelly BT, Jaschke A, Griffiths AD; *Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization*. Proc Natl Acad Sci U S A **2005**; 102: 16170-5

44. Levy M, Griswold KE, Ellington AD; *Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization*. Rna-a Publication of the Rna Society **2005**; 11: 1555-1562

45. Kelly BT, Baret JC, Taly V, Griffiths AD; *Miniaturizing chemistry and biology in microdroplets*. Chemical Communications **2007**: 1773-1788

46. Olsen MJ, Stephens D, Griffiths D, Daugherty P, Georgiou G, Iverson BL; *Function-based isolation of novel enzymes from a large library*. Nature Biotechnology **2000**; 18: 1071-1074

47. Varadarajan N, Gam J, Olsen MJ, Georgiou G, Iverson BL; *Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity*. Proceedings of the National Academy of Sciences of the United States of America **2005**; 102: 6855-6860

48. Taly V, Kelly BT, Griffiths AD; *Droplets as microreactors for high-throughput biology*. Chembiochem **2007**; 8: 263-272

49. Kintses B, van Vliet LD, Devenish SR, Hollfelder F; *Microfluidic droplets: new integrated workflows for biological experiments*. Curr Opin Chem Biol **2010**; 14: 548-55

50. Huebner A, Sharma S, Srisa-Art M, Hollfelder F, Edel JB, Demello AJ; *Microdroplets: a sea of applications?* Lab Chip **2008**; 8: 1244-54

51. Theberge AB, Courtois F, Schaerli Y, Fischlechner M, Abell C, Hollfelder F, Huck WT; *Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology*. Angew Chem Int Ed Engl **2010**; 49: 5846-68

52. Umbanhowar PB, Prasad V, Weitz DA; *Monodisperse emulsion generation via drop break off in a coflowing stream*. Langmuir **2000**; 16: 347-351

53. Thorsen T, Roberts RW, Arnold FH, Quake SR; *Dynamic pattern formation in a vesicle-generating microfluidic device*. Physical Review Letters **2001**; 86: 4163-4166

54. Lowe KC, Davey MR, Power JB; *Perfluorochemicals: their applications and benefits to cell culture*. Trends Biotechnol **1998**; 16: 272-7

55. Huque FTT, Jones K, Saunders RA, Platts JA; *Statistical and theoretical studies of fluorophilicity*. Journal of Fluorine Chemistry **2002**; 115: 119-128

56. Studer A, Hadida S, Ferritto R, Kim SY, Jeger P, Wipf P, Curran DP; *Fluorous synthesis: A fluorous-phase strategy for improving separation efficiency in organic synthesis*. Science **1997**; 275: 823-826

57. Link DR, Grasland-Mongrain E, Duri A, Sarrazin F, Cheng Z, Cristobal G, Marquez M, Weitz DA; *Electric control of droplets in microfluidic devices*. Angew Chem Int Ed Engl **2006**; 45: 2556-60

58. Zagnoni M, Baroud CN, Cooper JM; *Electrically initiated upstream coalescence cascade of droplets in a microfluidic flow*. Phys Rev E Stat Nonlin Soft Matter Phys **2009**; 80: 046303

59. Niu X, Gielen F, deMello AJ, Edel JB; *Electro-coalescence of digitally controlled droplets*. Anal Chem **2009**; 81: 7321-5

60. Baroud CN, de Saint Vincent MR, Delville JP; *An optical toolbox for total control of droplet microfluidics*. Lab Chip **2007**; 7: 1029-33

61. Verneuil E, Cordero M, Gallaire F, Baroud CN; *Laser-induced force on a microfluidic drop: origin and magnitude*. Langmuir **2009**; 25: 5127-34

62. Mazutis L, Baret JC, Griffiths AD; *A fast and efficient microfluidic system for highly selective one-to-one droplet fusion*. Lab Chip **2009**; 9: 2665-72

63. Woronoff G, El Harrak A, Mayot E, Schicke O, Miller OJ, Soumillion P, Griffiths AD, Ryckelynck M; *New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays in droplet-based microfluidic applications*. Anal Chem **2011**; 83: 2852-7

64. Frenz L, Blank K, Brouzes E, Griffiths AD; *Reliable microfluidic on-chip incubation of droplets in delay-lines*. Lab Chip **2009**; 9: 1344-8

65. Abbyad P, Dangla R, Alexandrou A, Baroud CN; *Rails and anchors: guiding and trapping droplet microreactors in two dimensions*. Lab Chip **2010**; 11: 813-21

66. Courtois F, Olguin LF, Whyte G, Theberge AB, Huck WT, Hollfelder F, Abell C; *Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays*. Anal Chem **2009**; 81: 3008-16

67. Mazutis L, Araghi AF, Miller OJ, Baret JC, Frenz L, Janoshazi A, Taly V, Miller BJ, Hutchison JB, Link D, Griffiths AD, Ryckelynck M; *Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis*. Anal Chem **2009**; 81: 4813-21

68. Pekin D, Skhiri Y, Baret JC, Le Corre D, Mazutis L, Salem CB, Millot F, El Harrak A, Hutchison JB, Larson JW, Link DR, Laurent-Puig P, Griffiths AD, Taly V; *Quantitative and sensitive detection of rare mutations using droplet-based microfluidics*. Lab Chip **2011**; 11: 2156-66

69. Baret JC, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, Rick C, Samuels ML, Hutchison JB, Agresti JJ, Link DR, Weitz DA, Griffiths AD; *Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity*. Lab Chip **2009**; 9: 1850-8

70. Huebner A, Olguin LF, Bratton D, Whyte G, Huck WT, de Mello AJ, Edel JB, Abell C, Hollfelder F; *Development of quantitative cell-based enzyme assays in microdroplets*. Anal Chem **2008**; 80: 3890-6

71. Mazutis L, Baret JC, Treacy P, Skhiri Y, Araghi AF, Ryckelynck M, Taly V, Griffiths AD; *Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme*. Lab Chip **2009**; 9: 2902-8

Objectives and strategy

Through evolution nature has developed exquisite enzymes precisely adjusted to their specific function by iterative rounds of mutation and selection. Directed evolution is used to mimic natural evolution at the laboratory scale in order to find enzymes presenting new properties, with high potential not only from the fundamental point of view, by allowing studying in detail the evolvability of natural enzymes and the structure-activity relationship; but also for industrial applications. Biocatalysis is a growing field with a strong potential in synthetic chemistry applications. As natural enzymes are generally not well suited for industrial applications, their activity or selectivity needs to be engineered.

This work aims to develop tools and methods for the directed evolution of enzymes by *in vitro* compartmentalization (IVC) which allows the directed evolution of enzymatic functions in a high throughput screening format. The technique consists of an aqueous phase dispersed in an oil phase. Each aqueous picoliter or nanoliter droplet will act as a microreator in which a separated reaction can be carried out. Emulsions generated by microfluidic techniques are highly monodisperse and every droplet can be easily manipulated, incubated, analyzed or sorted based on its intrinsic properties, such as fluorescence. The drastic reduction in volume consumption compared to classical screening methods together with the automation in droplets creation and handling should probably be the two major advantages of IVC in microfluidics. We take the

advantage of working with small volumes in order to develop assays that would be too expensive in classical formats.

Penicillin G acylase of *E. coli* (PAC) will be used as model enzyme to optimize the directed evolution process in IVC, before applying it to a real case of directed evolution, where we intend to modify the specificity of the enzyme for its substrate.

The general strategy is depicted in Figure I, and consists of the creation of a library of PAC mutants that will be expressed in droplets in the presence of a small collection of substrates, bearing various acyl chains. Variants showing an activity toward one or more substrates of the collection will be sorted based on the fluorescence present in the droplets, and their substrates specificity will be characterized. Each step of the process will have to be evaluated and optimized separately before performing the fully integrated directed evolution experiment. The DNA recovery after sorting will have to be studied in detail as the strategy implies the manipulation of very little amount of DNA molecules.

We will develop two fluorogenic assays compatible with droplet-based microfluidics. The first assay is a fluorogenic assay with a highly hydrosoluble coumarin as the leaving group, especially designed to circumvent a major limitation of enzymatic assays in droplets. Indeed, fluorescent leaving groups, usually used for classical enzymatic assays which are typically poorly soluble in water, have a high propensity to be transported between compartments, drastically impairing compartmentalization. The second assay consists of an "activity fed enzyme translation assay" ("AFET"), based on the auto amplification of the *in vitro* translation of the enzyme of interest. Active enzymes will liberate an amino acid that will be subsequently consumed for the *in vitro* translation of the enzyme itself, reflecting the auto amplification nature of the assay, and for the translation of a reporter gene used for the detection. We will use phenyl acetylated methionine as sole methionine source for the IVTT coupled to the translation of a reporter green fluorescent protein (GFP).

The efficiency of the sorting will be evaluated on a model experiment. Genes encoding for the wild type PAC will be mixed to genes encoding for a less active variant, with ratio of one PAC for 200 mutants.

After validation, the strategy will be employed to engineer the substrate specificity of penicillin G acylase of *E. coli* (PAC), an industrial biocatalyst used in the β-lactam antibiotic synthesis. Mutants libraries of the enzyme will be screened for searching variants exhibiting new substrate specificities. This study intends to provide enzymes with new potential applications in biocatalysis but should also help to gain insights in the structure to activity relationships of this particular enzyme by contributing to the understanding of the biochemical determinants of the specificity of the enzyme and its evolvability. Genetic engineering of penicillin acylase also hold great potential for application at the industrial level in the quest to create biocatalysts endowed with new specificities.

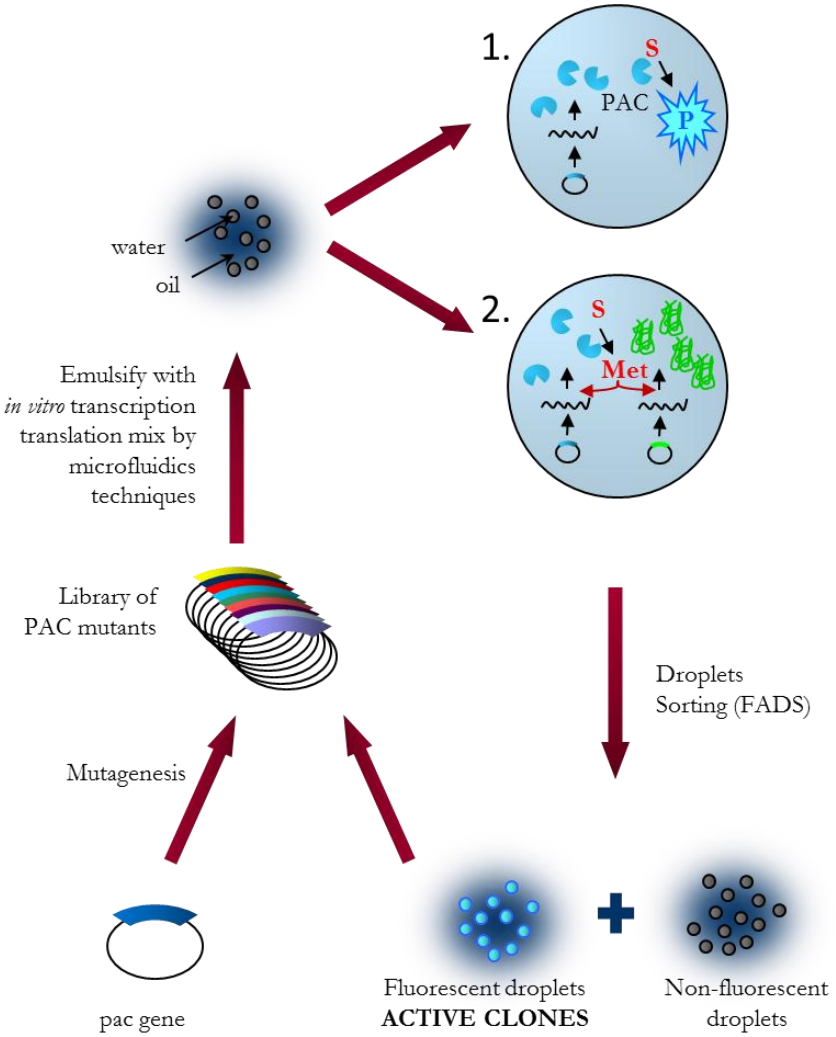


Figure I: Schematic representation of the strategy used. First, a library is created starting from the penicillin acylase gene. The library is emulsified with *in vitro* translation and transcription mix and with the substrate. The first assay requires the presence of a phenyl acetylated coumarin, a fluorogenic substrate (1.). The second assay requires the co-encapsulation of the gene of interest with a reporter gene, here a plasmid coding for GFP, and phenyl acetylated methionine substrate as sole methionine source (2.). Active enzymes will generate methionine that will be used in the translation of both the enzyme and the reporter, producing fluorescence in the droplet. Then, droplets will be sorted based on their fluorescence in order to recover the genes of interest that will be analyzed or reintroduced in a new process of selection and mutation.

RESULTS

Chapter 1: Expression vectors for penicillin acylase

Introduction

The *E. coli* Penicillin acylase (PAC) was previously studied in our laboratory and expressed using phage display vectors. Vanina Belis and Géraldine Pivato characterized PAC expression for its evolution in droplets. PAC was displayed either on the capsid of a phage or on the inner membrane of *E. coli*. Because of the low level of PAC expression and the poor genetic stability of the construction that impaired the reproducibility of the experiments, we decided to explore several expression systems and choose the most appropriate solution for directed evolution of PAC using *in vitro* compartmentalization (IVC). Four plasmids were constructed in order to evaluate *in vivo* and *in vitro* expression yields. The general scheme of our study for PAC production is outlined in Figure 1.1.

Concerning the *in vivo* expression, two vectors for PAC expression in *E. coli* were constructed and tested: "pBAD-PAC" for cytoplasmic expression and "pBAD-SRP-PAC" for periplasmic expression, for which the pac gene was cloned after an SRP-dependent signal sequence. This sequence will guide the protein to the co-translational Signal Regognition Particle (SRP) translocation pathway.[1] In *E. coli*, two translocation machineries have been described.[2,3] The first one, called SecYEG, is the most studied and involves the export of unfolded proteins through co- or post-translational translocation.[4-6] The SecYEG complex forms a hydrophilic channel. The pore is passive and requires partners to help the translocation: the ribosome for the co-translational system and the ATPase SecA for the post-translational system. In the co-translational mode, the SRP (Signal Recognition Particle) protein recognizes and binds the signal sequence emerging from the ribosome. The translation is stalled and the complex formed is transported to the translocation machinery. Translation is reactivated and transport through the translocase occurs simultaneously. This system is mainly used for the transport of hydrophobic membrane proteins, and only rarely for soluble proteins.[7] The second

translocation machinery is the Tat pathway (for "Twin Arginine Translocation").[8,9] This pathway transports proteins displaying an N-terminal "twin-arginine" leader motif (S/T)RRXFLK[10] and functions independently to other secretory systems of *E. coli*. The machinery transports folded proteins that were previously associated to their cofactors in the cytoplasm.[11] Penicillin G acylase is a periplasmic enzyme of *E. coli* which is naturally directed to the Tat pathway although its targeting peptide do not corresponds to the consensus sequence. It bears two arginine residues surrounding an asparagine.[12] We decided to test the SRP signal sequence for periplasmic expression of PAC because it was proven that compared to the post-translational system, it may enhance protein export in certain cases.[1,13,14]

Two vectors for *in vitro* expression were also constructed. *In vitro* expression systems offer new possibilities for recombinant protein production and engineering. They present significant advantages such as the complete production of protein starting from the gene in several hours; it also permits the production of proteins toxic for living cells and allows the direct manipulation of the reaction conditions.[15]

Several companies commercialize their own cell-free systems which combine the bacteriophage T7 RNA polymerase for transcription with an *E. coli* S30 extract that provides the translational machinery, enriched with limiting factors such as rare tRNAs. These systems require the presence of the gene of interest between a T7 promoter and a terminator region. The DNA can be a plasmid or a linear PCR product.[16] The pIVEX vectors family (Roche), especially optimized for cell-free protein synthesis, share an identical multiple cloning site surrounded by T7 promoter and terminator sequences and allow the use of a single cloning strategy to fuse the gene of interest with various tags at the N- or C-terminus while keeping the reading frame. Some investigations revealed the importance of the tag position that can greatly influence the yields of protein production.[17] The major problem of those systems comes from the

uncoupling of transcription and translation. T7 RNA polymerase transcription produces unprotected mRNAs that accumulate and may adopt stable secondary structures in the translation initiation region (TIR) that would decrease the rate of translation and limit the yield of produced protein.[18] It was demonstrated that some TIR can lead to high yields of protein production. We have tested the pIVEX2.3d and pIVEX2.4d, providing a His-tag fusion at the C or N-terminus.[19]

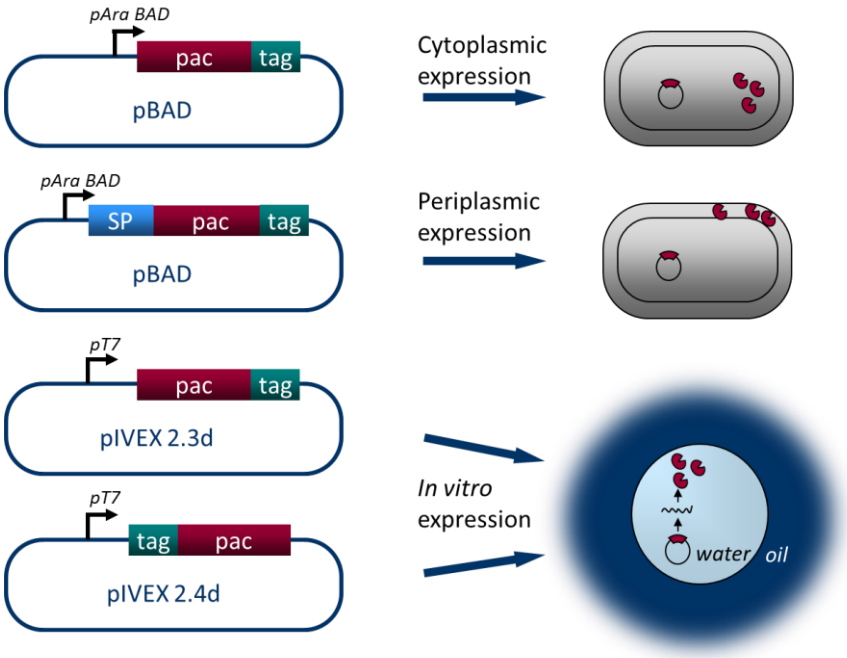


Figure 1.1: Four vectors for *in vivo* and in vitro expression of PAC were constructed and tested: pBAD-PAC for cytoplasmic expression in E. coli, pBAD-SRP-PAC where the gene is fused to an SRP-dependant signal sequence ("ss") to address the enzyme to the periplasm of E. coli, pIVEX2.3d-PAC and pIVEX2.4d-PAC, differing only by the His-tag position will allow in vitro expression of PAC, first in bulk, then in microdroplets of a water-in-oil emulsion.

Results and discussion

Penicillin acylase expression vectors construction

The *E. coli* penicillin acylase gene was amplified from DNA samples that were used in the laboratory to clone the gene into a filamentous phage vector.[20] An overlap PCR, described in detail in Figure 1.2A, was needed to remove the internal NcoI restriction site present in the gene; and NcoI, XbaI and XmaI restriction sites needed for subsequent cloning were introduced using tailed primers 1 and 2 (Figure 1.2). The removal of the internal NcoI site was confirmed by restriction analysis on the PCR fragments (Figure 1.3).

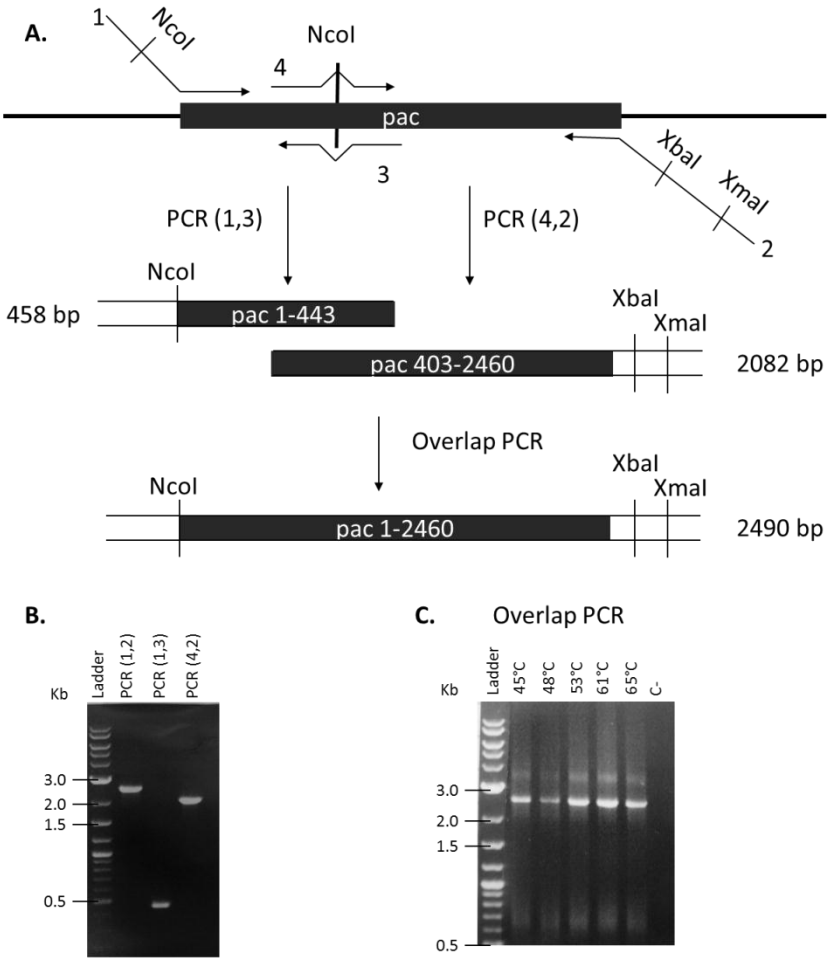


Figure 1.2: Pac gene amplification from PCR fragment of pac. A. Schematic view of the pac amplification strategy. Overlap PCR was made in order to remove the NcoI restriction site present in the pac gene. A first PCR with primers 1 and 3, and a second with primers 4 and 2 were performed. Primers 3 and 4 possess the silent mutation needed to remove the NcoI internal site (**cc**atgg → **ct**atgg, codon acc to act). As a control, a PCR with primers 1 and 2 was also conducted. PCR products were analysed on agarose gel and were of expected sizes (B). The overlap PCR was performed in the presence of primers 1 and 2 and PCR fragments "PCR(1,3)" and "PCR(4,2)" at several annealing temperatures (C). PCR products were analysed on agarose gel and presented higher yields at 53°C and 61°C. C- is a negative control PCR without DNA. The resulting amplicon possesses at its extremities the NcoI, XbaI and XmaI restriction sites needed for subsequent cloning, and the NcoI interal site was removed.

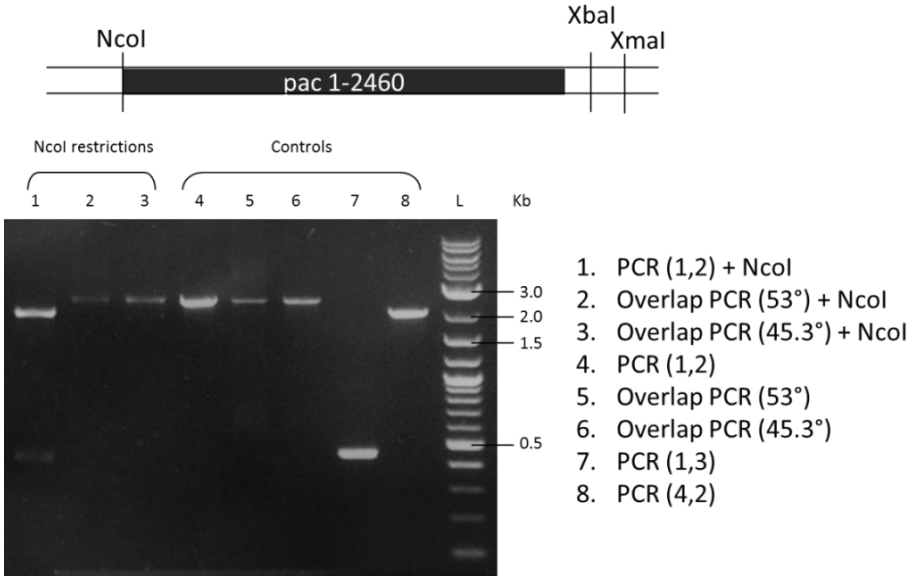


Figure 1.3: Restriction controls on the PCR fragment resulting from the overlap PCR. 1 to 3: PCR(1,2), overlap PCRs with annealing temperatures of 53°C and 45.3°C products after NcoI restriction; 4 to 6: same PCR products not digested; 7 and 8: PCR(1,3) and PCR(4,2) products not restricted.

The pac gene has then been inserted by TA cloning into pGEM-T vector and several clones were sequenced in order to identify an ORF without mutations (Figure 1.4). Following the general strategy outlined in Figure 1.4, four plasmids for *in vitro* and *in vivo* expression of PAC were successfully constructed: pIVEX 2.3d-PAC, pIVEX 2.4d-PAC, pBAD-PAC and pBAD-SRP-PAC. The last construct contain a DsbA signal sequence (named DsbAss), for the SRP-dependant translocation of the protein.[1]

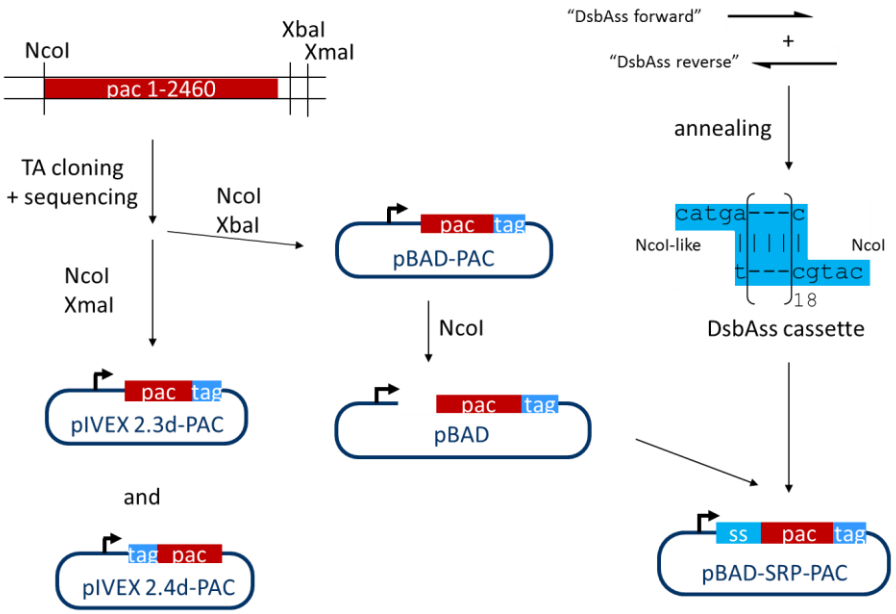


Figure 1.4: General construction strategy. First, PCR products obtained after the overlap PCR were cloned using TA-cloning and sequenced in order to start from a gene without unexpected mutations. In the first PCRs, two tailed primers (primers 1 and 2) were used to introduce NcoI restriction site upstream and XbaI and XmaI sites downstream of the gene. NcoI and XbaI were used to clone the gene into the pBAD vector for *in vivo* expression constructs, whilst NcoI and XmaI were used for the cloning into pIVEX vectors. Finally, the cloning of synthetic oligonucleotides in pBAD-PAC afforded the insertion of the SRP signal sequence. Two complementary oligonucleotides were annealed (called "DsbAss forward" and "DsbAss reverse"); the resulting cassette was purified on acrylamide gel and cloned into pBAD-PAC vector at the NcoI restriction site. The cassette possessed two 5' cohesive ends compatible with the cohesive ends generated by NcoI digestion. The cassette design is such that one NcoI site will be present after the signal sequence but not before, permitting the future cloning of other genes in the vector using NcoI and XbaI (see Figure 1.4 or the experimental section for sequence details). The cassette was ligated into the pBAD-PAC vector, and the ligation product was transformed in the Top10 E. coli strain (Invitrogen). 10 clones were picked up randomly; a PCR of the region of cloning was performed, followed by NcoI restriction in order to sequence clones with the DsbAss cassette inserted in the right position.

In vivo expression of penicillin acylase

pBAD vectors were electroporated in Top10 *E. coli* strain (Invitrogen). As shown in Figure 1.5, SDS page analysis of cytoplasmic and periplasmic extracts and insoluble fractions after arabinose induction confirmed the presence of mature penicillin acylase in the periplasmic extracts for the pBAD-SRP-PAC construct since the β subunit is visible (65 kDa, upper dot on Figure 1.5). A second band between 20 and 25 kDa should correspond to the α subunit (20.5kDa, lower dot on Figure 1.5). Penicillin acylase precursor (85 kDa) is present in insoluble fractions of both constructions. Because of the complexity of the cell composition, the presence of mature PAC in the cytoplasmic extracts is hardly detectable on gel.

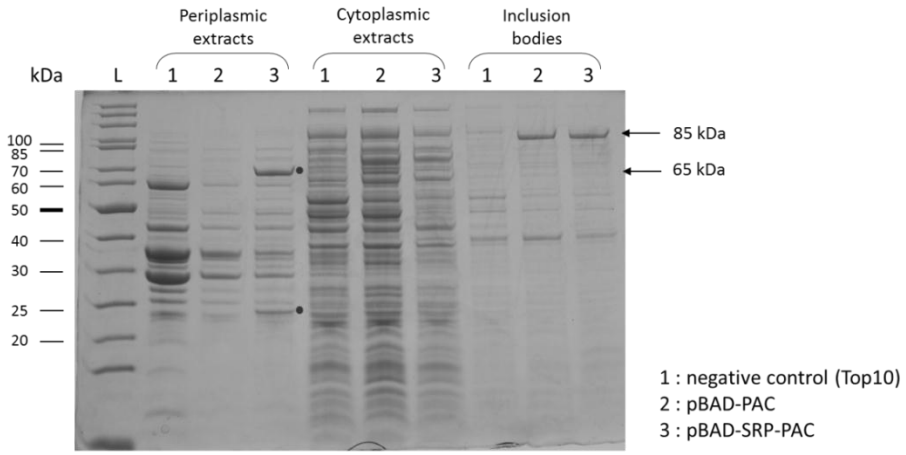


Figure 1.5: SDS-page gel of different proteic fractions (periplasmic and cytoplasmic extracts, inclusion bodies) of E. coli Top10 strain without expression vector (negative control, 1), with pBAD-PAC (2) and pPAD-SRP-PAC (3). Dots mark bands of interest (65 kDa and 25 kDa) corresponding to the a and β subunits of PAC.

Active penicillin acylase concentrations in soluble fractions and in the supernatant were evaluated using NIPAB hydrolysis in standard conditions. The initial velocity was used to determine the PAC concentration in the samples. Results obtained are presented in Figure 1.6 and confirmed that, for the cytoplasmic expression construct, active PAC is present only in the

cytoplasmic extracts whereas for the periplasmic expression construct, active PAC is detected in the periplasmic extracts and slightly in the supernatant.

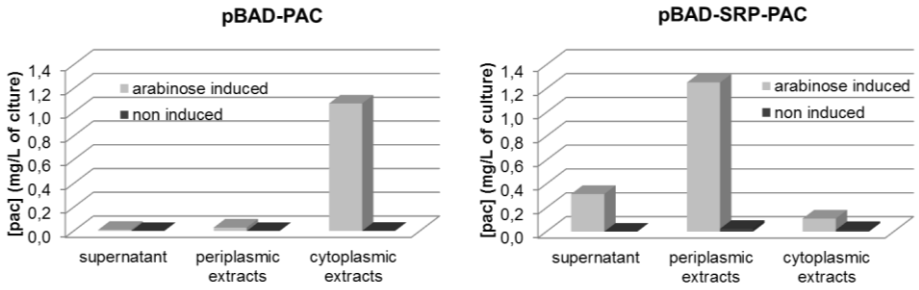
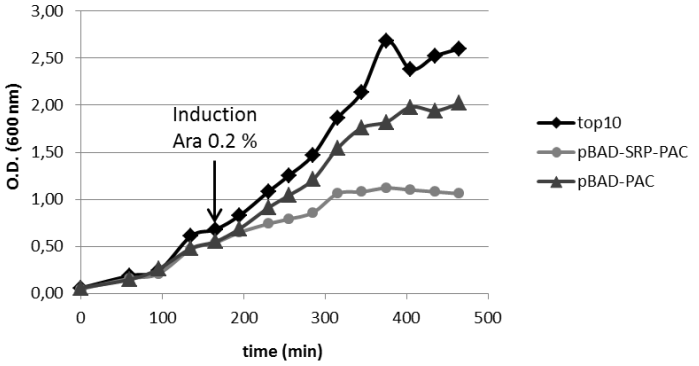


Figure 1.6: Quantification of PAC concentration in the different proteic fractions calculated by enzymatic assays on NIPAB.

The potential toxicity of the constructs was also evaluated. Liquid cultures were inoculated with pBAD-PAC or pBAD-SRP-PAC transformants and shaken at 37°C. Arabinose (0.2 %) was added when the OD reached 0.6 to induce PAC production. The OD at 600 nm of each culture was monitored during 8 hours. Results displayed in Figure 1.7 showed first that Top10-pBAD-SRP-PAC cells do not grow anymore after induction (spheres, light grey curve). The pBAD-PAC construct growth (triangles, dark grey curve) slowed down compared to the control culture (Top10 cells, squares, black curve) but continued to increase. In parallel, bacteria from the culture were plated on LB agar Tet-Ara (inducing medium) and Tet-Glu (repressing medium) just prior to and after 2.5 h arabinose induction. Prior to induction, the same level of cfu (colony forming unit) is observed in the three samples under inducing and repressing conditions on solid medium. Colonies growing under inducing conditions are however of small size. This suggests that the mortality on solid inducing medium is low but that the division rate is reduced. We cannot exclude that mortality is appearing in the colony a few hours after plating. However, after 2.5 h of induction in liquid media, an important mortality is observed for the bacteria transformed with the periplasmic construct: a 10-fold difference in cfu is observed between periplasmic and cytoplasmic constructs

on repressing solid medium, and a 1000-fold difference was observed on inducing solid medium.

In summary, a high level of toxicity is observed for the periplasmic expression system in liquid cultures that renders its application to directed evolution of PAC challenging because it could enhance unwanted biases during screening or selection processes. The toxicity is due to the SRP signal sequence as similar behavior (i.e. evolution of OD of liquid culture, small colonies on inducing plates) is observed with other proteins or peptides exported with the same signal sequence.



	Before induction		After 2.5h of induction	
	Tet-Ara (cfu/mL.OD)	*Tet-Glu (cfu/mL.OD)*	*Tet-Ara (cfu/mL.OD)*	*Tet-Glu (cfu/mL.OD)*
top10	$3.0.10^8$	$3.0.10^8$	$1.6.10^8$	$1.9.10^8$
pBAD-SRP-PAC	$5.4.10^8$	$6.0.10^8$	$1.6.10^5$	$1.1.10^7$
pBAD-PAC	$9.9.10^8$	$4.4.10^8$	$1.5.10^8$	$1.3.10^8$

Figure 1.7: Toxicity evaluation for pBAD-SRP-PAC and pBAD-PAC constructs. Upper part: growth curve of Top10 cells (control, squares, black), Top10 cells transformed with pBAD-SRP-PAC (preriplasmic expression, spheres, light grey) or pBAD-PAC (cytoplasmic expression, triangles, dark grey). OD (600 nm) was followed over 8 hours. Lower part: cell titration on LB agar plates (Tet-Ara as induction medium and Tet-Glu as repression medium). Samples of each culture were taken and titrated right before induction and 2.5 h after induction.

Our directed evolution strategy will require the presence of enzyme in microcompartments and synthetic fluorogenic substrate will have to access it. It

would therefore be more convenient to express the enzyme directly in the compartment or in the periplasm of a compartmentalized bacteria, since small substrate (<600 Daltons) can usually reach the periplasm freely. In order to perform the assay in good conditions, cytoplasmic expression would necessitate an additional step to lyse the cells, and will render the screening assay less straightforward. Consequently, considering the toxicity of periplasmic expression, *in vitro* expression would be the most appropriate solution to evolve penicillin acylase in droplets.

Penicillin acylase in vitro expression

The pIVEX vector family was designed by Roche for high-level *in vitro* transcription and translation applications. The *in vitro* expression yields reached by pIVEX2.3d-PAC and pIVEX2.4d-PAC vectors were evaluated using the RTS100 HY *E. coli* kit (Roche). IVTT mixes containing the plasmids in the recommended concentration (500 ng of DNA for 25 µL of reaction volume) were incubated at 37°C for four hours. Samples were analyzed by SDS-page together with the negative control (IVTT mix without DNA) and the positive control (IVTT of the control DNA coding for GFP provided by Roche). In the third lane of the gel (Figure 1.8), a band corresponding to GFP (27 kDa, dot) is observed. Concerning the two samples, the α and β subunits of PAC are not clearly observable. The β subunit (65 kDa) is slightly visible in the zone marked by the black dot.

Active PAC concentrations were estimated by chromogenic assay (using NIPAB). As shown in Figure 1.9, the yield of active PAC production is about three times higher with pIVEX2.4d-PAC (17 mg/µL) than with pIVEX2.3d-PAC(6 ng/µL). This first experiment indicates that pIVEX2.4d with the His-tag fusion at the N-terminus allows a better expression of PAC than its homolog with the C-terminal fusion.

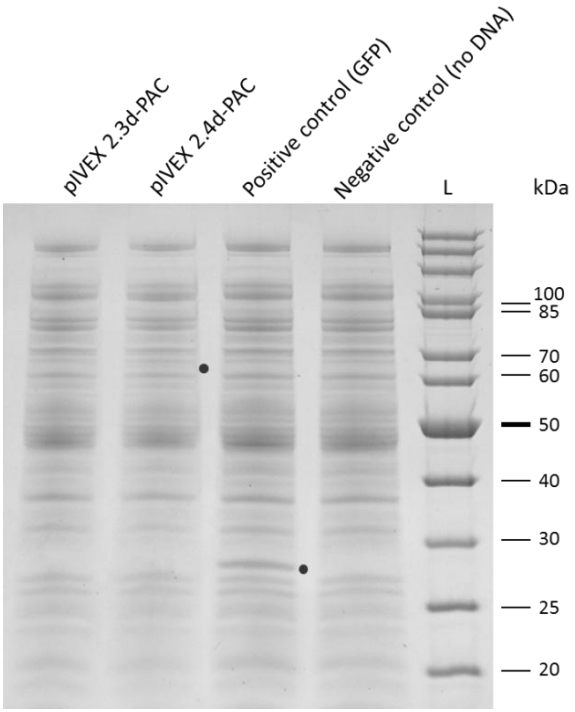


Figure 1.8: SDS-page gel of *in vitro* expression reaction mixes containing pIVEX 2.3d-PAC, pIVEX2.4d-PAC, plasmid coding for GFP (positive control provided with the kit) and without DNA (negative control). The α subunit is 20 kDa, the β subunit 65 kDa and the GFP is 27 kDa.

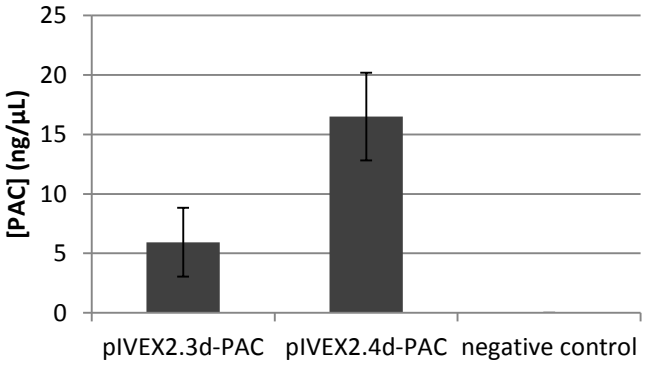


Figure 1.9: Quantification of PAC concentration in the different IVTT reactions calculated by enzymatic assays on NIPAB. Error bars represent the standard deviations from the mean value of three measurements.

Six other IVTT commercial kits were tested on pIVEX2.4d-PAC. All reactions were incubated 4 h at 37°C and PAC activity was then evaluated by NIPAB hydrolysis. As shown on Figure 1.10, the RTS100 HY *E. coli* kit from Roche reaches the highest yields of PAC expression together with the Qiagen and Invitrogen kits. We favored the RTS100 HY *E. coli* kit that provides the methionine separated from others amino acids. This will be useful for the development of a new enzymatic assay that will be described in detail in chapter 4.

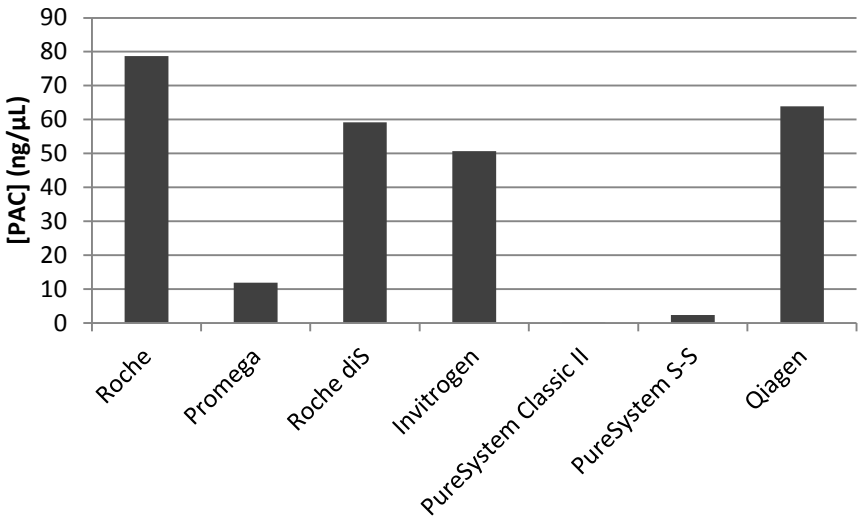


Figure 1.10: PAC concentration obtained from the IVTT of pIVEX2.4d-PAC with several commercial IVTT kits, determined kinetically on NIPAB.†

In order to follow the rate of expression of penicillin acylase in IVTT, a reaction was incubated at 37°C for 4 hours, and samples were taken after 5, 15, 30, 60, 150 and 240 minutes. Mature penicillin acylase quantities in each sample

† IVTT kits used: RTS100 HY *E. coli* kit (Roche), RTS100 HY *E. coli* Di Sulfide kit (Roche), 30S circular DNA kit (Promega), Express Way System (Invitrogen), PureSystem Classic II and PureSystem S-S (PureSystem), Easy Express (Qiagen).

were evaluated by chromogenic assay. By measuring initial slopes of product appearance, PAC concentrations were calculated and plotted on the graph bellow (Figure 1.11, light grey: initial slope). Enzymatic activity was observed after 15 minutes of incubation. Interestingly, we noticed that slopes corresponding to samples incubated for 15, 30 and 60 min became steeper upon time. A second calculation from those maximal slopes gave a second concentration value, plotted in dark grey on Figure 1.11. This indicates that the concentration of active PAC increases in the cuvette. This phenomenon could be explained by several hypotheses. IVTT could continue to synthesize the protein, even if the IVTT mix was diluted 40 times prior to measurement. Ribosomes that started the translation before dilution could also be able to continue the translation even if the tRNAs supply is slowed down. Finally, even if IVTT stopped completely upon dilution, the majority of the synthetized proteins can still be in the pre-protein form and need to maturate to give the active hetero-dimeric enzyme. The first hypothesis could be verified by adding RNAse into the cuvette. If the phenomenon is not present anymore, it will confirm that *in vitro* translation occurs during the chromogenic assay. The question could also be elucidated by following the IVTT reaction by Western Blot using an anti His-tag antibody or by incorporating radioactive [35S]-methionine in the IVTT. The appearance of the precursor and the mature enzyme could then be followed on gel.

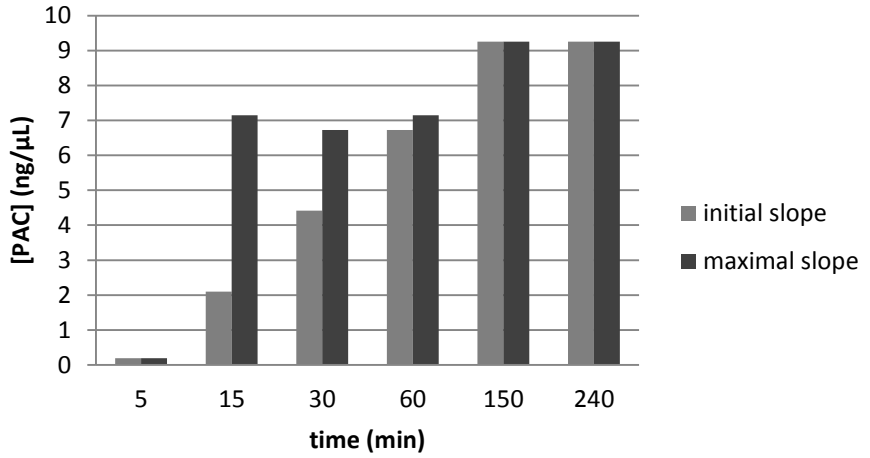


Figure 1.11: Quantification of PAC activity during incubation of IVTT reactions at 37°C for the specified time.

It should be mentioned that large variations in *in vitro* expression level were observed between experiments in this chapter and also in chapter 4. Those variations might be due to the fact that IVTT mixes vary from batch to batch and are very sensitive to temperature and handling. Even if kits are aliquoted and mixes are thawed no more than two times (first for aliquoting and second for the experiment), other variations such as preparation time of the samples could influence the quality of the mixes. In further experiments, attention has been given to always compare expression values between samples of the same set of experiments (same day, same master mix).

Conclusion and future work

Four expression vectors were successfully constructed and PAC expression was assayed both under *in vivo* and *in vitro* conditions.

Concerning *in vivo* expression, the cytoplasmic expression gave convincing results but remained the least convenient expression system for *in vitro*

compartmentalization applications. Indeed, to perform an enzymatic assay in droplets coupled with cytoplasmic expression, the bacteria would have first to be harvested in order to produce a sufficient amount of enzyme and then be lysed in order to release the enzyme in the media to perform the assay. In this case, a two-step experiment would be needed. The periplasmic expression would enable elimination of the lysis step but we observed toxicity upon expression. Toxicity should be avoided as it could enhance variations in expression levels between variants and generate bias in the experiments. It has been reported in the literature that the SRP signaling pathway is suitable for expression of insoluble proteins or membrane proteins[6] but might not be suitable for the over-expression of soluble proteins. It could be interesting to test other signal sequences that would address the protein to the periplasm via the post-traductional Sec pathway or the Tat pathway. Moreover, it has been reported that lowering the growth temperature to 17°C leads to a 5 fold increase in periplasmic levels of penicillin acylase, while increasing the temperature to 37°C promotes cell lysis upon over-expression (Alkema W. unpublished results, thesis). Hence, toxicity of the "SRP" periplasmic expression could have been tested at lower temperatures.

As an interesting area for future investigation, high quantities of proPAC were observed in both periplasmic and cytoplasmic constructs. Purification of the denaturated proPAC in guanidine hydrochloride could help in the study of the enzyme maturation, a process which is not yet fully understood.

The pac gene was cloned into two pIVEX vectors especially designed for *in vitro* expression of proteins. The N-terminal His-tag fusion gave the highest PAC expression yields. The vector pIVEX2.4d-PAC was chosen as an expression system for *in vitro* compartmentalisation applications. For simplicity, the chosen expression vector pIVEX2.4d-PAC will be named pPAC in the following chapters.

References

1. Steiner D, Forrer P, Stumpp MT, Pluckthun A; *Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display*. Nat Biotechnol **2006**; 24: 823-31

2. Driessen AJ; *SecB, a molecular chaperone with two faces*. Trends Microbiol **2001**; 9: 193-6

3. Fekkes P, Driessen AJ; *Protein targeting to the bacterial cytoplasmic membrane*. Microbiol Mol Biol Rev **1999**; 63: 161-73

4. Luirink J, Sinning I; *SRP-mediated protein targeting: structure and function revisited*. Biochim Biophys Acta **2004**; 1694: 17-35

5. Koch HG, Moser M, Muller M; *Signal recognition particle-dependent protein targeting, universal to all kingdoms of life*. Rev Physiol Biochem Pharmacol **2003**; 146: 55-94

6. Manting EH, van Der Does C, Remigy H, Engel A, Driessen AJ; *SecYEG assembles into a tetramer to form the active protein translocation channel*. EMBO J **2000**; 19: 852-61

7. Ulbrandt ND, Newitt JA, Bernstein HD; *The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins*. Cell **1997**; 88: 187-96

8. Robinson C, Bolhuis A; *Tat-dependent protein targeting in prokaryotes and chloroplasts*. Biochim Biophys Acta **2004**; 1694: 135-47

9. Fisher AC, DeLisa MP; *A little help from my friends: quality control of presecretory proteins in bacteria*. J Bacteriol **2004**; 186: 7467-73

10. Muller M, Klosgen RB; *The Tat pathway in bacteria and chloroplasts (review)*. Mol Membr Biol **2005**; 22: 113-21

11. Palmer T, Sargent F, Berks BC; *Export of complex cofactor-containing proteins by the bacterial Tat pathway*. Trends Microbiol **2005**; 13: 175-80

12. Ignatova Z, Hornle C, Nurk A, Kasche V; *Unusual signal peptide directs penicillin amidase from Escherichia coli to the Tat translocation machinery*. Biochem Biophys Res Commun **2002**; 291: 146-9

13. Deschuyteneer G, Garcia S, Michiels B, Baudoux B, Degand H, Morsomme P, Soumillion P; *Intein-mediated cyclization of randomized peptides in the periplasm of Escherichia coli and their extracellular secretion*. ACS Chem Biol **2010**; 5: 691-700

14. Thie H, Schirrmann T, Paschke M, Dubel S, Hust M; *SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in E. coli*. N Biotechnol **2008**; 25: 49-54

15. Spirin AS; *High-throughput cell-free systems for synthesis of functionally active proteins*. Trends Biotechnol **2004**; 22: 538-45

16. Betton JM; *Cloning vectors for expression-PCR products*. Biotechniques **2004**; 37: 346-7

17. Betton JM; *High throughput cloning and expression strategies for protein production*. Biochimie **2004**; 86: 601-5

18. Simmons LC, Yansura DG; *Translational level is a critical factor for the secretion of heterologous proteins in Escherichia coli*. Nat Biotechnol **1996**; 14: 629-34

19. Roge J, Betton JM; *Use of pIVEX plasmids for protein overproduction in Escherichia coli*. Microb Cell Fact **2005**; 4: 18

20. Shi YF, Soumillion P, Ueda M; *Effects of catalytic site mutations on active expression of phage fused penicillin acylase*. J Biotechnol **2009**; 145: 139-42

Chapter 2: New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays and directed evolution in droplet-based microfluidic applications

Gabrielle Woronoff, Abdeslam El Harrak, Estelle Mayot, Olivier Schicke, Oliver Miller, Patrice Soumillion, Andrew Griffiths and Michael Ryckelynck

(Published in *Analytical Chemistry*, 2011, **83**, 2852-7)

Abstract

Droplet-based microfluidics is a powerful tool for biology and chemistry as it allows the production and the manipulation of picoliter-size droplets acting as individual reactors. In this format, high-sensitivity assays are typically based on fluorescence, so fluorophore exchange between droplets must be avoided. Fluorogenic substrates based on the coumarin leaving group are widely used to measure a variety of enzymatic activities, but their application in droplet-based microfluidic systems is severely impaired by the fast transport of the fluorescent product between compartments. Here we report the synthesis of new amidase fluorogenic substrates based on 7-aminocoumarin-4-methanesulfonic acid (ACMS), a highly water-soluble dye, and their suitability for droplet-based microfluidics applications. Both substrate and product had the required spectral characteristics and remained confined in droplets from hours to days. As a model experiment, a phenylacetylated ACMS was synthesized and used as a fluorogenic substrate of *E. coli* penicillin G acylase. Kinetic parameters (k_{cat} and K_M) measured in bulk and in droplets on-chip were very similar, demonstrating the suitability of this synthesis strategy to produce a variety of ACMS-based substrates for assaying amidase activities both in microtiter plate and droplet-based microfluidic formats.

Introduction

Droplet-based microfluidics is becoming a powerful screening tool for both biology and chemistry. It allows the production of highly monodisperse picoliter droplets, each of which functions as an independent microreactor. In addition, serial operations can be performed on these droplets after their formation such as incubation, droplet fusion and fluorescence-based activated sorting (recently reviewed in[1]). The high frequencies (kHz) at which the operations can be performed together with the small volume of the reactors make this technology well-suited for high-throughput screening in fields as diverse as small molecules synthesis,[2] drug discovery[3,4] and directed evolution.[5] These applications typically use fluorescence-based assays, which necessitate the fluorescent product remaining confined in droplets until the assay readout. However, to date, two mechanisms challenging the confinement of small chemicals have been described[6,7]: the compound (i) can partition into and diffuse through the carrier oil[8,9] or (ii) can be exchanged between droplets through micellar transport.[10,11] In addition, one can predict that droplet size might influence compound leakage since the smaller the droplets, the higher the ratio of surface a to volume v ($a/v = 3/r$, where r is the radius of a spherical droplet). Nevertheless, leakage from droplets can be limited by the use of fluorinated oils, in which non-fluorinated molecules are highly insoluble[12,13] or by using additives such as bovine serum albumine (BSA), by coating the interface.[14]

Coumarins are widely used as leaving groups in fluorogenic assays of enzyme activities. Indeed, substitution of position 7 by either a hydroxyl or an amino group does not affect the fluorescence of the coumarin and allows it to be coupled through ester or amide bonds to a variety of compounds (e.g. sugars and amino-acids). The fluorescence of such conjugated coumarins is strongly diminished, but can be fully restored by freeing position 7 through the action of

enzymes such as esterases or amidases.[15-17] Therefore, enzymatic activity can be determined directly using fluorescence spectroscopy.

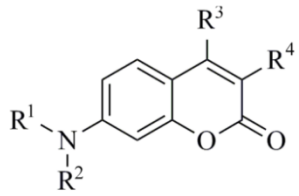
Here we describe the development of a fluorogenic assay for amidase activity based on an amino coumarin leaving group, which is compatible with both conventional microtiter plate and droplet-based microfluidic formats. Our data show that sulfonated coumarin stays confined in microdroplets under conditions (time and temperature) relevant for a typical biological assay. In addition, a set of sulfonated coumarin-based substrates were synthesized. One of them was used as a substrate of the *E. coli* penicillin G amidase and presented all the required features for use in droplet-based microfluidic applications: (i) presence of a single amide function, (ii) high signal/background ratio, (iii) good retention properties in droplets of both the substrate and the product and (iv) similar behavior in bulk and in droplet-based microfluidic experiments.

Results and discussion

Controlling the exchange of coumarin between compartments

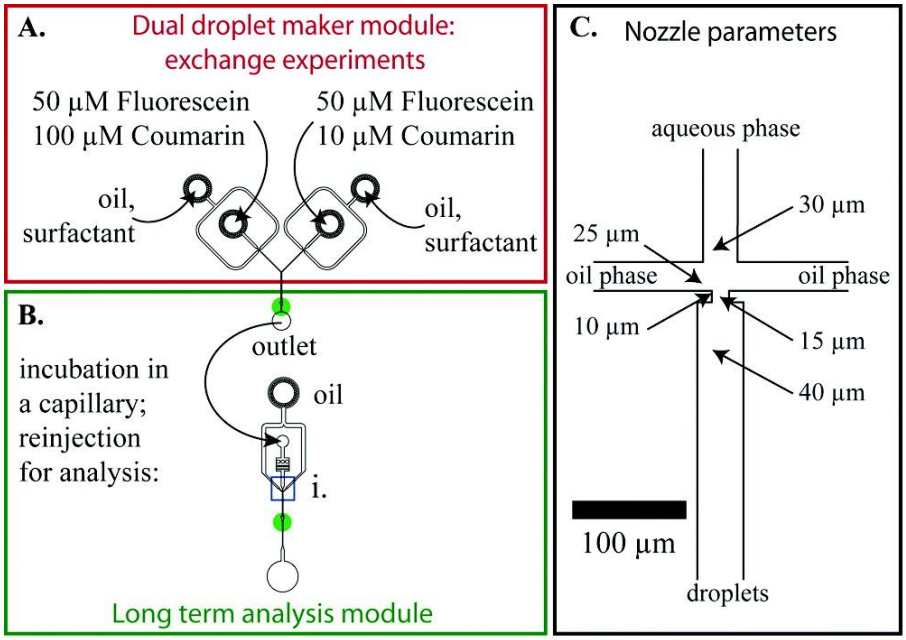
As a first experiment, we characterized the retention in droplets of 7-amino-4-methyl coumarin (**1**), a typical reaction product of amidase activity (Table 2.1). To this end, two populations of droplets containing a fixed concentration of 50 μM of fluorescein (an internal reference) and either 10 μM or 100 μM amino coumarin were produced on a dual droplet maker module (Figure 2.1A) using a fluorinated carrier oil containing EA surfactant (3 % w/w). The populations were mixed at the cross section showed in Figure 3.1 box iii. and the fluorescence intensities of individual droplets were monitored over short time ranging from 2.5 to 150 seconds using the short term analysis module (Figure 2.1E). While the two populations presented coumarin

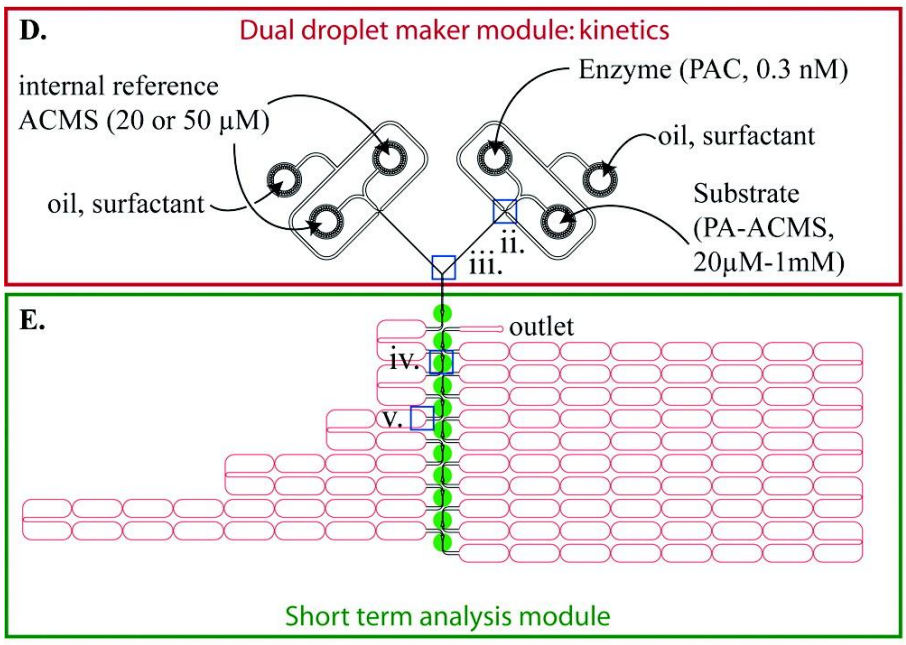
fluorescence signals differing by the expected order of magnitude directly after their formation (0 seconds on Figure 2.2A), this difference progressively decreased after 2.7 and 5.3 seconds and both populations of droplets were undistinguishable after only 10 seconds (Figure 2.2A), without significant change in the background fluorescence of the oil. Considering the low solubility of small organic molecules in fluorinated oils,[12,13] the exchange between droplets is likely be the result of a micellar transport of the dye.[10,11] The short retention time of the amino coumarin in droplets compromises its use for any biological assay in droplet-based microfluidic applications.



no.	R^1, R^2	R^3	R^4	logD	$t_{1/2}$
1	H, H	CH_3	H	1.14	2.1 ± 0.1 s
2	Et, Et	H	CO_2^-	-1.09	19.4 ± 6.4 s
3	H, H	CH_3	$CH_2CO_2^-$	-2.59	2.8 ± 1 day
4	H, H	$CH_2SO_3^-$	H	-4.82	> 5 day

Table 2.1: Characterization of the exchange of coumarin variants between droplets. Half-life of retention of coumarins in water-in-fluorinated oil emulsions. Distribution coefficient values (logD), were predicted at pH 7 and 25°C using Advanced Chemistry Development (ACD/Labs) Software V11.02 (© 1994-2011 ACD/Labs).


A.
Dual droplet maker module:
exchange experiments
50 µM Fluorescein
100 µM Coumarin
50 µM Fluorescein
10 µM Coumarin
oil,
surfactant
oil,
surfactant
B.
incubation in
a capillary;
reinjection
for analysis:
outlet
oil
i.
Long term analysis module
C.
Nozzle parameters
aqueous phase
30 µm
25 µm
oil phase
oil phase
10 µm
15 µm
40 µm
100 µm
droplets


D.
Dual droplet maker module: kinetics
internal reference
ACMS (20 or 50 µM)
Enzyme (PAC, 0.3 nM)
oil, surfactant
oil, surfactant
Substrate
(PA-ACMS,
20µM-1mM)
ii.
iii.
E.
outlet
iv.
v.
Short term analysis module

F.


i.
ii.
iii.
iv.
v.

Figure 2.1: Microfluidic designs schemes. **A.** For leakage experiments, two populations of droplets were created using the dual droplet maker module (details of the nozzle are shown in **C**). Emulsions were analyzed by using either the long-term analysis module **B** (in this case the emulsion is collected in a capillary, incubated and reinjected for analysis) or the short-term analysis module **E**. Depths of the channels were 20 and 75 μm (color-coded on the figure: 20 μm = black, 75 μm = red). Fluorescence intensities of the droplets were measured in the 20 μm-deep channels. Using the short-term analysis module **E**, fluorescence could be read at the beginning and during the incubation process, with the time of incubation being determined by the flow rates and the volume of the channels. In the case of the long-term incubations, fluorescence was read by using a reinjection device (long-term analysis module **B**). Measurement points where the laser was focused are indicated by green spots. **F**. Pictures of the droplet reinjection (i.), droplet formation (ii.) droplet mixing (iii.), measurement point (iv.) and droplet incubation (v.) Measurement points are shown as blue squares on **B**, **D** and **E**.

It has been previously reported[14] that, when using a hydrocarbon carrier oil, the rate of exchange of fluorescent compounds between droplets correlates inversely with their hydrophilicity. Even though this observation was not necessarily transposable to our system using fluorinated carrier oil, we decided to determine if a link existed between the hydrophilicity of compounds and their retention in droplets. The assay described above was repeated with coumarin variants of increasing hydrophilicity. In addition to the short term analysis, a long term analysis was performed by collecting the droplets in a glass-capillary and reinjecting them in the long term analysis module (Figure 2.1B). The partition coefficient (logD) is a measure of the differential solubility of a compound between two solvents (water and octanol) and therefore reflects the hydrophilicity of the compound. Interestingly, the addition of a carboxylic group on a coumarin (compound **2**) increased its retention time by an order of magnitude as the half-life of exchange was ~20 seconds (Table 2.1). Further increasing the hydrophilicity of the coumarin in compound **3** (7-amino-4-methyl-3-coumarinylacetic acid), featuring both a carboxylic group and a primary amine group, further increased the half-life of exchange of the coumarin (Table 2.1) to ~3 days at 37°C.

We therefore decided to synthesize and test the retention of the highly water-soluble 7-aminocoumarin-4-methanesulfonic acid (ACMS), initially

developed by Sato *et al.*.[18] The original protocol was modified according to known procedures[19-21] (Figure 2.3A). First, the addition of ethylchloroformate in diethyl ether to the amino phenol afforded the protected amino phenol **6** with 77 % yield. The coumarin **7** was then obtained by Pechmann condensation[22] of **6** with ethyl 4-chloroacetoacetate in concentrated sulfuric acid with 66 % yield. Sulfonation in a mixture of acetone/water (60/40) gave **8** with 94% yield. Finally, **8** was deprotected under acid conditions, yielding ACMS (**4**). The retention of ACMS in droplets was tested as described above over short (hours) and long (days) incubation periods, at both room temperature (23°C) and 37°C. Even after 5 days of incubation at 37°C, the fluorescence of both populations was essentially unchanged (Figure 2.2B) showing that no detectable exchange occurred between droplets over the course of the experiment. Moreover, the examination by fluorescence microscopy of an emulsion conserved 3 weeks at room temperature didn't reveal any significant leakage of the ACMS between droplets (Figure 2.2C). Interestingly, the retention of coumarin derivatives in droplets directly correlated with their distribution coefficient (logD), therefore their hydrophilicity (Table 2.1), despite the fact that LogD does not, *a priori*, give any information about the distribution coefficient between an aqueous and a fluorinated phase.

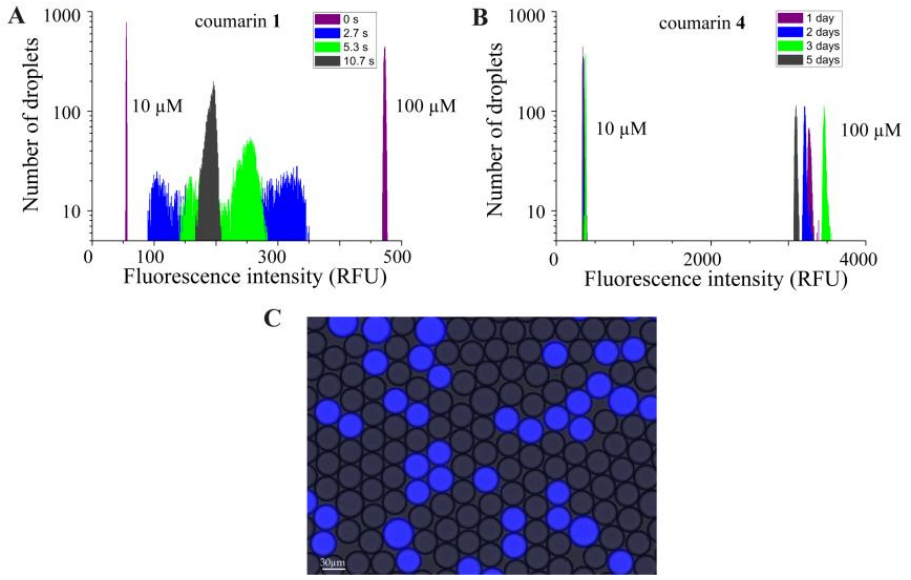


Figure 2.2: Characterization of the exchange of coumarin variants between droplets. **A.** An emulsion containing two populations of droplets of **1** at 10 or 100 µM was incubated and analyzed at 23°C on chip. The fluorescence intensities of 5000 droplets were analyzed at different measurement times. **B.** An emulsion containing two populations of droplets containing ACMS at 10 and 100 µM was incubated at 37°C and the fluorescence intensities of 5,000 droplets were analyzed after 1, 2, 3 and 5 days of incubation. The 11% variation of the fluorescent signal intensity is attributed to slight variations in the alignment of the optical setup from one day to another. **C.** Fluorescence microscopy image of an emulsion composed of droplets containing either 10 or 100 µM of ACMS. Two populations of droplets containing ACMS at 10 and 100 µM were generated, mixed and conserved for 3 weeks at room temperature prior imaging.

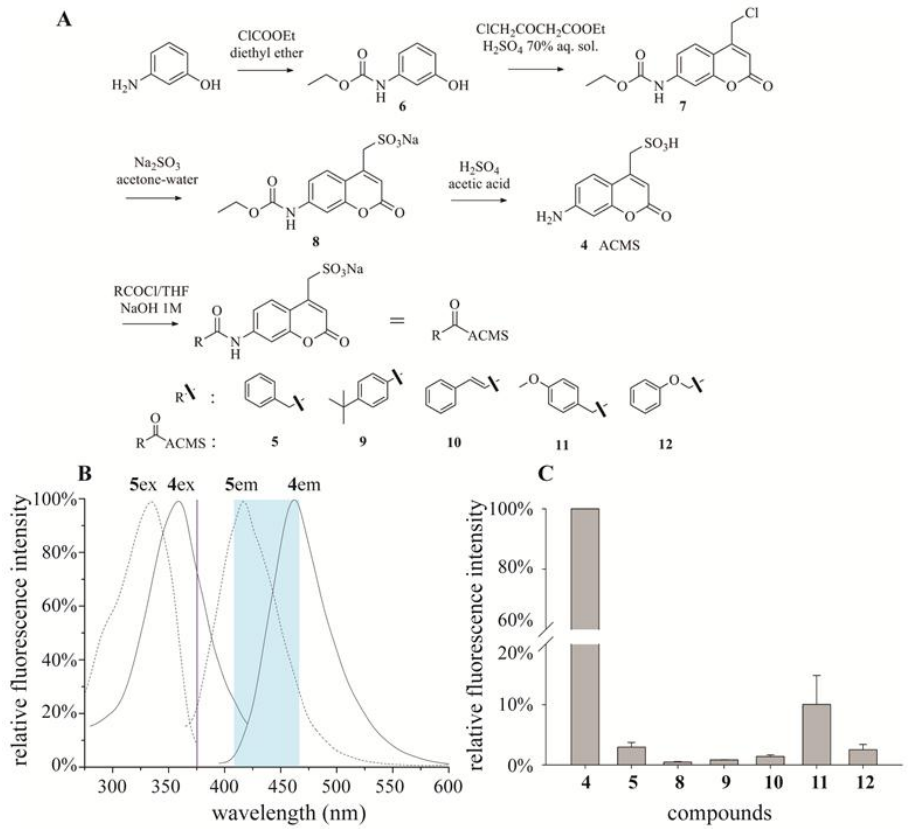


Figure 2.3: Synthesis and characterization of ACMS and its variants. **A.** Synthesis of the fluorophore 7-aminocoumarin-4-methane-sulfonic acid (ACMS, **4**) and related substrates (Phenylacetyl-ACMS **5**, 4-tert-butylbenzoyl-ACMS **9**, cynamoyl-ACMS **10**, 3-methoxyphenylacetyl-ACMS **11**, phenoxyacetyl-ACMS **12**). **B**. Fluorescence spectra of ACMS (**4**) and PA-ACMS (**5**). Excitation spectra of PA-ACMS (5ex) and ACMS (4ex) were obtained by monitoring the emission of light at 416 and 460 nm, respectively. Emission spectra of PA-ACMS (5ex) and ACMS (4ex) were obtained by monitoring emission following excitation at 336 and 360 nm, respectively. While the maximum excitation and emission wavelengths for the ACMS were found to be 362 and 462 nm, respectively, they shifted to 325 and 425 nm, respectively, for the conjugated substrates. The purple line at 375 nm represents the excitation wavelength of the UV laser used and the light blue region shows the wavelengths analyzed by the photomultiplier tubes on the microfluidic station. **C.** Relative fluorescence intensity of the substrates compared to ACMS (**4**). The molecules were excited at 375 nm and the light emitted at 450 nm was measured. The wavelengths were chosen to match the optical setup of the microfluidic station.

Synthesis and characterization of ACMS-based amidase substrates

Given the optimal retention properties of the sulfonated amino-coumarin, we decided to further explore its suitability for biological applications and devised the synthesis of new ACMS-based amidase substrates. The reactions were performed in aqueous solution and under basic conditions (Figure 2.3A). For each synthesis, the solution of acid chloride of the precursor dissolved in dry THF and a solution of sodium hydroxide were added dropwise simultaneously.[18,23] Using this procedure, phenylacetyl-ACMS (**5**), 4-tert-butylbenzoyl-ACMS (**9**), cynamoyl-ACMS (**10**), 3-methoxyphenylacetyl-ACMS (**11**), phenoxyacetyl-ACMS (**12**) were produced with yields ranging from 6.6 to 20 %. The production of this small collection of compounds (Figure 2.3A) in a single step demonstrated the suitability of the method to synthesize a variety of ACMS-based substrates.

A spectral analysis of ACMS revealed a shift of 40 nm toward the UV both in excitation and emission wavelengths shared by all conjugated molecules compared to the uncoupled ACMS (Figures 2.3B and 2.4). The combination of these shifts and the quenching induced by the coupling on the position 7 of the coumarin[15] led to strong extinction of fluorescence when excited at 375 nm: 1 to 3 orders of magnitude compared with the free ACMS (Figure 2.3C). Finally, additional experiments performed on substrate **5** in order to demonstrate that the molecule stayed perfectly confined for hours in droplets. Following the same scheme of experiments used to characterize the retention time of coumarins **1**, **2**, **3** and **4**, two populations of droplets were created, mixed and their fluorescence was analyzed over time. The first population contained 50 μM of fluorescein and 0.3 nM of PAC, while the second population contained 100 μM of **5**. No exchange has been observed after 5 hours, indicating that both the enzyme and its substrate stayed confined in their respective droplets. These results demonstrate that a variety of ACMS-based substrates can be synthesized with the optical and physical characteristics

necessary for fluorogenic amidase assays compatible with droplet-based microfluidics.

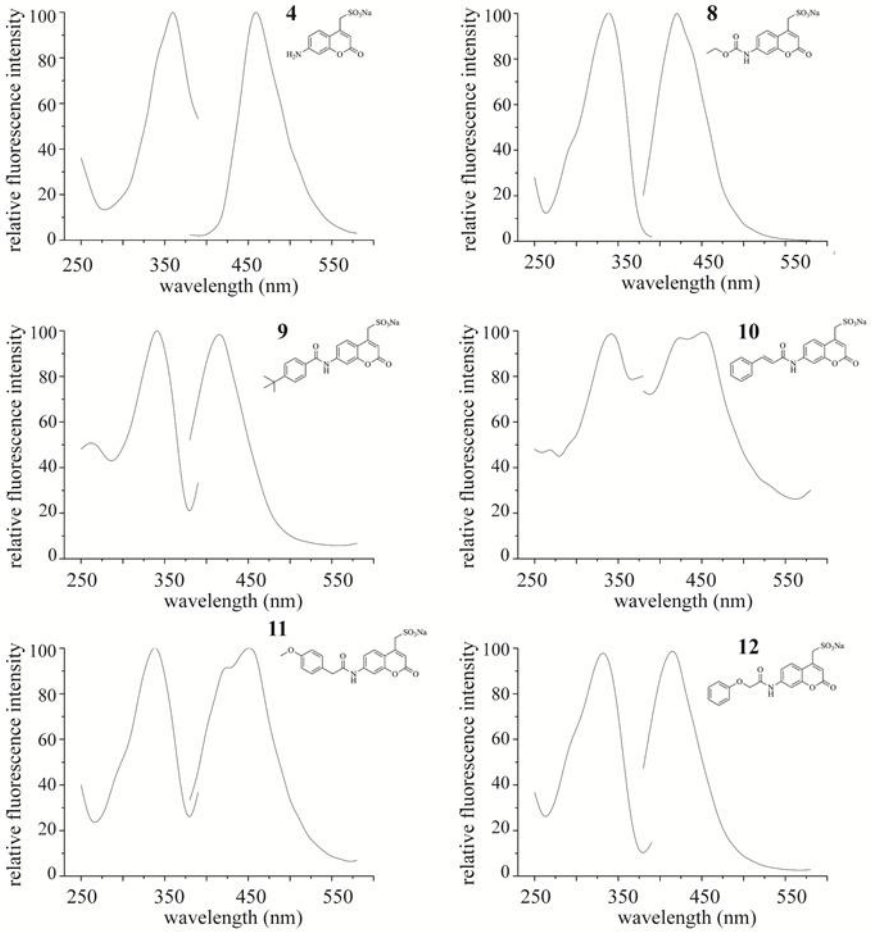


Figure 2.4: Fluorescence spectra of ACMS and ACMS-based substrates (**4**, **8**, **9**, **10**, **11** and **12**). A 10 μM solution of each compound in PBS was prepared and analyzed on a spectrofluorophotometer (SpectraMax M5, Molecular Devices). Excitation spectra (em. 416 nm) and emission spectra (ex. 336 nm) were normalized to 100 % of maximal intensity of excitation and emission.

Penicillin Acylase kinectics with PA-ACMS

Substrate **5** (phenyacetyl-ACMS or PA-ACMS) is ACMS substituted with a phenyl acetyl function. This function is the main determinant of specificity for *E. coli* penicillin G acylase (PAC).[24] While this enzyme is widely used in industry, its natural function is unknown and no fluorescence-based amidase assay compatible with droplet-based microfluidic was available. Since substrate **5** was predicted to be a fluorogenic substrate of the PAC, we monitored its hydrolysis both in a 96-well plate and in a droplet-based microfluidic device. The bulk analysis was performed by mixing 0.3 nM of PAC (sold commercially as penicillin amidase) and concentrations of PA-ACMS ranging from 20 μM to 1 mM, with the liberation of fluorescent product being monitored by a spectrofluorophotometer. As expected for a fluorogenic substrate, the transformation of PA-ACMS in fluorescent ACMS was observed only in the presence of PAC (data not shown). In addition, the kinetic constants k_{cat} and K_M (74 ± 2 s^{-1} and 67 ± 3 μM respectively) extracted from this experiment (Figure 2.5) were close to those for other known substrates of PAC: NIPAB (6-nitro-3-phenylacetamide benzoic acid), the most commonly-used chromogenic substrate, and amide-based substrates such as penicillin G and phenylacetamide for which the k_{cat} values are 15, 42 and 50 s^{-1}, and K_M values are 15, 7 and 180 μM respectively (Table 2.1).[25] We subsequently studied the suitability of substrate **5** for performing amidase assays in droplet-based microfluidic format. 0.3 nM of PAC and 20 μM to 1 mM of PA-ACMS substrate were combined by co-flow, segmented into droplets by a microfluidic nozzle, (Figures 2.1C and D), collected and incubated in a delay line (Figure 2.1E). Fluorescence intensity was measured at time points ranging from 0 to 140 s, and values were adjusted using a second population of droplets containing ACMS as an internal standard. Kinetic constants and the efficiency of the reaction were determined as above (k_{cat} = 135 ± 1 s^{-1}; K_M = 98 ± 5 μM; and k_{cat}/K_M = (1.10 ± 0.06).10^6 s^{-1} M^{-1}). Therefore, PA-ACMS can be used as an amidase fluorogenic substrate to perform assays both in microtiter plate and droplet-based microfluidic formats.

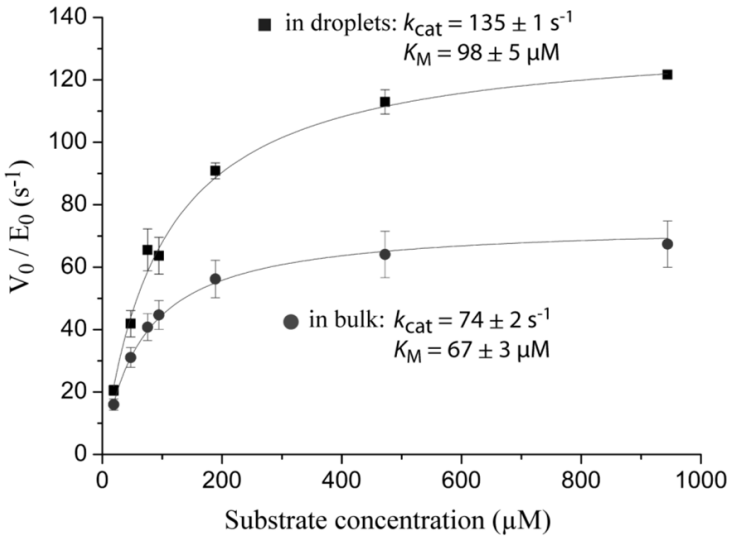


Figure 2.5: Michaelis-Menten representation of PA-ACMS hydrolysis by penicillin G Acylase in bulk (grey dots) and in droplets (black squares). V_0/E0 values (the initial velocity divided by the enzyme concentration, in s-1) were plotted as a function of substrate concentration (PA-ACMS, in µM). The non-linear fitting function is shown as a red line and the resulting constants kcat and KM are given for each format.

		k_{cat} (s^{-1})	K_M (µM)	k_{cat} / K_M (s^{-1}µM^{-1})
PA-ACMS	96-well plate	74 ± 2	67 ± 3	1.10 ± 0.06
PA-ACMS	In droplet	135 ± 1	98 ± 5	1.38 ± 0.06
NIPAB*		15	15	1
Penicillin G*		42	7	6
Phenylacetamide*		50	180	0.277

Table 1: Kinetic parameters of the hydrolysis of various substrates by penicillin acylase. Enzymatic assays for the hydrolysis of PA-ACMS (**5**, concentrations ranging from 20 µM to 1 mM have been performed in bulk (96-well plate, line 1) and in emulsion (line 2). * Kinetic parameters for the hydrolysis of penicillin acylase on NIPAB (common chromogenic substrate for penicillin acylase assays), Penicillin G and Phenylacetamide (lines 3 to 5) have been described in the work of Alkema et al.[25]

Conclusion

The work presented here demonstrates that new fluorogenic amidase substrates can be synthesized using the highly water-soluble sulfonated coumarin leaving group, ACMS. We have demonstrated the synthesis of fluorogenic substrates for penicillin G acylase (PAC) based on ACMS, but conjugating other functional groups to ACMS should also allow the production of substrates for other amidases (e.g. proteases) and the strategy could even be extended to generate fluorogenic assays for other activities with umbelliferone as fluorescent leaving group (e.g. esterase), or via an indirect release mechanism, where a primary product is converted to a fluorescent product (e.g. alcohol dehydrogenase, aldolase).[26, 27] Since there was no detectable exchange of ACMS between droplets over several days, this new class of molecules can be used for high-throughput screening in droplet-based microfluidic format with applications, for instance, in directed evolution and drug screening. This study also revealed that the exchange of small molecules between droplets in fluorinated carrier oil was, at least in case of coumarin variants, directly linked to their hydrophilicity. Therefore, it is likely that any direct (e.g. based on a fluorogenic substrate) or indirect enzyme assay could be performed in the microdroplets provided that the product of the enzymatic reaction(s) and the reporter molecules are sufficiently hydrophilic. Based on this study, functionalization of weakly hydrophilic molecules with highly dissociative salts like sulfonate or phosphate could prove to be a generic strategy to prevent exchange of molecules between droplets and render them compatible with droplet-based microfluidic systems.

Acknowledgments

This work was supported by the F.R.I.A. (Fond pour la Formation la Recherche dans l'Industrie et l'Agriculture, studentship to G.W.), the Université catholique de Louvain, the FRS-FNRS (short-term fellowships to G.W.) and by the EU FP7 ENEFP Marie Curie Initial Training Network. The authors want to thank Professor Olivier Riant, Dave van Ditmarsch, Thomas Mangeat and Christian Rick for advice and technical assistance and Raindance Technologies for providing the surfactant.

References

1. Theberge AB, Courtois F, Schaerli Y, Fischlechner M, Abell C, Hollfelder F, Huck WT; *Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology*. Angew Chem Int Ed Engl **2010**; 49: 5846-68

2. Theberge AB, Whyte G, Frenzel M, Fidalgo LM, Wootton RC, Huck WT; *Suzuki-Miyaura coupling reactions in aqueous microdroplets with catalytically active fluorous interfaces*. Chem Commun (Camb) **2009**: 6225-7

3. Brouzes E, Medkova M, Savenelli N, Marran D, Twardowski M, Hutchison JB, Rothberg JM, Link DR, Perrimon N, Samuels ML; *Droplet microfluidic technology for single-cell high-throughput screening*. Proc Natl Acad Sci U S A **2009**; 106: 14195-200

4. Miller OJ, Harrak AE, Mangeat T, Baret JC, Frenz L, Debs BE, Mayot E, Samuels ML, Rooney EK, Dieu P, Galvan M, Link DR, Griffiths AD; *High-resolution dose-response screening using droplet-based microfluidics*. Proc Natl Acad Sci U S A **2012**; 109: 378-83

5. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA; *Ultrahigh-throughput screening in drop-based microfluidics for directed evolution*. Proc Natl Acad Sci U S A **2010**; 107: 4004-9

6. Hai M, Magdassi S; *Investigation on the release of fluorescent markers from w/o/w emulsions by fluorescence-activated cell sorter*. Journal of Controlled Release **2004**; 96: 393-402

7. Sela Y, Magdassi S, Garti N; *Release of markers from the inner water phase of W / O / W emulsions stabilized by silicone based polymeric surfactants*. Journal of Controlled Release **1995**; 33: 1-12

8. Bjerregaard S, Pedersen H, Vedstesen H, Vermehren C, Söderberg I, Frokjaer S; *Parenteral water/oil emulsions containing hydrophilic compounds with enhanced in vivo retention: formulation, rheological characterisation and study of in vivo fate using whole body gamma-scintigraphy*. International Journal of Pharmaceutics **2001**; 215: 13-27

9. Pays K, Giermanska-Kahn J, Pouligny B, Bibette J, Leal-Calderon F; *Double emulsions: how does release occur?* Journal of Controlled Release **2002**; 79: 193-205

10. Cheng J, Chen J-F, Zhao M, Luo Q, Wen L-X, Papadopoulos KD; *Transport of ions through the oil phase of W1/O/W2 double emulsions*. Journal of Colloid and Interface Science **2007**; 305: 175-182

11. Wen L, Papadopoulos KD; *Visualization of water transport in W1/O/W2 emulsions*. Colloids and Surfaces A: Physicochemical and Engineering Aspects **2000**; 174: 159-167

12. Huque FTT, Jones K, Saunders RA, Platts JA; *Statistical and theoretical studies of fluorophilicity*. Journal of Fluorine Chemistry **2002**; 115: 119-128

13. Studer A, Hadida S, Ferritto R, Kim SY, Jeger P, Wipf P, Curran DP; *Fluorous synthesis: A fluorous-phase strategy for improving separation efficiency in organic synthesis*. Science **1997**; 275: 823-826

14. Courtois F, Olguin LF, Whyte G, Theberge AB, Huck WT, Hollfelder F, Abell C; *Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays*. Anal Chem **2009**; 81: 3008-16

15. Seixas de Melo JS, Becker RS, Macanita AL; *Photophysical Behavior of Coumarins as a Function of Substitution and Solvent: Experimental Evidence for the Existence of a Lowest Lying 1(n,.pi.*) State*. The Journal of Physical Chemistry **1994**; 98: 6054-6058

16. Goddard JP, Reymond JL; *Enzyme assays for high-throughput screening*. Curr Opin Biotechnol **2004**; 15: 314-22

17. Goddard JP, Reymond JL; *Recent advances in enzyme assays*. Trends Biotechnol **2004**; 22: 363-70

18. Sato E, Matsuhisa A, Sakashita M, Kanaoka Y; *Organic Fluorescent Reagents .13. New Water-Soluble Fluorogenic Amine - 7-Aminocoumarin-4-Methanesulfonic Acid (Acms) and Related Substrates for Proteinases*. Chemical & Pharmaceutical Bulletin **1988**; 36: 3496-3502

19. Atkins RL, Bliss DE; *Substituted coumarins and azacoumarins. synthesis and fluorescent properties*. J. Org. Chem. **1978**; 43: 1975-1980

20. Pianowski ZL, Winssinger N; *Fluorescence-based detection of single nucleotide permutation in DNA via catalytically templated reaction*. Chemical Communications **2007**: 3820-3822

21. Kanaoka Y, Kobayashi A, Sato E, Nakayama H, Ueno T, Muno D, Sekine T; *Organic Fluorescent Reagents .10. Multifunctional Cross-Linking Reagents .1. Synthesis and Properties of Novel Photoactivable, Thiol-Directed Fluorescent Reagents*. Chemical & Pharmaceutical Bulletin **1984**; 32: 3926-3933

22. v. Pechmann H; *Neue Bildungsweise der Cumarine. Synthese des Daphnetins. I*. Berichte der deutschen chemischen Gesellschaft **1884**; 17: 929-936

23. Heltweg B, Dequiedt F, Marshall BL, Branch C, Yoshida M, Nishino N, Verdin E, Jung M; *Subtype selective substrates for histone deacetylases*. Journal of Medicinal Chemistry **2004**; 47: 5235-5243

24. Alkema WB, Hensgens CM, Snijder HJ, Keizer E, Dijkstra BW, Janssen DB; *Structural and kinetic studies on ligand binding in wild-type and active-site mutants of penicillin acylase*. Protein Eng Des Sel **2004**; 17: 473-80

25. Alkema WB, de Vries E, Floris R, Janssen DB; *Kinetics of enzyme acylation and deacylation in the penicillin acylase-catalyzed synthesis of beta-lactam antibiotics*. Eur J Biochem **2003**; 270: 3675-83

26. Wahler D, Reymond J-L; *Novel methods for biocatalyst screening*. Current Opinion in Chemical Biology **2001**; 5: 152-158

27. Reymond JL, Fluxa VS, Maillard N; *Enzyme assays*. Chem Commun (Camb) **2009**: 34-46

Chapter 3: Droplet-based microfluidic screening of penicillin acylase variants for altered substrate specificities

Introduction

In the early 50's, it was demonstrated that protein synthesis does not require the integrity of the cell. Proteins can be synthesized by using crude extracts of cells containing all of the required machinery for translation.[1] Commonly used cell free expression systems consist of extracts from *E. coli*, rabbit reticulocytes or wheat germ. Extracts contain all molecular components for translation of RNA (70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthases, initiation, elongation and termination factors, …) and are supplemented with amino acids, ATP, GTP, energy regeneration systems (creatine phosphate and creatine phosphokinase for eukaryotic systems, or phosphoenol pyruvate and pyruvate kinase for prokaryotic systems), and other cofactors (K^+, Mg^{2+}, …).

Combined transcription-translation systems allow the protein production starting from DNA coding sequences. Typically, such systems use prokaryotic phage RNA polymerases (T7 or SP6) introduced into eukaryotic or prokaryotic extracts to produce mRNA and translated it *in situ*.[1] Since T7 RNA polymerase is much faster than the translation machinery, no real coupling between transcription and translation occurs.

In vitro expression systems present significant advantages over standard *in vivo* expression tools in protein synthesis strategies, especially for directed evolution purposes. Cloning, transformation and expression in cells can be laborious, while *in vitro* expression allows better expression of toxic proteins and reduces degradation by proteolysis. *In vitro* expression can be made starting from PCR products, requiring only the presence of a promoter (T7 promoter for example) and ribosome binding site on the DNA. *In vitro* expression systems can therefore easily be combined with all main PCR based mutagenesis techniques such as deletion or insertion of sequences, point or random mutagenesis, and fusion of different DNA fragments.[2] The rapid expression of

proteins in only several hours starting directly from a PCR fragment is one of the major advantages of such systems. Moreover, toxicity problems related to *in vivo* expression are circumvented. One important limitation of *in vitro* systems is their high sensitivity to heat. A high variation in protein production yields from one sample to the other can also be present. The rapid depletion of some essential substrates can lead to the inactivation of the system after a short incubation time (i.e. 1 or 2 hours), it can nevertheless be overcome by continuous supply of those substrates, and continuous removal of the products.[3]

Compartmentalization is believed to be essential for evolution through natural selection, and an indispensable feature in the emergence of early life on Earth. Compartments enable the linkage between the genotype and the phenotype, i.e. between the nucleic acid that can be replicated and the encoded functional trait, such as a catalytic activity. In nature, genotype and phenotype are linked by the cell. By mimicking natural cells, aqueous droplets dispersed in an oil phase can act as *in vitro* compartments and will provide the physical link needed for directed evolution of enzymes. *In vitro* compartmentalization (IVC) proposes a totally *in vitro* process that can be focused on a single gene or set of genes under controlled conditions, which is impossible in living cells.[4]

Emulsions composed of millions of aqueous droplets can be created and handled using microfluidic techniques. Droplets can be mixed, fused, incubated at various temperatures from few seconds to hours or days, thermocycled, analyzed to measure size and fluorescence, and sorted based on their fluorescence.[5,6] Microfluidic systems can be used to analyze complex and sequential biological processes in droplets. Single DNA molecules can be encapsulated in droplets and translated using *in vitro* transcription and translation mixtures (IVTT). Courtois et al. followed the *in vitro* production of GFP in droplets based microfluidics. The high yield of GFP *in vitro* expression allowed the production of GFP from single plasmids.[7]

Thanks to advances made in droplet manipulation with microfluidic devices, reactions that require several steps can be envisioned. As some proteins are hardly expressed from single genes, a DNA amplification step could be necessary before IVTT. A first emulsion will contain single copies of a gene of interest and a DNA amplification machinery (e.g. PCR or rolling circle reagents). Droplets can be incubated or thermocycled and then fused to droplets containing IVTT. Then, proteins are synthesized and an enzymatic assay can be performed in the droplet. Mazutis and coworkers developed a method for isothermal amplification of single plasmidic DNA in droplets and their fusion to droplets containing IVTT and a fluorogenic substrate to assay β-galactosidase activity.[8] When the enzymatic assay is not compatible with the IVTT conditions, such as the cotA laccase for example, an experiment with first conducting IVTT in droplets, and then fusing those droplets with droplets containing reagents for the enzymatic assay is required.[9]

In order to mimic natural evolution, compartments containing the most interesting variants of the population should be selected. Fluorogenic assays are commonly used to assay enzymatic activities in droplets. Sorting devices can be integrated into microfluidic platform, allowing the selective recovery of the most fluorescent droplets.[10]

In this work, we developed a system for the directed evolution of the penicillin acylase of *E. coli* by selection of droplets that contain enzymes produced *in vitro* from a single copy of the pac gene (Figure 3.1). Phenylacetylated 7-aminocoumarin-4-methanesulfonic acid (PA-ACMS), a fluorogenic substrate developed for microfluidic applications was initially used to design a model sorting experiment based on the activity of the wild type PAC.

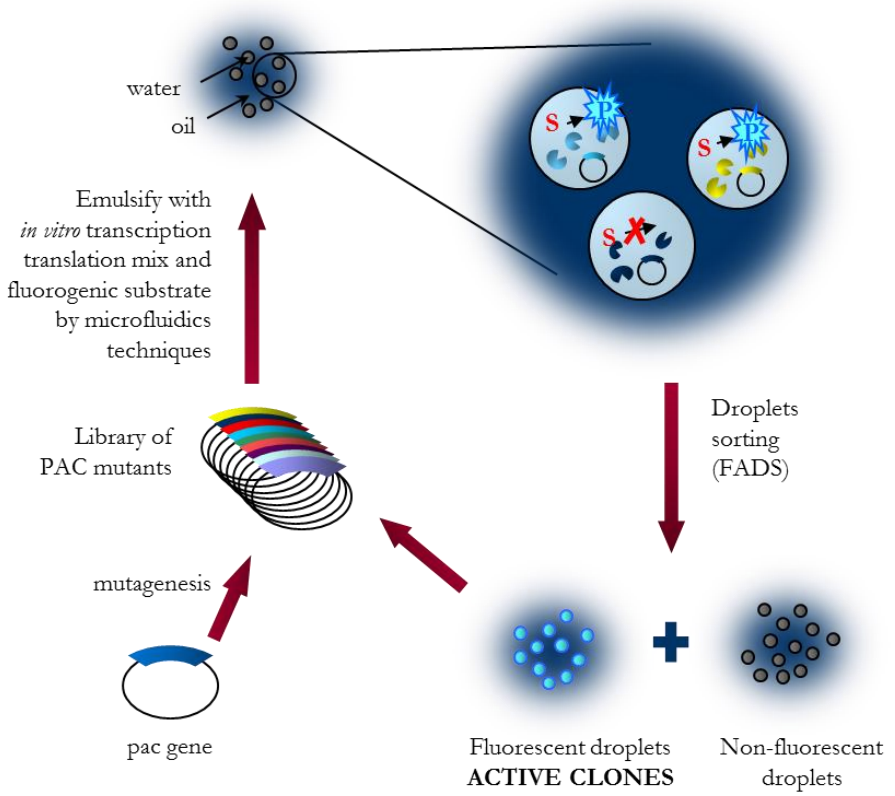


Figure 3.1: General strategy for the directed evolution of penicillin acylase. Starting with pPAC (pac gene cloned into pIVEX 2.4d) at the lower left corner of the scheme, a library of mutants is created by mutagenesis. Genes are encapsulated in microcompartments of an emulsion with IVTT mix and fluorogenic substrate. Enzymes are synthesized and assayed towards the fluorogenic substrate. Fluorescent droplets containing active variants will be sorted using fluorescent activated droplet sorting techniques (Fluorescent Activated Droplet Sorting, or FADS). The DNA of the selected variants will be recovered, amplified and analyzed. Several steps of mutagenesis and screening for activity can be performed if necessary.

The validation of the strategy requires the optimization of each step of the process. First, the *in vitro* expression and the enzymatic assay of the enzyme were characterized. As droplets containing only one DNA molecule, DNA recovery was also a key point and several methods were tested. Then the full sorting process was tested on model experiments: starting from mixtures of plasmids encoding for the wild type PAC (called pPAC) and a less active PAC mutant, F146L (called pF146L).

After validation of our strategy, libraries of variants of PAC expressed *in vitro* in droplets were tested towards a small collection of ACMS based substrates featuring various acyl groups (**8**, **9**, **10**, **11** and **12** on Figure 2.3 in chapter 2) in order to find variants that possess new substrates specificities. Since the objective was to select PAC variants that possess a widen specificity for the acyl function of the substrate, the substrates were mixed together in a single experiment. Selected variants were then characterized in detail.

Results and discussion

Penicillin acylase expression in droplets

As described in detail in chapter 1, the combination of penicillin acylase gene cloned in pIVEX 2.4d as expression vector and *in vitro* translation and transcription (IVTT) expression system "RTS 100 *E. coli* HY kit" from Roche was the most optimal system for efficient production of enzymatically active penicillin acylase.

Plasmid coding for penicillin Acylase (pPAC) was encapsulated into 22 pL droplets containing IVTT mixture using the coflow-based microfluidic device depicted in Figure 3.2A and B. The droplets were formed by mixing two solutions at the same flow rate (total flow rate of the aqueous phase: 100 μL/h). One contained the *E. coli* lysate and methionine, whilst the second contained the amino acids, the DNA template, the fluorogenic substrate (PA-ACMS, 100 μM) and the "reaction mix". We assumed that latter should at least contain the ribosomes, tRNAs, translation factors and T7 RNA polymerases. The coflow approach allows keeping the reagents unmixed until droplet formation. Indeed, even if liquids meet each other before the nozzle, the laminar flow regime adopted by the liquids prevents mixing to occur. Droplets were

collected in a capillary and incubated at 37°C for 4 hours before reinjection for analysis in another microfluidic device (Figure 3.2C).

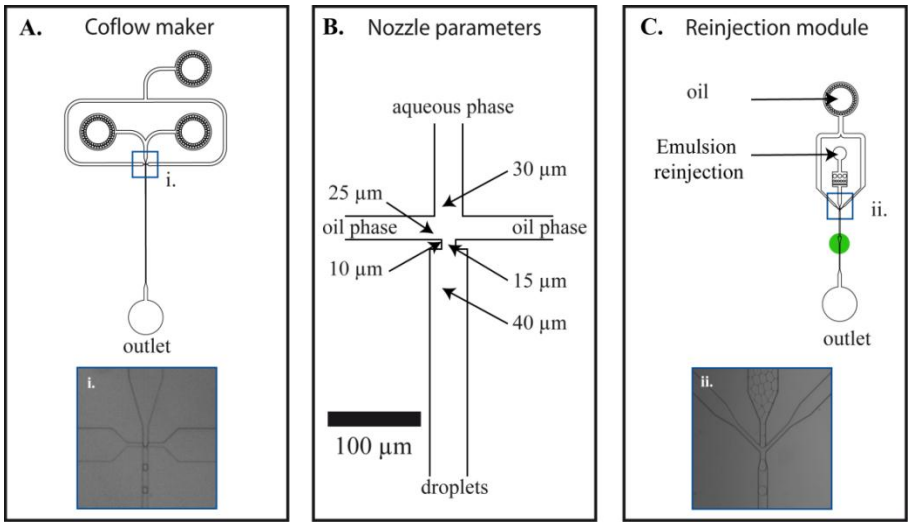


Figure 3.2: Microfluidic device scheme: (A) Encapsulation of DNA and IVTT is made using a coflow-based droplet maker module (picture of the module in i., details of the nozzle shown in part B). (C) Reinjection device for emulsion analysis. Measurement point where the laser is focused is indicated by the green spot. ii. Picture of an emulsion reinjection.

Three emulsions were collected. The first one was the negative control without pPAC. Upon reinjection, only one narrow population of droplets with fluorescence intensity around 40 rfu was observed (blue population ($\lambda = 0$) at Figure 3.3). Two emulsions containing different concentrations of DNA were also collected. The mean number of DNA molecules (called λ) per droplets was 0.2 and 2 respectively. The percentage of droplets containing a single copy of the template is calculated by Poisson distribution statistics:

$$P\,(X = k) = \frac{e^{-\lambda}}{k!}\lambda^{k}$$

Where $P\,(X = k)$ is the probability to have k DNA molecules per droplet at a mean number of molecules per droplets, λ. At concentrations of DNA of $\lambda = 0.2$, 81.9 % of the droplets will be empty, 16.4 % will contain a single copy

of DNA, and 1.7 % of the droplets will contain 2 or more DNA copies. At λ = 2, 13.5 % of the droplets will be empty, 27.1 % of the droplets will contain one DNA copy, 27.1 % will contain 2 copies, and 32.3 % will contain 3 or more copies. After 4 hours of incubation at 37°C the enzymatic assay has reached the plateau (see chapter 1), and emulsions are reinjected for analysis of their fluorescence. The graph below shows the distribution of droplets fluorescence intensities; 8,000 droplets were analysed for each emulsion. At λ = 0.2 (green populations on Figure 3.3), the positive population represents 18 % of the total emulsion; at λ = 2, it represents 86 %. The values are close to what was expected demonstrating that the method is sensitive enough to detect single molecules of DNA. The slight variation is mainly due to errors in dilution of the DNA.

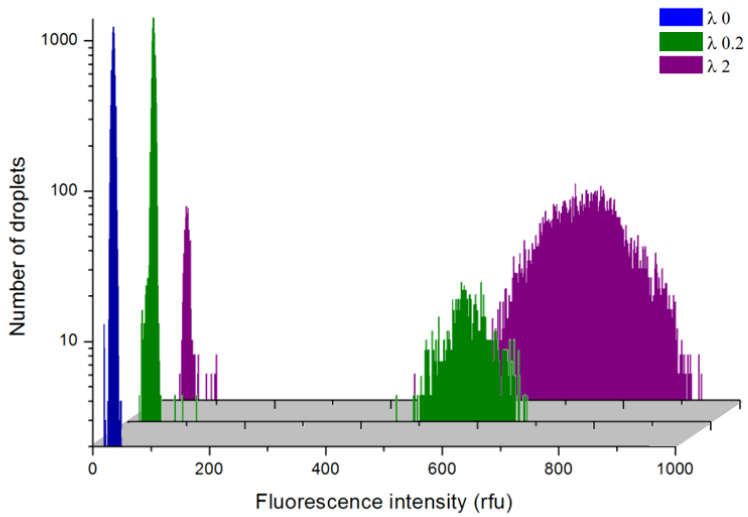


Figure 3.3: Characterization of penicillin acylase *in vitro* synthesis in droplets from single plasmidic genes. Three emulsions have been collected, incubated over night at 37°C and reinjected for analysis. The graph represents the fluorescence intensity analysis of 8,000 droplets of three different emulsions: firstly, negative control without DNA (blue), secondly, emulsions containing 0.2 pPAC or 2 pPAC per droplets (λ=0.2: green; λ=2: purple).

Recovering of few plasmid DNA molecules after sorting

Before starting the screening experiment on a model system, an effective protocol to recover and amplify DNA molecules collected after the sorting has to be set up. Several methods were tested.

Firstly, direct electroporation of the DNA recovered after sorting was considered. Dilutions of pPAC plasmid into total RNA as carrier were electroporated into electrocompetent cells (1 µL of DNA in 20 µL of MegaX DH10B Electrocomp Cells, Invitrogen). Transformation efficiency was evaluated by titration on LB-Amp plates. As shown in Table 3.1, around 2 % of the DNA is efficiently transformed. Since the Fluorescent Activated Droplet Sorter (FADS)[10] screens at rates around 200 Hz, a 4 h sorting experiment will analyze 3 x 10^6 droplets. Taking into account the electroporation efficiency, it means that the effective number of analyzed clones will be around 6 x 10^4. This direct transformation approach was therefore abandoned.

electroporated molecules	10^6	10^5	10^4	10^3	10^2
Number of transformants	21450	2382	234	20	1
% of transformed DNA	2,2%	2,4%	2,3%	2,0%	1,0%

Table 3.1: Electroporation of 5 different amounts of DNA corresponding to 100, 10^3, 10^4, 10^5, and 10^6 molecules of pPAC (from 4.5 x 10^{-7} to 4.5 x 10^{-3} ng/µL). 1 µL of DNA in solution in total RNA (carrier) was electroporated into 20 µL of electrocompetant cells (Invitrogen). Electroporation conditions: 2.5 kV, 25 µF, 200 Ω, 2 mm cuvette. Cells were suspended into 1 mL of SOC medium and incubated at 37°C for 1 hour before plating on LB-Amp plates for titration. The percentage of transformed DNA is the total number of transformants divided by the number of molecules present in the electroporated medium.

Secondly, PCR was performed on several dilutions of pPAC and was reported not to be sensitive enough since it was impossible to amplify less than 1,000 molecules in a single reaction. This was mainly attributed to an inhibition of the PCR by the presence of small amounts of IVTT mix, and a purification step would have been necessary.

Since for one transcripted molecule of DNA, at least a few molecules of RNA should be present in droplets,[8] Reverse Transcription PCR (RT-PCR) was tested using the RT-PCR kit from MP biomedicals. The PCR after the reverse transcription did not show encouraging results with the Taq Polymerase provided with the kit, but amplification of the gene was possible in bulk with the Fast Start Polymerase (Roche). IVTT reactions on different plasmids concentrations (from 500 to 3 x 10^{-4} ng/25 μL, corresponding to $\lambda = 1$ in 22 pL droplets) was performed, purified by gel filtration (Chromaspin 400, Clonetech), and reverse transcription was carried out on 10 μL (volume of approximately 45,000 droplets of 22 pL). As shown on the gel at Figure 3.4, amplification occurs at $\lambda \geq 10$ but not at $\lambda = 1$. Typically, a 2 hours sort corresponds generally to 500 to 2,000 droplets, approximately 100 times below the detection level reported here.

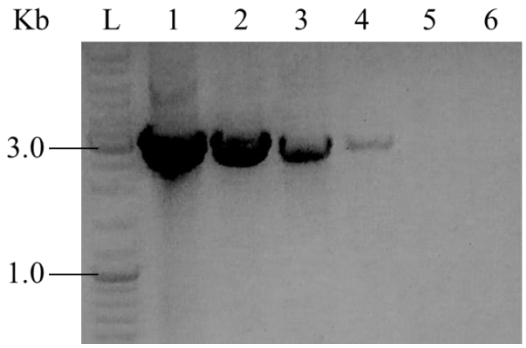


Figure 3.4: Agarose gel of RT-PCR products (1 % in TAE). IVTT on several dilutions of DNA were done, followed by reverse transcrition (RT-PCR kit from MP biomedicals) and PCR (Faststart polymerase, Roche). Lane L: ladder (O'GeneRuler DNA Ladder Mix, Fermentas), lane 1: positive control, DNA amplified from IVTT reaction on 500 ng/μL of pPAC; lane 2 to 5: DNA amplified from IVTT reaction on concentrations of pPAC corresponding to $\lambda = 1000$, 100, 10 and 1. Lane 6: RT-PCR negative control.

Rolling circle (RC) amplification was also tested as directed evolution experiments we intend to perform are based on the IVTT of PAC cloned into a plasmid. RC was first performed on dilution of plasmids (pPAC diluted with total RNA as carrier) from 4 ng (= 6.10^{8} molecules) to 4.10^{-9} ng per reaction

(= 0.6 molecule, therefore 1 or 0 molecule). Amplification products were restricted with Eco RI (single cutting enzyme) and put on agarose gel (Figure 3.5A). A 6,000 bp band corresponding to pPAC is present in every sample except for the negative control, where only smear was observed. Then a PCR on RC products confirmed the amplification of the PAC gene (Figure 3.5B). Experimental details are given in the legend of Figure 3.5.

Then, the same amplification process was tested on DNA recovered from sorted droplets of a model experiment (that will be described later in this chapter). After sorting, the selected droplets are collected in a tube; the emulsion was then broken using one volume of emulsion breaker (Raindance) and the aqueous phase, supplemented with total RNA in water, was recovered. DNA was precipitated and rolling circle was performed on the DNA pellet. Eco RI was added to linearize the resulting concatemers; DNA was washed by ethanol precipitation, suspended in water, and analyzed on agarose gel (Figure 3.5C). The band corresponding to the linearized plasmid is not clearly visible. A PCR is performed on each sample and on the RC negative control with suitable primers in order to attest the presence of the pac gene. The gel, presented in Figure 3.5D, confirmed the presence of the pac gene in samples 2 and 3, but not in sample 1. Negative controls from the RC and the PCR are negative. Details of the experiment are given in the legend of the figure.

Concerning sorting experiments made on libraries of PAC variants, concatemers will be restricted using a single cutting enzyme and ligated. Ligation products will be electroporated into electrocompetent cells; cells will be plated on big LB-Amp agarose plates and incubated over night at 37°C. Colonies will be collected, suspended in liquid medium (LB-Amp, 50 mL) and incubated at 37°C for 4h. Finally, plasmids will be extracted using midi-prep kits (Qiagen) and be ready for a new round of selection.

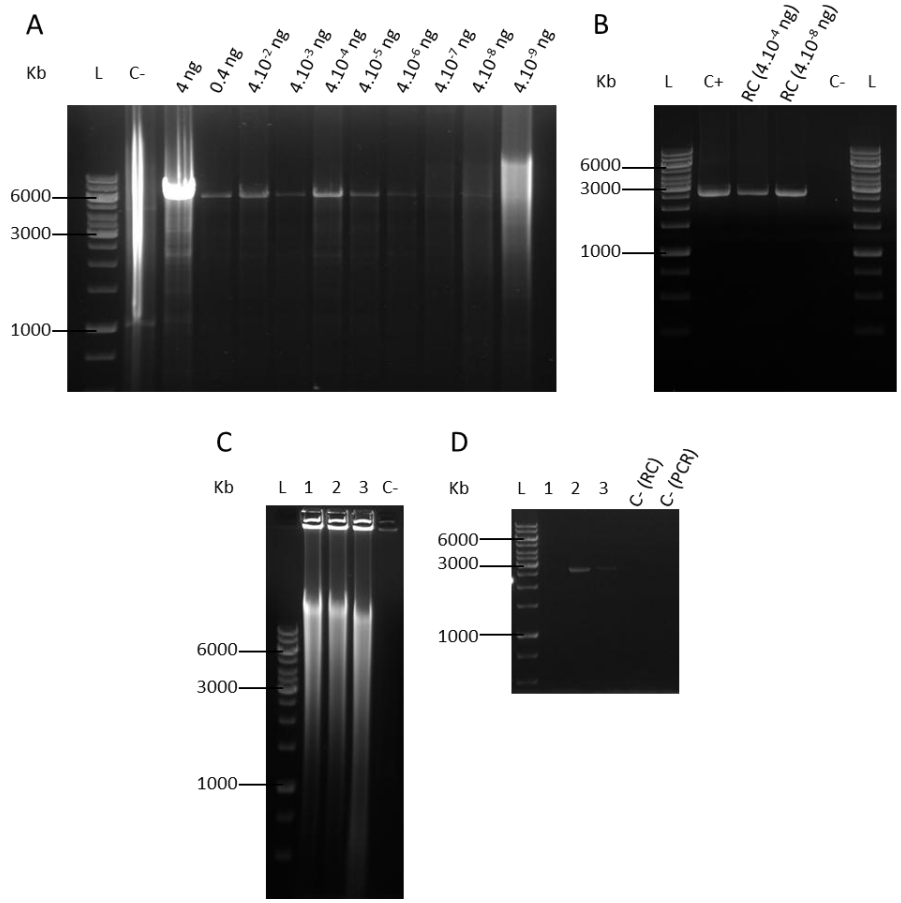


Figure 3.5: Set up of the amplification of plasmids at low concentrations by rolling circle (RC) and PCR. **DNA amplification from DNA dilutions: A.** Rolling circle was performed on dilutions of pPAC from 4 ng (= 6.10^8 molecules) to 4.10-9 ng (= 0.6 molecule, therefore 1 or 0), C-: negative control, rolling circle without DNA. **B.** PCR was performed on two rolling circle products (4.10-4 ng and 4.10-8 ng). C+: positive control, PCR on 1 ng of pPAC, C-: negative control, PCR without DNA. The expected band size is 2,500 bp and corresponds to the PAC gene. **DNA amplification from sorted droplets (model experiment): C.** Rolling circle was performed on DNA coming from a sorting experiment of droplets at $\lambda = 0.2$. Sorted droplets contained 1 pPAC per droplet. Sample 1: 1300 droplets (≈ 1,300 pPAC molecules), sample 2: 2,500 droplets (≈ 2,500 pPAC molecules), Sample 3: unsorted droplets (high DNA quantity), C-: negative control, RC without DNA. **D.** PCR was performed on each RC product (samples 1 to 3 and the negative control of the RC). Expected band size: 2,500 bp. C-(PCR): negative control, PCR without DNA.

Model sorting experiment

In order to test the efficiency of the droplet sorting in microfluidics, a model experiment was conducted. Plasmids encoding for the wild type penicillin acylase (named pPAC) or a less active mutant that carries the F146L mutation (named pF146L)[11] were mixed and encapsulated into droplets containing IVTT mix and a fluorogenic substrate of Wt PAC (PA-ACMS, 100 μM). Droplets were collected and incubated in a glass capillary at 37°C for 3 hours. Plasmid concentrations are calculated to reach a mean number of plasmid per droplets (λ) of 0.2 in order to avoid double encapsulation. After incubation, droplets are injected in a sorting device. Schemes of the device and pictures are presented in Figure 3.6. Droplets are spaced out with surfactant-free fluorinated oil before reaching the sorting junction. Because of lower hydraulic resistance, droplets flow by default through the negative arm. The laser is focused at the analysis point just before the sorting junction (depicted by the green spot on the scheme). Droplets of interest can be conducted toward the positive arm using dielectrophoresis (DEP). DEP consists in an electrical force that results from the difference in electric polarizability of the aqueous droplet and the oil phase. As water is more conductive than the oil, the induced dipole will align with the field and droplets will be attracted toward the region of higher field strength (positive DEP).[12,13] The application of an AC electric field of 1.5 kV across the electrodes (see Figure 3.6) will drive the droplets into the positive arm. The field is piloted using a Labview program that allows the analysis of the droplets fluorescence and the sorting of the desired population based on fluorescence intensities and size of each droplet. The optical set up is the same as used before and is depicted in Figure 5.1 (experimental section).

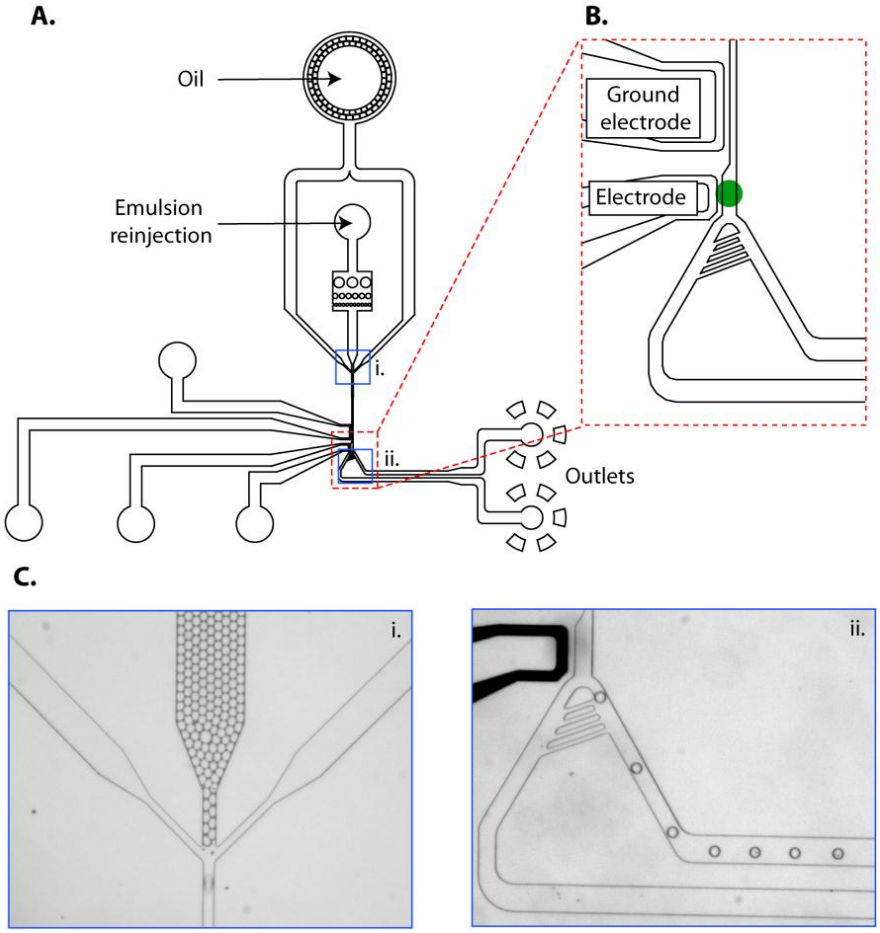


Figure 3.6: Sorter device scheme and pictures. For sorting experiments, droplets were collected in capillaries and reinjected in the sorter microfluidic module depicted in part A. Reinjected droplets are spaced-out with surfactant-free fluorinated oil (C. i.). At the sorting junction (scheme in B. and picture in C. ii.), droplets flow by default into the negative arm (due to lower hydraulic resistance). Droplets fluorescence intensities measurements point where the laser was focused is indicated with the green spot (B). When an AC electric field is applied across electrodes (30 kHz, 1.5 kV), droplets are deflected into the positive arm.

The selected droplets were collected in a tube; after sorting, the emulsion is broken, the aqueous phase is supplemented with total RNA in water. Each sort fraction contained 500 to 10,000 droplets, and therefore 500 to 10,000 molecules of DNA. The DNA was isolated and amplified using the rolling circle protocol described above (see previous section). Since a sequence

encoding a His-tag is positioned respectively at the beginning and the end of the gene for pPAC and pF146L, the two plasmids can be discriminated by PCR. Primers annealed before the start codon and around the 400th base pair of the PAC gene. Sizes of the resulted amplicons are of 520 bp for pPAC and 445 bp for pF146L (Figure 3.7). Intensities of the DNA bands obtained on agarose gel (2 % in TBE) will be analyzed.

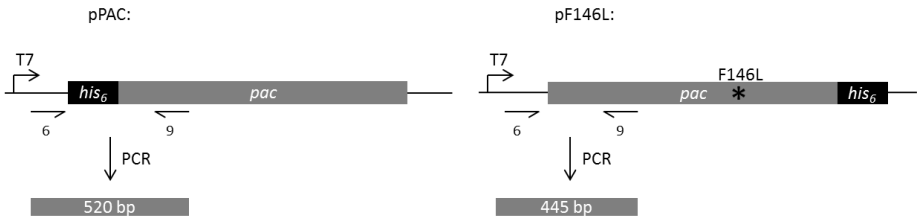


Figure 3.7: Scheme of the two plasmids used in the model experiment: pPAC, pac gene coding for the wild type penicillin acyalse cloned into pIVEX 2.4d and pF146L, pac gene carrying the F146L mutation cloned into pIVEX 2.3d. Plasmids differ by the position of the His-tag and can be discriminated by PCR with primers 6 and 9. Fragments generated by PCR will be of 520 bp with pPAC as template and 445 bp with pF146L.

Five different sorting experiments were carried out. Droplets contained a mix of 1 pPAC for 199 pF146L. Sorted droplets were collected into separated tubes and DNA was amplified. After PCR, the DNA is put on gel for analysis (2 % agarose in TBE, Figure 3.8). As a control, the same recovery process was made on the starting DNA mix (named 1.0 and 2.0).

For sorts 1 to 4, bands corresponding to pF146L are not observed after sorting. We estimated that the percentage of pF146L genes is less than 10 % which means that the enrichment factor is at least ~2,000 fold. On the contrary, concerning the sort 5, a weak band corresponding to pF146L is present on the agarose gel; we can roughly estimate bands intensities calculations that the ratio pPAC/pF146L is around 10. The enrichment factor can be estimated to be at least around 2,000 fold (ratio after sorting/ratio before sorting: 10 divided by 1/200). Using Equation 3.1,[10] which takes into

account the co-encapsulation probability of pPAC with one or more pF146L plasmids when working at $\lambda = 0.2$, the maximum enrichment factors that can be obtained should be around 1,000 fold. Enrichment factors obtained for the model sorting experiments are in the expected ranges.

$$\eta = \frac{1}{1 - e^{-\varepsilon_0 \lambda/(1+\varepsilon_0)}}$$

Equation 3.1: Theoretical enrichment factor η, archived after one round of selection. ε_0 is the ratio of positive to negative genes before selection; λ is the initial mean number of genes per droplets.[10]

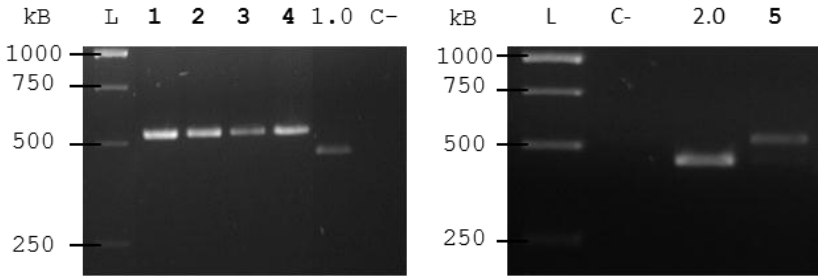


Figure 3.8: Agarose gels (2 %, TBE) showing the PCR products of DNA fractions collected before and after sorting experiments. Left agarose gel: lane 1: ladder, lanes 2 to 5: amplified DNA from sorts 1 to 4, lane 6: amplified DNA from DNA mix before sorting, lane 7: negative control of the amplification process. Right agarose gel: lane 1: ladder, lane 2: negative control of the amplification process, lane 3: amplified DNA from DNA mix before sorting, lane 4: amplified DNA from sort 5. Sizes of expected fragments: PCR on pPAC: 520 bp, PCR on pF146L: 445 bp. (GenRuler 1kb DNA ladder, Fermentas).

Libraries of PAC mutants

The *in vitro* expression system has been chosen for its potential in droplet-based microfluidic applications. We intend to develop high throughput screening methods for directed evolution of enzymes by applying IVTT to the evolution of penicillin acylase specificity; this topic will be addressed in the following chapters. Screening methods will first be evaluated on a model system before being applied to the screening of libraries of mutants. Two methods for the creation of molecular diversity were considered. In the two following

sections, we will describe the construction of error prone PCR libraries and of a library that displayed 7 randomized residues at specific positions in the phenyl acetyl binding pocket.

Error prone PCR libraries

First, error prone PCR (epPCR) was chosen to prepare libraries with low and high mutational loads. This technique does not necessitate a deep knowledge of the enzyme's structure as it targets the whole gene; thus, this technique is versatile, with straightforward implementation. Nevertheless, due to the degeneracy of the genetic code and the amplification inherent to the process, the epPCR technique possesses several drawbacks. As it is very rare to have two mutations at the genetic level in the same codon, only some mutations at the proteic level will be represented. Moreover, biases in error types could be present. For example, transitions (purine → purine or pyrimidine → pyrimidine) are more frequent than transversions (pyrimidine ↔ purine). In particular, the g ↔ c mutation is very rare using standard epPCR protocols.[14]

Error prone PCR methods are based on the intrinsic error rates of DNA polymerases, enhanced by altered reaction conditions,[15] and were conducted in the presence of increased concentration of Mg^{2+}, dNTPs at unbalanced stoichiometry, and in the presence of Mn^{2+} to favour mutations. Two Mn^{2+} concentrations were used to tune the mutational load to higher and lower rates (named LR and HR for low and high rate libraries). Amplified DNA fragments were cloned back into the pIVEX2.4d vector (scheme outlined in Figure 3.9). The diversity of the two libraries was evaluated at 3.39×10^7 and 1.88×10^7 variants. Several clones were picked up for sequence analysis. As shown in Table 3.2, the low rate and high rate libraries displayed a mean of 4.3 and 8.6 mutations per gene respectively. Although the statistical relevance is rather poor because too few sequences were analyzed, relatively equal proportions of

transitions and transversions are observed. Moreover, 1 and 3 indels were observed in the LR and HR libraries. Since these indels are generally single nucleotide insertion or deletion causing frame shifting, it means that a significant proportion of the libraries is useless (42 and 74 % for LR and HR libraries). Anyhow, we should remain cautious with those numbers since only 4,600 and 5,500 bp from 6 different clones for the LR library and 7 for the HR library were analyzed, and high variations in mutations frequency between variants within the same library is well-attested.[16]

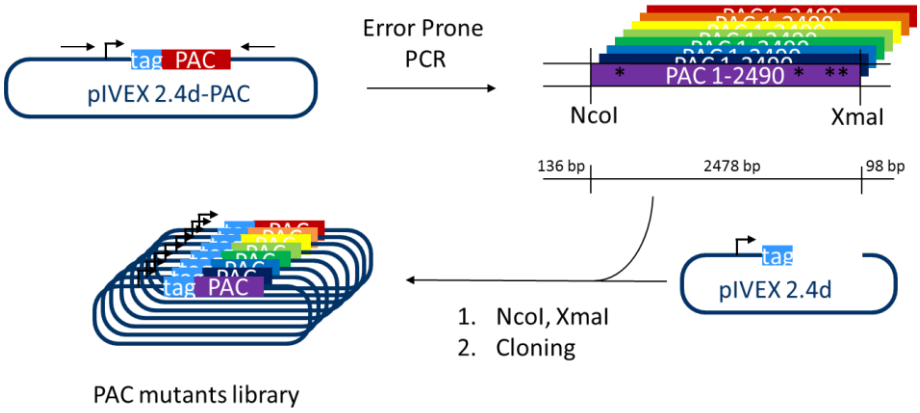


Figure 3.9: Strategy for the construction of the error prone PCR libraries. First, an epPCR is performed on pIVEX2.4d-PAC with primers 5 and 6. The PCR products are then ligated into the pIVEX2.4d between the NcoI and XmaI restriction sites.

	Diversity	Transitions	%	Transversions	%	Insersions	Deletions	Mutations	Analysed bp	Mutations /gene	Indels /gene
LR	$3.4.10^7$	4	50%	4	50%	0	1	8	4613	4.3	0.5
HR	$1.9.10^7$	8	42%	11	58%	1	2	19	5498	8.6	1.4

Table 3.2: Sequence analysis of variants from the low rate (LR) and high rate (HR) libraries.

Targeted mutations library

A way to address the problems generated via standard epPCR methods is to consider saturation mutagenesis. This method consists of the randomization of defined sites in the enzyme with the creation of focused libraries. Such a method requires information about the enzyme structure to accurately choose mutagenesis sites. By reducing considerably the extent of the protein space where mutations are concentrated, we probably enhance the quality of the library and limit the diversity that will have to be screened.[17]

Seven amino acids are described in the literature as being involved in substrate specificity and have been chosen for randomization: PheB24, ThrB32, ProB49, ValB56, AlaB69, TrpB154 and IleB177 (Figure 3.10).[18-20] Details about those residues and their role in substrate specificity and binding are given in chapter 1 of the Introduction.

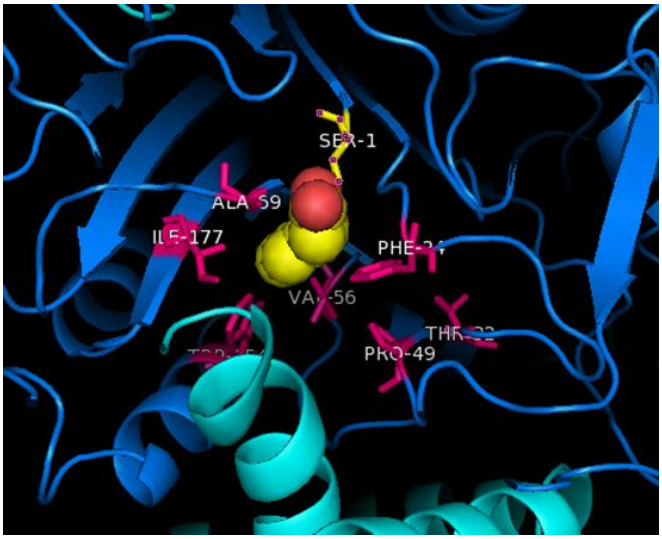


Figure 3.10: Scheme of the seven residues to be saturated in PAC 3D structure. The seven residues from the specificity pocket to be degenerated in the TM library are displayed in pink (PheB24, ThrB32, ProB49, ValB56, AlaB69, TrpB154 and IleB177). The α chain is in light blue and the β chain is in dark blue. The enzyme is complexed with the 3,4-dihydroxyphenylacetic acid, a substrate analog, showed as spheres. The catalytic serine is displayed in yellow. (PDB code: 1ai4[21])

A DNA cassette of 980 base pairs was designed and synthesized. This cassette possesses 7 degenerated codons (NNK) at positions mentioned above, and two unique restriction sites BstXI and Bsu36I. A signature has also been introduced as a silent mutation in order to allow the discrimination between variants coming from the Targeted Mutation Library (TM Library) and contaminants coming from the wild type gene. The second Bsu36I site present in the wild type sequence of PAC was mutated (silent mutation) in order to allow the direct cloning of the cassette into the vector pIVEX2.4d-PAC. The cloning strategy is detailed at Figure 3.11.

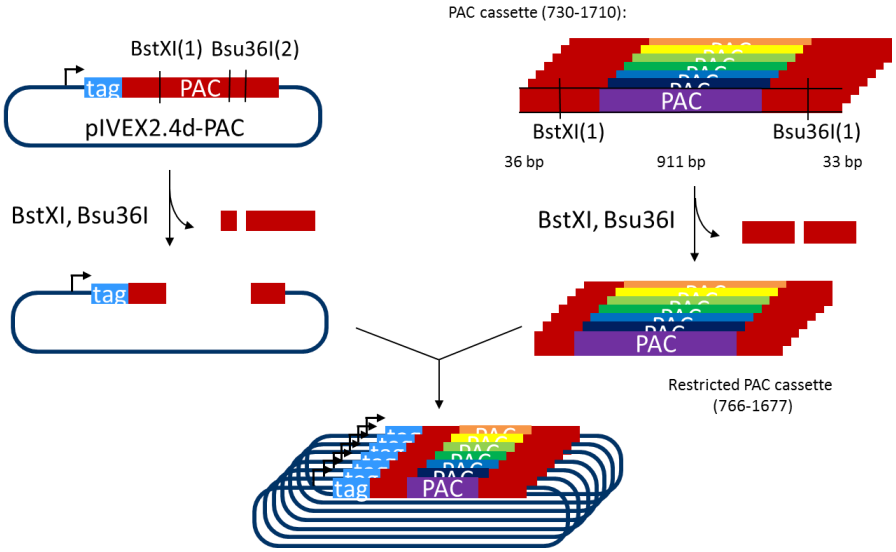


Figure 3.11: Construction strategy of the Focussed Library. A 911 bp cassette bearing seven degenerated codons is inserted in pIVEX2.4d-PAC, between BstXI and Bsu36I restriction sites present in the pac gene.

First, the cassette was amplified by PCR using the Expand High Fidelity enzyme (Roche) and primers 7 and 8. After restriction of the plasmid, SDS (0.1 %) and proteinase K (1 %, NEB) were added to the reaction and incubated for 30 min at 37°C. The DNA was purified on anion-exchange column and dephosphorylated. After ligation of the two fragments, the resultant product was transformed into Top10 *E. coli* strains and plated on LB-agar. The diversity

of the library was evaluated at 2.27 x 10^7 variants. Knowing that the maximum proteic diversity is 1.28 x 10^9(= 20^7), further work could have been done to increase the library size, but our microfluidic screening strategy would not enable the screening of a significantly larger library. Several clones were picked and sequenced. Sequence analysis is presented at Table 3.3. Except for clone 4, clones are well constructed and all 7 residues are randomized. One should notice that some other mutations and deletions are also present and may result from the PCR amplification of the cassette. DNA was then prepared for *in vitro* experiments by conventional techniques (Maxi prep kits, Qiagen).

WT residue:	F24	T32	P49	V56	A69	W154	I177	extra mutations
clone 1	aag (Lys)	gtt (Val)	cgt (Arg)	atg (Met)	tat (Tyr)	tat (Tyr)	ctg (Leu)	deletion of 2bp
clone 2	tgg (Trp)	cgt (Arg)	aag (Lys)	aat (Asn)	ttg (Leu)	cct (Pro)	cat (His)	1 mutation (c-t)
clone 3	aag (Lys)	tct (Ser)	aag (Lys)	gat (Asp)	atg (Met)	cag (Gln)	acg (Thr)	1 mutation (g-a)
clone 4	gtt (Val)	aag (Lys)	aat (Asn)	gtg (Val)	-	-	-	only 200 bp aligned
clone 5	cgt (Arg)	gtg (Val)	tgt (Cys)	ggt (Gly)	gag (Glu)	cct (Pro)	cat (His)	-

Table 3.3: Sequence analysis of 5 randomly picked clones of the Targeted Mutations Library.

Screening of libraries for enzymes with altered specificities

A model strategy for sorting penicillin acylase variants based on their enzymatic activity and recovering of the DNA material after sorting has been described above. It demonstrated the potential of this strategy for selection of new activities or specificities in libraries of PAC mutants.

Libraries that will be screened are High Rate Error Prone PCR Library (HR-EPL) with a total diversity of 2 x 10^7 variants and Targeted Mutations Library (TML) with a 2.3 x 10^7 diversity (see chapter 1 for more detail). In the first library, mutations are present all over the gene with approximately 8.6

mutations per gene, whereas the TML library has been obtained by the cloning of a cassette where 7 amino acids playing a key role in substrate specificity are randomized. These two libraries were screened to look for variants with altered substrate specificity. Variants were expressed *in vitro* into droplets containing a collection of various substrates: amino coumarin based substrates with diverse side chains (molecules **8**, **9**, **10**, **11** and **12** on Figure 2.3, at 20 μM each). The wild type enzyme is not active on substrates **8**, **9** and **10**, but possesses a slight activity on **11** and **12**. Screening for the activity towards five substrates present together in the droplet would enable to find variants exhibiting a "generalist profile" by being able to hydrolyze several substrates in the same time. On the other hand, variants displaying a high specificity for one substrate and no inhibition by the other substrates will also be selected. Those variants present an interest for the study of the substrate specificity of PAC. Droplets were incubated at 37°C over night. For the first round, the average number of plasmids per droplet (λ) was of 10. Since only between 100 and 200 droplets are screened per second, a first round of screening with 10 genes per droplet will broaden the analysis of the initial diversity present in libraries to around 8 x 10^6 variants in 2 hours. After sorting, DNA will be recovered and amplified by rolling circle, concatemers will be digested by Eco RI (single cutting enzyme) and the linearized plasmids will be ligated and electroporated into *E. coli* cells. DNA will then be extracted from harvested cells and be used for a second round of selection. At the second round, the mean number of DNA molecules per droplet will be 0.1 or 0.2 in order to minimize multiple encapsulation events.

Screening on high rate epPCR library

Firstly, two consecutive rounds of selection were performed on the high rate epPCR library (HR-EPL). The first round of sorting at $\lambda = 10$ allowed the selection of the 0.3 % most active droplets. 2,600 positive droplets were selected out of 810,000 analyzed droplets, which corresponds to 8.1×10^6 clones. As the library size is 2×10^7, we estimated that approximately 35 % of the total diversity of the library has been explored. At the end of the recovery process of the 26,000 selected genes (2,600 selected droplets with 10 genes per droplet), only 16,000 colonies were obtained, so a maximum of only 45 % of the sorted clones were recovered. The second round of screening was performed at $\lambda = 0.1$, starting with the variants from the first round. The 0.12 % most fluorescent droplets were selected. 600 colonies were obtained after transformation of the amplified DNA material. Altogether, 0.12 % of droplets containing a maximum of 16,000 different variants (selected variants from the first round) at $\lambda = 0.1$ should represent maximum 192 different variants. The 600 colonies obtained after the second round of selection will then be representative of most of the selected clones.

Enzymatic activities contained in the selected fractions were evaluated in droplets by creating three different emulsions analyzed on the same day. Emulsions of HR-EPL, 1st round variants and 2nd round variants with a $\lambda = 0.2$ occupancy, were created as before with IVTT and the 5 substrates and incubated at 37°C over night. Distribution histograms obtained after analysis of each emulsion are shown in Figure 3.12. The mean level of activity seems to increase at each round of selection.

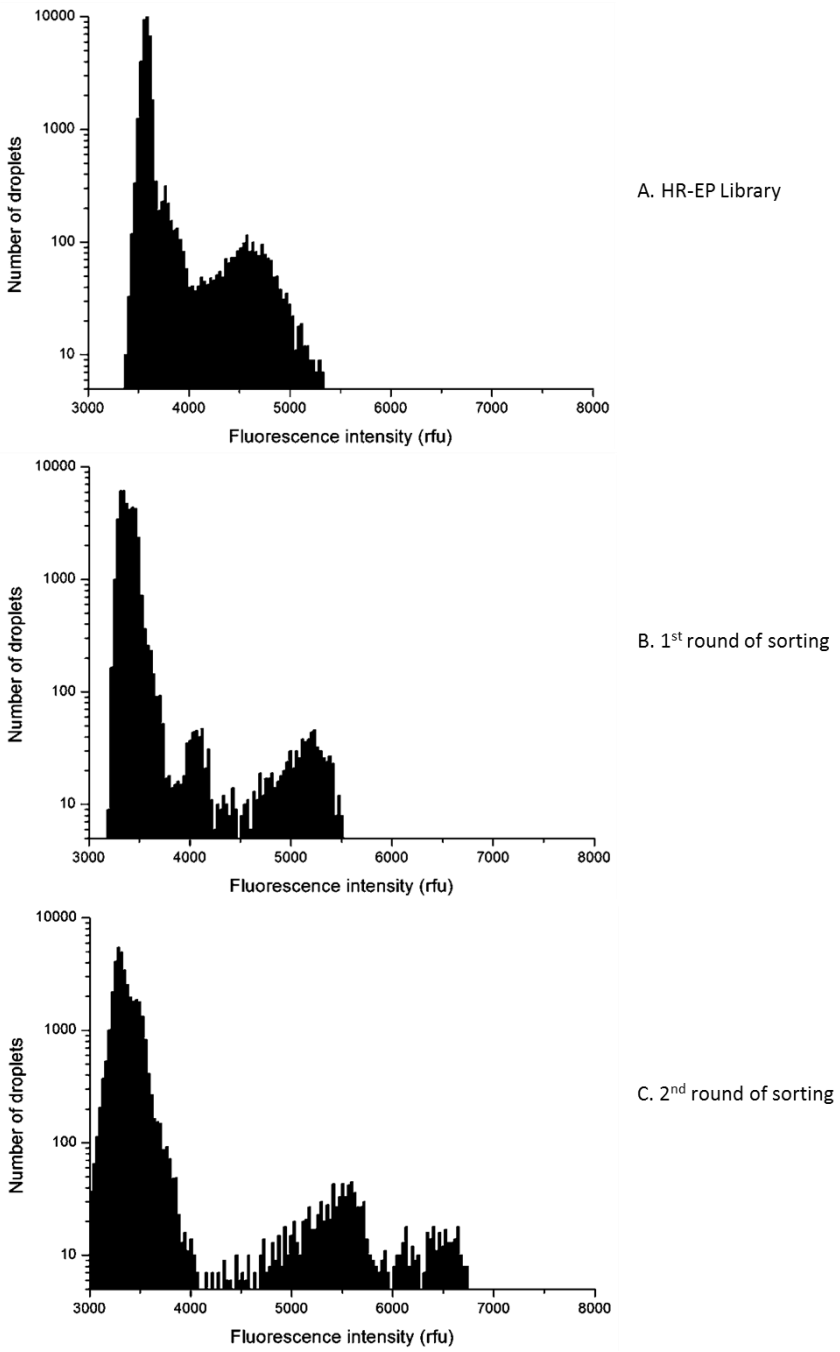


Figure 3.12: Droplet-based analysis in droplets of the activity of the HR-EP Library on the collection of substrates. (A) and the selected DNA after 1 and 2 rounds of selection (B and C, respectively). Emulsions were made at $\lambda = 0.2$. 40,000 droplets were analyzed per emulsion.

38 individual clones were picked and analyzed individually. 9 plasmids of the expected size (6,000 bp) were sequenced. 7 clones possess the wild type pac sequence, and only two variants displayed mutations: clones 7 and 12. Clone 7 bears 8 coding mutations and clone 12 possess 1 deletion and 5 mutations. These clones were expressed *in vitro* in bulk and assayed in microtiter plates towards the wild type substrate PA-ACMS and the collection of substrates (compounds **5**, **8**, **9**, **10**, **11** and **12** on Figure 2.3 chapter 2). Unfortunately, none of the clones carrying mutations showed activity toward any substrates. As the screening was made while the reaction has already reached the plateau, no positive selective pressure was applied in order to discriminate mutants that would be more active than the wild type from the wild type. Based on these results, we can hypothesize that the library may not contain any interesting variants. The wild type sequence was the most active. All mutants displayed too many mutations and were not active anymore. Their presence in the selected population might be explained either because of false positives coming from mistakes during the sorting process or because they were co-encapsulated with a wild type contaminant. Working at high mutational loads can be too deleterious for the protein. However, when working at high mutational loads, selected clones can carry silent and neutral mutations, which was not the case here. It might be possible that the wild type sequence was selected because of its strong presence in the starting library. The library was characterized by sequencing only 10 clones, which may not be enough to statistically attest that the wild type sequence was not over represented. Another hypothesis would be that a contamination by the wild type sequence occurred during the recovery and amplification process of the selected DNA.

Error prone PCR libraries present several advantages. Their implementation is easy since deep knowledge of the enzyme function, mechanism or structure is not necessary, mutations are spread over the gene randomly. But this methodology possesses some significant drawbacks. Mutations are likely to be silent mutations, or at a position in the scaffold of the

protein that will not have any effect on the property to evolve. Moreover, at high mutational loads, stop codons and unfolded enzymes have to be expected in a significant portion of the library.

Screening on Targeted Mutations Library

A second set of screening was performed on the Targeted Mutations Library (TML) toward the same collection of substrates. This library possesses seven point mutations at positions reported as key residues in substrate specificity. As for the selection on the EP Library, a first round of selection was accomplished at $\lambda = 10$, and the 1 % most fluorescent droplets were selected. 26,700 droplets were sorted on a total of 2.5 x 10^6 analyzed droplets. Therefore, at $\lambda = 10$, it corresponds to approximately 2.7 x 10^6 genes. The diversity of the library being 2.3 x 10^7, approximately 66 % of the diversity was explored (taking Poisson distribution into account). After amplification and transformation of the recovered DNA, the selected variants were introduced in a second round of selection, at $\lambda = 0.2$. The 0.5 % most fluorescent droplets were selected and 48 individual selected clones were analyzed. 36 out of the 48 clones showed the expected profile upon restriction and analysis on gel of the plasmidic DNA (Figure 3.13). Those clones were translated *in vitro* at 37°C over night and were assayed toward each ACMS substrates separately (PA-ACMS and molecules **5**, **8**, **9**, **10**, **11** and **12**).

None of the clones showed a significantly different activity profile compared to the wild type PAC which is active on PA-ACMS and less active on **11** (12 fold) and **12** (45 fold). Enzymatic assays results of the analyzed clones are presented in Table 3.4. The activity pattern of each clone is rated compared to the wild type enzyme: the same level of activity as the wild type (+++), less active (++ or +), or inactive (-). While analyzing the obtained results we should keep in mind that *in vitro* translation yields could vary from one sample to the other. Clones that showed an even slight activity were sequenced. Only the

region of the gene that is supposed to present the mutations was sequenced (approximately 1,000 bp, 40 % of the gene). A silent mutation was inserted in the library as a signature to distinguish clones of the library from potential wild type gene contamination. The TM Library signature is present in only two of the selected clones (16 and 18), while all other clones possess the wild type signature, and none of them possess mutations at the 7 randomized positions. 17 clones possess no or silent mutations, and 5 clones bear one amino acid mutation in the analyzed amino acid sequence.

Clone 16 has the TM Library signature, located at the beginning of the inserted cassette, but does not possess any mutations at the genetic level, even though with the NNK degeneracy, some silent mutations would have been observed for encoding wild type residues. Hence, this clone is probably the result of a recombination event between a variant of the TM Library and a wild type sequence.

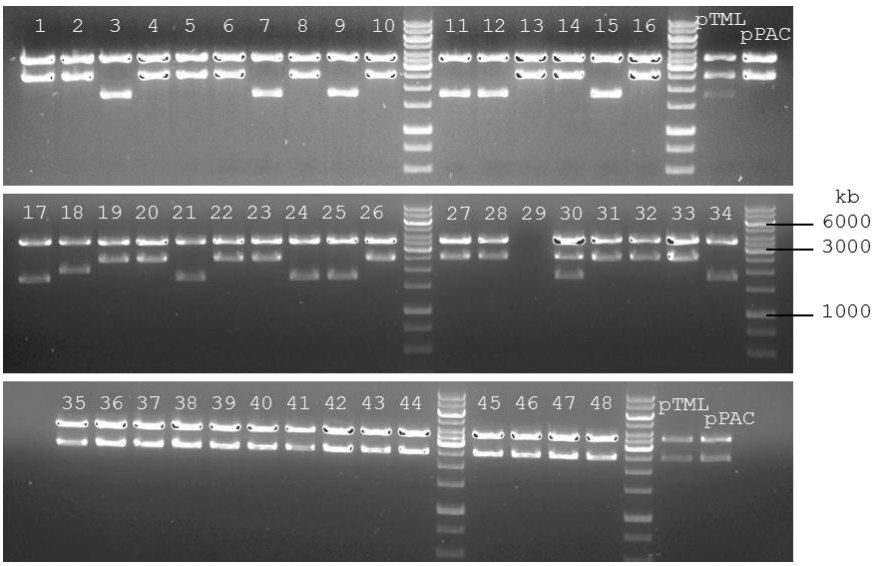


Figure 3.13: Agarose gel of plasmidic DNA from 48 sorted clones randomly picked after the second round of sorting. Plasmids were restricted with NcoI and BamHI. Bands of 3,500 and 2,500 kb corresponding to the vector and the pac insert respectively. "pPAC" and "pTML" (plasmids from the library) were taken as controls. A band of ~1,800 kb is already present in the library.

Only clone 18 possesses the targeted mutations: F24R, T32L, P49V, V56L, A69G which corresponds to 5 out of the 7 degenerated residues introduced in the library. S67P mutation is also present and must have been introduced by PCR. It shows a deletion of approximately 450 bp just after the point mutations, and without frame shifting. Figure 3.14 represents the penicillin acylase structure with the deleted sequence depicted in yellow. This region contains two out of the seven mutated residues (TrpB154 and IleB177) and AsnB241, one of the residues implicated in the catalytic mechanism and forming the oxyanion hole. This part of the enzyme possesses a β sheet that is taking part in β-β interactions with the rest of the protein. It seems impossible that this clone still possess some residual activity. This deletion should result in a strong rearrangement of the protein structure since the two residues before and after the deletion are not located in the same region in the wild type enzyme structure.

Sequencing data from both selections lead us to conclude that the pac wild type sequence is highly present after sorting. It could come from a significant presence of wild type in the libraries at the beginning, which can be massively present after screening if any clones of the libraries showed a significantly higher activity on the collection of substrates than the wild type. Another hypothesis would be that the wild type sequence could have been introduced during the DNA recovery and amplification process after sorting. The whole process could also be subjected to contaminations with the wild type plasmid even if strong precautions were taken to avoid it, especially before the rolling circle amplification, or at the electroporation step.

Clone	WT activity	signature	Sequence from residue 280 to 584	silent mutations
1	-			
2	-			
4	-			
5	-			
6	+	WT	-	1
8	-	WT	L152P	-
10	-			
13	+++	WT	-	3
14	-			
16	+++	TML	-	-
18	+	TML	F24R, T32L, P49V, V56L, A69G/ S67P, deletion (450 pb, conserved phase)	-
19	+++	WT	-	-
20	-			
22	+++	WT	-	1
23	-			
26	+	WT	-	-
27	-			
28	+	WT	-	-
29	+++	WT	-	-
30	-			
31	-	WT	-	-
32	+++	WT	-	-
33	+++	WT	Q12H	-
35	-			
36	+	WT	L91S	1
37	-			
38	++	WT	-	1
39	++	WT	-	1
40	++	WT	-	1
41	++	WT	-	1
42	-			
43	++	WT	I63V	1
44	++	WT	-	1
45	++	WT	K8E	1
46	++	WT	-	1
47	++	WT	-	1
48	-	WT	-	2

Table 3.4: Analysis of selected clones from TM Library. 48 clones have been picked and their DNA analyzed by restriction. 36 of them showed the good profile. Those 36 clones have been translated *in vitro* and analyzed toward the collection of substrates and PA-ACMS. None of the clones showed a significantly different activity pattern from the wild type PAC. The second column shows the activity of each clone compared to the WT PAC. The three last columns: first, the presence of the library or the WT signature (silent mutation), then the presence of residues mutations or silent mutations. (AA_{280} to AA_{584} were sequenced, which corresponds to the whole inserted cassette, clones showing no activity were not sequenced.)

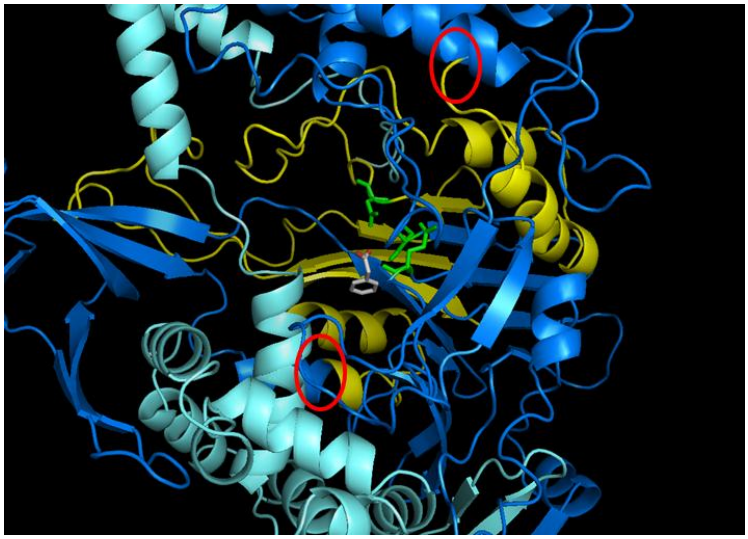


Figure 3.14: Penicillin acylase structure (PDB code: 1PNL). The part depicted in yellow represents the deletion in the β domain observed in clone 18. The rest of the β domain is represented in dark blue. The α domain is in light blue. Lateral chains of amino acids implicated in the catalysis (SerB1, AsnB241 and GlnB23) are represented in green. Red circles highlight the beginning and the end of the deleted region.

Conclusion and future work

This chapter presents an *in vitro* strategy for directed evolution of enzymatic activities in droplet. We initially developed a method for the *in vitro* translation of PAC in droplets starting from a single plasmid. Nevertheless, working at low DNA concentrations could lead to drawbacks that should be considered. One gene per droplet renders the experiment sensitive to biological variance (variance in mRNA synthesis or degradation, in protein synthesis), moreover, IVTT yields also vary from one experiment to the other. An additional variance effect could also come from small variations in droplets sizes upon formation. Indeed, 10 % variation in droplet's diameter leads to a 30 % variation in droplet's volume and therefore a variation in amount of reagents available in the droplet.

Several methods for DNA recovery after sorting were evaluated. Electroporation allows the recovery of about 2 % of the plasmids. PCR and RT-PCR have been also tested, but we were not able to amplify DNA starting from very low concentrations of plasmids. Rolling circle amplification followed by traditional PCR allows the recovery of a few molecules of DNA. However, as discussed in this chapter, sorting experiments might have been contaminated either at the origin of the process via the libraries, or during the DNA recovery process.

In the future, several solutions may help to circumvent the contamination problem. First, nested PCR could be tested. This technique is known for small DNA quantities amplification and consists in two consecutive PCRs with two different sets of primers. The second PCR is made on a small fraction of the first PCR with a second set of primers which are annealed into the first amplicon.[22] Secondly, we can consider direct electroporation of the DNA material contained into selected droplets without any *in vitro* amplification step. Even if only 2 % of the material can effectively be transformed, the process is shorter and the control of the experiment is much better since the number of transformants will represent directly the number of recovered plasmids. A third strategy would be the amplification of the DNA in droplets before IVTT, as working with higher DNA quantities should increase the DNA recovery yield and impair further contaminations. This approach has proved to be efficient for the co-expression of PAC and GFP in the same droplet (see chapter 4 for more detail).

Construction of two types of libraries of PAC mutants was described. As libraries will be screened via droplet-based microfluidic methods, relatively large libraries can be screened. We first opted for error prone PCR libraries since they are easy to implement. Two libraries were successfully obtained. The high rate library presents 8.6 mutations per gene in average, while the low rate one possesses 4.3 mutations per gene. Indels were also observed and even if the

number of sequenced based pairs is too low to statistically determine their frequencies, one should keep in mind that a significant part of the libraries could encode truncated proteins. A deeper sequencing analysis could be necessary for a better characterization of the libraries if needed in the future.

Additionally, we constructed the TM Library specifically designed to broaden the specificity of the enzyme for the acyl part of the substrate. Seven residues from the pocket of specificity of the enzyme have been degenerated by the cloning of a cassette. Even though the theoretical diversity of this library is of 1.28×10^9 distinct variants, the current screening techniques that we used did not permit to screen the total diversity. The library contains 2.27×10^7 different variants that correspond to the range of what could be effectively screened using microfluidic techniques.

The screening and sorting strategy was successfully validated on a model experiment. Wild type PAC was selected from mixes of wild type and F146L PAC with enrichment factors about 2,000 fold, which corresponds to the theoretical one. We focused then on the screening of libraries. HR-EP Library was firstly screened but we were not able to recover any interesting clones. From analysis of selected clones, we could only conclude that a contamination by the wild type gene was present and sufficiently active to be selected.

We should keep in mind the intrinsic drawbacks of such libraries. First, since multiple substitutions in a single codon are very rare, it biases the distribution of mutations type, and not all amino acid substitutions will be present in the library. Moreover, as the pac gene is quite large (2,500bp), the likelihood of advantageous mutations appearance is extremely low compared to neutral or deleterious ones. Error prone PCR also introduces at a non-negligible rate of non-synonymous mutations that are insertions, deletions or stop codons. Indeed, our chances of finding a mutant with an altered substrate specificity using error prone libraries might probably be quite low. Moreover, droplet screening in microfluidics do not reach substantially high throughput

levels. Only around 700,000 droplets can be screened in two hours. However, if days would be required to screen the entire diversity of such libraries at low DNA molecule levels in droplets, high enrichment factors enable to work at high plasmid number per droplet in a first screening round and get through large libraries in shorter time.

A second set of screening was performed on the TM Library. After two rounds of selection, sequence analysis also revealed the presence of the wild type sequence. Two selected clones bear the TML signature. One of them probably came from a recombination event with the wild type sequence, and the second possesses a 150 amino acids deletion in its sequence, and therefore could more than probably not be well folded and active. The existent enzymatic activity could come from a contaminant plasmid that could be discarded by transformation in order to ensure that the clone is not active anymore.

The maturation of the enzyme can be strongly impaired by mutations near the active site or in the specificity pocket. The α and the β subunits can be expressed separately, without the peptide linker. An ATG codon is inserted at the beginning of the β subunit sequence, and the resulting N-terminal Met will be removed autocatalytically and revealed the catalytic SerB1 residue. This strategy has been used for the optimal expression of PAC[23] but also for cephalosporin acylase.[23-25] Moreover, our strategy implies an *in vivo* DNA amplification step where "generalist" enzymes may be toxic and sufficiently expressed to be counter selected, whereas deleted clones could become predominant.

In the future, in order to favor screening of the most variety towards hydrophobic moieties, we could envision using VYS codons‡ instead of NNK. VYS encodes for seven hydrophobic amino acids (Leu, Pro, Ile, Met, Ala, Thr,

‡IUPAC nucleotide code: V = A or C or G, Y = C or T, S = G or C, K = G or T, N representing any base.

Val). The theoretical diversity of this library is of 823,000 clones, which can be easily screened in droplet based microfluidics. Moreover, a first screening for PAC native activity using a fluorogenic substrate specific for the wild type activity could be useful in order to remove all variants that are not properly folded or that do not maturate. Subsequently, a second screening for new activities can be performed.

In conclusion, the sorting strategy seems to work in regard of the model experiments. Libraries construction strategies could be modified in order to eradicate the presence of wild type sequence in too high proportion. Additionally, DNA recovery and amplification were extensively analyzed, but remain highly random and not efficient enough since the number of transformants was sometimes lower than the number of sorted droplets. In order to avoid any potential contamination during the amplification process, an amplification step in droplets before fusion with IVTT and substrates[8] could be a good alternative (see chapter 4). Since more material will be present in the droplets, DNA contained in the sorted droplets will be more easily recovered and its amplification will be less sensitive to contamination. Moreover, DNA amplification should strongly diminish the biological variance that can result from expression of single genes. PCR can be carried out in small droplets and subsequently fused to IVTT droplets. This process has been evaluated for other purposes and is described in more detail in chapter 4.

References

1. Spirin SA; *Cell-free translation systems*, Springer, **2002**

2. Swartz JR; *Cell-free protein expression*, Springer, **2003**

3. Spirin AS; *High-throughput cell-free systems for synthesis of functionally active proteins*. Trends Biotechnol **2004**; 22: 538-45

4. Tawfik DS, Griffiths AD; *Man-made cell-like compartments for molecular evolution*. Nature Biotechnology **1998**; 16: 652-656

5. Theberge AB, Courtois F, Schaerli Y, Fischlechner M, Abell C, Hollfelder F, Huck WT; *Microdroplets in microfluidics: an evolving platform for discoveries in chemistry and biology*. Angew Chem Int Ed Engl **2010**; 49: 5846-68

6. Kintses B, van Vliet LD, Devenish SR, Hollfelder F; *Microfluidic droplets: new integrated workflows for biological experiments*. Curr Opin Chem Biol; 14: 548-55

7. Courtois F, Olguin LF, Whyte G, Bratton D, Huck WT, Abell C, Hollfelder F; *An integrated device for monitoring time-dependent in vitro expression from single genes in picolitre droplets*. Chembiochem **2008**; 9: 439-46

8. Mazutis L, Araghi AF, Miller OJ, Baret JC, Frenz L, Janoshazi A, Taly V, Miller BJ, Hutchison JB, Link D, Griffiths AD, Ryckelynck M; *Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis*. Anal Chem **2009**; 81: 4813-21

9. Mazutis L, Baret JC, Treacy P, Skhiri Y, Araghi AF, Ryckelynck M, Taly V, Griffiths AD; *Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme*. Lab Chip **2009**; 9: 2902-8

10. Baret JC, Miller OJ, Taly V, Ryckelynck M, El-Harrak A, Frenz L, Rick C, Samuels ML, Hutchison JB, Agresti JJ, Link DR, Weitz DA, Griffiths AD; *Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity*. Lab Chip **2009**; 9: 1850-8

11. Alkema WB, Dijkhuis AJ, De Vries E, Janssen DB; *The role of hydrophobic active-site residues in substrate specificity and acyl transfer activity of penicillin acylase*. Eur J Biochem **2002**; 269: 2093-100

12. Pethig R; *Dielectrophoresis: using inhomogeneous AC electrical fields to separate and manipulate cells.* Critical Reviews in Biotechnology **1996**; 16: 331-348

13. Ahn K, Kerbage C, Hunt TP, Westervelt RM, Link DR, Weitz DA; *Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices*. Appl. Phys. Lett. **2006**; 88: 024104

14. Brakmann S, A. S; *Evolutionary Methods in Biotechnology: Clever Tricks for Directed Evolution*, Susanne Brakmann, Andreas Schwienhorst, **2005**

15. Rasila TS, Pajunen MI, Savilahti H; *Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment*. Analytical Biochemistry **2009**; 388: 71-80

16. Drummond DA, Iverson BL, Georgiou G, Arnold FH; *Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins*. Journal of Molecular Biology **2005**; 350: 806-816

17. Reetz MT, Kahakeaw D, Lohmer R; *Addressing the numbers problem in directed evolution*. Chembiochem **2008**; 9: 1797-1804

18. Alkema WB, Hensgens CM, Snijder HJ, Keizer E, Dijkstra BW, Janssen DB; *Structural and kinetic studies on ligand binding in wild-type and active-site mutants of penicillin acylase*. Protein Eng Des Sel **2004**; 17: 473-80

19. Oinonen C, Rouvinen J; *Structural comparison of Ntn-hydrolases*. Protein Sci **2000**; 9: 2329-37

20. Duggleby HJ, Tolley SP, Hill CP, Dodson EJ, Dodson G, Moody PC; *Penicillin acylase has a single-amino-acid catalytic centre*. Nature **1995**; 373: 264-8

21. Done SH, Brannigan JA, Moody PC, Hubbard RE; *Ligand-induced conformational change in penicillin acylase*. J Mol Biol **1998**; 284: 463-75

22. Griffiths AD, Tawfik DS; *Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization*. Embo Journal **2003**; 22: 24-35

23. Oh B, Kim M, Yoon J, Chung K, Shin Y, Lee D, Kim Y; *Deacylation activity of cephalosporin acylase to cephalosporin C is improved by changing the side-chain conformations of active-site residues*. Biochem Biophys Res Commun **2003**; 310: 19-27

24. Kim S, Kim Y; *Active site residues of cephalosporin acylase are critical not only for enzymatic catalysis but also for post-translational modification*. J Biol Chem **2001**; 276: 48376-81

25. Yoon J, Oh B, Kim K, Park J, Han D, Kim KK, Cha SS, Lee D, Kim Y; *A bound water molecule is crucial in initiating autocatalytic precursor activation in an N-terminal hydrolase*. J Biol Chem **2004**; 279: 341-7

Chapter 4: Activity Fed Enzyme Translation Assay "AFET", a new *in vitro* detection strategy for applications in droplets

Gabrielle Woronoff, Michaël Ryckelynck, Olivier Schicke, Andrew Griffiths, Patrice Soumillion

Manuscript in preparation

Introduction

Fluorogenic assays are widely used for enzyme screening. For example, coumarin or rhodamine based esters or amides are common for detecting esterase or amidase activity.[1-3] However, such assays can lead to undesirable drawbacks. Some enzymes may not accommodate easily a large aromatic fluorophore. Also, their use can lead to the selection of variants bearing a good activity for the fluorogenic substrate but not for the native substrate, as the chemical nature of the fluorescent leaving group promotes the instability of the amide or ester bond. To avoid the problem, coupled assays requiring several successive chemical or biochemical reactions have been developed. For example, Reymond et al. have modified the release chemistry of several umbelliferone based substrates: after the enzymatic cleavage, the product undergoes a β-elimination reaction in the presence of BSA which releases the fluorescent umbelliferone. Moreover, the cascade reaction moves the fluorophore away from the active site, rendering the substrate more appropriate to deeply buried active sites, or could be more similar to native substrates. Bornscheuer et al. use product-reactive 7-chloro or 7-hydrazino 4-nitrobenzofurazane (7-CNBF or 7-HNBF) to assay lipase, esterase or acylase activities. 7-CNBF reacts with amines resulting from amide hydrolysis to form a fluorescent product. Similarly, 7-HNBF reacts with acetaldehyde released from hydrolysis of vinyl ester by lipase or esterase.[4,5] Some activities can be followed by fluorescent pH indicator as it is the case for dehalogenase activity.[6] However, it is worth noting that the use of these tests might be limited for high throughput screening technologies such as conventional IVC in bulk mainly because substrates are often too hydrophobic[7,8] or reaction media cannot be easily modified when successive reactions are needed.

We propose here an auto amplification enzyme assay based on the liberation of an amino acid that can be used for the generation of both the enzyme itself and a fluorescent signal (Figure 4.1). The general strategy consists

in *in vitro* translation of a very small amount of enzyme whose activity on a specific substrate will generate an amino acid that will subsequently be used in the translation reaction for producing more enzymes and a fluorescent reporter protein. The detection reaction system we choose in this project is based on *in vitro* transcription and translation of the penicillin acylase (PAC) coupled to the signal of GFP, where the source of methionine is a phenylacetylated methionine.

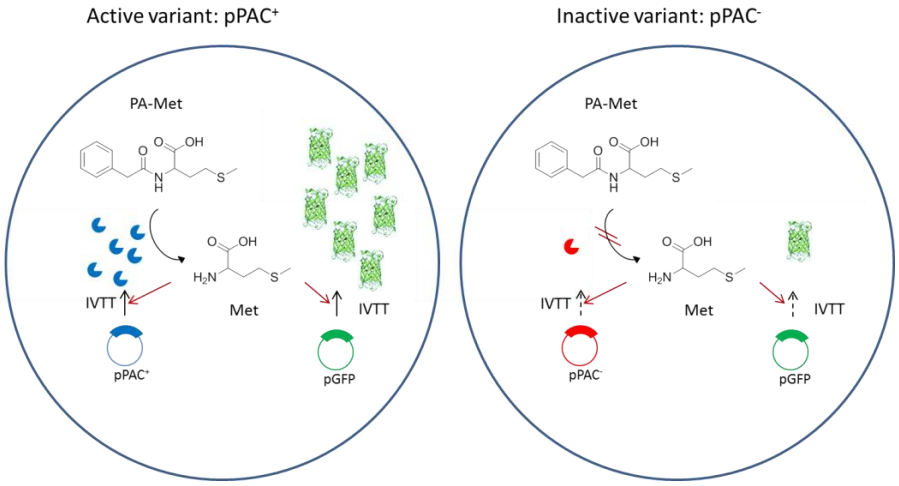
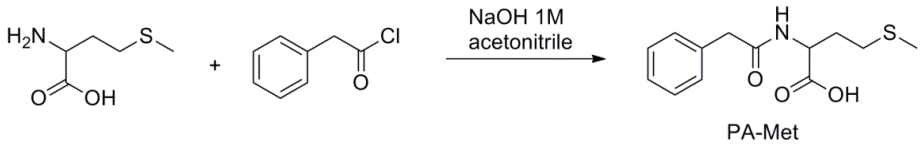


Figure 4.1: General strategy scheme. A. In the droplet containing an active variant, basal amount of methionine allows the synthesis of few PAC molecules, that hydrolyses PA-Met in Met which will be used for the translation of both PAC and the reporter gene GFP, resulting in an auto amplification of PAC and a concomitant increase of the green fluorescence in the droplet. Conversely, if the PAC variant is not able to hydrolyze the substrate (i.e. PA-Met), no methionine will be delivered to the IVTT mixture and no green fluorescence will be detected in the droplet.

Results and Discussion

Substrate synthesis

Phenylacetylated methionine (PA-Met) was synthesized under Schotten-Baumann conditions. Addition of phenylacetyl chloride and acetonitrile to L-methionine in 1M NaOH afforded, after few steps of purification, phenylacetylated methionine (PA-Met) in 14 % yield (Scheme 4.1). PA-Met will be used as substrate for penicillin acylase in *in vitro* transcription and translation (IVTT) experiments.



Scheme 4.1: Synthesis of the phenylacetylated methionine under Schotten-Baumann conditions.

Assays in solution

Preliminary tests

The IVTT feeding by the PAC was tested under various conditions. The production of PAC or GFP by IVTT with or without methionine, with or without PA-methionine was initially followed. Assays were carried out in the presence of either a plasmid bearing the pac gene (pPAC), the gfp gene (pGFP), or both plasmids. The concentrations of PAC obtained under the three conditions of methionine supply are reported on the upper graph of Figure 4.2. As expected, the addition of Met (2 mM) led to a significant increase in the production of PAC. When the reaction was performed in the presence of PA-Met (2 mM) a significant amount of PAC was also obtained. In a second set of experiments we used the GFP as reporter. In the presence of pGFP only (Figure 4.2, lower graph, light grey), significant green fluorescence is only

present when Met is added. The addition of pPAC to the reaction (Figure 4.2, lower graph, dark grey) enables the production of Met and its use for translation of GFP. These two experiments demonstrate that Met can be produced through PAC activity and use for translation of both proteins. More generally, even though the reproducibility of the IVTT experiments has to be improved, the feasibility of an indirect detection system based on a fluorescent reporter was demonstrated.

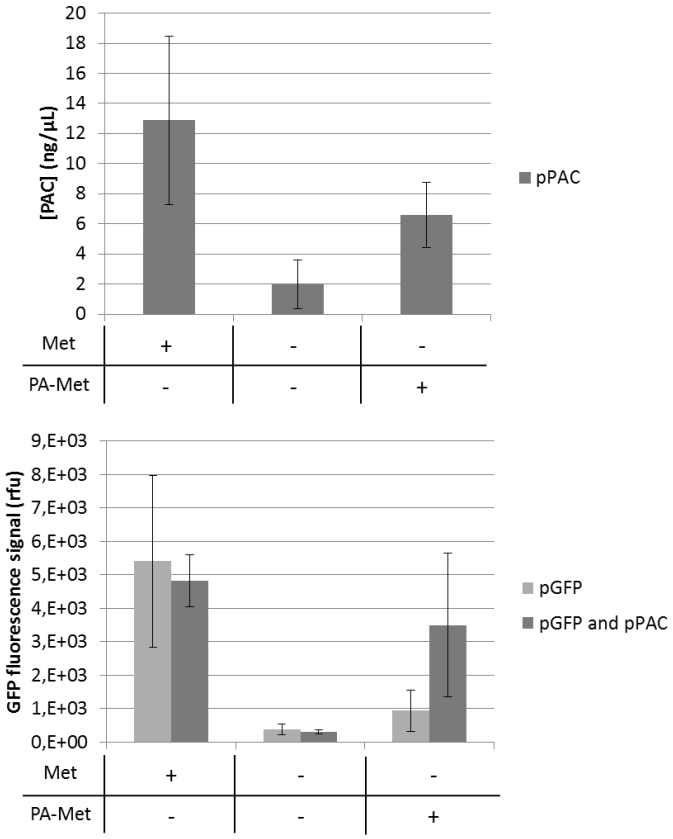


Figure 4.2: Translation efficiency using different methionine sources. Upper graph: PAC concentrations produced by IVTT on pPAC with or without Met or PA-Met, assayed by enzymatic activity on NIPAB. Lower graph: GFP fluorescence produced by IVTT in presence of pGFP only (light grey) or pGFP and pPAC (dark grey), with or without Met or PA-Met. Error bars represent the standard deviations from the mean value of three measurements.

The kinetic of PAC synthesis was also monitored over time in the presence of Met or PA-Met and the results are displayed in Figure 4.3. In both cases, a production of PAC was observed, but around twice the time was needed to produce the same amount of penicillin acylase with PA-Met than with Met reflecting the auto feeding process. The maximal level of PAC production is also lower in the reaction with PA-Met. This difference might be attributed to the fact that IVTT is subjected to heat and time inactivation.

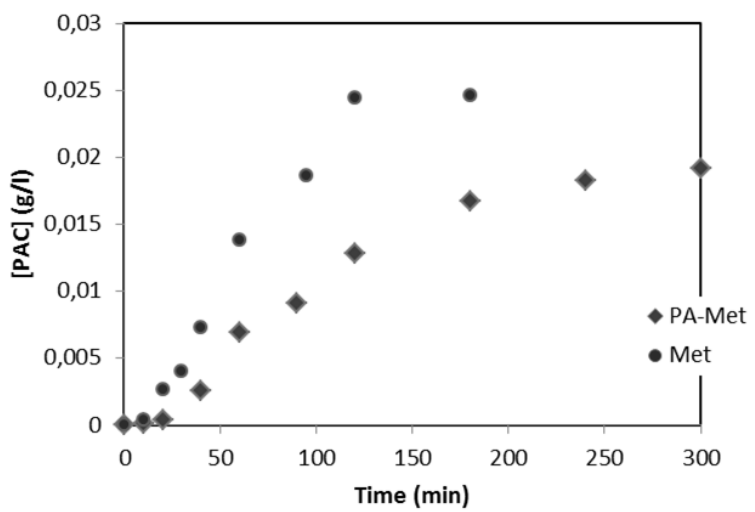
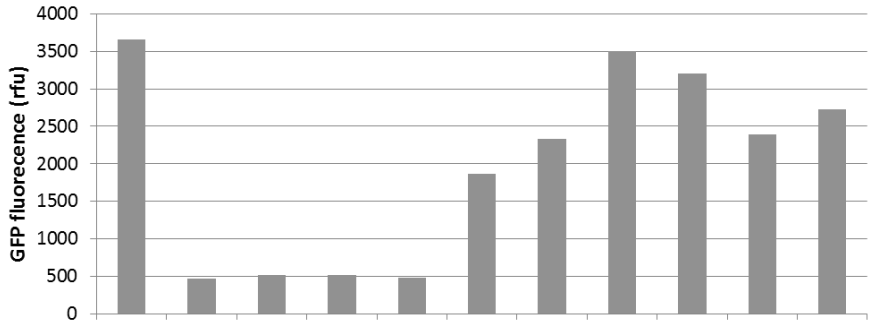


Figure 4.3: Appearance of PAC activity in IVTT in the presence of PA-Met (squares) or methionine (circles). Initial slopes resulting from the hydrolysis of NIPAB were used to calculate PAC concentration.

Indirect assay set up for PAC activity

We next optimized the conditions to perform the indirect assay of PAC activity. In the presence of two different genes, competition for transcription and/or translation may occur.[9] Therefore, the ratio of pPAC/pGFP has to be optimized. IVTT reactions were performed with different ratios of pPAC/pGFP plasmids and the GFP fluorescence was measured in a fluorimeter (Figure 4.4). A fixed concentration of pGFP (250 ng/25 μL of total volume corresponding to 4.8 x 10^4 molecules per 22 pL droplets) was mixed

with an increasing concentration of pPAC ranging from 7.4 x 10^{-3} ng/25 µL (corresponding to one gene per 22 pL droplet) to 250 ng/25 µL (3.4 x 10^4 molecules/22 pL droplet). In addition, three controls samples were performed where an IVTT/pGFP mixture was not supplemented with methionine source or was supplemented with methionine, PA-Met. All samples were incubated over night at 37°C before measuring the fluorescence signal of GFP using a fluorimeter (excitation/emission wavelengths: 488/520 nm). As shown on Figure 4.4, the presence of Met gave a fluorescence 7 times higher than without methionine source or with PA-Met. Concerning the two plasmids assays, GFP was synthesized from PA-Met with pPAC concentrations of at least 7.4 x 10^{-1} ng/25 µL, corresponding to 100 genes per 22 pL droplets.



	1	2	3	4	5	6	7	8	9	10	11
Met	+	-	-	-	-	-	-	-	-	-	-
PA-Met	-	-	+	+	+	+	+	+	+	+	+
pPAC (ng/25µL)	0	0	0	7.4 x 10^{-3}	7.4 x 10^{-2}	0.74	7.4	10	50	74	250
λ (gene /droplet)	0	0	0	1	10	100	1000	1400	6800	10^4	3.10^4

Figure 4.4: *In vitro* assay for transcription and translation of GFP and PAC simultaneously, with increasing amounts of pPAC in presence of phenyl acetylated methionine (PA-Met). pPAC quantities are given either in ng in 25 µL of reaction volume, or in gene per 22 pL droplets (λ) (theoretical calculation). 1 is the positive control (pGFP + methionine) and 2 the negative control (pGFP without extra methionine supply). From samples 3 to 11, increasing quantities of pPAC was added to a fixed quantity of pGFP (250 ng/25 µL).

The GFP fluorescence was also followed *in situ* in real time on mixtures containing 50 or 250 ng of pGFP and 3, 10 or 50 ng of a plasmid encoding the wild type PAC (pPAC) or a less active mutant (pF146L,[10] 0.5 % of activity compared to the wild type)(Figures 4.6 and 4.7). Control experiments are presented in Figure 4.5. As a general trend (Figure 4.6), a biphasic GFP signal appearance over time is observed with a slow phase followed by a faster phase that progressively slow down for reaching a plateau. The initial slow phase was attributed to the GFP produced from the basal Met present in the IVTT mixture, while the fast phase was attributed to the protein produced from Met released from PA-Met after PAC action. The intersection between the tangents extrapolated from these two first phases was used to assign the response time. Response times are given in Table 4.1. Three major traits were observed: first, response times increased with decreasing amounts of PAC plasmid (pPAC or pF146L). Second, response times are shorter with the wild type PAC than with the mutant, suggesting that the assay could be suited for libraries screening purposes where different variants have to be distinguished according to their activities. The effect was more pronounced for slower reactions which reflects the autoamplificative process that may be slowed down by the IVTT inactivation over time. Indeed, IVTT is subjected to inactivation that can results from heat sensitivity of the IVTT, substrate depletion, inhibition by products, etc. Third, response times are shorter with 50 ng of pGFP than with 250 ng. This was attributed to a competition between pGFP and pPAC, an effect that could be used for controlling the time response of the system.

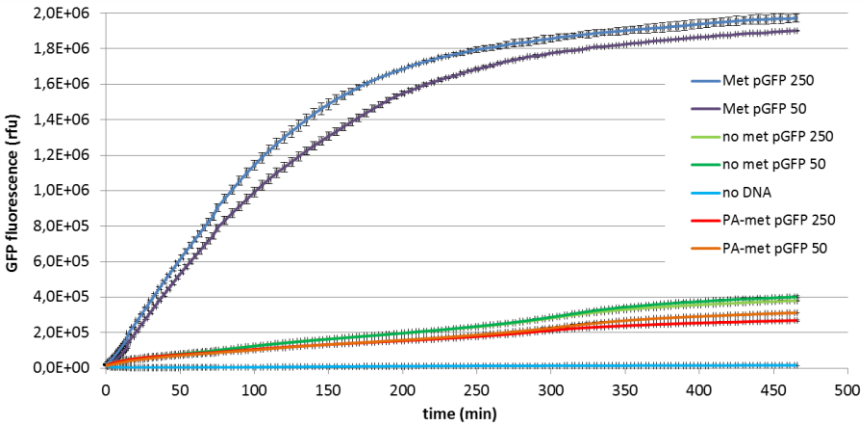
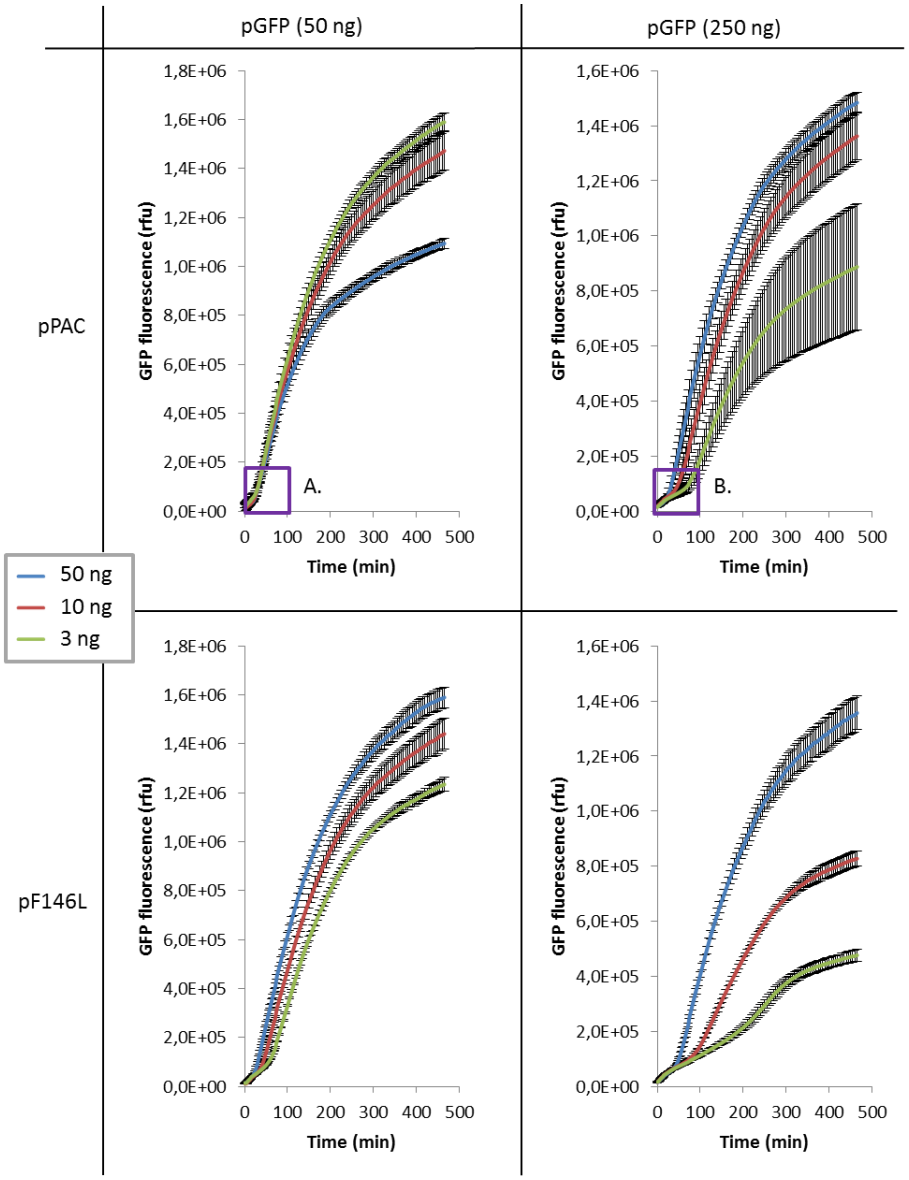


Figure 4.5: Controls of the experiment demonstrating the time response fine-tuning of the indirect assay. GFP fluorescence signal measured over time. 25 µL IVTT reaction containing: 1.dark blue: 250 ng of pGFP and Met, 2. purple: 50 ng of pGFP and Met, 3. light green: 250 ng of pGFP, 4. green: 50 ng of pGFP, 5. light blue: no DNA, 6. red: 250 ng of pGFP and PA-Met, 7. orange: 50 ng of pGFP and PA-Met. Error bars represent the standard deviations from the mean value of three measurements.

		Response time (min)					
		pPAC			**pF146L**		
		3 ng	10 ng	50 ng	3 ng	10 ng	50 ng
pGFP	50 ng	29	24	22	44	29	17
	250 ng	75	45	25	177	87	50

Table 4.1: Response times in minutes calculated from the curves presented at Figure 4.6.


pGFP (50 ng)
pGFP (250 ng)
pPAC
pF146L
50 ng
10 ng
3 ng
GFP fluorescence (rfu)
Time (min)
A.
B.
1,8E+06
1,6E+06
1,4E+06
1,2E+06
1,0E+06
8,0E+05
6,0E+05
4,0E+05
2,0E+05
0,0E+00
0
100
200
300
400
500

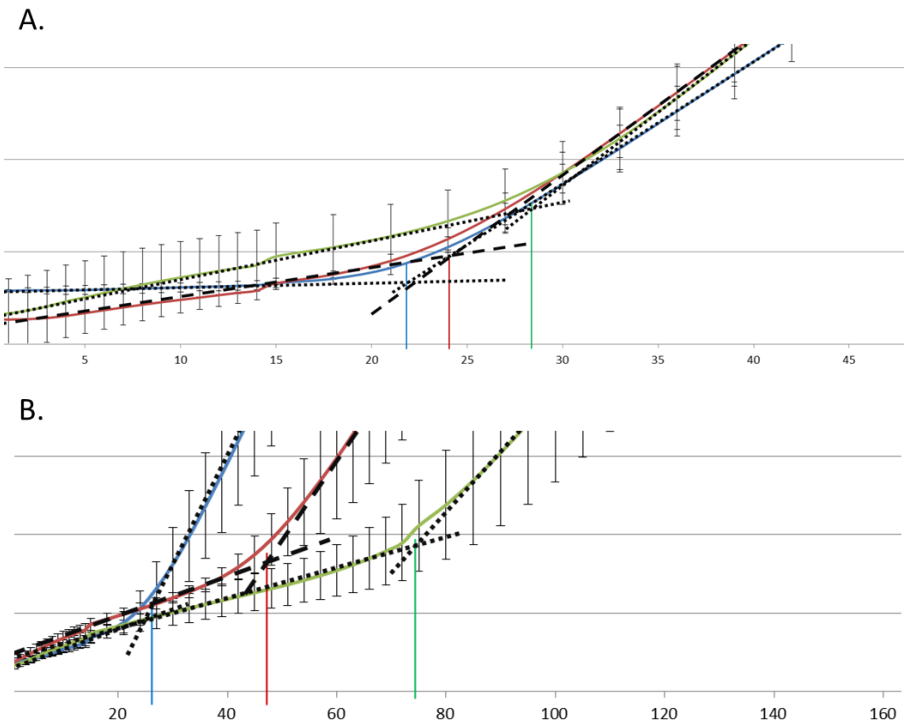


Figure 4.6: GFP fluorescence appearance in IVTT assay of PAC and GFP with PA-Met as methionine source. GFP signal was monitored over time in IVTT reactions (25 µL) containing pGFP (250 or 50 ng) and pPAC (3, 10 or 50 ng) or pF146L, coding for F146L mutant of PAC (3, 10 or 50 ng). Details of the response times determination from boxes A. and B. are shown in the second part of the figure. Error bars represent the standard deviations from the mean value of three measurements.

The fluorescence intensity reached at the plateau seems to decreases for slower reactions, an observation that could be explained by the sensitivity of IVTT to heat and time inactivation. Indeed, a correlation exists between the reached plateau and the response time (Figure 4.6). However, results obtained with the set of experiments with of 50 ng of pGFP with 3, 10 or 50 ng of pPAC were more ambiguous. In this case, response times were much closer to each other, therefore an IVTT inactivation cannot be invoke to explain variations in plateau intensities. Especially for the reaction with 50 ng of pPAC (circled star in Figure 4.6). A hypothesis is that an important PAC activity may have an inactivating effect on the IVTT (such as hydrolyzing the Phe-tRNA or

inhibition by phenylacetic acid). We reached the limit where auto feeding is no more rate limiting.

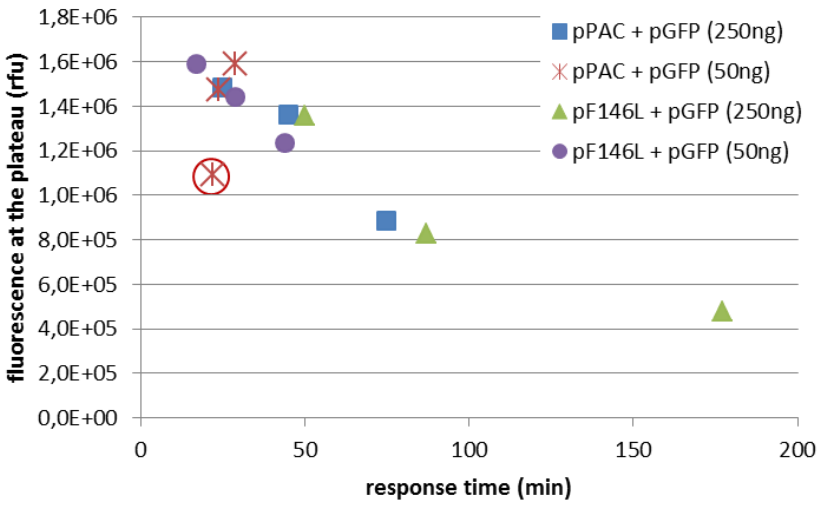
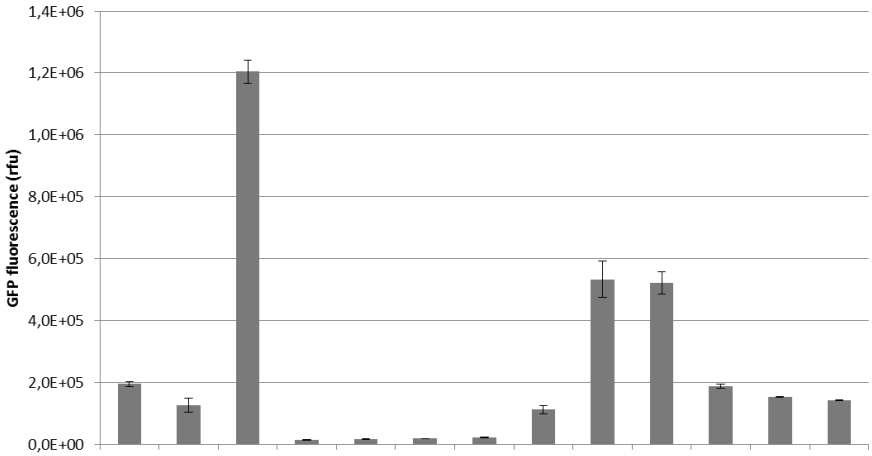


Figure 4.7: Green fluorescence signal at the plateau (465 min) in function of the response time, data coming from the experiment detailed in Figure 4.6.

Having established that our indirect fluorescent assay worked and can be optimized by adjusting plasmids relative concentrations, we wondered if the assay was still efficient at low plasmid concentrations in order to adapt the assay at a droplet-based microfluidic format. To work at single gene per droplet of 22 pL, concentration of pGFP and pPAC have to be dropped to 0.2 pg/µl and 0.3 pg/µl respectively. As before, IVTT experiments were performed without methionine supply, with Met or with PA-Met and GFP fluorescence was monitored. However, no significant difference between negative and the positive controls was observed. Since raising the concentration of pGFP increases the competition with pPAC and working with very low pGFP concentrations is not generating fluorescence signal, we reasoned that an intermediate pGFP concentration may allow the autoamplification process with a concentration of pPAC corresponding to a single gene per droplet.

IVTT with a fixed concentration of 0.3 pg/μL pPAC and increasing concentrations of pGFP in the presence of PA-Met were incubated at 37°C and GFP fluorescence was measured over 8 hours. Only two samples reached fluorescence intensities higher than the negative control (IVTT reaction with PA-Met and pGFP (250 ng) without pPAC). Significant fluorescence signals were observed with 2 and 5 ng of pGFP (λ values of 400 and 1,000 respectively) but not with lower or higher concentrations. Hence, with the optimized amount of reporter gene, the assay should be functional in droplets at one enzyme gene per compartment.



	1	2	3	4	5	6	7	8	9	10	11	12	13
Met	-	-	+	-	-	-	-	-	-	-	-	-	-
PA-Met	-	+	-	+	+	+	+	+	+	+	+	+	+
pPAC (λ)	0	0	0	1	1	1	1	1	1	1	1	1	1
pGFP (ng/25μL)	250	250	250	0	2×10^{-3}	0.01	0.02	0.2	2	5.1	51	145	250
pGFP (λ)	4.8×10^4	4.8×10^4	4.8×10^4	0	0.4	2	4	40	400	1000	10^4	2.8×10^4	4.8×10^4

Figure 4.8: *In vitro* assay for transcription and translation of GFP and PAC simultaneously, with increasing amounts of pGFP in presence of phenyl acetylated methionine (PA-Met) and very low pPAC concentration (λ = 1; 3 x 10-4 ng/μL). pGFP quantities are given either in ng in 25 μL of reaction volume, or in gene per 22 pL droplets (λ) (theoretical calculation). Error bars represent the standard deviations from the mean value of three measurements. Error bars represent the standard deviations from the mean value of three measurements.

In conclusion, even though previous works reported that GFP was efficiently expressed from one gene in 0.6 to 1.1 pL droplets using the same plasmid and the same IVTT kit (RTS 100 E. coli HY kit, Roche),[11] production of GFP at $\lambda = 1$ was not detected in our case. Nevertheless, we demonstrated that adjusting the ratio pGFP and pPAC allows controlling the output and optimizing the reaction for microfluidic experiments at one enzyme gene per droplet.

Experiments in droplet format

pGFP expression in droplets

After its validation in bulk, the enzymatic assay was evaluated in droplet-based microfluidics. Droplets production was performed on a microfluidic device made of two aqueous inlets, one oil inlet and one outlet (Figure 3.2). The IVTT mix was split into two solutions, the first composed of the *E. coli* lysate, buffer, and methionine supply and the second contained the reaction mix, amino acids (all but Met) and the DNA (diluted in total RNA if low concentrations are used, i.e. one gene per droplet). We first performed *in vitro* expression of GFP (10ng/μL of pGFP) either without methionine supply, or with PA-Met or with Met. After an overnight incubation at 37°C in capillaries, emulsions were reinjected in a reinjection device for analysis and the fluorescence (GFP) content of each droplet was measured (Figure 4.9). Data obtained were consistent with bulk experiment, with an efficient production of GFP obtained only in the presence of Met (~40 x increase in the fluorescence signal compared with negative control).

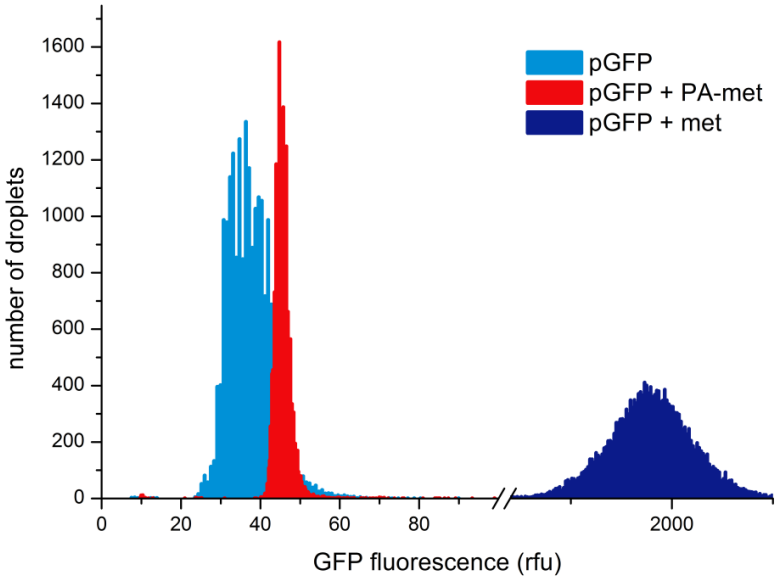


Figure 4.9: Analysis of populations of droplets containing pGFP with methionine (dark blue), or phenyacetylated methionine (red) or no methionine supply (light blue). 18,000 droplets were analyzed per emulsion.

GFP and PAC expression in droplets at the single gene level

The co-expression of PAC and GFP in the same droplet, previously tested in bulk, was transferred to droplets. First, we tested Met generation from PA-Met and its incorporation in the *in vitro* translation process in the presence of the pPAC vector. An emulsion containing 0.67 molecules of pPAC and 30 molecules of pGFP per 22 pL droplet was created using the coflow device described above. DNA concentrations were calculated so that each droplet contained ~30 pGFP, and 50 % of the droplets were occupied by 1 or more pPAC molecule. The aqueous phase was supplemented with PA-Met (2 mM) as sole methionine source and PA-ACMS (100 mM) as fluorogenic probe for penicillin acylase activity. After 4 hours of incubation at 37°C in a capillary, the emulsion was reinjected and the blue fluorescence of droplets (ACMS fluorescence) was analyzed and used to assay PAC activity. Assuming a Poisson distribution, only 48 % of the droplets should contain one or more pPAC gene

when a λ value of 0.67 is used. As expected, the analysis of the emulsion reveals the presence of two populations of droplets: one initially containing a molecule of pPAC or not, the former representing the expected 48 % (Figure 4.10). This demonstrates that the autoamplification process is effective at one enzyme gene per droplet. The green fluorescence of the emulsion was not analyzed since an important cross-talk exists between the detection channels of GFP and ACMS fluorophores (the blue fluorescence is strongly detected in the green fluorescence measurement channel).

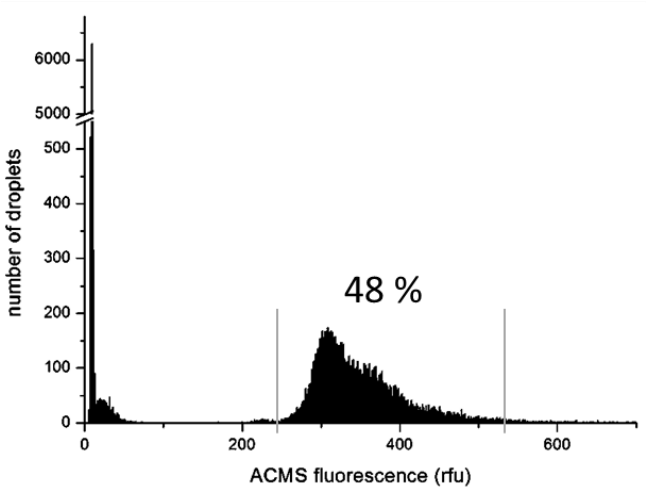


Figure 4.10: Blue fluorescence analysis of an emulsion containing 0.67 pPAC and 27 pGFP, IVTT, PA-Met and PA-ACMS.

Next, the most optimal ratios of pPAC/pGFP deduced from the previous bulk experiments (Figure 4.8) were transferred into droplets. Several amounts of pGFP (λ = 200, 400 and 5,000) were encapsulated with pPAC (λ = 0.3), IVTT mix and PA-Met. Droplets were incubated over night at 37°C, and reinjected for analysis of the green fluorescence. Results are presented at Figure 4.11. A positive population of about 20 % is present in emulsions containing 200 and 400 pGFP per droplet, as expected since pPAC is present at λ = 0.3. On the contrary, no positive population is present with 5,000 pGFP per droplet. The obtained results are consistent with the bulk experiment

presented at Figure 4.8 where the best pGFP concentrations seemed to be around of 400 and 1,000 genes per droplet.

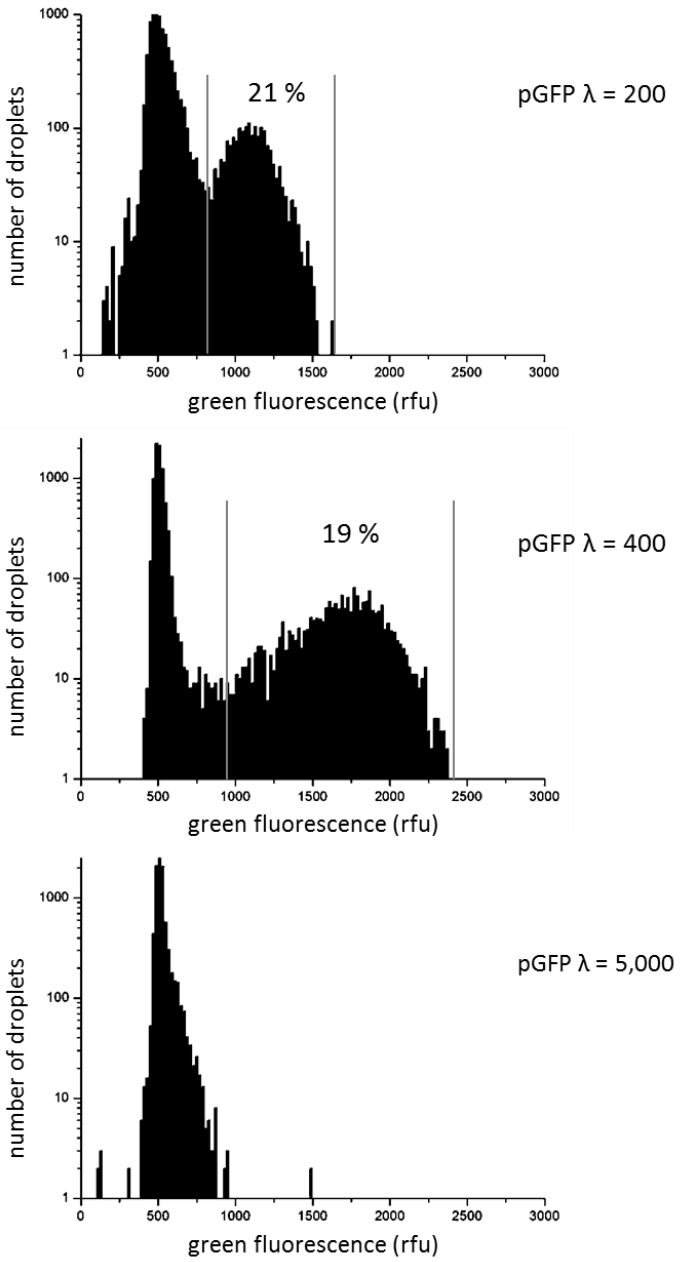


Figure 4.11: Analysis of populations of 22 pL droplets containing IVTT, PA-Met, 0.3 pPAC per droplet and increasing quantities of pGFP (200, 400 and 5,000 genes per droplet). Emulsions were incubated over night at 37°C and reinjected for analysis.

Uncoupling IVTT expression of PAC and GFP

The pGFP plasmid was provided by Dr. F. Courtois who previously used it for *in vitro* translation of GFP in 0.6 to 1.1 pL droplets using the same IVTT kit as we used (RTS 100 HY *E. coli* kit, Roche). It was shown that in droplets, one GFP gene produce in average about 10,000 molecules of GFP.[11] However, in our case, we didn't find conditions allowing for detectable GFP expression while working at 1 pGFP per 22 pL droplet. Previous section describes the successful co-expression of PAC and GFP at single pPAC gene per droplet that was archived thanks to the screening in bulk of several ratios of genes. An alternative approach was also explored where two consecutive IVTT reactions were performed. A first one aimed at producing PAC through the autoamplification process using PA-Met as methionine source. This first mixture was then used to "inoculate" a second IVTT mixture aiming at producing GFP using Met release through the action of PAC produced in the first reaction.

The first IVTT (pPAC-based) was performed in small droplets (~2 pL) using a reaction mixture supplemented with PA-Met, 0.45 pPAC molecule per droplet (1.4×10^{-3} ng/μL) and Texas Red (Sulforhodamine 101, hydrolsoluble fluorohore) as tracing agent of small droplets, and the emulsion was incubated for 1 hour at 37°C. Penicillin acylase was only produced in those compartments containing initially one or more pPAC molecule; representing approximately 35 % of the total population, assuming a Poisson distribution. After incubation, small droplets were reinjected into a fusion device and fused in an electro-coalescence device (Figure 4.15C), with 20 pL droplets containing fresh IVTT mix, PA-Met and the reporter gene (pGFP, 10 ng/μL, λ = ~50,000). Images analysis revealed that approximately 70 % of the droplets were fused one to one, less than 2 % were double fused (2 small droplets with one big) and ~30 % of the 20 pL droplets stayed unfused. The fresh emulsion was collected in a capillary and incubated 4 h at 37°C. The emulsion was then reinjected and

red and green fluorescence of each droplet was monitored using the optical set up described on Figure 5.2 (experimental section). The Texas Red present in the 2 pL droplets was used to quantify the fusion efficiency.

Analysis of the emulsion after incubation revealed that 32 % of the droplets were found to be unfused (Figure 4.12 box i.) as they had low red fluorescence (no Texas Red coming from a small droplet) and, as expected, did not present significant GFP fluorescence (green fluorescence). Amongst de remaining fused droplets, 69 % possessed a similar level of green fluorescence while having a significantly higher red fluorescence indicating that this population corresponds to large droplets fused with 2 pL droplets (high red fluorescence) that did not contained a pPAC molecule since no GFP was produced, therefore no PA-Met was converted into Met. A second population (~31 % of the fused population) of single-fused droplets with an identical red fluorescence but a significant increase of the green fluorescence was observed. These droplets resulted from the fusion of large droplets with 2 pL droplets that contained one or more pPAC molecule.

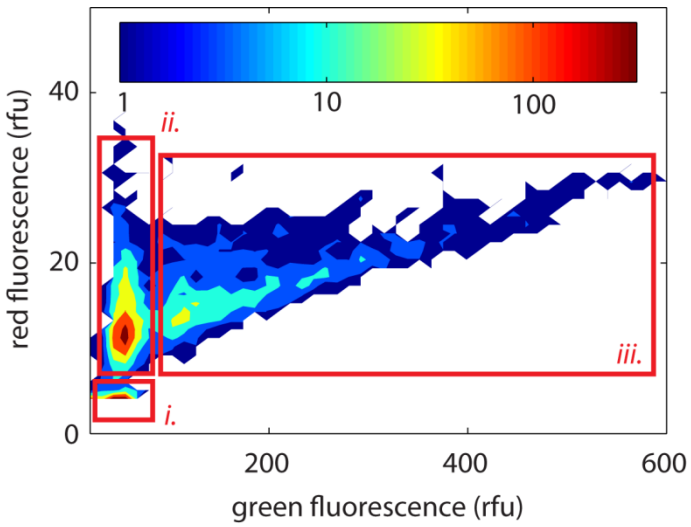


Figure 4.12: Green and red fluorescence analysis of a fused emulsion. of 2 pL droplets containing pPAC (λ = 0.67), IVTT, PA-Met, and Texas red (red fluorophore) incubated 1 hour at 37°C were fused 20 pL droplets containing IVTT, pGFP and PA-Met, and subsequently incubated at 37°C for 4 hours before reinjection for analysis. Droplets are plotted on the graph in function of their green fluorescence (GFP signal, x axis, divided in 50 bins) and red fluorescence (Texas red fluorescence, y axis, divided in 200 bins). The color code indicated the number of droplets per bin. i. Population of unfused 20 pL droplets (32 % of the analyzed droplets). ii. Population of fused 22 pL droplets without PAC activity (44 %). iii. Population of fused 22 pL droplets with PAC activity (20 %).

Despite this encouraging result, it should be noted that this approach was technically challenging because the formation of precipitate during the incubation of small droplets was often observed. This led to obstruction of channels of the microfluidic device and/or promoted coalescence between the small droplets upon reinjection, therefore leading to a significant failure rate. The experiment presented here was the sole successful one in eight attempts. To overcome this limitation, we explored another approach where DNA was amplified in droplets prior to protein expression.

DNA amplification in droplets

The alternative strategy consists in an intitial amplification of few DNA molecules of pPAC and pGFP in small droplets, then fusion of these droplets with larger droplets containing IVTT mix and PA-Met, followed by the incubation and analysis of the droplets.

PCR conditions were first tested in bulk format using the Dream Taq PCR system (Fermentas) supplemented with Pluronic F68 0.1 %, a non-ionic biocompatible surfactant (Sigma-Aldrich), helps to stabilize the emulsion and prevent coalescence of droplets during thermocycling (Faith Coldren, personal communication). A first set of experiments was done in bulk in order to control the amplification biases between the two plasmids that will be present at the same time in the droplet. Indeed, both target genes have respective sizes of 3,000 bp (PAC) and 1,500 bp (GFP). To limit a bias in the favor of the smallest molecule, elongation steps were lasted for 6 minutes. pPAC and pGFP were diluted such as 2 pL volumes contained no more than one pPAC molecule (3 pg/μL) while containing ~25 pGFP molecules (56 pg/μL).

Starting with pGFP concentration of 25 molecules per 2 pL ($\lambda = 25$) allowed for producing amplification product of the expected 1,500 bp size (Figure 4.13A, lanes 1 and 2). PCR amplifications were also performed on pPAC using starting concentrations of 1,000 (lane 3 on Figure 4.13A) or 1 molecule (lane 4 on Figure 4.13A) per 2 pL. In both cases, a band with the expected 3,000 bp was obtained. Therefore, PCR has the required sensitivity to amplify DNA starting from single pPAC or 25 pGFP in 2 pL droplets. Finally, a PCR was performed starting from a mixture of both molecules in a ratio of 1 pPAC for 25 pGFP (lane 5 on Figure 4.13A). Both bands were obtained with the expected size and a rough estimate of the ratio was in the expected range, indicating that no significant bias occurred during the amplification. However, unwanted lower intensity bands were also present. Such parasitic bands could

arise from non-specific primer annealing but testing various annealing temperatures ranging from 55°C to 69°C did not change the result (data not shown).

The PCR was then performed in droplets. pPAC and pGFP were encapsulated into 2 pL droplets (λ = 0.35 and 25 respectively) in the presence of a DNA polymerase, dNTPs, primers and Pluronic F68 0.1 % (picture of the emulsion creation at Figure 4.14A). The emulsion was collected in a PCR tube designed by Pekin et al.[12] to efficiently collect, thermocycle and reinject the emulsion while ensuring a good thermostability of the droplets (see Figure 4.15B). Part of the emulsion was then broken and analyzed on gel (Figure 4.13B). Both bands were observed and the pGFP/pPAC ratio was roughly estimated at a ratio pGFP/pPAC of 100. This fitted our expectations since, for a λ value of 0.35 and assuming a Poisson distribution of the pPAC molecule, a theoretical occupancy of 29 % was expected. Therefore, the ratio between pGFP and pPAC has to be 25/0.29 = 186. Interestingly, extra bands due to non-specific amplification were no more observed when the amplification was performed in droplets, a phenomenon already observed in the laboratory with other systems.

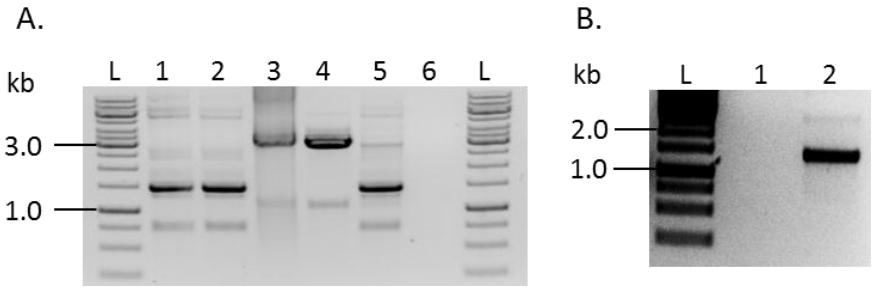


Figure 4.13: A. Agarose gel of a bulk PCR. Lane 1: pGFP, 2.5ng of template; lane 2: pGFP, 1.4 ng (equivalent concentration to lambda 25 in 2 pL droplets); lane 3: pPAC, 80 ng; lane 4: pPAC, 0.081 ng (eq. cc. to lambda 1 in 2 pL droplets); lane 5: 1.4 ng of pGFP and 0.081 ng of pPAC (competition conditions); lane 6: negative control, PCR without template. B. Lane 1: negative control (PCR without DNA sample), lane 2: PCR in emulsion containing pGFP (λ = 25) and pPAC (λ = 0.3)

In vitro PAC synthesis yield starting from PCR fragments was compared to IVTT yields starting from plasmids and was evaluated to be about 4 times lower. Knowing that PAC activity is detected from IVTT at one plasmid per droplets, production from multiple PCR copies should easily allow PAC detection.

We next mimicked the fusion process where 2 pL PCR droplets should be fused with 20 pL IVTT droplets. To do so, 25 µL of IVTT reactions were performed in bulk using 2.5 µL of PCR products (GFP only or a mixture GFP/PAC) from the experiment described above either in the absence of additional methionine supply or in the presence of Met or PA-Met (Figure 4.14). Reactions were incubated 4 hours at 37°C and their GFP content was determined by measuring its fluorescence. Controls with and without methionine showed ratios of 7 and 11 in presence of GFP genes or mix of PAC and GFP genes respectively. In the presence of PA-Met and PAC genes, the amount of produced GFP was ~4 times higher than without methionine supply (Figure 4.14), while the fluorescence was identical to the negative control in the absence of PAC. This demonstrates that samples can be discriminated as a function of their PAC activity, confirming the feasibility of the assay using PCR products.

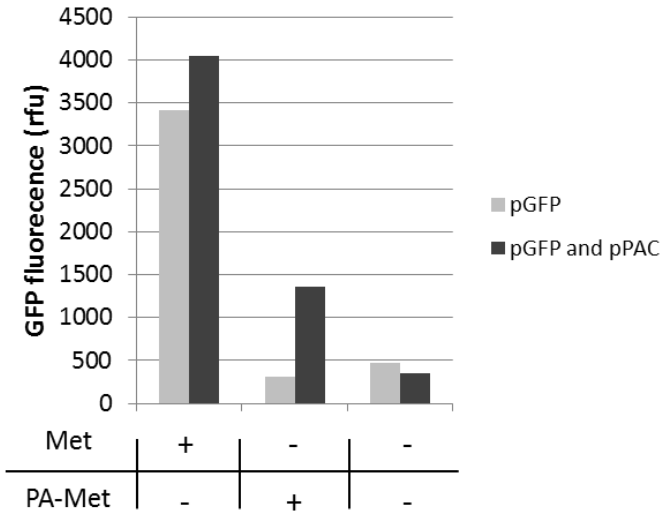


Figure 4.14: GFP fluorescence signal coming from IVTT in bulk on PCR products.

Fusion of thermocycled droplets to IVTT and analysis

Small droplets (2 pL) containing initially 25 pGFP and 0 or 1 pPAC (λ = 0.35: 70 % of empty droplets, 25 % of droplets containing 1 plasmid, and 5 % containing 2 or more plasmids, therefore occupancy = 30 %) were produced and collected in a PCR tube (as shown on Figure 4.15B).[12] After thermocycling, the emulsion is reinjected into the synchronization/fusion microfluidic device (Figure 4.15 C to F) and fused to 20 pL droplets containing IVTT mixture and PA-Met. To optimize droplet pairing, we maintained a small excess of large droplets (~10 %) by adjusting the flow rates and used a voltage of 450 V. Interestingly, from image analysis revealed that even if double paired droplets (one large and two small droplets) pass between electrodes, only one of the small droplets fuse with the large one, and the second small droplet remains unfused. The general scheme of the experiment is depicted in Figure 4.16.

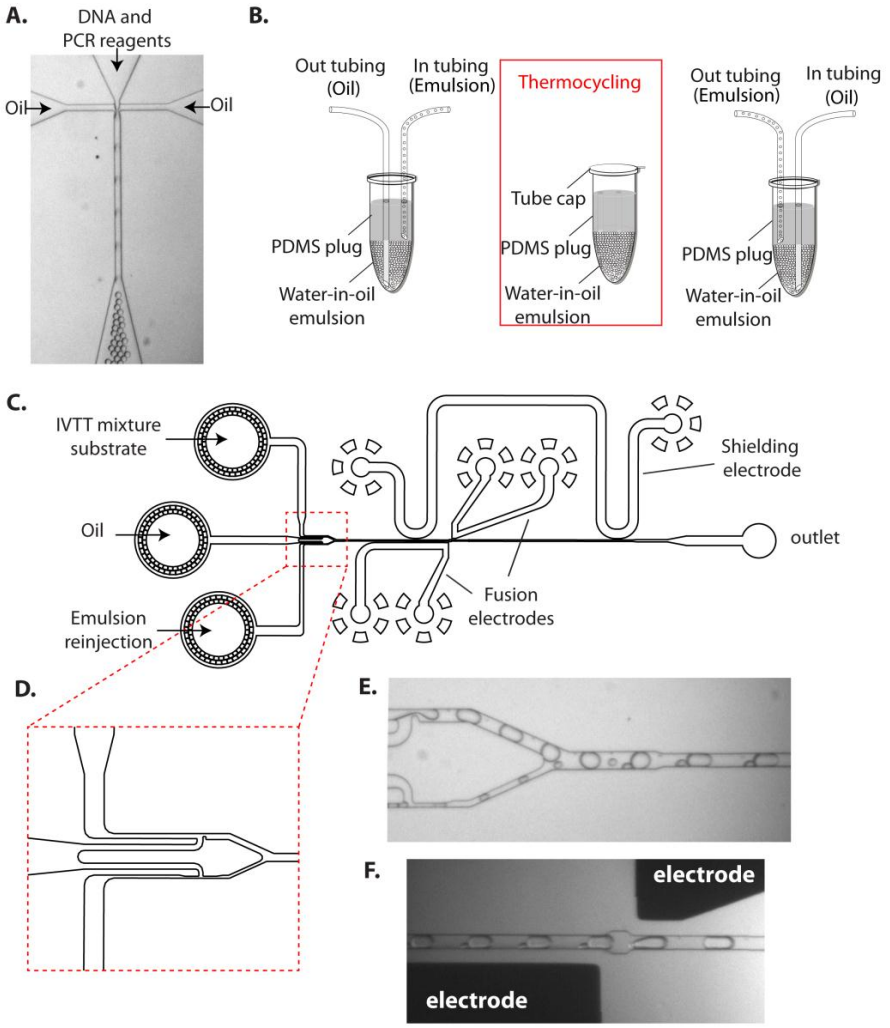


Figure 4.15: Overview of the microfluidic system. A. The aqueous phase containing the DNA and PCR reagents is emulsified I a microfluidic device. B. The emulsion is collected in a PDMS-sealed PCR tube and thermocycled.[12] C. Scheme of the fusion microfluidic device. IVTT mix and fluorinated oil + surfactant are injected at the upper and middle entries, while the thermocycled emulsion is reinjected at the lower entry. The region described in D. show a scheme of the emulsification and droplets pairing region (Picture of this region in E.). A field is applied at the fusion electrodes (C.) and fusion between coupled droplets occurs in the fusion chamber (picture in F.).

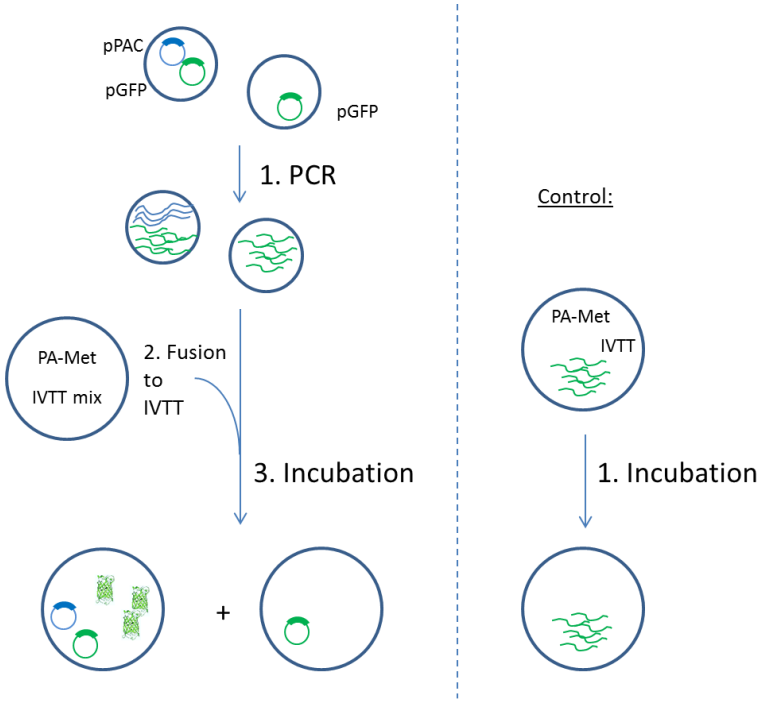


Figure 4.16: General scheme of the model experiment. A first emulsion of 2.3 pL droplets were created and contains 0.35 pPAC and 30 pGFP per droplet; plus all components required for emulsion PCR. After thermocycling, small droplets are fused to bigger droplets containing IVTT and PA-Met as methionine source; and will result after incubation in two different droplets populations: one with GFP signal and one without, corresponding to droplets containing the PAC gene or not. The negative control emulsion containing IVTT, PCR amplicons of the GFP gene and PA-Met and is not supposed to have any fluorescent signal.

In addition, a control experiment where small droplets contained only PCR amplicons of pGFP and large droplets contained IVTT mixture supplemented with PA-Met) was performed. Emulsions (assay and control) were collected in capillaries, incubated over night at 37°C; and subsequently reinjected for GFP fluorescence analysis. Droplets fluorescence distributions are presented in Figure 4.17 (A: control, B: sample).

As expected for the control, the use of PA-Met as only source of methionine did not allowed a significant synthesis of GFP (Figure 4.17A).

Analysis of 10,000 droplets revealed a single and narrow population of droplets with a weak green fluorescence.

In comparison, the fused emulsion analysis revealed two populations of droplets (Figure 4.17B). A first sharp and weakly fluorescent population was composed of unfused droplets and IVTT droplets fused with a PCR droplet containing no pPAC template. The second population was more fluorescent and corresponded to droplets fused with a PCR droplet containing one or more pPAC template. This experiment shows that the indirect PAC activity assay we developed allows for unambiguously distinguish between droplets containing a gene coding for an active PAC and empty ones. In addition, the distribution of both active and inactive populations was relatively sharp. This is important, since it indicates that the assay could enable the discrimination between variants with different levels of activity.

The positive population represents 25 % of the total emulsion. Plasmid encoding for penicillin acylase was diluted to reach a lambda value of 0.35; according to Poisson distribution, 29 % of the 2.3 pL droplets contain 1 or more pPAC molecules. Taking a fusion efficiency value of 90 %, one should expect 26 % occupancy.

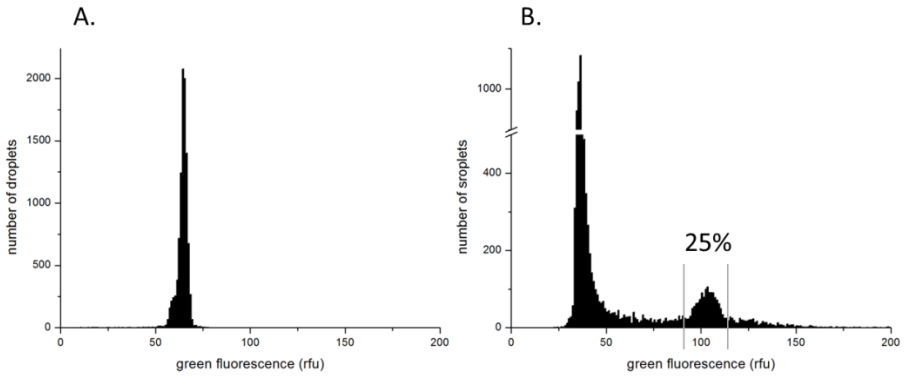


Figure 4.17: Fusions experiments and controls results. Green fluorescence distribution of the droplets in the PCR-to-IVTT fused emulsion (B) and control (A). 10,000 droplets were analyzed for both emulsions. See in text for more details.

Conclusion

This chapter presents a strategy for indirect enzyme autoamplification assays in bulk or picoliter droplet formats, based on the production of a specific amino acid in the medium that will be subsequently incorporated by IVTT synthesis into the enzyme itself and/or a reporter protein. The assay can therefore be applicable to many enzymatic activities that will generate amino acids or other products that can be further transformed into amino acids using a coupled enzyme. The assay, consisting in an auto-amplification of the *in vitro* translation, has proven to be reliable in bulk and in droplets in the presence of pPAC only or coupled to a reporter gene. GFP was used here but can be replaced by any other reporter genes encoding a protein that can generate a detectable signal suitable with the experiment, making the assay highly versatile. We have demonstrated that varying the two plasmids quantities allows tuning of the signal apparition time by taking the advantage of the competition between them. Assays conditions could therefore be adjusted as desired.

Directed evolution of enzymes in droplets necessitates single molecules of plasmids encapsulated per droplet. The assay has shown encouraging results in bulk at pPAC concentrations corresponding to one gene per droplet volume, and has been successfully transferred into droplets. In cases of low *in vitro* protein translation at one gene per droplets, we have also shown that single DNA molecules can be amplified *in situ* by PCR in 2 pL droplets and be subsequently fused to IVTT containing droplets to successfully perform the assay.

The reporter gene could be the GFP as described in this work or any other protein that would either be able to give a fluorescent or colorimetric signal or interact directly with the encoding DNA or RNA, such as a polymerase, or restriction or methylation enzymes (see Figure 4.18B).[13,14] Furthermore, the system can be used without any reporter gene but with a second enzyme

substrate (for ex. a fluorogenic substrate), enabling the detection of two different activities of one single enzyme (Figure 4.18A). This strategy can be potentially used to evolve a promiscuous activity while maintaining a selection pressure on the native activity: the native activity is assayed via a methionine-based substrate; only enzymes that retained the native activity will be massively translated compared to the others. In the meanwhile, an increased promiscuous activity can be assayed concomitantly with a fluorogenic substrate. Fluorescent signal will correspond to enzymes that possess both native and promiscuous activities.

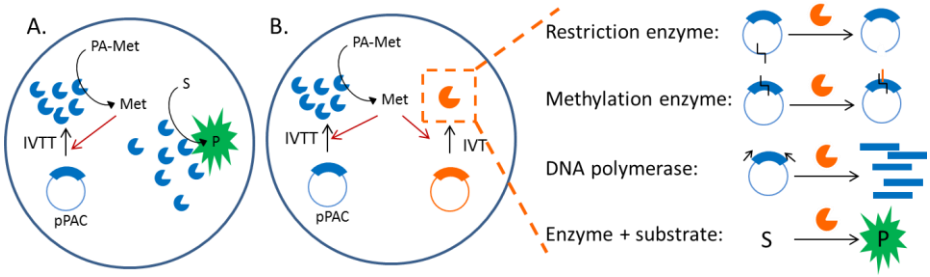


Figure 4.18: Various strategies can be used with our system. A. In the presence of only one gene, if the variant is active on the amino acid based substrate (here PA-Met), there will be sufficient amount of variant produced in the droplet, and if the variant is active on a second fluorogenic or chromogenic substrate S, detection of two activities can be done simultaneously. B. Other reported genes can be envisioned such as restriction or methylation enzymes that will directly mark the DNA contained in the droplet, or DNA polymerase that will enable in a second time the gene amplification, or an enzyme plus its specific chromogenic or fluorogenic substrate.

Following the same general strategy, dNTPs-based substrates could be synthetized and used for "Darwinian" evolution of enzymatic activities in droplets without the need of screening and sorting. If the enzyme variant present in the compartment is active, dNTPs will be released in the medium and a subsequent step of amplification by PCR in droplets will allow the selective replication of the genes encoding active enzymes.

References

1. Seixas de Melo JS, Becker RS, Macanita AL; *Photophysical Behavior of Coumarins as a Function of Substitution and Solvent: Experimental Evidence for the Existence of a Lowest Lying 1(n,.pi.*) State*. The Journal of Physical Chemistry **1994**; 98: 6054-6058

2. Goddard JP, Reymond JL; *Enzyme assays for high-throughput screening*. Curr Opin Biotechnol **2004**; 15: 314-22

3. Goddard JP, Reymond JL; *Recent advances in enzyme assays*. Trends Biotechnol **2004**; 22: 363-70

4. Henke E, Bornscheuer UT, Schmid RD, Pleiss J; *A molecular mechanism of enantiorecognition of tertiary alcohols by carboxylesterases*. Chembiochem **2003**; 4: 485-93

5. Konarzycka-Bessler M, Bornscheuer UT; *A high-throughput-screening method for determining the synthetic activity of hydrolases*. Angew Chem Int Ed Engl **2003**; 42: 1418-20

6. Zhao H; *A pH-indicator-based screen for hydrolytic haloalkane dehalogenase*. Methods Mol Biol **2003**; 230: 213-21

7. Woronoff G, El Harrak A, Mayot E, Schicke O, Miller OJ, Soumillion P, Griffiths AD, Ryckelynck M; *New generation of amino coumarin methyl sulfonate-based fluorogenic substrates for amidase assays in droplet-based microfluidic applications*. Anal Chem **2011**; 83: 2852-7

8. Courtois F, Olguin LF, Whyte G, Theberge AB, Huck WT, Hollfelder F, Abell C; *Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays*. Anal Chem **2009**; 81: 3008-16

9. Elbaz Y, Steiner-Mordoch S, Schunldiner S, Danieli T; *In vitro Expression of the Multidrug Transporter EmrE Using the Rapid Translation System (RTS)*. Biovhemica **2004**; 4: 23-25

10. Alkema WB, Hensgens CM, Kroezinga EH, de Vries E, Floris R, van der Laan JM, Dijkstra BW, Janssen DB; *Characterization of the beta-lactam binding site of penicillin acylase of Escherichia coli by structural and site-directed mutagenesis studies*. Protein Eng **2000**; 13: 857-63

11. Courtois F, Olguin LF, Whyte G, Bratton D, Huck WT, Abell C, Hollfelder F; *An integrated device for monitoring time-dependent in vitro expression from single genes in picolitre droplets*. Chembiochem **2008**; 9: 439-46

12. Pekin D, Skhiri Y, Baret JC, Le Corre D, Mazutis L, Salem CB, Millot F, El Harrak A, Hutchison JB, Larson JW, Link DR, Laurent-Puig P, Griffiths AD, Taly V; *Quantitative and sensitive detection of rare mutations using droplet-based microfluidics*. Lab Chip **2011**; 11: 2156-66

13. Lee YF, Tawfik DS, Griffiths AD; *Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation*. Nucleic Acids Research **2002**; 30: 4937-4944

14. Ong JL, Loakes D, Jaroslawski S, Too K, Holliger P; *Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide*. Journal of Molecular Biology **2006**; 361: 537-550

Conclusion

First, we studied *in vivo* and *in vitro* protein expression systems to choose the most suitable for high throughput screening of penicillin G acylase libraries in microfluidics. *In vivo* expression in the periplasm and cytoplasm of *E. coli* were evaluated. Periplasmic constructs showed some toxicity upon expression which can promote undesirable bias during the selection process. On the other hand, in the cytoplasmic expression system, the enzyme is less accessible to the fluorogenic substrate that is in the culture medium, lysis of the cell would therefore have been needed before the assay. Finally, *in vitro* expression of the enzyme afforded suitable yields. Genes were separately encapsulated with the IVTT mixture and the substrate(s) into aqueous droplets which provides a physical link between the gene, the encoded enzymes and the generated products. It enables high throughput screening applications using droplet-based microfluidics.

In vitro compartmentalization requires appropriate enzymatic substrates. Some fluorophores like 7-amino-4-methylcoumarins for example, do not remain well confined in droplets. The exchange rate of various coumarins has been shown to be correlated to their hydrophilicity. Sulfonated coumarin-based substrates were therefore synthetized, characterized and used in high throughput screening applications. No transport of the sulfonated coumarin (ACMS) between compartments was observed after weeks. Therefore, sulfonation appeared to be a convincing tool to improve retention of chemical compounds in droplets. And has also been successfully applied to

umbeliferones, or to the creation of compatible substrate for the alcohol deshydrogenase in droplets, replacing the classical PMS into 5-(3-Sulfonatopropyl)phenazinium (Yosr Skhiri, PhD thesis).

We developed a second alternative and innovative assay called "activity fed enzyme translation" ("AFET") based on a non fluorogenic substrate and linking the enzymatic activity to the *in vitro* expression of a reporter gene. The leaving group of the substrate is an amino acid that is incorporated in the translation of the penicillin acylase and of GFP as reporter gene. The auto amplificative nature of the assay permits the discrimination between different enzymatic efficiencies. Response times can be finely tuned by adjusting both plasmid concentrations, and was demonstrated to be applicable with one gene of enzyme per droplet.

As a more general perspective, it would be of great interest to apply the system to substrates with dNTPs or primers as leaving groups. Therefore, in a PCR mix lacking the corresponding dNTP or primer, the genes of interest could be amplified *in situ* only if the enzyme is active; and the number of amplification cycles could be proportional to the number of turnovers. The process will allow the selection of the active clones instead of screening, and should greatly simplify the microfluidic manipulation. Droplets will have to be generated and massive emulsion preparations can be envisioned. Even though the technique used in this work allows the production of 100 µL of aqueous droplets per hour (5×10^6 droplets of 20 pL per hour), other microfluidic techniques producing droplets at rates around 100 mL/h have been described.[1] More diverse libraries could then be envisioned.

High throughput screening was then used to find variants bearing new substrate specificities in libraries of mutants of PAC. Screening efficiency was first evaluated on a model system. The wild type gene was recovered from a mixture containing the wild type and a less active mutant sequence in a 1 to 200 proportion. Enrichments factors about 2,000 fold were obtained.

Two libraries of mutants of penicillin acylase were successfully constructed, and screened using droplet based microfluidics for their activity towards the small collection of ACMS-based substrates. The first one is an epPCR library with a high mutational load (8.6 mutations per gene), presenting a good diversity but potentially high levels of indels. In the second library, seven degenerated residues have been inserted at position implicated in the specificity toward the phenyl acetyl moiety. Unfortunately, screening of both libraries did not lead to the selection of mutants with altered specificities of substrates. Two major hypotheses could explain this failure.

First, the clones with altered specificities might not be present in the libraries. The epPCR library possesses only about 10 % of clones without indels or stop codons. A wider screening might be necessary to find few variants of interest. We may also avoid the wild type contamination present in our libraries by modifying the cloning strategy. Maturation could also be drastically impaired by mutations in the specificity pocket or near the active site and consequently decreases considerably the quality of the libraries. Oh and coworkers[2] mutated several residues from the specificity pocket and the active site of the enzyme, and expressed the α and β side chains of PAC separately without linker peptide in order to avoid impairment of the clones because of inefficient maturation. This strategy was also applied to cephalosporin C acylase[3,4] and might be a good alternative in the directed evolution of PAC specificity experiments.

The second hypothesis is that the wild type contamination might arise from our DNA recovery protocol. Indeed, because of the very small amount of DNA material (500 to 2,000 plasmid molecules) that is recovered after sorting, the experiment might be very sensitive to external contamination. Strong bias could also occur during the amplification and some of the clones might be over represented.

Nevertheless, the work presented in this thesis represents significant progress toward developing a high throughput droplet-based platform for

directed evolution of enzymes. The tools developed in this work span all aspects of platform development including optimized IVTT conditions, novel fluorogenic substrates and assays for assessing enzymatic activity, and refined microfluidic modules for high throughput detection and sorting. These tools and the scientific insight gained from initial screening experiments lay a solid foundation for future work in the directed evolution of penicillin acylase and other enzymes of interest.

References

1. Nisisako T, Torii T; *Microfluidic large-scale integration on a chip for mass production of monodisperse droplets and particles*. Lab Chip **2008**; 8: 287-93

2. Oh B, Kim K, Park J, Yoon J, Han D, Kim Y; *Modifying the substrate specificity of penicillin G acylase to cephalosporin acylase by mutating active-site residues*. Biochem Biophys Res Commun **2004**; 319: 486-92

3. Oh B, Kim M, Yoon J, Chung K, Shin Y, Lee D, Kim Y; *Deacylation activity of cephalosporin acylase to cephalosporin C is improved by changing the side-chain conformations of active-site residues*. Biochem Biophys Res Commun **2003**; 310: 19-27

4. Kim S, Kim Y; *Active site residues of cephalosporin acylase are critical not only for enzymatic catalysis but also for post-translational modification*. J Biol Chem **2001**; 276: 48376-81

Experimental section

Molecular biology standard methods

Strains

Top10 (Invitrogen): F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-

MegaX DH10B T1R Electrocomp Cells (Invitrogen): F- mcrA Δ(mrr-hsdRMS-mcrBC) 80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL nupG tonA

Culture media

LB medium: 20 g/L of LB Broth Base (Lennox L Broth Base) powder (Invitrogen).

Solid medium: 32 g/L of LB Agar (Lennox L Agar) powder (Invitrogen).

Ampicillin stock solution (50 mg/mL, in water, filtered). Final concentration: 100 μg/mL.

Tetracyclin stock solution (5 mg/mL, in ethanol). Final concentration: 7.5 μg/mL.

SOC medium: 500 μL of $MgCl_2$ (2 M, filtered) and 500 μL of glucose (2 M, filtered) were added to 50 mL of sterile LB medium.

Electrocompetent cells preparation

A preculture of bacteria was started in 5 mL of LB medium and incubated overnight at 37°C, with shaking at 180 rpm. 1 mL of the preculture was transferred into 1 L of fresh LB medium incubated at 37°C, 180 rpm, until the optical density (600 nm) reached 0.6.The culture was then centrifugated at 3000 x g for 10 min, 4°C. The pellet was resuspended in 1 L of cold miliQ water, and again centrifugated and successively resuspended (after each centrifugation) in 500 mL of cold milliQ water, 2 mL of 10% glycerol and finally in 2 mL of 10 % glycerol. All steps following the culture were carried out with cold and sterile solutions, on ice or at 4°C. Competent cells were aliquoted into 20 x 100 μL volumes. The sterile tubes were directly frozen into ethanol/dry ice mix and kept at -80°C.

General cloning methods

Ligation were made with T4 DNA ligase from Roche or New England Biolabs, restrictions with enzymes from New England Biolabs, dephosphorylation with phosphatase from Roche, TA-cloning using the vectors from the "pGEM-T and pGEM-T Easy Vector Systems" kit from Promega. Reactios were conducted following the protocols provided by the manufacturers.

Synthetic oligonucleotides were provided by Eurogentec (Liège).
Cells electroporation was made using the "Cell Jet Basic" (Eurogentec), 2500 V, followed by 1.5 h incubation at 180 rpm in SOC medium, then plated on LB agar plate and incubated over night at 37°C.

Site-directed mutagenesis
The F146L mutant penicillin acylase was prepared using the Quick Change Site-Directed Mutagenesis kit (Stratagene), the pIVEX2.3d-PAC plasmid and primers.
F146L-for: ctatggcaaaccgcttatctgatagcactagcg
F146L-rev: cgctagtgctatcagataagcggtttgccatag

PCR methods
Classical PCRs were performed using either the "Go Taq DNA polymerase" (Promega) or the "FastStart High Fidelity PCR System" (Roche) when a high fidelity polymerase was needed, following protocols provided by manufacters.

Error prone PCR
Error prone PCRs were conducted on 0.1 pmol of DNA in the presence of increased concentration of Mg^{2+} for stabilising noncomplementary base pairs (6.5 mM of $MgCl_2$), dNTPs at unbalanced stoichiometry for forcing misincorporation (200 μM of dATP and dGTP and 1 mM of dTTP and dCTP), in the presence of $Mn^{2+}$for reducing the base pair specificity. Two Mn^{2+} concentrations were used to tune the mutational load to higher and lower rates (0.2 mM or 0.1 mM of $MnCl_2$). (ref= book) Amplified DNA fragments were restricted by NcoI and XmaI and cloned back into the pIVEX2.4d vector (scheme outlined in Figure 1.12). Ligation product was transformed into Top10 electrocompetent cells and plated on solid media.
Conditions:
0.1 mM (low rate library) or 0.2 mM (high rate library) $MnCl_2$, 6.5 mM $MgCl_2$, 0.2 mMdATP, 0.2 mMdGTP, 1 mMdTTP, 1 mMdCTP. 0?5 μM of primers 7 and 8. 370 ng of DNA per 50 μL reaction.

Cycles:

1	95°C	2 min	
2	95°C	1 min	
3	60°C	1 min	
4	72°C	3 min	Go to step 4, repeat 29 times
5	72°C	5 min	
6	4°C		

PCR made on the synthetic DNA fragment (TM Library)
PCR was conducted on 150 ng of synthetic DNA fragment (~100 x the diversity), using the "Expand High Fidelity PCR system" (Roche), with primers 7 and 8 (300 nM each), dNTPs (200 μM each), Buffer containing 1.5 mM of $MgCl_2$.

Cycles:

1	94°C	2 min	
2	94°C	15 s	
3	45.7°C	30 s	
4	72°C	2 min	Go to step 4, repeat 9 times
5	72°C	7 min	
6	4°C		

The PCR product was thenpurified, restricted with BstXI and Bsu36I and cloned into pIVEX-2.4d-PAC, previously restricted using the same restriction endonucleases.

Phenol chloroform extraction and ethanol precipitation

(1) Phenol chloroform purification. An equal volume of Phenol-Chloroform-Isoamyl alcohol (25-24-1 volumes of each) was added to DNA solutions. The mix was vortexed 10-15 seconds and then centrifuged for 10 min at 15,000 x g. The aqueous upper phase was recovered in a new tube and the step repeated once. (2) Ethanol precipitation. One tenth volume of 3M NaOAc pH 5.2 (Sigma-Aldrich), 1 μL of LPA (GeneLute LPA Linear Polyacrylamide, Sigma-Aldrich) and three volumes of 100 % EtOH were added to the initial DNA solution. The solution was then immediately centrifuged at 20,000 x g for 30 min, 4°C. The pellet was washed twice with 70 % EtOH. Centrifugation at 20,000 x g for 30 min was carried out. The pellet was air dried and resuspended in milliQ water.

Periplasmic extracts preparation

Prepared from bacterial pellets with osmotic shock method. 10 mL of culture was centrifuge 10 min at 4,000 rpm (Centrifuge ALC pK120, rotor O-E24), then pellet were suspended in 330 μL of sucrose 20 %, tris-HCL 20 mM pH 8. 6.6 μL of EDTA (5 M pH 8) were added. Samples were incubated at room temperature for 10 min, then centrifuge 10 min at 6,000 rpm (Centrifuge Beckman, microfuge 18). Supernatant was discarded. 500 μL of $MgSO_4$ 5 mM (cold) was added, suspension was incubated for 20 min at 4°C, then centrifuged for 10 min at 10,000 rpm. Periplasmic extract are in the supernatant.

Cytoplasmic extracts preparation

Cytoplasmic extracts were obtained starting from the pellet recovered at the end of the periplasmic extracts preparation.the pellet was suspended in 200 μL of lysis buffer (tris-HCl 50 mM pH 8, 1 mM EDTA and 100 mM NaCl), and kept at -20°C aver night. Samples were thawed at 37°C. 8 μL of lysozyme (10 mg/mL) was added and samples were incubated 20 min at room temperature, then 45 min at 37°C. 16 μL of DNAse (1 mg/mL) were added and samples were incubated 15 min at room temperature, and then centrifuged 15 at 10,000 rpm. Cytoplasmic extracts are in the supernatant. The final pellet can be suspended in 500 μL of PBS 1x to give the insoluble fraction.

PBS 1L (10 x)
80 g of NaCl (1.4 M)
2 g of KCl (0.03 M)
11.5 g of $Na_2HPO_4.7H_2O$
2 g of KH_2PO_4
Adjust the pH at 7.4 (or 7.0)
Adjust the volume to 1 L with milliQ water
Autoclave 20 min and keep at 4°C

IVTT assays
IVTT assays were conducted following recommendations from the manufacturer, using the "RTS 100 E. coli HY Kit" (Roche).

For "AFET" assays conducted in the presence of PA-Met (phenyl acetyl methionine), 1 µL of PA-Met solution at 100 mM in DMF was added to a total volume of 50 µL of IVTT reaction. The GFP fluorescence was measured either using a spectrophotometer (SpectraMax M5, Molecular Devices) in the case of end points measurements. GFP signal appearance followed over time upon incubation at 37°C was performed in a Real Time PCR machine (StepOne Real Time PCR System, Applied Biosystems).
Assays in the presence of PA-ACMS were conducted at 100 µM (stock solution: 10 mM in DMSO or H_2O), or towards the small collection of ACMS based substrates were performed at concentrations of 20 µM for each substrates (**3**, **6**, **7**, **8** and **9**, stock solutions at 10 mM in DMSO). The total concentration of ACMS-based substrates was therefore 100 µM.
Texas red (1 mg/µL of 70,000 kDa dextran-Texas Red conjugate, Invitrogen) was added as tracking agent for the fusion experiment were 2 pL IVTT droplets were fused to 20 pL IVTT droplets (chapter 4).

Electrophoresis (SDS-PAGE)
Acrylamide gels were prepared using a separation gel and a concentration gel. Proteins samples were prepared by adding one volume of loading buffer 2x to one volume of sample; samples were heated for 10 min at 95°C and loaded on gels. Electrophoresis was perfomed in a cuve containing the cathode and anodes buffers at 100-150 V. Then, gels were put in the staining solution, heated in the microwave until boiling. Gels were subsequently rinse using the destaining solution.

Separation gel
5 mL acrylamide/bisacrilamide 30 %
5 mL Tris-Cl/SDS, pH 8.5
3 mL H_2O
1.6 mL glycerol
100 µL ammonium persulfate 1 %
7.5 µL TEMED

Concentration gel
2 mL acrylamide/bisacrilamide 30 %
3 mL Tris-Cl/SDS, pH 8.5
8 mL H_2O
100 µL ammonium persulfate 1 %
15 µL TEMED

Tris-Cl/SDS 500 mL
182 g of Tris base
1.5 g of SDS
Adjust pH at 8.45 with HCl
Adjust the volume to 500 mL with H_2O and keep at 4°C

Cathode Buffer 1L (1x)
12.11 g Tris-base
17.92 g Tricine
1 g SDS
Adjust the volume to 1 L with H_2O and keep at 4°C

Anode Buffer 1L (5x)
121.1 g Tris-base
500 mL H2O
Adjust pH to 8.9 with HCl
Adjust the volume to 1 L with H_2O and keep at 4°C

Loading buffer 10 mL (2x)
1 mL 1M Tris-Cl pH6.8
0.4 g SDS
2 ml glycerol
0.02 g bromophenol blue
Adjust the volume to 10 mL with H_2O
Add 31 mg of DTT per 1 mL of loading buffer prior using.
Molecular weight marker:
PageRuler™ Prestained Protein Ladder, Fermentas

Staining solution 500 mL
1 g Coomassie blue
180 mL methanol
40 mL acetic acid
180 mL H_2O

Destaining solution 1L
200 ml methanol
25 ml acetic acid
775 ml H_2O

Synthesis

The synthesis of 7-aminocoumarin-4-methanesulfonic acid (ACMS), initially developed by Sato *et al.*.[1], was modified according to known procedures.[2-5]

Reagents and characterization. Products and solvents were purchase from Sigma-Aldrich and used without further purification. NMR spectra were recorded either on Bruker Advance-400 or Brucker AC-300 Advance II instruments and calibrated using residual undeuterated solvent as an internal reference. Coupling constants are reported and expressed in Hz, splitting patterns are designated as s (singlet), d (doublet), dd (double doublet), q (quartet), m (multiplet).

3-carboethoxyaminophenol (1). (Figure 2.3A) A solution of ethyl chloroformate (20g, 184 mmol) in 100ml of diethyl ether was added dropwise to a stirred suspension of 3-aminophenol (10 g, 92 mmol) in a mixture of dry diethyl ether and dry THF (500 and 100 ml, respectively). A white precipitate (the amine hydrochloride) formed immediately. The reaction mixture was stirred an additional 2 hours at room temperature. The hydrochloride was removed by filtration. The filtrate was evaporated under vacuum to give an orange oil. The oil was solubilised in 400 ml of diethyl ether and the resulting organic phase was washed with solutions of 1N HCl, saturated $NaHCO_3$ and water. The organic phase was dried on $MgSO_4$ and the solvent was removed under vacuum. Recrystallization from chloroform (20ml) gave 12.9 g (77% yield) of colourless needles. ^{1}H NMR (300MHz, $CDCl_3$) *δ*: 1.3 (3H, t, OCH_2CH_3), 4.21 (2H, q, OCH_2CH_3), 6.59 (4H, m, N*H*, O*H*, 4-*H* and 6-*H*), 7.06 (1H, t, 7-*H*), 7.30 (1H, s, 2-*H*)

4-chloromethyl-7-carbethoxyaminocoumarin (2). (Figure 2.3A) Ethyl 4-chloroacetoacetate (12.67 g, 72.9 mmol) was added dropwise to a stirred suspension of 3-hydroxyohenylurethane **1** (12 g, 66.2 mmol) in sulfuric acid 70% (330 ml) on ice. The mixture was stirred overnight at room temperature. The mixture was then poured on ice water (480 g) and stirred for 30 minutes. The precipitate was filtered and washed with chilled diethyl ether. Recrystallization in acetone (1.8 l) gave 12.67 g (66% yield) of yellow needles. ^{1}H NMR (300MHz, DMSO-*d*$_6$) *δ*: 1.26 (3H, t, OCH_2CH_3), 4.17 (2H, q, OCH_2CH_3), 4.97 (2H, s, CH_2Cl), 6.52 (1H, s, 3-*H*), 7.42 (1H, dd, 2.0 Hz, 6-*H*), 7.59 (1H, d, 8-*H*), 7.76 (1H, d, 5-*H*), 10.21 (1H, s, N*H*).

Sodium 7-carbethoxyaminocoumarin-4-methanesulfonate (3). (Figure 2.3A) A suspension of **2** (8 g, 28.4 mmol) and Na_2SO_3 (17.9 g, 142 mmol) in a mixture of acetone and water (140 and 100 ml respectively) was stirred under reflux for 24 hours. After cooling, solvents were evaporated under vacuum to leave a pale yellow powder. The residue was suspended in hot water and

filtrated while warm. The filtrate was evaporated and the residue was suspended in hot methanol and filtrated while warm. The filtrate was evaporated to give a yellow powder (9.32 g, 94% yield). ^{1}H NMR (300MHz, DMSO-*d*$_6$) *δ*: 1.26 (3H, t, OCH$_2$C*H*$_3$), 3.98 (2H, s, C*H*$_2$SO$_3$Na), 4.15 (2H, q, OC*H*$_2$CH$_3$), 6.24 (1H, s, 3-*H*), 7.34 (1H, dd, 6-*H*), 7.55 (1H, d,8-*H*), 7.83 (1H, d, 5-*H*), 10.16 (1H, s, N*H*).

7-aminocoumarin-4-methanesulfonic acid (4). (Figure 2.3A) Concentrated sulfuric acid (5 ml) and acetic acid (5 ml) were added to **3** (2.5 g, 7.3 mmol), and the mixture was heated with stirring at 100°C overnight. After cooling, the solution was poured into ethanol (100 ml) and gave a red precipitate. The solid was filtered off, washed with ethanol to give a red powder (1.8 g, 96% yield). ^{1}H NMR (300MHz, DMSO-*d*$_6$) *δ*: 3.86 (2H, s, C*H*$_2$SO$_3$Na), 5.94 (1H, s, 3-*H*), 6.43 (1H, d, 8-*H*), 6.55 (1H, dd, 6-*H*), 7.57 (1H, d, 5-*H*).

Amide formation with acyl chlorides: general procedure

This method was used to synthesize products **5** to **9**. **4** (300 mg, 1.2 mmol) was solubilised in water and the solution was neutralized with 1M NaOH. NaOH (1M, 1 mmol/mmol of **4**, 1.18 ml) and water (total volume was adjusted to 10.5 mL) were added to the mixture and stirred on ice. A solution of acid chloride (4.4 eq, 5.2 mmol) solubilised in anhydrous THF (1 mL/mmol of **4**, 5.2 mL) and a second solution of 1M NaOH (1 mmol/mmol of **4**, 5.2 mL) were added dropwise and simultaneously. pH was monitored during the addition and maintained at 7. The reaction was stirred at room temperature overnight. The reaction mixture was then dried under vacuum, the residue was suspended in 1M HCl (11 mL) and ethyl acetate (11 mL); and the resulting mixture was stirred for 10 minutes. The precipitate was filtered, and washed successively with ethyl acetate, 1M HCl and water. Recrystallisation in EtOH-water 60-40 (2 mL) gave the desired product.

7-phenylacetylamidocoumarin-4-methanesulfonate (5). (Figure 2.3A) White crystals (73.8 mg, 9% yield). ^{1}H NMR (400MHz, DMSO-*d*$_6$) *δ*: 3.72 (2H, s, phenyl-C*H*$_2$CO-NH), 4.02 (2H, s, C*H*$_2$SO$_3$Na), 6.28 (1H, s, 3-*H*), 7.35 (6H, m, phenyl and 6-*H*), 7.78 (1H, d, 8-*H*), 7.90 (1H, d, 5-*H*), 10.59 (1H, s, N*H*). ^{13}C NMR (400MHz, DMSO-*d*$_6$) *δ*: 170.32, 160.72, 154.37, 150.37, 142.53, 135.96, 129.67 (2C), 128.79 (2C), 128.12, 127.08, 115.20, 115.05, 114.57, 105.74, 53.58, 43.75.

7-(4-tert-butylbenzoyl)amidocoumarin-4-methanesulfonate (6). (Figure 2.3A) White crystals (55mg, 6.5% yield). ^{1}H NMR (400MHz, DMSO-*d*$_6$) *δ*: 1.34 (9H, s, tert-butyl), 4.05 (2H, s, C*H*$_2$SO$_3$Na), 6.31 (1H, s, 3-*H*), 7.70 (7H, m, phenyl, 5-*H*, 6-*H* and 8-*H*), 10.55 (1H, s, N*H*). ^{13}C NMR (400MHz, DMSO-*d*$_6$) *δ*: 166.45, 160.77, 155.33, 154.27, 150.41, 142.74, 132.22, 128.16 (2C), 127.92, 125.71 (2C), 116.15, 115.28, 114.70, 106.68, 53.59, 35.19, 31.37 (3C).

7-cynamoylamidocoumarin-4-methanesulfonate (7). (Figure 2.3A) Yellow crystals (158mg, 20% yield). ^{1}H NMR (400MHz, DMSO-*d*$_6$) *δ*: 4.09 (2H, s, C*H*$_2$SO$_3$Na), 6.29 (1H, s, 3-*H*), 6.91 (10H, m, phenyl, C*H*=C*H*, 5-*H*, 6-H and 8-*H*), 10.58 (1H, s, NH). ^{13}C NMR (400MHz, DMSO-*d*$_6$) *δ*: 164.56, 160.67, 154.37, 150.15, 142.58, 141.65, 135.06, 130.41, 129.45 (2C), 128.43 (2C), 127.96, 122.15, 115.51, 115.08, 114.67, 105.92.

7-(3-methoxyphenylacetyl)amidocoumarin-4-methanesulfonate (8). (Figure 2.3A) Brown crystals (55mg, 6.6% yield). ^{1}H NMR (400MHz, DMSO-*d*$_6$) *δ*: 3.63 (2H, s, phenyl-C*H*$_2$CO-NH), 3.74 (3H, s, C*H*$_3$Ophenyl), 4.02 (2H, s, C*H*$_2$SO$_3$Na), 6.28 (1H, s, 3-*H*), 6.90 (2H, d, phenyl), 7.27 (2H, d, phenyl) 7.42 (1H, dd, 6-*H*), 7.78 (1H, d, 8-*H*), 7.89 (1H, d, 5-*H*), 10.54 (1H, s, N*H*). ^{13}C NMR (400MHz, DMSO-*d*$_6$) *δ*: 170.68, 160.73, 158.54, 154.37, 150.36, 142.60, 130.66 (2C), 128.09, 127.85, 115.19, 115.00, 114.54, 114.23 (2C), 105.70, 55.48, 53.58, 42.88.

7-phenoxyacetylamidocoumarin-4-methanesulfonate (9). (Figure 2.3A) Yellow crystals (55g, 6.8% yield). ^{1}H NMR (400MHz, DMSO-*d*$_6$) *δ*: 4.03 (2H, s, C*H*$_2$SO$_3$Na), 4.76 (2H, s, OC*H*$_2$CO), 6.28 (1H, s, 3-*H*), 7.15 (5H, m, phenyl), 7.50 (1H, dd, 6-*H*), 7.78 (1H, d, 8-*H*), 7.89 (1H, d, 5-*H*), 10.50 (1H, s, N*H*). ^{13}C NMR (400MHz, DMSO-*d*$_6$) *δ*: 167.81, 160.67, 158.19, 154.27, 150.31, 141.75, 129.99 (2C), 128.14, 121.71, 115.65, 115.39, 115.14 (2C), 114.84, 106.33, 67.53, 53.58.

Phenylacetylated methionine. (Scheme 4.1) L-methionine (200 mg, 1.34 mmol) was solubilized in 1M NaOH (5 mL). The solution was stirred on ice. Phenyl acetyl chloride (2 eq, 414 mg or 356 µL) and acetonitrile (3 mL) were added. The solution was stirred at room temperature overnight. The solution was acidified at pH 1 with a solution of 5M HCl. The mixture was extracted with dichloromethane, dried under vaccum, the residue was purified on silica flash column (eluent: diethyl ether: dichloromethane 1:1). Pure fractions (determined by TLC rf 2/6) were collected and dried under vacuum. Recrystallisation in chloroform gave the desired product (51 mg, 14% yield). ^{1}H NMR (300 MHz, $CDCl_3$) δ: 2.00 (3H, s, C*H*$_3$), 2.10 (2H, m, C*H*$_2$-CH$_2$-S) 2.43 (2H, t, C*H*$_2$-S), 3.64 (2H, s, C*H*$_2$-phenyl), 4.69 (1H, q, C*H*-COOH), 6.56 (1H, d, N*H*), 7.30 (5H, m, phenyl).

Microscopes set up

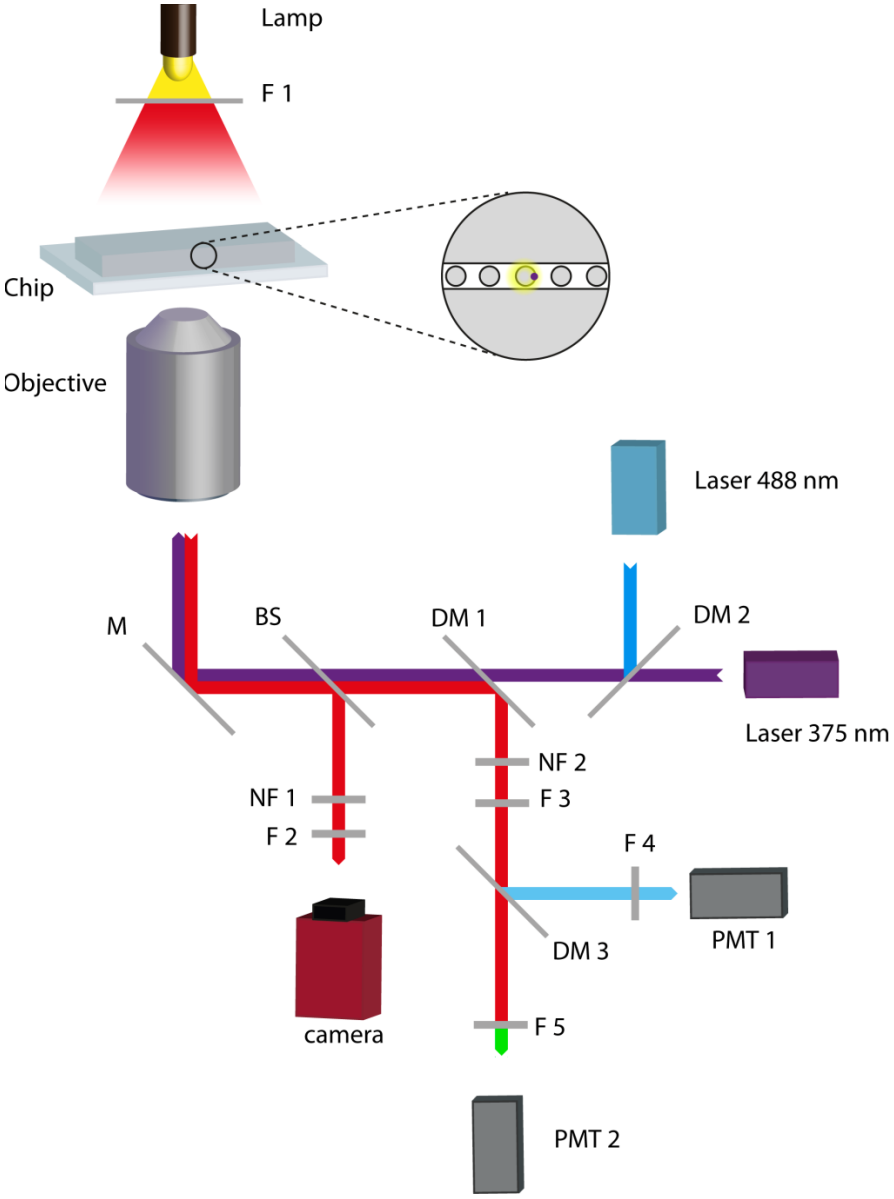


Figure 5.1: Schematic representation of the "UV and Blue" optical setup.

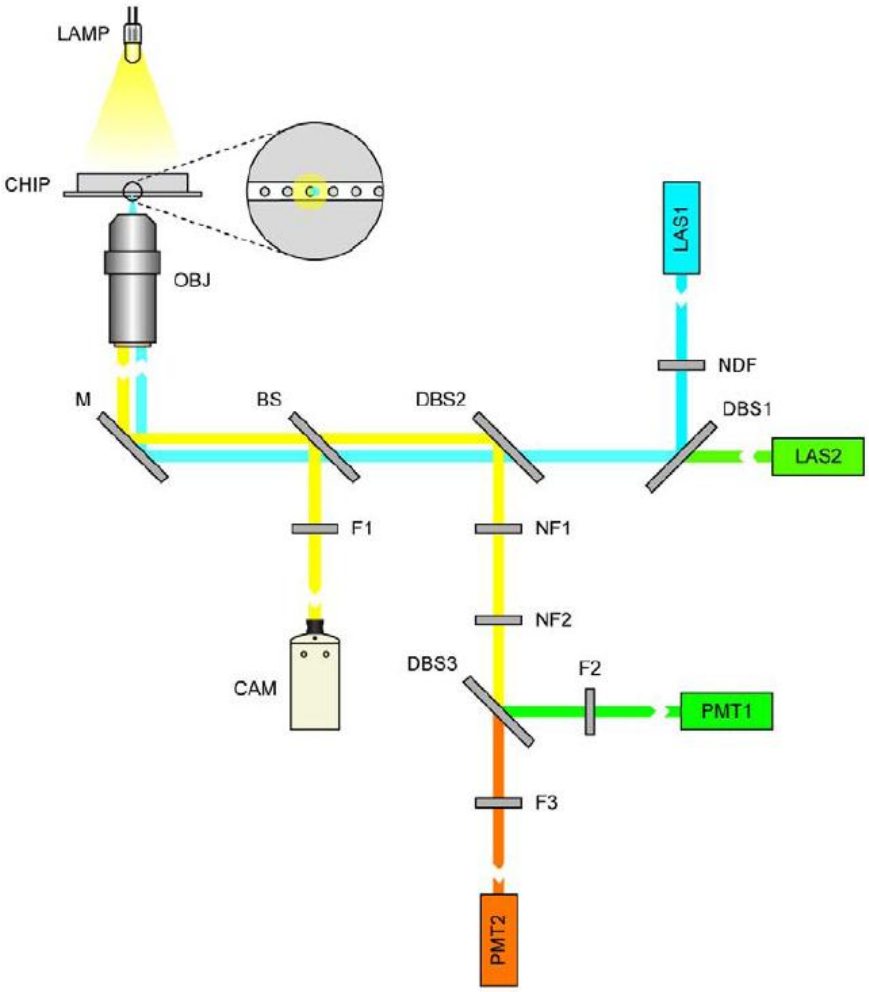


Figure 5.2: Schematic representation of the "Blue and Green" optical setup. Scheme taken from Mazutis et al.[6]

Figure 5.1. The optical setup comprised a Nikon Eclipse Ti-S inverted microscope (Nikon) mounted on a vibration-dampening platform. A 375 nm laser (LAS, Newport-Spectraphysics) and a 488 nm laser beams were combined using a dichroic mirror "DM2" (Z375BCM, Semrock Inc.) and the resulting beam was directed in the microscope objective using a mirror "M" and focused in the middle of the channel of the device at the detection point. The emitted fluorescence was collected by the microscope objective and separated from the laser beam by a first dichroic mirror "DM1" (Z375/rde-xt, Semrock Inc.). A notch filter "NF2" (NF488, Semrock Inc.) and a long pass filter "F3" (GG400, Schott) eliminated potentially damaging reflections of the lasers. Coumarin 1/ACMS fluorescence (around 460 nm) was resolved from the fluorescein

fluorescence (525nm) by a second dichroic mirror "DM3" (480 DCPL-AR, Semrock Inc.) reflecting ≤480nm and transmitting other wavelengths. Emitted fluorescence was detected by photomultiplier tubes (PMT, Hamamatsu Photonics KK) equipped with bandpass filters "F4" (FF01-460/84 and FF02-447/60) or "F5" (FF01-510/84) for coumarin and fluorescein detection, respectively. Finally, the output signal from the PMTs was analyzed using a PCI-7831R Multifunction Intelligent DAQ card (National Instruments Corporation,) executing a program written in LabView 8.2 (FPGA module, National Instruments Corporation). The data acquisition rate for the system was 100 kHz. The full process was monitored by redirecting part of the emitted light using a beam splitter "BS" towards a CCD camera (Guppy, Allied Vision Technologies) equipped with a notch filter "NF1" (NF03-488E-25, Semrock Inc.) and a long pass filter "F2" (FF01-409/LP-25, Semrock Inc.), eliminated potentially damaging reflections of the lasers.

Figure 5.2. Light from a 488 nm, 25 mW solid-state laser ('LAS1'; Newport-Spectraphysics), attenuated 1000 times with a neutral density filter 3OD ('NDF'; Thorlabs GmbH), was combined with light from a 532 nm, 25 mW solid-state laser ('LAS2'; Newport-Spectraphysics) using a single-edge dichroic beamsplitter ('DBS1'; LM01-503-25 from Semrock Inc.). The combined laser beams were transmitted through a multi-edge dichroic beam splitter ('DBS2'; Di-T488/532/638-25x36x5.0; Semrock Inc.) to the side camera port of an Axiovert 200 inverted microscope (Carl Zeiss SAS). Inside the microscope the laser light passed through a beam splitter ('BS') and was reflected up into the objective ('OBJ'; LD Plan Neofluar 40×/0.6 from Carl Zeiss SAS) by a conventional mirror ('M'). The combined laser beams were focused to a ~20 μm-diameter spot inside a channel in the microfluidic chip ('CHIP') where it excited droplets one at a time as they flowed past. The fluorescent emission from each droplet passed back along the path of the laser beam, but was reflected by the dichroic beam splitter ('DBS2') to the PMTs (both H5784-20; Hamamatsu Photonics KK). Two notch filters ('NF1' and 'NF2'; NF01-488-25 and NF01-532U-25, respectively; Semrock Inc.) eliminated potentially damaging reflections from the lasers and the light was split between the two PMTs by a single-edge dichroic beamsplitter ('DBS3'; FF562-Di02-25x36; Semrock Inc.). Light emitted by fluorescein or DNA-bound RDT-D1 intercalating dye was measured by one PMT ('PMT1') after filtering through a bandpass filter ('F2'; FF02-510/20-25 from Semrock Inc.). The remaining PMT ('PMT2') measured the light emitted by Texas Red or Resorufin after filtering through another bandpass filter ('F3'; FF01-617/73-25; Semrock Inc.). Light from the microscope's halogen lamp ('LAMP') illuminated the channels and droplets, allowing the droplets to be observed by a Phantom v4.2 high-speed digital camera ('CAM'; Vision Research) when fluorescence measurements were not being taken. (Description taken from Mazutis et al.[6])

Microfuidics

Microfluidic device fabrication

Microfluidic devices were obtained through classic replica molding process as described by Frenz *et al.*[7] Briefly, molds were first prepared by standard photolithography methods. SU8-2015 or SU8-2075 photoresist (MicroChem Corp.) were used to pattern 20 and 75 μm-deep channels onto silicon wafers (Siltronix). Microfluidic devices were then fabricated by using conventional soft lithography methods[8]. The channels of the devices were passivated with a solution of 1% (v/v) 1H, 1H, 2H, 2H-perfluorodecyltrichlorosilane (97 %, ABCR GmbH and Co) in HFE7100 (3M) and subsequently flushed with compressed nitrogen gas.

Microfluidic device operation

Liquids were pumped into microfluidic devices using standard-pressure infuse/withdraw PHD 22/2000 syringe pumps (Harvard Apparatus Inc., Holliston, MA). Syringes (Omnifix- F®, BRAUN) were connected to the microfluidic device using 0.6 x 25 mm Neolus needles (Terumo Corporation) and PTFE tubing with an internal diameter of 0.56 mm and an external diameter of 1.07 mm (Fisher Bioblock Scientific). The fluorinated oil for droplet production was HFE-7500 oil (3M) containing 3 % (w/w) EA surfactant (RainDance Technologies), a PEG-PFPE amphiphilic block copolymer[9]. Droplets were obtained using the flow rates detailed in the experimental section; their production and characterization were carried out on a Nikon Eclipse Ti-S inverted microscope (Nikon).

2 pL droplets production

2 pL IVTT droplets were created using the coflow device depicted at Figure 5.3. The oil phase (with 3 % w/w surfactant) is injected at 120 μL/h. The IVTT phase is injected at 80 μL/h (40 μL/h for each solution. One contained the *E. coli* lysate and methionine source, whilst the second contained the "reaction mix", amino acids, the DNA template).

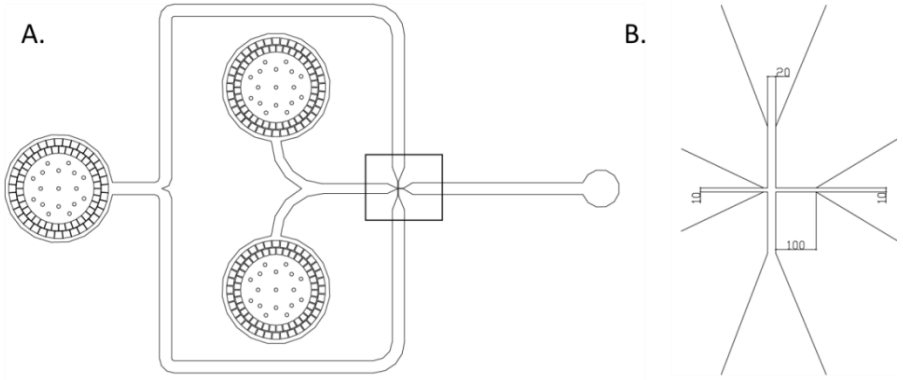


Figure 5.3: schematic representation of the device used to produce 2 pL IVTT droplets. A. general view, B. zoom on the nozzle area.

2 pL PCR droplets were created using the single flow device depicted at Figure 5.4. The oil phase (with 3 % w/w surfactant) is injected at 330 µL/h. The IVTT phase is injected at 100 µL/h.

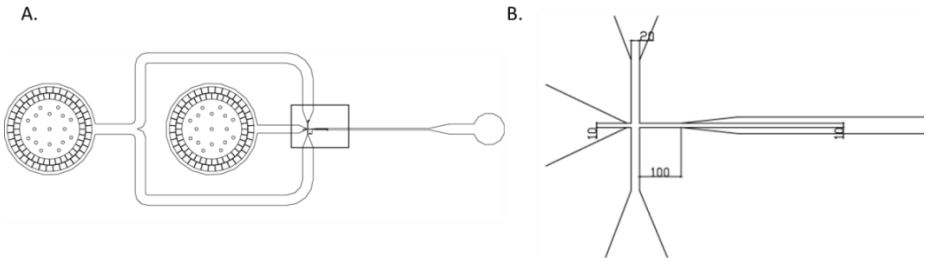


Figure 5.4: schematic representation of the device used to produce 2 pL PCR droplets. A. general view, B. zoom on the nozzle area.

22 pL IVTT droplets production

Scheme of the device at Figure 3.2A and B.

22 pL IVTT droplets were created using a coflow device. The oil phase (with 3 % w/w surfactant) is injected at 300 µL/h. The IVTT phase is injected at 100 µL/h (50 µL/h for each solution. One contained the *E. coli* lysate and methionine source, whilst the second contained the "reaction mix", amino acids, the DNA template and if necessary, the fluorogenic substrate (PA-ACMS)).

Droplets reinjection for analysis

Scheme of the device at Figure 3.2C.

The oil phase (HFE 1500, without surfactant) is injected at flow rates around 100 to 300 µL/h. The droplets are reinjected at flow rates around 20 to 50 µL/h. flow rates are adjusted in order to provide a good spacing between the reinjected droplets at the analysis point (depicted by a green spot on the figure).

Droplet fusion

Scheme of the device at Figure 4.15.

Small droplets were produced and collected either in a PCR tube (Figure 4.14B) or a capillary. After thermocycling or incubation, the emulsion is reinjected into the synchronization/fusion microfluidic device (Figure 4.14 C to F) via the reinjection inlet at ~40 μL/h. The aqueous phase (IVTT, DNA and PA-met or PCR mix and DNA) and the oil phase (3 % w/w EA surfactant in HFE7500) were injected in their respective inlets at flow rates of ~90 μL/h for the dispersed aqueous phase and ~180 μL/h for the continuous oil phase. Electrocoalescence requires an optimal droplet pairing prior to fusion, so that droplets are in contact when they reach the electrodes. In our case, the pairing was eased by the size difference of both droplets (Figure 4.14E-F). Small droplets travel faster than large droplets, and therefore remain blocked at the back of the large droplets before them. To optimize droplet pairing we maintained a small excess of large droplets (~10 %) by adjusting the flow rates and used a voltage of 450 V. Fused droplets were collected in a capillary.

Droplet sorting

Scheme of the device at Figure 3.6.

After incubation in a capillary, droplets are reinjected in the sorting device at flow rates around 10 to 30 μL/h. Droplets are spaced out with surfactant-free fluorinated oil before reaching the sorting junction, injected at flow rates around 150 to 400 μL/h. Flow rates were adjusted in order to avoid false positives (negative droplets flowing through the selection arm). Because of lower hydraulic resistance, droplets flow by default through the negative arm. Laser is focused at the analysis point just before the sorting junction (green spot on the scheme). An AC electric field of 1.5 kV is applied across the electrodes to conduct the droplets into the positive arm. The field is piloted using a Labview program that allows the analysis of the droplets fluorescence and the sorting of the desired population based on fluorescence intensities and size of each droplet. The sorted droplets are collected in a PCR tube.

Retention time characterization

Coumarins. 7-amino-4-methylcoumarin, 7-amino-4-methyl-3-coumarinylacetic acid and 7-(diethylamino)coumarin-3-carboxylic acid were purchased from Sigma-Aldrich. The retention of these molecules and ACMS in droplets was determined by preparing solutions of 10 and 100 μM, supplemented with 50 μM fluorescein (Fluka), as internal standard, and 100 mM phosphate-buffered saline (PBS; pH 7.00). Each aqueous solution was infused at 75 μL/h in a microfluidic device consisting of a dual droplet maker (Figure 2.1A). HFE-7500 fluorinated oil (3M) containing 3% (w/w) EA surfactant (Raindance Technologies), a PEG-PFPE amphiphilic block copolymer[9], was infused at 300 μL/h to produce 150 pL droplets by flow-focussing at a 15 μm nozzle (Figure 2.1C). Depending on the time scale of the analysis, the dual droplet maker was either combined with the short term analysis module (0-140 s) and the

fluorescence measured on-chip at 23°C (Figure 2.1E) or, for longer incubations, the emulsions were collected off-chip using a glass capillary-based device[6] and incubated at 23°C or 37°C for hours to days. Droplets were reinjected at 20 μL/h into the analysis device (Figure 2.1B), spaced with 100 μL/h of surfactant-free fluorinated oil and their fluorescence intensities were monitored by exciting (375 and 488 nm for coumarins and fluorescein, respectively) and detecting the emitted light (446/60 nm and 510/84 nm for coumarins and fluorescein, respectively) using the optical set up detailed in Figure 5.1.

Substrate 5. The same general procedure as for the retention time characterization of coumarins **1**, **2**, **3** and **4** was used with changes to the droplet contents: the first population of droplets contained 50 μM of fluorescein (Fluka) and 0.3 nM of penicillin amidase solution from *Escherichia coli* (Sigma-Aldrich), while the second population contained 100 μM of **5** in 100 mM phosphate buffered saline (pH 7.00). The fluorescence of the droplets was monitored for several hours.

Enzymatic kinetics measurements with NIPAB

Kinetic experiments were carried out with cellular extracts, culture supernatants or IVTT reaction mixtures. The hydrolysis of 2 mM of 6-nitro-3-phenylacetamido benzoic acid (NIPAB) in PBS (pH 7.0) was followed by measuring the absorbance at 405 nm (ε = 9090 $M^{-1}.cm^{-1}$) as a function of time in PBS buffer. All the measurements were performed at 25°C using the Cary 3Bio UV-visible spectrophotometer Varian. Enzymatic concentrations were determined using initial velocity values and a k_{cat} value of PAC on NIPAB of $15 s^{-1}$.[10]

Enzymatic kinetic measurements with PA-ACMS (chapter 2)

Bulk enzymatic assays were performed in a 96-well microtiter plate by mixing 25 μL of substrate at various concentrations (20 μM to 1 mM) in 100 mM PBS (pH 7.00) with 25 μL of a 0.3 nM solution of penicillin amidase from *Escherichia coli* (Sigma-Aldrich) containing 100 μg/mL BSA (New England Biolabs). Fluorescence (λ_{ex} = 375 nm/λ_{em} = 450 nm) was monitored for 3 hours at room temperature using a spectrofluorophotometer (SpectraMax M5, Molecular Devices). K_M and k_{cat} were derived from experimental data using non-linear fitting methods.

Droplet-based microfluidic enzymatic assays were performed by simultaneously producing two populations of droplets using a dual droplet maker. The first population resulted from a co-flow of enzyme and substrate solutions infused at a total flow rate of 100 μL/h. The second population was made of a 20 or 50 μM of ACMS solution (depending on the substrate concentration) infused at 100 μL/h and used as an internal reference. 70 pL droplets were obtained by flow-focussing (Figure 2.1D) with 300 μL/h HFE-7500 fluorinated oil containing 3% (w/w) of EA surfactant. The enzyme and substrate concentrations used were the same as in the microtiter plate experiments and

fluorescence was measured at detection points corresponding to incubation times of 2.5, 5, 10, 20, 40, 60, 80, 100, 120 or 140 s (Figure 2.1E).

Enzymatic kinetic measurement with PA-ACMS (chapters 3 and 4)
Bulk enzymatic assays were performed in a 96-well microtiter plate to determine PAC concentration in IVTT samples. IVTT samples were diluted in 100 mM PBS (pH 7.00) with PA-ACMS at 50 μM (or while assaying PAC and selected mutants activity toward the small collection of ACMS-based substrates at 50 μM). Fluorescence (λ_{ex} = 375nm/λ_{em} = 450nm) was monitored for 3 hours at room temperature using a spectrofluorophotometer (SpectraMax M5, Molecular Devices). Wild type PAC concentration was deduced from initial velocities using the k_{cat} value calculated in chapter 1.

DNA amplification after sorting
After sorting, the selected droplets are collected in a tube; the emulsion is then broken upon addition of 1 volume of emulsion breaker (Raindance) and the aqueous phase, supplemented with total RNA (50 μL) in water is recovered. The aqueous phase is washed by chloroform extraction, as described above, and DNA is recovered by ethanol precipitation with LPA (GeneLute LPA Linear Polyacrylamide, Sigma-Aldrich, an efficient neutral carrier for precipitation of pictogram amounts of nucleic acids with ethanol). Then, rolling circle amplification (RCA) is performed: the pellet is suspended in a suitable buffer with degenerated oligonucleotides, heated at 95°C for 5 minutes, then 10 μL of rolling circle buffer, phi 29 polymerase and dNTPs are added and the reaction is incubated at 37°C for 7h (Illustra GenomiPhi V2 DNA Amplification Kit, GE Healthcare), then at 65°C for 20 minutes to denature the polymerase. Eco RI is added and samples are incubated at 37°C for 2 h, and 65°C for 20 min. The resulted DNA is cleaned (chloroform extraction, ethanol precipitation) and resuspended in water. Starting from this step, different protocols have been set up depending on the type of experiment. For the model sorting experiment, a PCR is carried out on part of the amplified DNA with primers designed to discriminate the wild type and the mutant of PAC. For sorting experiments made on libraries of PAC variants, the linearized DNA plasmids are circularized by ligation (T4 DNA ligase, New England Biolabs). Ligation product is electroporated into electrocompetent cells (MegaX DH10B T1R Electrocomp Cells, Invitrogen); cells are plated on big LB-Amp agarose plates and incubated over night at 37°C. Colonies are collected, suspended in liquid medium (LB-Amp, 50 mL) and incubated at 37°C for 4 h. Finally, plasmids are extracted using midi-prep kits (Qiagen) and are ready for a new round of selection.

Oligonucleotides sequences

1. tcacatgccaccatggagcagtcgtcaagtgagataaa
2. ttgaatgcacccgggatctagacctctctgaacgtgcaacactt
3. tcagagaagcggtttgccatagtgcccacaaatatcatcgc
4. gcgatgatatttgtgggcactatggcaaaccgcttctctga
5. gctagttattgctcagcgg
6. gagaccacaacggtttccc
7. tgtcgctgcttgcagagtagg
8. ttgcacaatttgcacagggtgg
9. ttcccagcgcttaggagtaa

DsbAss-for: catgaaaaagatttggctggcgctggctggtttagttttagcgtttagcgcatcggcggc

DsbAss-rev:catggccgccgatgcgctaaacgctaaaactaaaccagccagcgccagccaaatcttttt

Genetic constructions

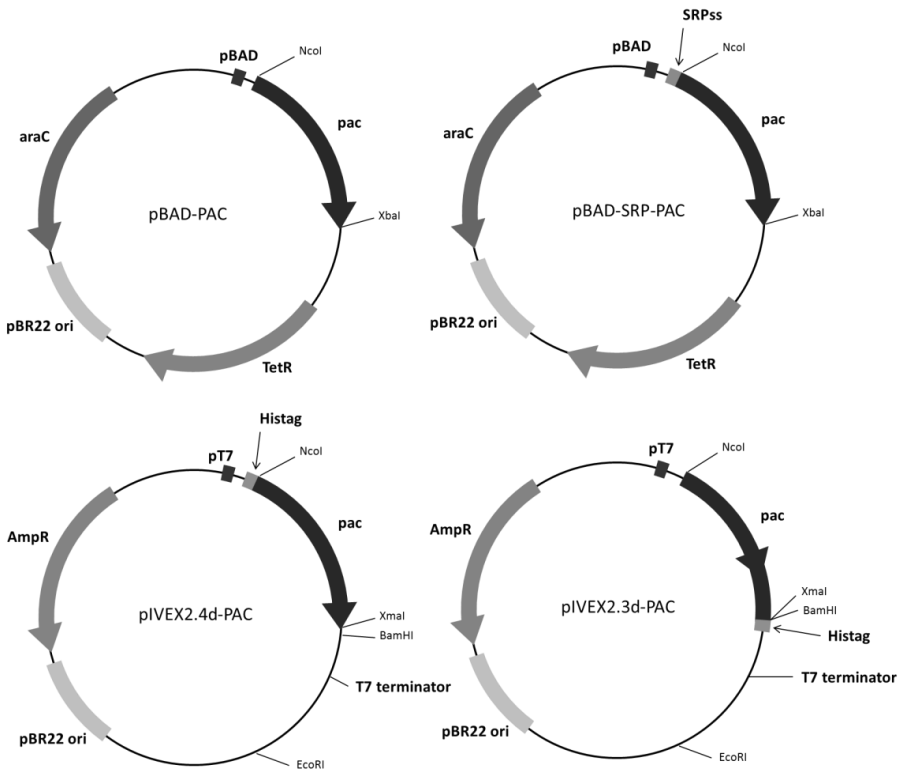


Figure 5.5: constructed plasmids that were tested for PAC expression.

Sequences

Pac gene translated from pIVEX2.4d-PAC. The propetide is in bold.

```
1     atg tct ggt tct cat cat cat cat cat cat agc agc ggc atc gaa   45
1     Met Ser Gly Ser His His His His His His Ser Ser Gly Ile Glu   15
                                          NcoI
46    ggc cgc ggc cgc tta att aaa cat atg acc atg gag cag tcg tca   90
16    Gly Arg Gly Arg Leu Ile Lys His Met Thr Met Glu Gln Ser Ser   30

91    agt gag ata aag att gtt cgc gat gaa tac ggc atg ccg cat att   135
31    Ser Glu Ile Lys Ile Val Arg Asp Glu Tyr Gly Met Pro His Ile   45

136   tat gcc aat gat aca tgg cac cta ttt tat ggc tat ggc tat gta   180
46    Tyr Ala Asn Asp Thr Trp His Leu Phe Tyr Gly Tyr Gly Tyr Val   60

181   gta gca caa gat cgc ctt ttt cag atg gaa atg gca cgt cgc agt   225
61    Val Ala Gln Asp Arg Leu Phe Gln Met Glu Met Ala Arg Arg Ser   75

226   act caa ggg act gtc gcg gaa gtg ctt ggc aaa gat ttt gtg aaa   270
76    Thr Gln Gly Thr Val Ala Glu Val Leu Gly Lys Asp Phe Val Lys   90

271   ttt gat aaa gat atc cgt cgt aac tac tgg ccg gat gct atc cgg   315
91    Phe Asp Lys Asp Ile Arg Arg Asn Tyr Trp Pro Asp Ala Ile Arg   105

316   gcg caa att gct gcc ctt tcc cca gag gat atg tcc att ctg caa   360
106   Ala Gln Ile Ala Ala Leu Ser Pro Glu Asp Met Ser Ile Leu Gln   120

361   ggc tac gct gat gga atg aat gcc tgg att gat aag gta aat acc   405
121   Gly Tyr Ala Asp Gly Met Asn Ala Trp Ile Asp Lys Val Asn Thr   135

406   aat cca gag acg ctc tta cca aaa cag ttt aat aca ttt ggc ttt   450
136   Asn Pro Glu Thr Leu Leu Pro Lys Gln Phe Asn Thr Phe Gly Phe   150

451   act cct aag cgc tgg gaa ccg ttt gat gtc gcg atg ata ttt gtg   495
151   Thr Pro Lys Arg Trp Glu Pro Phe Asp Val Ala Met Ile Phe Val   165

496   ggc act atg gca aac cgc ttc tct gat agc act agc gaa att gat   540
166   Gly Thr Met Ala Asn Arg Phe Ser Asp Ser Thr Ser Glu Ile Asp   180

541   aat ctg gca ctg cta acg gct tta aaa gat aaa tat ggt gta tca   585
181   Asn Leu Ala Leu Leu Thr Ala Leu Lys Asp Lys Tyr Gly Val Ser   195

586   caa ggc atg gcg gta ttt aat cag ttg aaa tgg ctg gta aac cca   630
196   Gln Gly Met Ala Val Phe Asn Gln Leu Lys Trp Leu Val Asn Pro   210

631   tca gcg cca acc act att gcc gta caa gag agt aac tac cca ctt   675
211   Ser Ala Pro Thr Thr Ile Ala Val Gln Glu Ser Asn Tyr Pro Leu   225

676   aaa ttt aat cag caa aac tcg caa aca gca gct ctg ttg cca cgc   720
226   Lys Phe Asn Gln Gln Asn Ser Gln Thr Ala Ala Leu Leu Pro Arg   240

721   tac gat tta cct gca cca atg ctt gac cga cca gca aaa ggg gcg   765
241   Tyr Asp Leu Pro Ala Pro Met Leu Asp Arg Pro Ala Lys Gly Ala   255

766   gat ggc gca ctg ctg gcg tta aca gca ggg aag aac cgg gaa act   810
256   Asp Gly Ala Leu Leu Ala Leu Thr Ala Gly Lys Asn Arg Glu Thr   270
                                              BstXI
811   att gtt gca caa ttt gca cag ggt ggt gcc aat ggt ctg gcg ggg   855
271   Ile Val Ala Gln Phe Ala Gln Gly Gly Ala Asn Gly Leu Ala Gly   285

856   tat cca acg acc agc aat atg tgg gtg atc ggc aaa agc aaa gcc   900
286   Tyr Pro Thr Thr Ser Asn Met Trp Val Ile Gly Lys Ser Lys Ala   300
```

```
901   cag gat gcg aaa gca atc atg gta aat ggt ccg cag ttt ggc tgg   945
301   Gln Asp Ala Lys Ala Ile Met Val Asn Gly Pro Gln Phe Gly Trp   315

946   tat gcg cct gcg tat act tat ggt att ggt ctg cac ggt gct ggt   990
316   Tyr Ala Pro Ala Tyr Thr Tyr Gly Ile Gly Leu His Gly Ala Gly   330

991   tat gat gtc act ggc aat aca cca ttt gcc tat cct ggg ctg gtt   1035
331   Tyr Asp Val Thr Gly Asn Thr Pro Phe Ala Tyr Pro Gly Leu Val   345

1036  ttt ggt cat aat ggt gtg att tcc tgg gga tca acg gca ggt ttc   1080
346   Phe Gly His Asn Gly Val Ile Ser Trp Gly Ser Thr Ala Gly Phe   360

1081  ggc gat gat gtc gat att ttt gct gaa cgg ctg tcg gca gag aaa   1125
361   Gly Asp Asp Val Asp Ile Phe Ala Glu Arg Leu Ser Ala Glu Lys   375

1126  cca ggc tac tac ttg cat aat ggt aag tgg gtg aaa atg tta agc   1170
376   Pro Gly Tyr Tyr Leu His Asn Gly Lys Trp Val Lys Met Leu Ser   390

1171  cgt gag gaa acc att acg gtg aaa aat ggt cag gca gag acc ttt   1215
391   Arg Glu Glu Thr Ile Thr Val Lys Asn Gly Gln Ala Glu Thr Phe   405

1216  act gtc tgg cgt acg gtg cat ggc aac att ctc caa act gac cag   1260
406   Thr Val Trp Arg Thr Val His Gly Asn Ile Leu Gln Thr Asp Gln   420

1261  acg aca caa acg gct tac gct aaa tcc cgc gca tgg gat ggt aaa   1305
421   Thr Thr Gln Thr Ala Tyr Ala Lys Ser Arg Ala Trp Asp Gly Lys   435

1306  gag gtg gcg tct ttg ctg gcc tgg act cat cag atg aag gcc aaa   1350
436   Glu Val Ala Ser Leu Leu Ala Trp Thr His Gln Met Lys Ala Lys   450

1351  aat tgg cag gag tgg aca cag cag gca gcg aaa caa gca ctg acc   1395
451   Asn Trp Gln Glu Trp Thr Gln Gln Ala Ala Lys Gln Ala Leu Thr   465

1396  atc aac tgg tac tat gct gat gta aac ggc aat att ggt tat gtt   1440
466   Ile Asn Trp Tyr Tyr Ala Asp Val Asn Gly Asn Ile Gly Tyr Val   480

1441  cat act ggt gct tat cca gat cgt caa tca ggc cat gat ccg cga   1485
481   His Thr Gly Ala Tyr Pro Asp Arg Gln Ser Gly His Asp Pro Arg   495

1486  tta ccc gtt cct ggt acg gga aaa tgg gac tgg aaa ggg cta ttg   1530
496   Leu Pro Val Pro Gly Thr Gly Lys Trp Asp Trp Lys Gly Leu Leu   510
                          Bsu36I
1531  cct ttt gaa atg aac cct aag gtg tat aac ccc cag tcg gga tat   1575
511   Pro Phe Glu Met Asn Pro Lys Val Tyr Asn Pro Gln Ser Gly Tyr   525

1576  att gct aac tgg aac aat tct ccc caa aaa gat tat ccc gct tca   1620
526   Ile Ala Asn Trp Asn Asn Ser Pro Gln Lys Asp Tyr Pro Ala Ser   540

1621  gat ctg ttt gcc ttt ttg tgg ggt ggt gca gat cgc gtt acg gag   1665
541   Asp Leu Phe Ala Phe Leu Trp Gly Gly Ala Asp Arg Val Thr Glu   555

1666  atc gac cga ctg ctt gag caa aag cca cgc tta act gct gat cag   1710
556   Ile Asp Arg Leu Leu Glu Gln Lys Pro Arg Leu Thr Ala Asp Gln   570
                                                      Bsu36I
1711  gca tgg gat gtt att cgc caa acc agt cgt cag gat ctt aac ctg   1755
571   Ala Trp Asp Val Ile Arg Gln Thr Ser Arg Gln Asp Leu Asn Leu   585

1756  agg ctt ttt tta cct act ctg caa gca gcg aca tct ggt ttg aca   1800
586   Arg Leu Phe Leu Pro Thr Leu Gln Ala Ala Thr Ser Gly Leu Thr   600

1801  cag agc gat ccg cgt cgt cag ttg gta gaa aca tta aca cgt tgg   1845
601   Gln Ser Asp Pro Arg Arg Gln Leu Val Glu Thr Leu Thr Arg Trp   615

1846  gat ggc atc aat ttg ctt aat gat gat ggt aaa acc tgg cag cag   1890
616   Asp Gly Ile Asn Leu Leu Asn Asp Asp Gly Lys Thr Trp Gln Gln   630
```

```
1891  cca ggc tct gcc atc ctg aac gtt tgg ctg acc agt atg ttg aag   1935
631   Pro Gly Ser Ala Ile Leu Asn Val Trp Leu Thr Ser Met Leu Lys   645

1936  cgt acc gta gtg gct gcc gta cct atg cca ttt gat aag tgg tac   1980
646   Arg Thr Val Val Ala Ala Val Pro Met Pro Phe Asp Lys Trp Tyr   660

1981  agc gcc agt ggc tac gaa aca acc cag gac ggc cca act ggt tcg   2025
661   Ser Ala Ser Gly Tyr Glu Thr Thr Gln Asp Gly Pro Thr Gly Ser   675

2026  ctg aat ata agt gtt gga gca aaa att ttg tat gag gcg gtg cag   2070
676   Leu Asn Ile Ser Val Gly Ala Lys Ile Leu Tyr Glu Ala Val Gln   690

2071  gga gac aaa tca cca atc cca cag gcg gtt gat ctg ttt gct ggg   2115
691   Gly Asp Lys Ser Pro Ile Pro Gln Ala Val Asp Leu Phe Ala Gly   705

2116  aaa cca cag cag gag gtt gtg ttg gct gcg ctg gaa gat acc tgg   2160
706   Lys Pro Gln Gln Glu Val Val Leu Ala Ala Leu Glu Asp Thr Trp   720

2161  gag act ctt tcc aaa cgc tat ggc aat aat gtg agt aac tgg aaa   2205
721   Glu Thr Leu Ser Lys Arg Tyr Gly Asn Asn Val Ser Asn Trp Lys   735

2206  aca cct gca atg gcc tta acg ttc cgg gca aat aat ttc ttt ggt   2250
736   Thr Pro Ala Met Ala Leu Thr Phe Arg Ala Asn Asn Phe Phe Gly   750

2251  gta ccg cag gcc gca gcg gaa gaa acg cgt cat cag gcg gag tat   2295
751   Val Pro Gln Ala Ala Ala Glu Glu Thr Arg His Gln Ala Glu Tyr   765

2296  caa aac cgt gga aca gaa aac gat atg att gtt ttc tca cca acg   2340
766   Gln Asn Arg Gly Thr Glu Asn Asp Met Ile Val Phe Ser Pro Thr   780

2341  aca agc gat cgt cct gtg ctt gcc tgg gat gtg gtc gca ccc ggt   2385
781   Thr Ser Asp Arg Pro Val Leu Ala Trp Asp Val Val Ala Pro Gly   795

2386  cag agt ggg ttt att gct ccc gat gga aca gtt gat aag cac tat   2430
796   Gln Ser Gly Phe Ile Ala Pro Asp Gly Thr Val Asp Lys His Tyr   810

2431  gaa gat cag ctg aaa atg tac gaa aat ttt ggc cgt aag tcg ctc   2475
811   Glu Asp Gln Leu Lys Met Tyr Glu Asn Phe Gly Arg Lys Ser Leu   825

2476  tgg tta acg aag cag gat gtg gag gcg cat aag gag tcg cag gaa   2520
826   Trp Leu Thr Lys Gln Asp Val Glu Ala His Lys Glu Ser Gln Glu   840
                                          XmaI
2521  gtg ttg cac gtt cag aga ggt cta gat ccc ggg atc cgg   2559
841   Val Leu His Val Gln Arg Gly Leu Asp Pro Gly Ile Arg
```

Sequence of the inserted DNA fragment for the TM Library construction.

Ordered from GenScript, *de novo* gene synthesis from oligonucleotides.

```
          BstXI                                              TML signature
1     gcc aat ggt ctg gcg ggg tat cca acg acc agc aat atg tgg gtt   45
1     Ala Asn Gly Leu Ala Gly Tyr Pro Thr Thr Ser Asn Met Trp Val   15

46    atc ggc aaa agc aaa gcc cag gat gcg aaa gca atc atg gta aat   90
16    Ile Gly Lys Ser Lys Ala Gln Asp Ala Lys Ala Ile Met Val Asn   30

91    ggt ccg cag nnk ggc tgg tat gcg cct gcg tat nnk tat ggt att   135
31    Gly Pro Gln NNN Gly Trp Tyr Ala Pro Ala Tyr NNN Tyr Gly Ile   45

136   ggt ctg cac ggt gct ggt tat gat gtc act ggc aat aca nnk ttt   180
46    Gly Leu His Gly Ala Gly Tyr Asp Val Thr Gly Asn Thr NNN Phe   60

181   gcc tat cct ggg ctg nnk ttt ggt cat aat ggt gtg att tcc tgg   225
61    Ala Tyr Pro Gly Leu NNN Phe Gly His Asn Gly Val Ile Ser Trp   75
```

```
226   gga tca acg nnk ggt ttc ggc gat gat gtc gat att ttt gct gaa   270
76    Gly Ser Thr NNN Gly Phe Gly Asp Asp Val Asp Ile Phe Ala Glu   90

271   cgg ctg tcg gca gag aaa cca ggc tac tac ttg cat aat ggt aag   315
91    Arg Leu Ser Ala Glu Lys Pro Gly Tyr Tyr Leu His Asn Gly Lys   105

316   tgg gtg aaa atg tta agc cgt gag gaa acc att acg gtg aaa aat   360
106   Trp Val Lys Met Leu Ser Arg Glu Glu Thr Ile Thr Val Lys Asn   120

361   ggt cag gca gag acc ttt act gtc tgg cgt acg gtg cat ggc aac   405
121   Gly Gln Ala Glu Thr Phe Thr Val Trp Arg Thr Val His Gly Asn   135

406   att ctc caa act gac cag acg aca caa acg gct tac gct aaa tcc   450
136   Ile Leu Gln Thr Asp Gln Thr Thr Gln Thr Ala Tyr Ala Lys Ser   150

451   cgc gca tgg gat ggt aaa gag gtg gcg tct ttg ctg gcc nnk act   495
151   Arg Ala Trp Asp Gly Lys Glu Val Ala Ser Leu Leu Ala NNN Thr   165

496   cat cag atg aag gcc aaa aat tgg cag gag tgg aca cag cag gca   540
166   His Gln Met Lys Ala Lys Asn Trp Gln Glu Trp Thr Gln Gln Ala   180

541   gcg aaa caa gca ctg acc nnk aac tgg tac tat gct gat gta aac   585
181   Ala Lys Gln Ala Leu Thr NNN Asn Trp Tyr Tyr Ala Asp Val Asn   195

586   ggc aat att ggt tat gtt cat act ggt gct tat cca gat cgt caa   630
196   Gly Asn Ile Gly Tyr Val His Thr Gly Ala Tyr Pro Asp Arg Gln   210

631   tca ggc cat gat ccg cga tta ccc gtt cct ggt acg gga aaa tgg   675
211   Ser Gly His Asp Pro Arg Leu Pro Val Pro Gly Thr Gly Lys Trp   225

676   gac tgg aaa ggg cta ttg cct ttt gaa atg aac cca aag gtg tat   720
226   Asp Trp Lys Gly Leu Leu Pro Phe Glu Met Asn Pro Lys Val Tyr   240

721   aac ccc cag tcg gga tat att gct aac tgg aac aat tct ccc caa   765
241   Asn Pro Gln Ser Gly Tyr Ile Ala Asn Trp Asn Asn Ser Pro Gln   255

766   aaa gat tat ccc gct tca gat ctg ttt gcc ttt ttg tgg ggt ggt   810
256   Lys Asp Tyr Pro Ala Ser Asp Leu Phe Ala Phe Leu Trp Gly Gly   270

811   gca gat cgc gtt acg gag atc gac cga ctg ctt gag caa aag cca   855
271   Ala Asp Arg Val Thr Glu Ile Asp Arg Leu Leu Glu Gln Lys Pro   285

856   cgc tta act gct gat cag gca tgg gat gtt att cgc caa acc agt   900
286   Arg Leu Thr Ala Asp Gln Ala Trp Asp Val Ile Arg Gln Thr Ser   300
                        Bsu36I
901   cgt cag gat ctt aac ctg agg   921
301   Arg Gln Asp Leu Asn Leu Arg
```

DsbA signal peptide, cloned in a NcoI restriction site

```
2     atg aaa aag att tgg ctg gcg ctg gct ggt tta gtt tta gcg ttt   46
1      M   K   K   I   W   L   A   L   A   G   L   V   L   A   F    15

47    agc gca tcg gcg gcg atg  64
16     S   A   S   A   A   M   20
```

References

1. Sato E, Matsuhisa A, Sakashita M, Kanaoka Y; *Organic Fluorescent Reagents .13. New Water-Soluble Fluorogenic Amine - 7-Aminocoumarin-4-Methanesulfonic Acid (Acms) and Related Substrates for Proteinases*. Chemical & Pharmaceutical Bulletin **1988**; 36: 3496-3502

2. Atkins RL, Bliss DE; *Substituted coumarins and azacoumarins. synthesis and fluorescent properties*. J. Org. Chem. **1978**; 43: 1975-1980

3. Pianowski ZL, Winssinger N; *Fluorescence-based detection of single nucleotide permutation in DNA via catalytically templated reaction*. Chemical Communications **2007**: 3820-3822

4. Kanaoka Y, Kobayashi A, Sato E, Nakayama H, Ueno T, Muno D, Sekine T; *Organic Fluorescent Reagents .10. Multifunctional Cross-Linking Reagents .1. Synthesis and Properties of Novel Photoactivable, Thiol-Directed Fluorescent Reagents*. Chemical & Pharmaceutical Bulletin **1984**; 32: 3926-3933

5. Heltweg B, Dequiedt F, Marshall BL, Branch C, Yoshida M, Nishino N, Verdin E, Jung M; *Subtype selective substrates for histone deacetylases*. Journal of Medicinal Chemistry **2004**; 47: 5235-5243

6. Mazutis L, Araghi AF, Miller OJ, Baret JC, Frenz L, Janoshazi A, Taly V, Miller BJ, Hutchison JB, Link D, Griffiths AD, Ryckelynck M; *Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis*. Anal Chem **2009**; 81: 4813-21

7. Frenz L, Blank K, Brouzes E, Griffiths AD; *Reliable microfluidic on-chip incubation of droplets in delay-lines*. Lab Chip **2009**; 9: 1344-8

8. Xia YN, Whitesides GM; *Soft lithography*. Annual Review of Materials Science **1998**; 28: 153-184

9. Holtze C, Rowat AC, Agresti JJ, Hutchison JB, Angile FE, Schmitz CH, Koster S, Duan H, Humphry KJ, Scanga RA, Johnson JS, Pisignano D, Weitz DA; *Biocompatible surfactants for water-in-fluorocarbon emulsions*. Lab Chip **2008**; 8: 1632-9

10. Alkema WB, de Vries E, Floris R, Janssen DB; *Kinetics of enzyme acylation and deacylation in the penicillin acylase-catalyzed synthesis of beta-lactam antibiotics*. Eur J Biochem **2003**; 270: 3675-83