

Université du Québec
INRS-Institut Armand-Frappier-Microbiologie et biotechnologie

Analyse du protéome de *Streptomyces coelicolor* M145

Stéphanie Trudel

Mémoire présenté pour l'obtention du grade de
Maître ès sciences (M.Sc.) en microbiologie appliquée

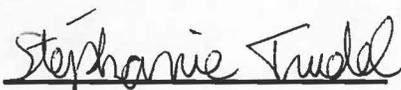
Jury d'évaluation

Examineur interne
Examineur externe
Directeur de recherche

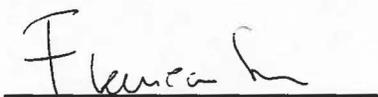
François Lépine, Microbiologie et biotechnologie
Carole Beaulieu, Université de Sherbrooke
François Shareck, Microbiologie et biotechnologie

RÉSUMÉ

Les streptomycètes sont des bactéries filamenteuses retrouvées dans le sol. Ils sont parmi les microorganismes les plus importants au niveau industriel, produisant de nombreux métabolites secondaires et sécrétant une grande quantité d'enzymes. Le séquençage du génome de 8 Mb de *Streptomyces coelicolor* A3(2) a été entrepris au Sanger Centre et l'analyse complète du génome devrait être terminée incessamment. La prochaine étape consiste donc à assigner une protéine à chacun des cadres de lecture nouvellement identifiés. En d'autres termes, l'étude du complément protéique exprimé par son génome, le protéome, sera réalisée. Nous nous sommes d'abord penché sur l'étude du protéome exprimé lorsque la souche est cultivée avec le xylane comme source de carbone. En effet, étant donné l'intérêt commercial des xylanases dans l'industrie des pâtes et papier, le groupe de recherche en biotechnologie des streptomycètes a acquis au cours des dernières années de nombreuses connaissances sur la dégradation du xylane. Durant le présent travail, les différentes techniques permettant l'analyse du protéome ont été mises au point et optimisées avec les souches *S. lividans* 1326 et IAF6. Par la suite, l'identification par spectrométrie de masse des protéines intracellulaires les plus abondantes sur gel en deux dimensions de *S. coelicolor* a débuté. Quatre protéines ont été identifiées, soit une méthylase qui confère la résistance au thiostrepton, deux isoformes d'une protéine de 20,3 kDa (Q9KXW4) ainsi qu'une protéine conservée (Q9KY22).



Étudiant



Directeur de recherche

REMERCIEMENTS

Au cours de mon séjour à l'Institut Armand-Frappier, j'ai été entourée de gens qui ont su m'aider et me supporter, tant au niveau scientifique que personnel, lorsque cela s'est avéré nécessaire. Je tiens donc à les remercier aujourd'hui, ils ont tous une part plus ou moins grande dans la réussite de mon projet.

Tout d'abord, je remercie la Fondation Armand-Frappier, la Power Corporation ainsi que mon directeur de recherche, François Shareck, pour les bourses d'études qu'ils m'ont accordées. Merci François de m'avoir accueillie dans ton laboratoire et de m'avoir fait confiance pour le projet du protéome. Merci pour les conseils judicieux que tu m'as donnés pendant mon travail de laboratoire mais aussi pendant ma rédaction. Tu as été un bon directeur pour moi, disponible mais surtout patient, étant donné mon caractère parfois difficile. Un merci spécial à François Lépine, avec qui j'ai appris énormément. Merci du temps passé à m'enseigner les rudiments de la spectrométrie de masse.

Il me faut aussi remercier les membres du groupe des Streptomycètes, vous m'avez tous aidée à un moment ou à un autre. Plus que des collègues vous avez été des amis avant tout et grâce à vous j'ai passé de belles années à l'Institut. Je garde un très bon souvenir de vous tous. Un merci tout particulier à mes compères de laboratoire: Lise, qui n'a jamais été avare de ton temps, Valérie, Marco et Maïté pour l'esprit d'équipe que vous avez réussi à créer.

Au moment de déposer mon mémoire, j'ai une pensée spéciale pour mes parents, Luc et Guylaine, qui m'encouragent et m'épaulent depuis toujours. Je vous suis reconnaissante des sacrifices que vous avez fait pour moi. C'est grâce à vous si je suis devenue ce que je suis aujourd'hui. Julie et Véronique, merci tout simplement d'être mes sœurs.

Je ne peux oublier Sébastien, qui à l'époque m'a encouragée à entreprendre ma maîtrise. Merci d'avoir été à mes cotés tout au long de mes études et merci de l'être encore.

Nathalie et Christine, vous avez été des oreilles attentives pour moi. Merci les filles de m'avoir changé les idées lorsque j'en avais besoin.

TABLE DES MATIÈRES

RÉSUMÉ	II
REMERCIEMENTS	III
TABLE DES MATIÈRES	V
LISTE DES FIGURES	VIII
LISTE DES TABLEAUX	X
LISTE DES ABRÉVIATIONS	XII
INTRODUCTION	1
1. REVUE BIBLIOGRAPHIQUE.....	3
1.1 DE GÉNOMES À PROTÉOMES	4
1.1.1 La révolution du génome	4
1.1.2 Le protéome	7
1.2 PRINCIPALES ÉTAPES DE L'ÉTUDE DU PROTÉOME.....	8
1.2.1 Séparation des protéines: l'électrophorèse en deux dimensions	8
1.2.2 Visualisation des protéines et analyse de l'image	11
1.2.3 Identification des protéines	11
1.2.3.1 Spectrométrie de masse.....	12
1.2.3.2 Autres techniques d'identification des protéines	16
1.2.3.3 Recherche dans les bases de données	16
1.2.4 Bases de données de gels 2-D.....	17
1.3 LES STREPTOMYCÈTES.....	20
1.3.1 Généralités et importance industrielle	20
1.3.2 Streptomyces coelicolor: un modèle chez les streptomycètes.....	21
1.3.2.1 Éléments génétiques variés	21
1.3.2.2 Outils permettant la manipulation génétique	21
1.3.2.3 Banque ordonnée de cosmides.....	22
1.3.3 Séquençage et analyse du génome de S. coelicolor M145	25
1.3.4 Analyse du protéome.....	25
1.3.5 La dégradation du xylane par S. lividans	26
1.3.5.1 Le xylane.....	26
1.3.5.2 Le système xylanolytique	26
2. APPROCHE EXPÉRIMENTALE	30
2.1 LISTE DES PRODUITS CHIMIQUES UTILISÉS	31
2.2 SOUCHES BACTÉRIENNES.....	33
2.3 PLASMIDES	33
2.4 MILIEUX DE CULTURES	35

2.4.1 Milieu solide pour la conservation des souches	35
2.4.2 Milieu liquide riche pour la production d'inoculum.....	36
2.4.3 Milieu liquide minimal pour la croissance cellulaire	36
2.5 CONDITIONS DE CULTURES	37
2.5.1 Production d'inoculum	37
2.5.2 Culture en fermenteur	37
2.6 PRÉPARATION DES ÉCHANTILLONS PROTÉIQUES	37
2.7 DOSAGE DES PROTÉINES.....	38
2.8 ÉLECTROPHORÈSE EN DEUX DIMENSIONS	38
2.8.1 Incorporation des protéines par réhydratation des languettes.....	38
2.8.2 La première dimension: la focalisation isoélectrique.....	39
2.8.3 Équilibrage des languettes.....	40
2.8.4 La deuxième dimension: l'électrophorèse en gel de polyacrylamide en présence de SDS.....	40
2.9 VISUALISATION DES PROTÉINES PAR COLORATION AU NITRATE D'ARGENT COMPATIBLE AVEC LA DIGESTION TRYPTIQUE.....	41
2.10 ANALYSE DES GELS 2-D.....	42
2.11 DIGESTION DES PROTÉINES AVEC LA TRYPSINE	42
2.12 ANALYSE DES PEPTIDES PAR SPECTROMÉTRIE DE MASSE	44
2.12.1 Concentration et purification de l'échantillon de peptides.....	44
2.12.2 Injection dans le spectromètre de masse	45
2.12.3 Recherche dans les bases de données.....	45
3. RÉSULTATS	49
3.1 MISE AU POINT DES DIFFÉRENTES ÉTAPES DE L'ANALYSE DU PROTÉOME	50
3.1.1 L'électrophorèse en deux dimensions et détection des protéines.....	50
3.1.2 Analyse des gels: estimation du point isoélectrique et du poids moléculaire	51
3.1.3 Analyse des peptides par spectrométrie de masse: profil de la masse des peptides.....	51
3.1.4 Recherche dans les bases de données avec le logiciel PeptIdent	56
3.1.5 Analyse des peptides par spectrométrie de masse: fragmentation et recherche de la séquence en acides aminés	60
3.2 ANALYSE DU PROTÉOME DE <i>S. COELICOLOR</i> M145	65
3.2.1 Séparation par électrophorèse en deux dimensions des protéines possédant des pI acides	65
3.2.2 Analyse des gels: identification des protéines les plus abondantes.....	65
3.2.3 Analyse par spectrométrie de masse des protéines les plus abondantes: profil de la masse des peptides	72
3.2.3.1 Protéine SC69 (122).....	72
3.2.3.2 Protéine SC34 (71).....	78
3.2.3.3 Protéine SC124 (68).....	78
3.2.3.4 Protéine SC117	85
3.2.3.5 Protéine SC121	85
3.2.4 Analyse des peptides par spectrométrie de masse: fragmentation et recherche de la séquence en acides aminés	91

3.2.4.1 Protéine 124 (68).....	91
3.2.4.2 Protéine 121	91
4. DISCUSSION.....	96
4.1 MISE AU POINT DES ÉTAPES D'UNE ANALYSE PROTÉOMIQUE.....	97
4.1.1 L'électrophorèse en deux dimensions et la détection des protéines.....	97
4.1.2 Analyse des peptides par spectrométrie de masse: profil de la masse des peptides.....	98
4.1.3 Recherche dans les bases de données avec le logiciel PeptIdent	99
4.1.4 Analyse des peptides par spectrométrie de masse: fragmentation des peptides et recherche de la séquence en acides aminés	101
4.2 ANALYSE DU PROTÉOME DE <i>S. COELICOLOR</i> M145	101
4.2.1 Profil de la masse des peptides des protéines SC69 (122) et SC34 (71).....	101
4.2.2 Profil de la masse des peptides des protéines SC124 (68), SC117 et SC121	102
4.2.2.1 Protéine SC124 (68).....	102
4.2.2.2 Protéine SC117.....	104
4.2.2.3 Protéine SC121.....	104
4.2.3 Comparaison des protéines SC124 (68) et SC121	108
5. CONCLUSION.....	114
BIBLIOGRAPHIE.....	117

LISTE DES FIGURES

	Page
Figure 1: La focalisation isoélectrique	9
Figure 2: L'électrophorèse en gel de polyacrylamide en présence de SDS	10
Figure 3: Le profil de la masse des peptides	14
Figure 4: Fragmentation d'un peptide par spectrométrie de masse	15
Figure 5: Image de gel retrouvée dans la base de données SWISS-2DPAGE	18
Figure 6: Carte physique du cosmide SuperCos 1	23
Figure 7: Carte physique et génétique du chromosome de <i>S. coelicolor</i> M145	24
Figure 8: Modèle hypothétique de la dégradation du xylane par <i>S. lividans</i>	29
Figure 9: Carte physique du plasmide IAF6	34
Figure 10: Exemple d'un ion possédant une charge et deux charges	47
Figure 11: Gel 2-D des protéines cytoplasmiques de <i>S. lividans</i> IAF6 et de <i>S. lividans</i> 1326	52
Figure 12: Image numérisée du gel 2-D de pH 4-7 de la souche <i>S. lividans</i> IAF 6	53
Figure 13: Spectre de masse de la digestion tryptique de la protéine SL82.	55
Figure 14: Spectre de masse de la fragmentation de l'ion d'un m/z de 515	61
Figure 15: Spectre de masse de la fragmentation de l'ion d'un m/z de 580	62
Figure 16: Spectre de masse de la fragmentation de l'ion d'un m/z de 690	63

Figure 17: Spectre de masse de la fragmentation de l'ion d'un m/z de 679	64
Figure 18: Image numérisée du gel 2-D de pH 3,5-4,5 de la souche <i>S. coelicolor</i> M145	67
Figure 19: Image numérisée du gel 2-D de pH 4,0-5,0 de la souche <i>S. coelicolor</i> M145	68
Figure 20: Spectre de masse de la digestion tryptique de la protéine SC69 (122)	75
Figure 21: Spectre de masse de la digestion tryptique de la protéine SC69 (122) des valeurs du rapport m/z de 500 à 680 et de 700 à 1120	76
Figure 22: Spectre de masse de la digestion tryptique de la protéine SC34 (71)	80
Figure 23: Spectre de masse de la digestion tryptique de la protéine SC34 (71) pour les valeurs du rapport m/z de 470 à 680 et de 700 à 1200	81
Figure 24: Spectre de masse de la digestion tryptique de la protéine SC124 (68)	82
Figure 25: Spectre de masse de la digestion tryptique de la protéine SC117	87
Figure 26: Spectre de masse de la digestion tryptique de la protéine SC121	88
Figure 27: Spectre de masse de la fragmentation de l'ion d'un m/z de 607	92
Figure 28: Spectre de masse de la fragmentation de l'ion d'un m/z de 762	93
Figure 29: Spectre de masse de la fragmentation de l'ion d'un m/z de 762	94
Figure 30: Spectre de masse de la fragmentation de l'ion d'un m/z de 629	95

LISTE DES TABLEAUX

	Page
Tableau 1: Identification des gènes d' <i>H. influenzae</i> , de <i>M. genitalium</i> et de <i>M. jannaschii</i>	6
Tableau 2: Exemples de logiciels pour l'identification des protéines	17
Tableau 3: Informations textuelles disponibles dans la base de données SWISS-2DPAGE	19
Tableau 4: Souches bactériennes utilisées	33
Tableau 5: Description des plasmides	35
Tableau 6: Conditions d'électrophorèse utilisées lors de la dernière étape de la focalisation isoélectrique	40
Tableau 7: Paramètres utilisés lors de la recherche dans les bases de données avec le logiciel PeptIdent	48
Tableau 8: Liste des protéines cytoplasmiques de la souche IAF6	54
Tableau 9: Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SL82.	58
Tableau 10: Description détaillée de la comparaison entre le profil de la protéine 1 versus le profil de la protéine SL82	59
Tableau 11: Liste des protéines cytoplasmiques séparées par gel 2-D de pH 3,5-4,5 de <i>S. coelicolor</i> M145 détectées par le logiciel ainsi que leur volume	69
Tableau 12: Liste des protéines cytoplasmiques de <i>S. coelicolor</i> M145 séparées par gel 2-D de pH 4,0-5,0 détectées par le logiciel ainsi que leur volume	70
Tableau 13: Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SC69 (122)	77

Tableau 14:	Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SC124 (68)	83
Tableau 15:	Description détaillée de la comparaison entre le profil de la protéine 1 versus le profil de la protéine SC124 (68)	84
Tableau 16:	Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SC121	89
Tableau 17:	Description détaillée de la comparaison entre le profil de la protéine 2 versus le profil de la protéine SC121	90
Tableau 18:	Résultat de la recherche par homologie de séquence de la protéine de 20,3 kDa	103
Tableau 19:	Description de la protéine TerE de <i>S. marcescens</i> disponible dans la base de données EMBL	104
Tableau 20:	Résultat de la recherche par homologie de séquence de la protéine conservée	105
Tableau 21:	Description de la protéine TerD de <i>S. marcescens</i> disponible dans la base de données EMBL	106
Tableau 22:	Résultat de l'alignement de séquences des protéines SC124 et SC121	109
Tableau 23:	Digestion tryptique théorique de la protéine hypothétique de 20,3 kDa	112
Tableau 24:	Digestion tryptique théorique de la protéine hypothétique conservée	113

LISTE DES ABRÉVIATIONS

ADN	acide désoxyribonucléique
ARN	acide ribonucléique
BSA	albumine de sérum bovin
C	cytosine
CCD	charge coupled device
CHAPS	3-[(3-cholamidopropyl)dimethyl-ammonio]-1-propanesulfonate
°C	degré Celsius
CLO (s)	cadre (s) de lecture ouvert (s)
2-D	deux dimensions
DO	densité optique
DTT	dithiothréitol
EMBL	European Molecular Biology Laboratories
ESI	electrospray ionisation
G	guanine
g	gramme (s)
µg	microgramme (s)
IEF	focalisation isoélectrique
IPG	immobilized pH gradients
IT	ion trap
kb	kilopaire (s) de base (s)
kDa	kilodalton (s)
M	molaire
mM	millimolaire (s)
mA	milliampère (s)
Mb	mégapaire (s) de base (s)
MALDI	matrix assisted laser desorption ionisation
mg	milligramme (s)

ml	millilitre (s)
MS/MS	spectrométrie de masse en tandem
Mr	masse moléculaire relative
m/z	masse/charge
μl	microlitre (s)
N	normale
NCBI	National Center for Biotechnology Information
pb	paire (s) de base (s)
pI	point isoélectrique
p/v	poids/volume
psi	livres par pouce carré
rpm	révolution (s) par minute
SDS	sodium dodécyl sulfate
SDS-PAGE	sodium dodécyl sulfate polyacrylamide gel electrophoresis
TCA	acide trichloroacétique
TES	acide N-tris(hydroxyméthyl)méthyl-2-aminoéthanesulfonique
TFA	acide trifluoroacétique
TOF	time of flight
TQ	triple quadrupole
Tris	tris(hydroxyméthyl) aminométhane
TSB	bouillon tryptone de soya
V	volt (s)
Vh	volt-heures
v/v	volume/volume
W	watt (s)

INTRODUCTION

Les streptomycètes sont utilisés pour la production de la majorité des antibiotiques, d'herbicides et de plusieurs enzymes importantes pour l'industrie papetière. Ils sont uniques parmi les bactéries en ce qui a trait à leur cycle de vie, qui passe de la forme de mycélium à celle de spores et qui implique une régulation complexe de l'expression des gènes. Une bonne compréhension de la génétique, de la biologie moléculaire et de la physiologie des streptomycètes aiderait à leur utilisation accrue dans l'industrie mais procurerait aussi des connaissances fondamentales sur les procaryotes.

La recherche, dans ces différents domaines (génétique, biologie moléculaire et physiologie), se fait principalement avec l'organisme modèle *Streptomyces coelicolor* A3(2) (souche M145). Suite à la réalisation d'une carte physique de son chromosome et à la construction d'une banque ordonnée de cosmides, le séquençage du génome du 8 Mb a été entrepris. Étant donné que ce projet sera terminé d'ici la fin de l'an 2001, l'étape suivante consistera à étudier l'expression et la fonction de tous ces gènes nouvellement identifiés. Ainsi, l'analyse du protéome, les protéines exprimées par le génome de *S. coelicolor* A3(2) a été entreprise dans différents laboratoires. Chacun des groupes de recherche se concentre sur le protéome exprimé dans des conditions particulières: protéines exprimées lorsque la souche est cultivée dans un milieu minimal comparées à celles exprimées dans un milieu riche ou encore en présence de différentes sources de carbone.

L'habitat naturel des streptomycètes est le sol et le compost. Ces bactéries sont donc capables de dégrader la biomasse végétale, principalement composée de cellulose et d'hémicellulose dont la principale est le xylane. Elles ont donc la capacité de dégrader le xylane en un sucre simple, le xylose et l'utilisent comme source de carbone. Étant donné que le xylane est un substrat complexe, les streptomycètes sécrètent une panoplie d'enzymes afin de le dégrader complètement. Plusieurs de ces enzymes ont des

applications biotechnologiques. Ainsi, le groupe de recherche en biotechnologie des streptomycètes de l'institut Armand-Frappier s'intéresse depuis plusieurs années au système xylanolytique de *S. lividans*, un très proche parent de *S. coelicolor*. C'est pourquoi le protéome exprimé lorsque le xylane est utilisé comme source de carbone nous a d'abord intéressé.

Une étude du protéome implique la mise au point de techniques nouvelles et relativement complexes. L'objectif premier de ce projet a donc été de mettre au point et d'optimiser les différentes étapes d'une analyse protéomique, soit:

- la séparation des protéines par électrophorèse en deux dimensions
- l'analyse des gels avec des logiciels spécialisés
- la digestion des protéines en peptides
- l'analyse des peptides par spectrométrie de masse
- l'identification des protéines par des recherches dans les bases de données

Dans un deuxième temps, une fois les différentes techniques optimisées, l'analyse du protéome de *S. coelicolor* a été entreprise. Plus précisément au niveau des protéines cytoplasmiques exprimées lorsque la bactérie est cultivée avec le xylane comme source de carbone. Les protéines les plus abondantes sur gel en deux dimensions seront identifiées en premier afin d'augmenter les chances de détecter des peptides en spectrométrie de masse.

1. REVUE BIBLIOGRAPHIQUE

1.1 DE GÉNOMES À PROTÉOMES

1.1.1 La révolution du génome

L'automatisation du séquençage de l'ADN a permis à de nombreux projets de génomique de voir le jour. Ainsi, la séquence complète du génome du premier organisme vivant, *Haemophilus influenzae*, a été obtenue en 1995 (Fleishmann *et al.*, 1995), suivie de celui de *Mycoplasma genitalium* (Fraser *et al.*, 1995) et de *Methanococcus jannaschii* en 1996 (Bult *et al.*, 1996). Aujourd'hui, il est maintenant chose courante de séquencer le génome de microorganismes et à ce jour, 45 génomes microbiens sont complètement caractérisés et plus de 112 sont actuellement en cours (NCBI: www.ncbi.nlm.nih.gov).

L'emphase n'a pas été mise seulement sur l'analyse des génomes d'organismes simples (procaryotes et virus) puisqu'en 1996 le génome du premier eucaryote unicellulaire, *Saccharomyces cerevisiae*, a été complètement séquencé (Goffeau *et al.*, 1996), suivi en 1998 par celui du nématode *Caenorhabditis elegans*, un eucaryote pluricellulaire (The *C. elegans* sequencing consortium, 1998). À ce jour, plus de 172 projets de séquençage de génomes eucaryotes sont présentement en cours (NCBI).

L'analyse du génome d'un organisme comprend le séquençage, la définition de tous les cadres de lecture ouverts (CLO, régions codantes possibles) et leur identification par homologie avec les séquences disponibles dans les bases de données. Une fonction probable est assignée à un CLO lorsqu'il possède une homologie de séquence avec un gène codant pour une protéine déjà caractérisée (chez l'organisme ou chez d'autres espèces) ou avec l'une des 1379 familles de protéines classifiées dans la base de données PROSITE (www.expasy.ch/prosite). Une fonction hypothétique est attribuée au CLO lorsque l'homologie concorde avec une séquence retrouvée chez d'autres organismes mais dont le rôle de la protéine correspondante n'est pas encore défini. Finalement, une fonction inconnue est assignée lorsque le CLO ne possède aucune homologie avec les séquences disponibles ou avec les familles de protéines.

Cependant, l'analyse des premiers génomes séquencés, tels que ceux d'*H. influenzae*, de *M. genitalium* et de *M. jannaschii*, a révélé qu'entre 20 à 60% de leurs gènes n'ont pas de rôles connus ou déterminés (Tableau 1). Cette nouvelle donnée a donc soulevé la question suivante: comment s'assurer que ces gènes sont réellement exprimés et comment leur assigner une fonction biologique? De plus, l'idée primordiale du séquençage du génome de *M. genitalium* était de définir les gènes essentiels de base, puisque ce parasite possède le plus petit génome connu parmi tous les êtres vivants (580 kb, 470 CLOs). Ce génome minimal devrait contenir en majorité les gènes nécessaires à une existence indépendante et ils seraient donc conservés chez les organismes supérieurs. Comme 20% des protéines de *M. genitalium* possèdent une fonction inconnue, il est alors difficile de comprendre les processus biologiques de base et ceci laisse présager qu'il sera encore plus ardu de comprendre la biologie d'organismes possédant des génomes plus complexes.

L'analyse des génomes procure donc peu d'informations sur le fonctionnement individuel ou synergique des protéines afin de remplir leur rôle au sein de l'organisme, mais elle révèle également d'autres limites. En effet, la séquence d'ADN ne peut prédire à elle seule:

- 1) si et dans quelles conditions un gène est transcrit;
- 2) la concentration relative d'une protéine;
- 3) les modifications post-traductionnelles (clivage protéolytique, glycosylation, phosphorylation, prénylation, acylation, etc.);
- 4) les complexes protéines-protéines, protéines-ADN;
- 5) les effets multiples de la dislocation génétique ou de la surexpression d'un gène;
- 6) la présence et le rôle de petites protéines (moins de 100 acides aminés) puisque les CLOs de 300 pb et moins ne sont pas détectés lors de leur définition par les logiciels;
- 7) le phénotype des phénomènes multigéniques (prise de médicament, stress, sénescence, maladie, etc.).

Tableau 1. Identification des gènes d'*H. influenzae*, de *M. genitalium* et de *M. jannaschii*.

Organisme	Génome	Fonction probable	Fonction hypothétique	Fonction inconnue
<i>H. influenzae</i>	1,83 Mb			
	1743 CLOs	58%	20%	22%
<i>M. genitalium</i>	580 kb			
	470 CLOs	80%	-	20%
<i>M. jannaschii</i>	1,66 Mb			
	1682 CLOs	38%	6%	56%

Ainsi, de plus amples efforts doivent être déployés dans l'étude des protéines ou à plus grande échelle, des protéomes.

1.1.2 Le protéome

Afin de définir les analyses de l'expression génique basée sur l'étude des protéines, le terme "protéome" a été proposé (Wasinger *et al.*, 1995). Un protéome est simplement le complément protéique exprimé par le génome d'une cellule ou d'un tissu. Son concept est fondamentalement différent de celui du génome: alors que le génome est statique et bien défini pour un organisme, le protéome est dynamique puisqu'il peut être différent d'un tissu à un autre et qu'il change continuellement en réponse aux différentes conditions internes et externes de la cellule. De ce fait, un protéome reflète l'état d'un système biologique dans une condition donnée. Les étapes de différenciation cellulaire, les réponses au stress et aux changements de température, les maladies et les différentes pathologies représentent des états cellulaires particuliers auxquels correspondent des protéomes caractéristiques. Néanmoins, un protéome demeure toujours le produit direct du génome. Ainsi, les analyses du protéome, qui ont pour but d'identifier toutes les protéines présentes dans une cellule ou un tissu afin de les assigner à un gène, sont le complément nécessaire aux analyses du génome.

Les études du protéome de différents organismes ou tissus ont donc été entreprises. Les protéomes microbiens actuellement en cours incluent, par exemple ceux de *S. cerevisiae* (Hodges, Payne et Garrels, 1998), *Salmonella enterica* (O'Connor *et al.*, 1997), *Spiroplasma melliferum* (Cordwell, Basseal et Humphery-Smith, 1997), *Mycobacterium tuberculosis* (Urquhart *et al.*, 1997), *H. influenzae* (Link *et al.*, 1997) et *E. coli* (VanBogelen *et al.*, 1997). Les analyses du protéome d'organismes plus complexes comprennent entre autres celles de *C. elegans* (Bini *et al.*, 1997) et de différents tissus humains. Les récents développements dans le domaine de la biochimie des protéines au niveau des techniques et ressources disponibles ont rendu possible l'analyse du protéome.

1.2 PRINCIPALES ÉTAPES DE L'ÉTUDE DU PROTÉOME

1.2.1 Séparation des protéines: l'électrophorèse en deux dimensions

Afin d'analyser le protéome d'un organisme, il est nécessaire de pouvoir séparer et visualiser le plus grand nombre de protéines possible afin d'étudier l'expression de milliers de gènes. L'électrophorèse en deux dimensions (2-D) est une des méthodes les plus efficaces et les plus puissantes pour purifier des protéines en petite quantité. Cette technique permet de séparer les protéines selon leur charge ou point isoélectrique (pI) par une focalisation isoélectrique (IEF, première dimension, Figure 1) et selon leur masse moléculaire par une électrophorèse en présence de SDS (SDS-PAGE, seconde dimension, Figure 2). Même si la méthode a été décrite pour la première fois dans les années 70 (O'Farrell, 1975), certaines modifications ont dû être apportées avant qu'elle devienne la pierre angulaire des études du protéome.

La méthode originale de O'Farrell n'était pas très reproductible et ne permettait pas le chargement d'une grande quantité de protéines. En effet, la reproductibilité est essentielle pour comparer les gels provenant d'expériences différentes et ceux effectués dans d'autres laboratoires. Quant à la capacité de chargement, elle devait être augmentée de façon notable afin d'atteindre le seuil de sensibilité des techniques qui permettent l'identification des protéines. Le remplacement des ampholytes classiques pour former le gradient de pH de la première dimension par les gradients de pH immobilisés (IPG) (Bjellqvist *et al.*, 1982, Görg, Postel et Gunther, 1988) permet maintenant une meilleure résolution, une reproductibilité améliorée ainsi qu'une plus grande capacité de chargement. Un gradient de pH immobilisé est formé en liant de façon covalente les différents ampholytes à la matrice d'acrylamide du gel au moment où celui-ci est fabriqué. Il en résulte un gradient stable, comparativement à celui obtenu avec les ampholytes mobiles. Les gels IPG sont maintenant vendus commercialement avec un support de plastique sous forme de minces languettes, augmentant ainsi la reproductibilité entre les différents laboratoires.

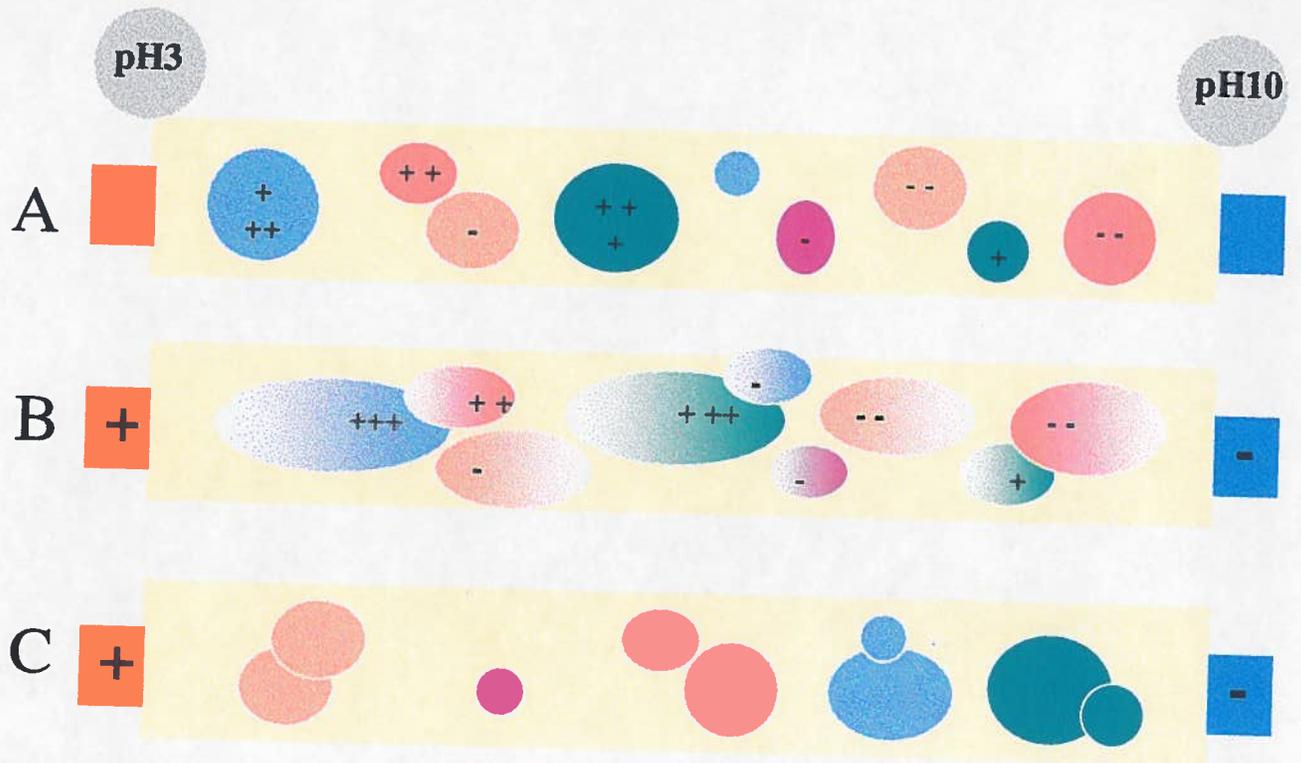


Figure 1. La focalisation isoélectrique. A) L'échantillon est incorporé dans le gel des languettes. B) Le courant électrique est appliqué et les protéines chargées migrent dans le gradient de pH immobilisé. C) Les protéines arrêtent leur migration lorsqu'elles atteignent le pH égal à leur pI (charge nette égale à zéro) et les protéines forment alors une zone distincte. Comme le gradient de pH est immobilisé dans le gel, la migration des protéines ne peut se poursuivre.

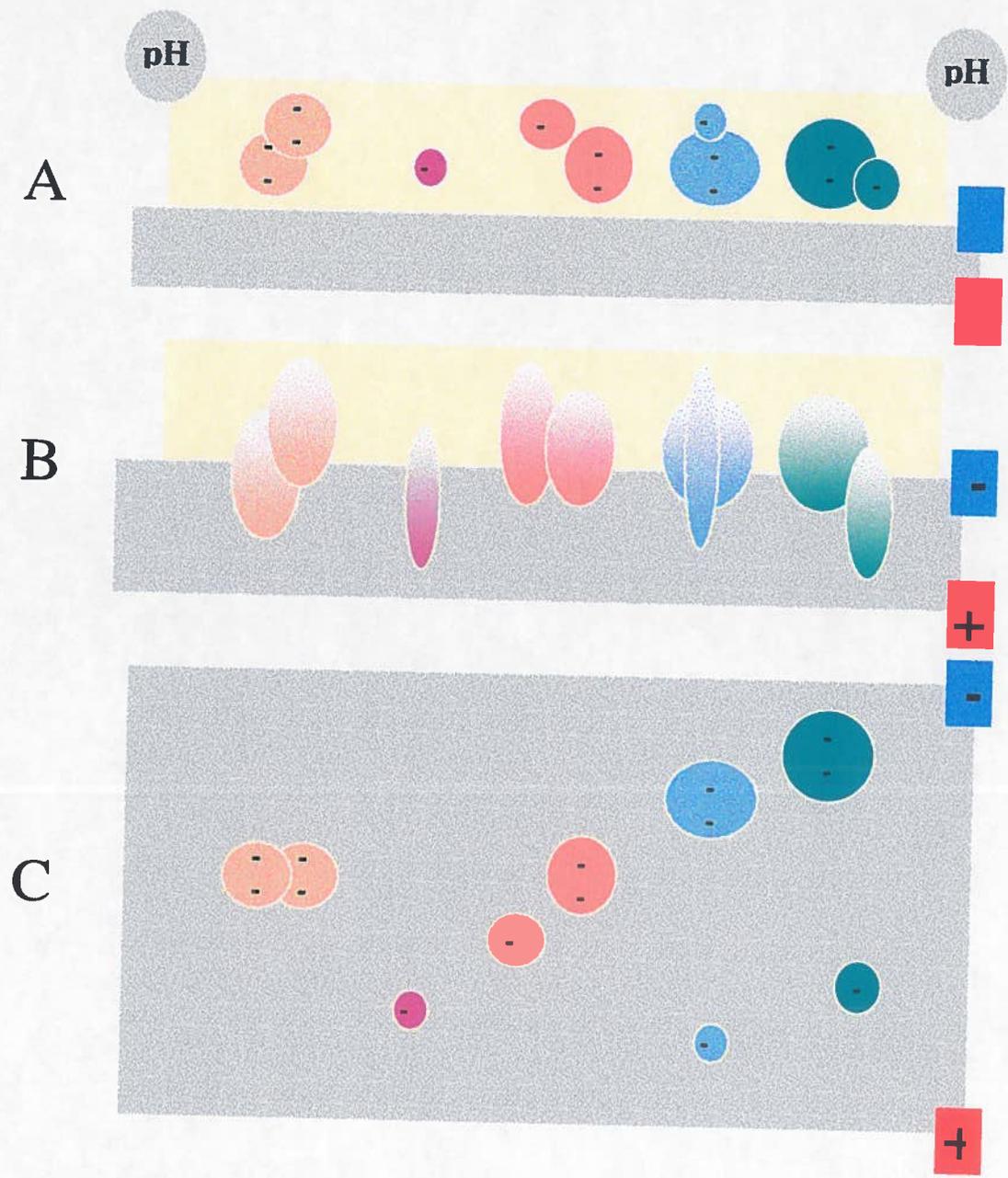


Figure 2. L'électrophorèse en gel de polyacrylamide en présence de SDS. A) Après avoir préalablement équilibré le gel IPG dans un tampon contenant du SDS, celui-ci est déposé sur le gel de polyacrylamide, à la cathode. B) Le courant électrique est appliqué et les protéines chargées négativement migrent vers l'anode. C) La vitesse à laquelle les protéines se déplacent dépend de leur grosseur: les plus petites se déplacent rapidement et les grosses, plus lentement.

1.2.2 Visualisation des protéines et analyse de l'image

Toutes les méthodes de révélation des protéines utilisées pour les gels de polyacrylamide sont applicables aux gels 2-D. L'autoradiographie est la technique la plus sensible mais nécessite l'utilisation de radioisotopes, tels que la ^{35}S -méthionine. Les méthodes non-radioactives les plus sensibles sont la coloration au nitrate d'argent et celles utilisant des colorants fluorescents (par exemple le SYPROTMRuby). De plus, certains protocoles de coloration au nitrate d'argent ne modifient pas de façon covalente les polypeptides, permettant ainsi leur analyse subséquente par spectrométrie de masse. La coloration au bleu de Coomassie peut également être utilisée mais sa sensibilité est plus de 50 fois inférieure à celle de la coloration au nitrate d'argent. Suite à la visualisation des protéines, les images de gels sont numérisées soit par un appareil à balayage numérique (scanner), un densitomètre au laser ou par une caméra CCD (charge-coupled device).

Avec l'amélioration de la technique d'électrophorèse en 2-D et des méthodes permettant l'identification des protéines (section suivante), des logiciels plus puissants pour l'analyse des gels 2-D ont été développés. Ces logiciels permettent d'analyser et de comparer des patrons complexes de protéines. Les protéines sont automatiquement détectées, leur densité est évaluée, le bruit de fond est éliminé de même que les stries verticales et horizontales, la masse moléculaire et le pI sont déterminés et les protéines peuvent être corrélées entre différents gels. Ces logiciels sont aussi capables de détecter les changements qualitatifs (protéines nouvelles ou absentes) et quantitatifs (niveau d'expression différent) entre plusieurs gels. De plus, des images de référence peuvent être créées à partir de plusieurs gels du même échantillon.

1.2.3 Identification des protéines

Traditionnellement l'identification de protéines séparées par électrophorèse en 2-D était difficile: la quantité de protéines présente était souvent limitante et les bases de données sur les protéines et les gènes n'étaient pas suffisamment développées. Ainsi,

l'identification de façon précise des centaines de protéines retrouvées dans un gel 2-D requerrait une technique rapide, sensible et peu dispendieuse, de même qu'un accès simplifié à des bases de données plus exhaustives.

1.2.3.1 Spectrométrie de masse

Toutes les méthodes classiques d'identification des protéines sont aujourd'hui supplantées par la spectrométrie de masse, puisque aucune autre technique n'est à la fois aussi rapide, précise et sensible.

La spectrométrie de masse est une technologie qui permet de mesurer de façon très précise la masse de n'importe quelle molécule ou de différentes molécules présentes dans un mélange. Avec l'avènement des techniques d'ionisation 'douces' telles que MALDI (matrix assisted laser desorption ionisation) et ESI (electrospray ionisation), il est maintenant possible d'ioniser les biomolécules de hautes masses (peptides et protéines). Les spectromètres de masse utilisés pour l'analyse des peptides et des protéines possèdent deux parties principales: une source qui permet d'introduire et d'ioniser les composantes de l'échantillon et un analyseur qui mesure la masse des ions introduits. Trois types de spectromètre de masse (TOF (time of flight), TQ (triple quadrupole) ou IT (ion trap) sont principalement utilisés, en raison de leur coût et de leur facilité d'utilisation, pour mesurer le rapport de la masse sur la charge (m/z) des ions produits. L'ESI a été adaptée aux instruments de type TQ et IT, tandis que le MALDI est généralement utilisée avec un appareil de type TOF.

À ce jour, trois applications de la spectrométrie de masse, soient le profil de la masse des peptides, le séquençage de peptides et la détection des modifications post-traductionnelles, peuvent être utilisées pour l'analyse des protéines. Le profil de la masse des peptides implique la génération de peptides à l'aide d'une endoprotéase spécifique à un résidu, la détermination de la masse des peptides par le spectromètre de masse et la comparaison des masses obtenues avec les masses théoriques prédites par les séquences des bases de données (Figure 3).

Il est aussi possible de fragmenter un peptide, de déterminer la masse des fragments résultants et d'obtenir de l'information sur sa séquence grâce au spectre de fragmentation généré. En effet, certains types de spectromètre de masse (TQ et IT), permettent de fragmenter les peptides en isolant l'ion 'parent' et en le fragmentant par collision avec un gaz inerte tel que l'argon. L'ion subit de multiples collisions avec le gaz et se fragmente de manière aléatoire, souvent au niveau d'un lien peptidique. Les ions 'filles' résultants sont détectés et la séquence en acides aminés est dérivée en soustrayant la masse de deux ions adjacents (Figure 4). Cependant, les données générées sont complexes et un certain niveau d'expertise est requis pour leur interprétation. Généralement, cette séquence, combinée avec le profil de la masse des peptides, permet d'identifier sans équivoque une protéine.

Finalement, une modification post-traductionnelle est détectée lorsque la masse observée d'un peptide identifié à l'aide d'une séquence obtenue par spectrométrie de masse ne correspond pas à la masse prédite. Le site et le type de modifications peuvent être déterminés par les différences entre les masses observées et les masses prédites des fragments du peptide.

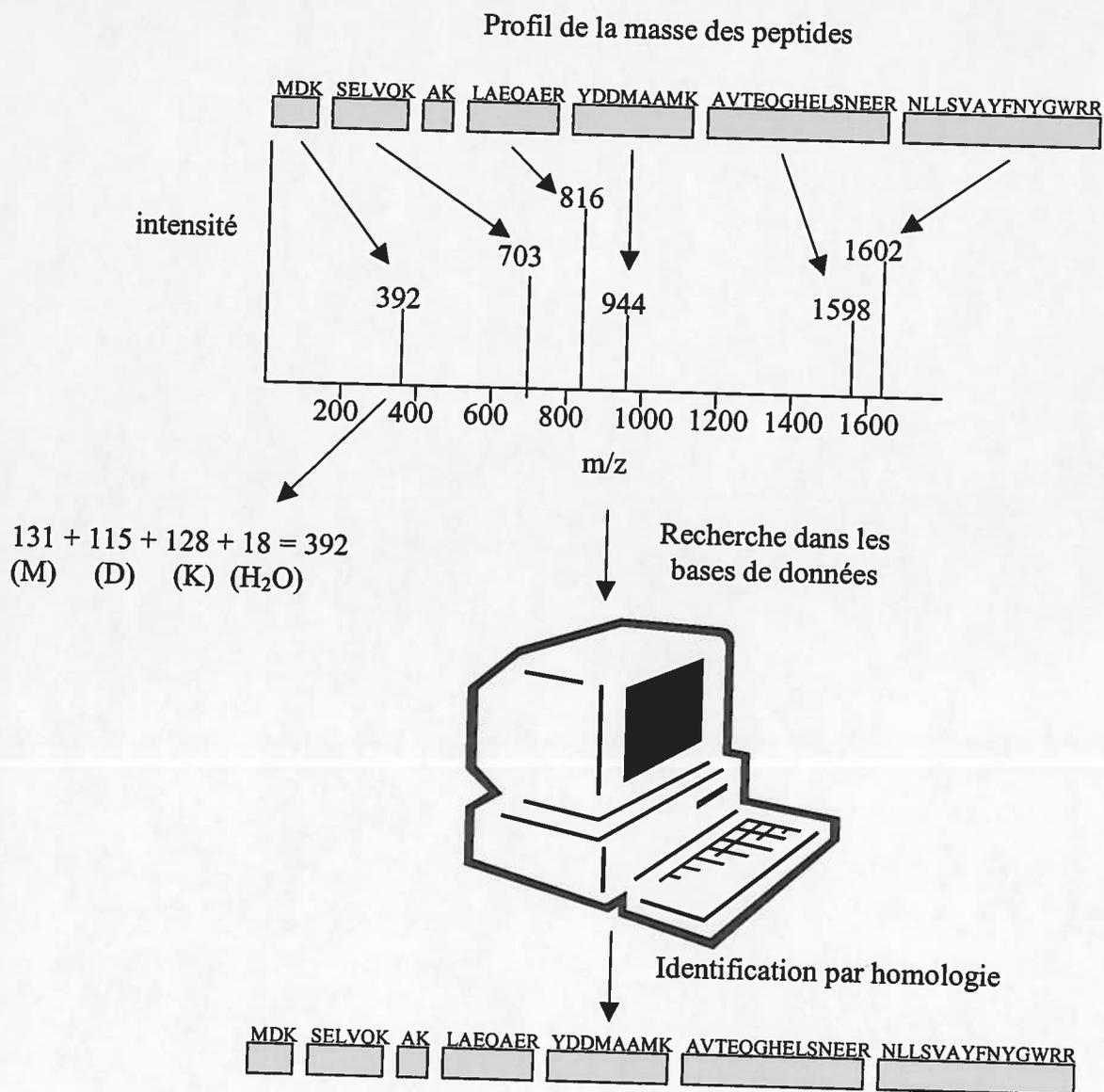


Figure 3. Profil de la masse des peptides. Un spectre de masse de l'intensité en fonction du rapport m/z, où chaque pic correspond à un peptide, permet d'obtenir la masse des peptides ionisés provenant de la digestion d'une protéine. Le profil est utilisé pour une recherche dans les bases de données afin de trouver le profil théorique qui possède le plus d'homologie avec celui obtenu.

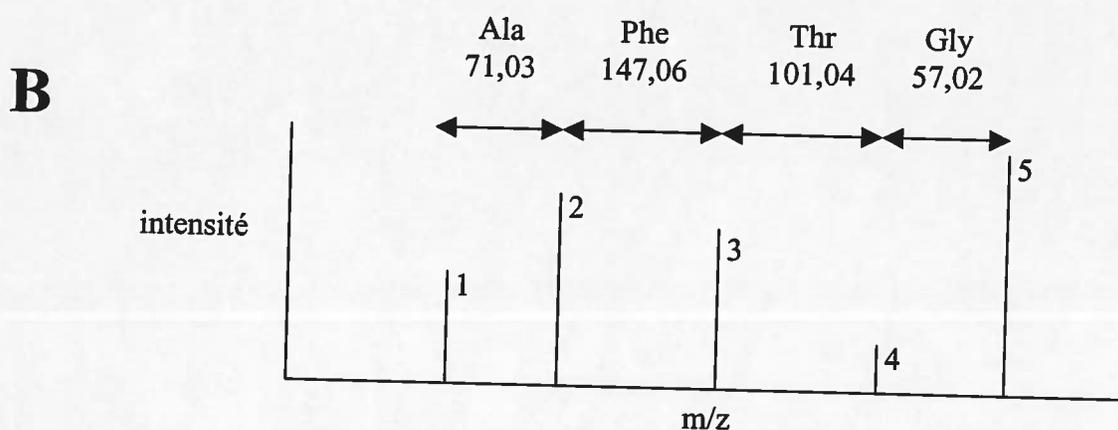
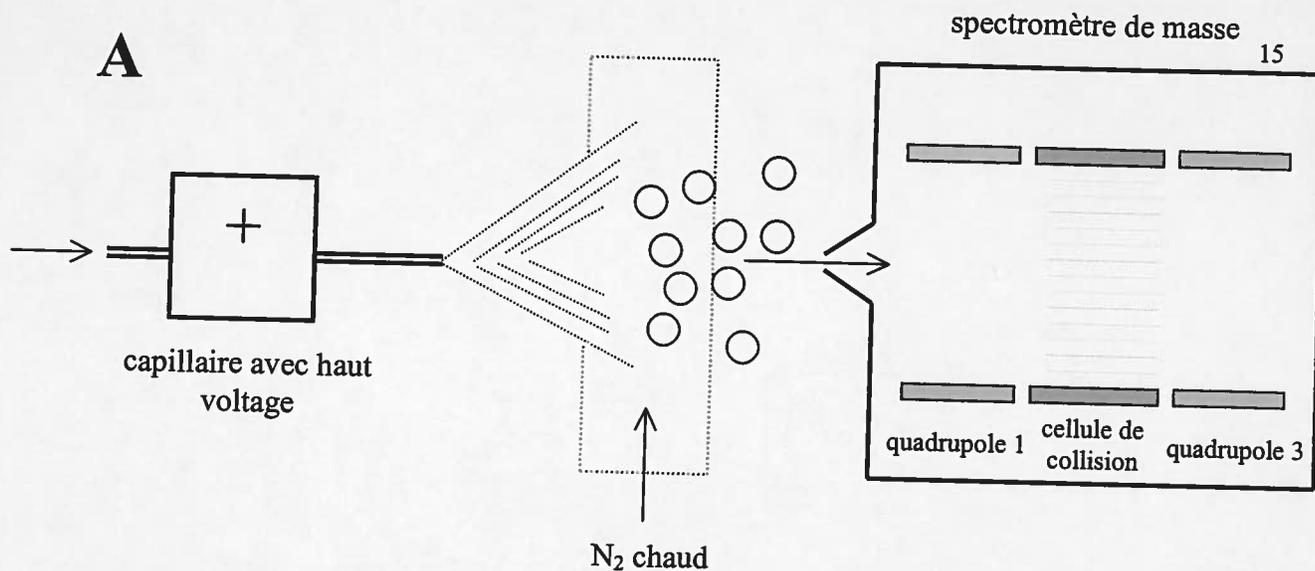


Figure 4. Fragmentation de peptides avec un spectromètre de masse triple quadrupole équipé d'une source d'ions 'electrospray'. A) L'échantillon liquide est déposé à l'extrémité d'un capillaire possédant un potentiel élevé où a lieu la formation de fines gouttelettes de solvant ayant une forte charge électrique et renfermant des peptides. Le solvant s'évapore grâce à l'azote gazeux chaud, libérant les peptides ionisés dans la phase gazeuse. Dans le premier quadrupole, la masse des peptides est mesurée et les peptides sélectionnés peuvent être transmis du premier quadrupole vers la cellule de collision pour être fragmentés avec un gaz inerte. La masse des fragments est mesurée dans le troisième quadrupole. B) Une séquence peptidique peut être dérivée lorsque la différence de masse entre deux fragments correspond à la masse d'un acide aminé. Dans cet exemple, le fragment 5 correspondrait à XXAFTG, le fragment 4 à XXAFT, le fragment 3 à XXAF, le fragment 2 à XXA et le fragment 1 à XX.

1.2.3.2 Autres techniques d'identification des protéines

Le séquençage de l'extrémité N-terminale d'une protéine par dégradation d'Edman et la composition en acides aminés sont aussi deux méthodes puissantes pour permettre l'identification d'une protéine, mais elles sont trop dispendieuses pour être utilisées à grande échelle (Wilkins *et al.*, 1996).

1.2.3.3 Recherche dans les bases de données

Il existe deux types de bases de données contenant des séquences de protéines: celles regroupant simplement les séquences en acides aminés traduites à partir de la séquence d'ADN, telle que TrEMBL (TRAnslation of EMBL: www.expasy.ch/sprot) et celles qui contiennent d'autres informations (annotations) en plus de la séquence en acides aminés. En effet, la description de la fonction de la protéine, ses domaines structuraux, sa structure secondaire, les modifications post-traductionnelles, les maladies reliées à la protéine ainsi que les références sont aussi disponibles. SWISS-PROT: www.expasy.ch/sprot est un exemple de ce type de base de données. TrEMBL est un supplément de SWISS-PROT et elle a été créée afin que les chercheurs aient accès à toutes les séquences connues le plus rapidement possible mais sans diluer la qualité de l'information fournie par SWISS-PROT en ajoutant des séquences qui n'avaient pas été analysées et annotées. De plus, la combinaison de SWISS-PROT et TrEMBL procure une base de données complète et pratiquement sans redondance comparativement aux bases de données telles que nrdb (Non-Redundant DataBase offerte par NCBI) et OWL (Wilkins *et al.*, 1997: 100-101).

Plusieurs logiciels permettant l'identification des protéines par une recherche dans ces bases de données sont accessibles à l'adresse suivante: www.expasy.ch/tools. Quelques exemples de ces logiciels ainsi que les données requises pour leur utilisation sont décrits au tableau suivant.

Tableau 2. Exemples de logiciels pour l'identification des protéines.

Logiciels	Caractéristiques des protéines
AACompIdent	Composition en acides aminés
MultiIdent	pI, Mr, composition en acides aminés, séquence en acides aminés, profil de la masse des peptides
PeptIdent	pI, Mr, profil de la masse des peptides
TagIdent	pI, Mr, séquence en acides aminés

1.2.4 Bases de données de gels 2-D

Récemment, des bases de données des protéines identifiées sur gels 2-D ont été développées et rendues disponibles à la communauté scientifique via l'Internet. Leur nombre augmente sans cesse et une liste à jour est retrouvée dans WORLD-2DPAGE: www.expasy.ch/ch2d/2d-index.html. Ce type de bases de données est formé de deux composantes: les images (Figure 5) et les informations textuelles (Tableau 3). L'image consiste en un ou plusieurs gels digitalisés provenant d'un échantillon biologique donné, soit un tissu, un liquide physiologique ou une cellule. Des images de référence, qui sont la combinaison de plusieurs gels où chacun couvre différentes régions de pI et de Mr, sont de plus en plus fréquentes. L'information sur chacune des protéines qui a été identifiée inclut généralement le point isoélectrique et la masse moléculaire apparents, le nom de la protéine, la méthode d'identification, les références bibliographiques et les références à d'autres bases de données, telle que SWISS-PROT. Dans un futur rapproché, la composition en acides aminés et les données de spectrométrie de masse seront également incluses. Il est possible d'accéder à ces informations par une liste des protéines identifiées ou en cliquant sur la protéine à partir de l'image.

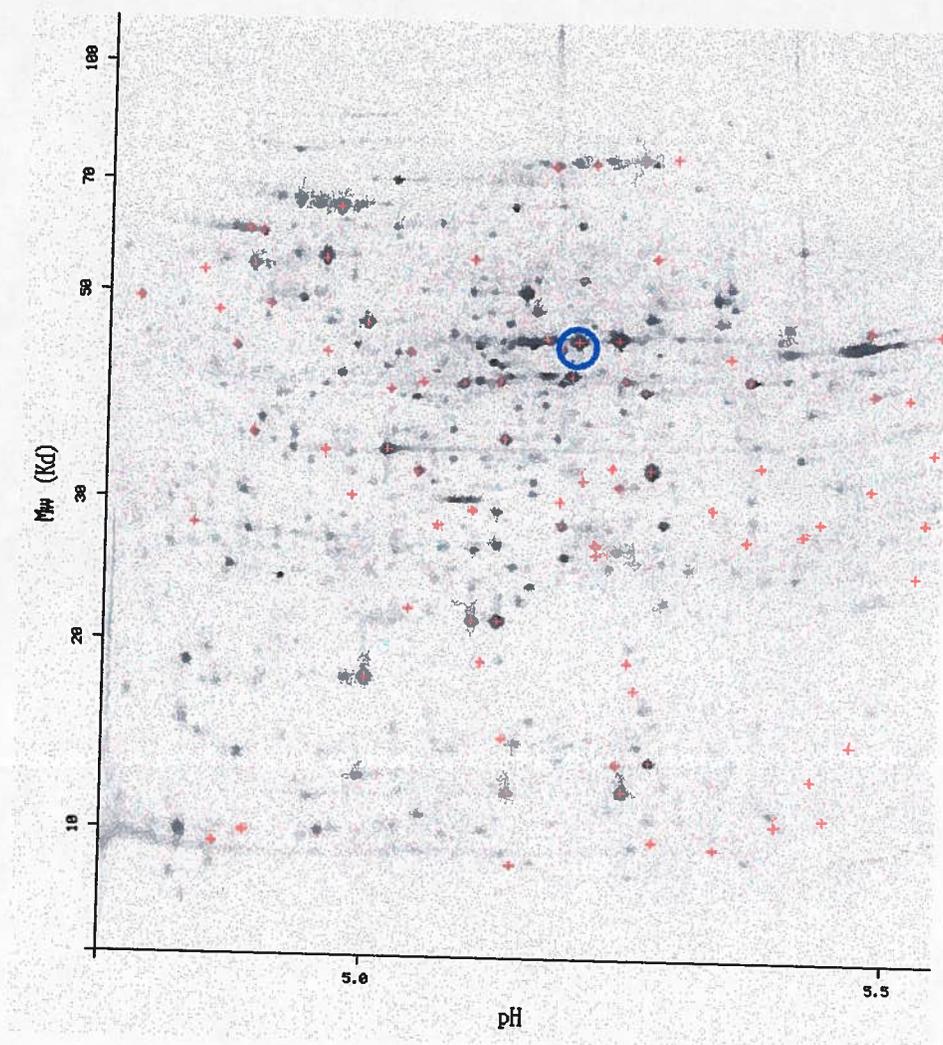


Figure 5. Exemple d'une image de gel numérisé retrouvée dans la base de données SWISS-2DPAGE. En sélectionnant "by clicking on a spot" comme mode d'accès à la base de données, l'image du gel de pH de 4,5-5,5 des protéines d'*Escherichia coli* a été choisie. Les protéines connues sont identifiées par une croix rouge. La protéine entourée d'un cercle bleu a été sélectionnée afin d'obtenir l'information relative à cette dernière (Tableau 3).

Tableau 3. Informations textuelles disponibles dans la base de données SWISS-2DPAGE pour la protéine encadrée en bleu (Figure 5).

<p>Entry name: IDH_ECOLI Primary accession number: P08200 (Entered in SWISS-2DPAGE in Release 02, August 1995 Last modified in Release 12, April 2000) Description: ISOCITRATE DEHYDROGENASE [NADP] (EC 1.1.1.42) (OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDH) (IDP). From <i>Escherichia coli</i>. Taxonomy Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; <i>Escherichia</i>.</p>
<p>[1]MAPPING ON GEL MEDLINE; 96314059. [NCBI, ExPASy, Israel, Japan] Pasquali C., Frutiger S., Wilkins M.R., Hughes G.J., Appel R.D., Bairoch A., Schaller D., Sanchez J.-C., Hochstrasser D.F. "Two-dimensional gel electrophoresis of <i>Escherichia coli</i> homogenates: the <i>Escherichia coli</i> SWISS-2DPAGE database." Electrophoresis 17:547-555(1996).</p>
<p>[2]MAPPING ON GEL Vanbogelen R.A., Abshire K.Z., Pertsemliadis A., Clark R.L., Neidhardt F.C.; "Gene-protein database of <i>Escherichia coli</i> K-12, edition 6". (In) Neidhardt et al. (eds.). <i>Escherichia coli</i> and <i>Salmonella</i>: Cellular and Molecular Biology (2nd ed.), pp.2067-2117, ASM Press, Washington DC (1996).</p>
<p>[3]MAPPING ON GEL. MEDLINE; 81069671. [NCBI, ExPASy, Israel, Japan] Phillips T.A., Bloch P.L., Neidhardt F.C. "Protein identifications on O'Farrell two-dimensional gels: locations of 55 additional <i>Escherichia coli</i> proteins." J. Bacteriol. 144:1024-1033(1980).</p>
<p>[4]MAPPING ON GEL. Tonella L., Sanchez J.-C., Appel R.D., Bairoch A., Hoogland C., Binz P.-A., Hochstrasser D.F. Submitted (MAR-2000) to the SWISS-2DPAGE database.</p>
<p>SUBUNIT: MONOMER.</p>

1.3 LES STREPTOMYCÈTES

1.3.1 Généralités et importance industrielle

La famille des actinomycètes, auquel le genre *Streptomyces* appartient, sont des bactéries qui ont pour la plupart la particularité de croître en forme de mycélium. Elles sont Gram positives, aérobies et leur habitat naturel est généralement restreint au sol. Les streptomycètes font donc partie de la biomasse capable de dégrader la matière organique retrouvée dans le sol en produisant une variété d'enzymes hydrolytiques extracellulaires (McCarthy et Williams, 1992). Plusieurs de ces enzymes sont employés au niveau industriel, particulièrement dans l'industrie alimentaire. Étant donné leur capacité à sécréter de grandes quantités d'enzymes, les streptomycètes sont amplement utilisés pour l'expression et la sécrétion de protéines homologues (provenant d'un streptomycète) et hétérologues (provenant de d'autres genres) (Gusek et Kinsella, 1992).

Les streptomycètes font preuve d'une grande diversité biochimique puisqu'ils produisent une panoplie de métabolites secondaires, incluant plus de la moitié de tous les antibiotiques connus (Miyadoh, 1993). Plusieurs de ces composés sont utilisés en médecine humaine et vétérinaire (agents antibactériens, antitumoraux et antifongiques) ainsi qu'en agriculture (herbicides, fongicides et agents antiparasitiques).

Les streptomycètes sont uniques parmi les bactéries en ce qui a trait à leur cycle de vie. Celui-ci comprend trois principales formes: le mycélium végétatif, le mycélium aérien et les spores. Étant donné que la production d'enzymes et de métabolites d'intérêt commercial coïncide avec cette différenciation morphologique, la physiologie et le développement des streptomycètes sont devenus d'importants champs d'étude. En effet, une meilleure compréhension du métabolisme primaire et secondaire et de la régulation génétique qui influence la différenciation morphologique permettrait éventuellement l'amélioration des souches productrices d'enzymes et de métabolites secondaires.

1.3.2 *Streptomyces coelicolor*: un modèle chez les streptomycètes

L'actinomycète le plus caractérisé au niveau génétique et dont le développement morphologique a été largement étudié est *Streptomyces coelicolor* A3(2) (Bibb, 1996). Plusieurs de ses caractéristiques ont contribué à faire de cet organisme le modèle par excellence chez les streptomycètes. Tout d'abord, contrairement à la plupart des procaryotes cet organisme ne possède pas un chromosome circulaire, mais linéaire (Lin *et al.*, 1993). Ensuite, son génome est très riche en guanine et cytosine (G + C 72%) (Volff et Altenbuchner, 1998) et on remarque chez cet organisme la présence d'éléments génétiques variés (1.3.2.1). Finalement, la disponibilité d'outils de biologie moléculaire (1.3.2.2) ainsi qu'une banque ordonnée de cosmides (1.3.2.3) ont récemment facilité l'étude de cet organisme.

1.3.2.1 Éléments génétiques variés

S. coelicolor possède au moins deux séquences d'insertion (IS110, IS117, Bibb, 1996), un large plasmide linéaire (SCP1, 350 kb) qui porte les gènes pour la production d'un antibiotique (méthylénomycine, Kirby et Hopwood, 1977). Il possède également un élément d'intégration SLP1 (Bibb *et al.*, 1981) qui est stable de façon générale mais qui est capable d'excision, de conjugaison et de réplication autonome chez d'autres streptomycètes. *S. coelicolor* possède également un plasmide circulaire conventionnel SCP2 (Bibb et Hopwood, 1981) qui est auto transmissible et capable de mobiliser le chromosome.

1.3.2.2 Outils permettant la manipulation génétique

La manipulation génétique de *S. coelicolor* a été grandement facilitée par le développement de vecteurs de clonage versatiles, de systèmes efficaces permettant l'introduction de gènes chez les streptomycètes, comme par exemple la transformation à l'aide de protoplastes (Bibb, Ward et Hopwood, 1981). Il est aussi possible d'insérer des gènes clonés au niveau du chromosome selon différentes procédures, ce qui a entre autre

l'avantage de produire des recombinants qui ne requièrent pas de sélection avec un antibiotique afin de maintenir le gène (Baltz, 1997). Finalement, la recombinaison homologe (recombinaison entre des séquences d'ADN qui sont partiellement homologues) est très utile pour obtenir des souches hybrides qui possèdent des caractéristiques nouvelles. La disponibilité de ces outils et méthodes permet aujourd'hui d'exploiter la ressource génétique des streptomycètes.

1.3.2.3 Banque ordonnée de cosmides

En 1992, le chromosome de *S. coelicolor* a été cartographié physiquement par Kieser et ses collaborateurs. Ainsi, l'électrophorèse en champ pulsé (PFGE) a permis de construire une carte de restriction avec deux enzymes dont les sites de reconnaissance sont peu fréquents chez les streptomycètes, soit *AseI* (5'ATTAAT3') et *DraI* (5'TTTAAA3') et de relier la carte physique et génétique. L'étape suivante fut la préparation d'une librairie en utilisant le cosmide SuperCos (Figure 6) de l'ADN chromosomique de *S. coelicolor* M145 (Redenbach *et al.*, 1996). Cette souche, provenant de la souche sauvage A3(2), est prototrophe et ne contient aucun plasmide. Ainsi, 320 cosmides ont été sélectionnés afin de former une banque de clones se chevauchant et couvrant la totalité du chromosome, à l'exception de trois régions discontinues. Par hybridation, plus de 170 gènes, groupes de gènes et autres éléments génétiques ont été localisés sur des cosmides spécifiques et une carte physique et génétique détaillée en a résulté (Figure 5). Cette banque est rapidement devenue une ressource inestimable pour la communauté scientifique œuvrant avec des streptomycètes en aidant à la cartographie et à l'isolement de gènes. Mais de façon encore plus importante, elle est devenue le point de départ du projet de séquençage du génome.

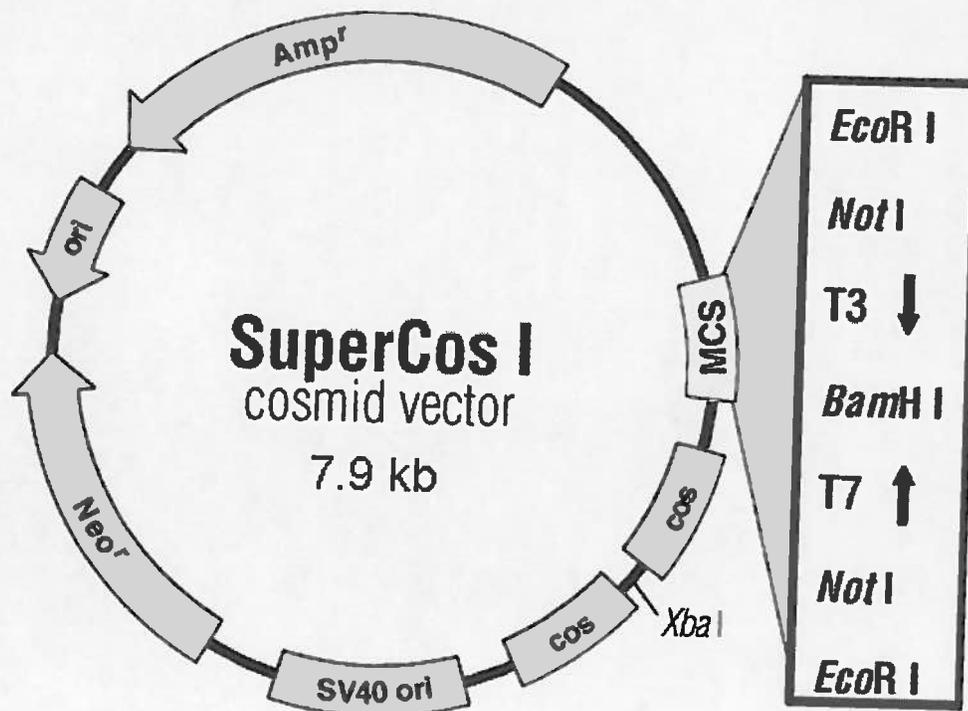


Figure 6. Carte physique du cosmide SuperCos 1 utilisé lors de la préparation de la banque cosmique de l'ADN chromosomique de *S. coelicolor* M145. Certains sites de restriction de la cassette de clonage (MCS) sont indiqués. *Neo^r*, gène de la résistance à la néomycine; *Amp^r*, gène de la résistance à l'ampicilline; ori, origine de répliation du cosmide chez les procaryotes; SV40 ori, origine de répliation chez les eucaryotes; T3 et T7, promoteurs de bactériophages.

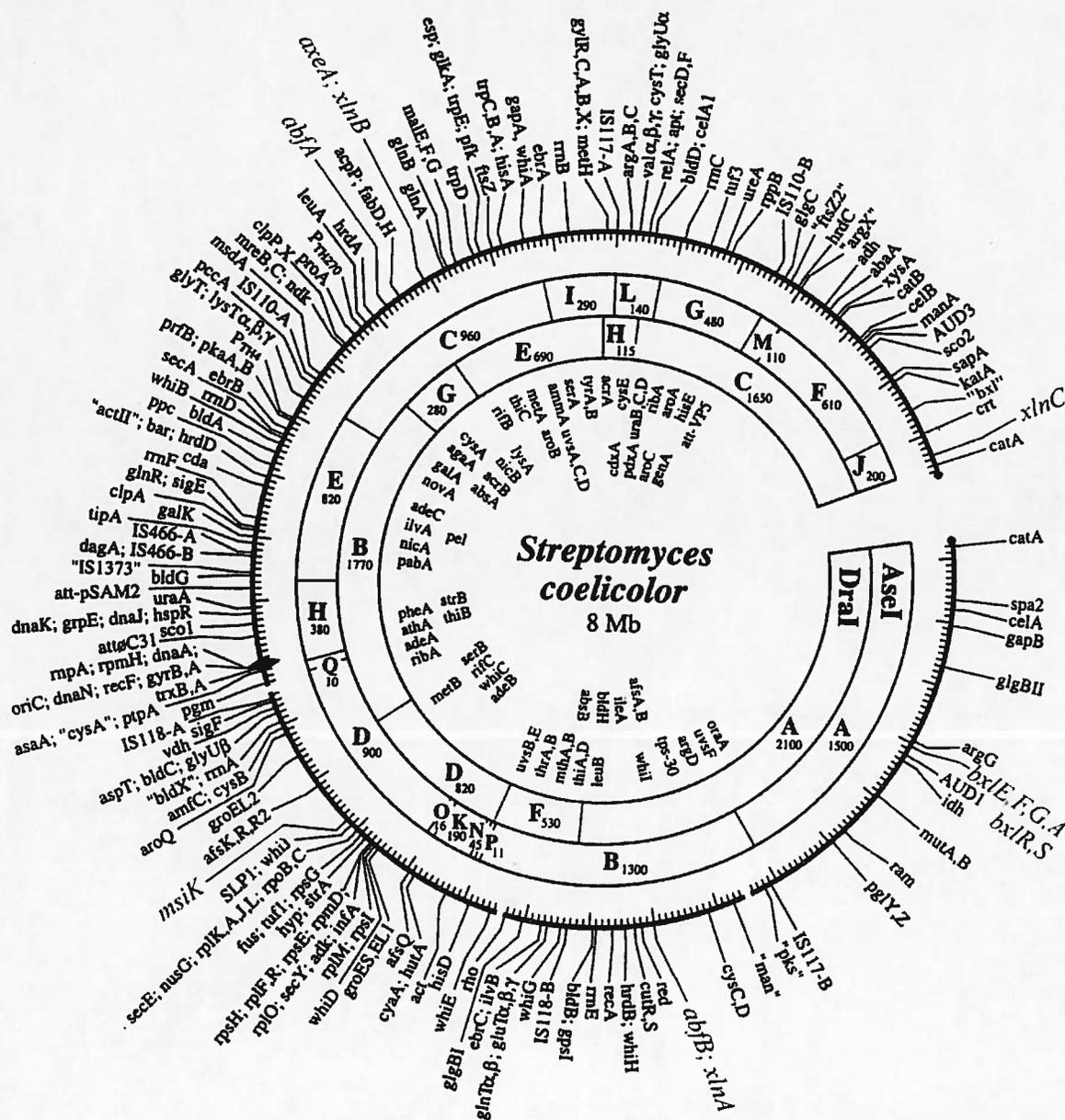


Figure 7. Carte physique et génétique du chromosome de *S. coelicolor* M145. La position des marqueurs à l'extérieur du cercle a été déterminée physiquement par hybridation à la banque de cosmides. Les marqueurs à l'intérieur du cercle ont été localisés de façon génétique seulement et leur position approximative a été déterminée par interpolation par rapport aux marqueurs cartographiés physiquement. La grosseur des fragments *DraI* et *AseI* sont en kb, ●, télomères du chromosome, ◆, origine de réplication bidirectionnelle du chromosome. Tiré de Redenbach *et al.*, 1996.

1.3.3 Séquençage et analyse du génome de *S. coelicolor* M145

Le projet de séquençage du chromosome de 8 Mb de *S. coelicolor* a débuté le premier août 1997 au Wellcome Genome Campus du Sanger Centre à Hinxton, Cambridge. Chaque cosmide est d'abord vérifié au John Innes Centre avant d'être transmis au Sanger Centre. Le séquençage des cosmides commence par une phase aléatoire, où les extrémités d'un nombre suffisant de fragments sélectionnés au hasard sont séquencées et assemblées afin d'obtenir le génome complet. Les régions discontinues dans la séquence sont complétées par marche chromosomique à partir d'amorces. Finalement, les inserts sont annotés (définition et identification des cadres de lecture ouverts) avec le logiciel Artemis et les séquences annotées sont immédiatement déposées dans la base de données EMBL. La séquence complète devrait être terminée d'ici la fin de l'an 2001 et l'état actuel du projet est disponible à l'adresse suivante: www.sanger.ac.uk/Projects/S_coelicolor/. À ce jour, plus de 87% des cosmides ont été séquencés, soit 7,9 Mb, mais seulement 82% de ces cosmides ont été annotés et déposés dans la base de données EMBL.

1.3.4 Analyse du protéome

Comme le projet de séquençage tire à sa fin, il faut maintenant s'intéresser à l'expression de ces gènes nouvellement identifiés, de même qu'à leur fonction. Deux groupes de recherche sur les streptomycètes (entre autres celui du John Innes Centre et celui de l'Institut Armand-Frappier) ont donc entrepris l'analyse du protéome de *S. coelicolor* M145. Le groupe dirigé par le Prof. M. Bibb comparera les protéomes exprimés après croissance en milieu minimal et en milieu riche (communication personnelle), tandis que nous nous attarderons à l'étude des différents protéomes obtenus après croissance avec diverses sources de carbone.

1.3.5 La dégradation du xylane par *S. lividans*

1.3.5.1 Le xylane

Le constituant majeur de l'hémicellulose (polysaccharide non-cellulosique) est le xylane, un polymère complexe constitué d'unités D-xylose reliées par des liens β -1,4. L'hémicellulose, avec la cellulose (polymère d'unités D-glucosyl reliées par des liens β -1,4) et la lignine (un composé phénolique complexe) composent la paroi cellulaire des plantes et du bois dur (Wong, Tan et Saddler, 1988). L'hémicellulose se retrouve à leur interface et relie la cellulose et la lignine. Afin de pouvoir utiliser les fibres de cellulose, l'industrie papetière doit d'abord éliminer la coloration causée par la présence de lignine. Elle procède par conséquent au blanchiment de la pâte de papier à l'aide du chlore, un polluant ayant des effets très néfastes sur l'environnement.

Grâce à leur spécificité de substrat, des enzymes comme les hémicellulases peuvent enlever les composants hémicellulosiques de la pâte de papier et en libérer ainsi les fibres de cellulose. Ceci permet d'éliminer ou de réduire considérablement l'utilisation du chlore et constitue un procédé de blanchiment beaucoup plus écologique. Étant donné leur potentiel au niveau du bioblanchiment dans les industries des pâtes et papier, les hémicellulases, particulièrement les xylanases, ont alors attiré l'attention. Le groupe de recherche en biotechnologie des streptomycètes de l'Institut Armand-Frappier s'est donc intéressé à la production des xylanases par *S. lividans*.

1.3.5.2 Le système xylanolytique

Jusqu'à ce jour, sept protéines du complexe xylanolytique ont été purifiées et caractérisées, soit trois endo- β -1,4-xylanases (XlnA-B-C) (Morosoli *et al.*, 1986; Kluepfel *et al.*, 1990; Kluepfel *et al.*, 1992), deux α -L-arabinofuranosidases (AbfA-B) (Manin *et al.*, 1994; Vincent *et al.*, 1997), une acétyl xylane estérase (AxeA) (Dupont *et al.*, 1996), une protéine liant l'ATP et impliquée dans l'importation des sucres (MsiK) (Penzès, 1995) et une β -xylosidase (BxlA) (Lam, 1999). Deux autres enzymes ont été

partiellement purifiées et caractérisées dans nos laboratoires: une protéine de liaison aux sucres (BxlE) et une protéine de régulation (BxlR). Finalement, les trois gènes suivants ont été clonés et séquencés: deux perméases (BxlF-G) (Larocque, 1998) et une autre protéine de régulation (BxlS) (Larocque, 1998). Une α -glucuronidase serait également nécessaire pour dégrader complètement le xylane, mais cette enzyme n'a pas encore été identifiée chez *S. lividans*.

Tel qu'illustré à la Figure 8, les liens β -1,4 du xylane sont d'abord clivés par les trois xylanases, générant des oligoxylosides plus courts et de différentes longueurs. Les chaînes latérales du xylane (arabinosyl, acide glucuronique et groupements acétyles) sont débranchées par l'arabinofuranosidase B et l'acétyl xylane estérase A. Ces oligoxylosides sont alors plus facilement assimilables et peuvent maintenant pénétrer à l'intérieur de la cellule pour servir de source de carbone. Ils sont d'abord reconnus et fixés par la protéine BxlE afin de traverser le pore formé par les deux perméases BxlF et BxlG. L'ATP hydrolysé par la protéine MsiK fournit l'énergie au système.

Dans le cytoplasme, les oligoxylosides possédant encore des chaînes latérales sont débranchées par l'arabinofuranosidase A. Finalement, la β -xylosidase permet la dégradation complète du xylane en hydrolysant le xylotriose et le xylobiose en xylose. Quant à la protéine BxlR, elle semble avoir un rôle dans la régulation du système xylanolytique puisqu'elle se fixe à un palindrome à proximité des gènes du complexe (Larocque, 1998). Cependant, la position du palindrome par rapport aux régions promotrices (en amont ou en aval), de même que la distance entre les deux, laisse croire que BxlR agirait sur certaines enzymes du complexe comme un répresseur et sur d'autres comme un activateur. Cette observation a été confirmée par d'autres résultats obtenus dans nos laboratoires.

Quant à la fonction précise de la protéine codée par le cadre de lecture *bxlS*, elle reste à déterminer puisqu'elle n'a pas encore été purifiée et donc aucune expérience de liaison à l'ADN n'a pu être effectuée.

Afin de mieux comprendre le rôle de ces deux protéines au sein de la cellule, une souche surexprimant *bxlR* et *bxlS*, *S. lividans* IAF6, a été construite. Cette dernière possède un plasmide à copies multiples contenant les gènes *bxlR* et *bxlS*. À la lumière des résultats obtenus dans nos laboratoires, il s'est avéré nécessaire d'éclaircir la fonction de BxlR au niveau de la régulation du complexe. Un moyen d'approfondir le rôle de cette protéine serait l'identification des protéines induites et/ou réprimées par la surexpression de BxlR en comparant les protéomes de la souche sauvage et de la souche IAF6. Ainsi, l'optimisation des différentes étapes d'une analyse protéomique a été réalisée dans cette optique.

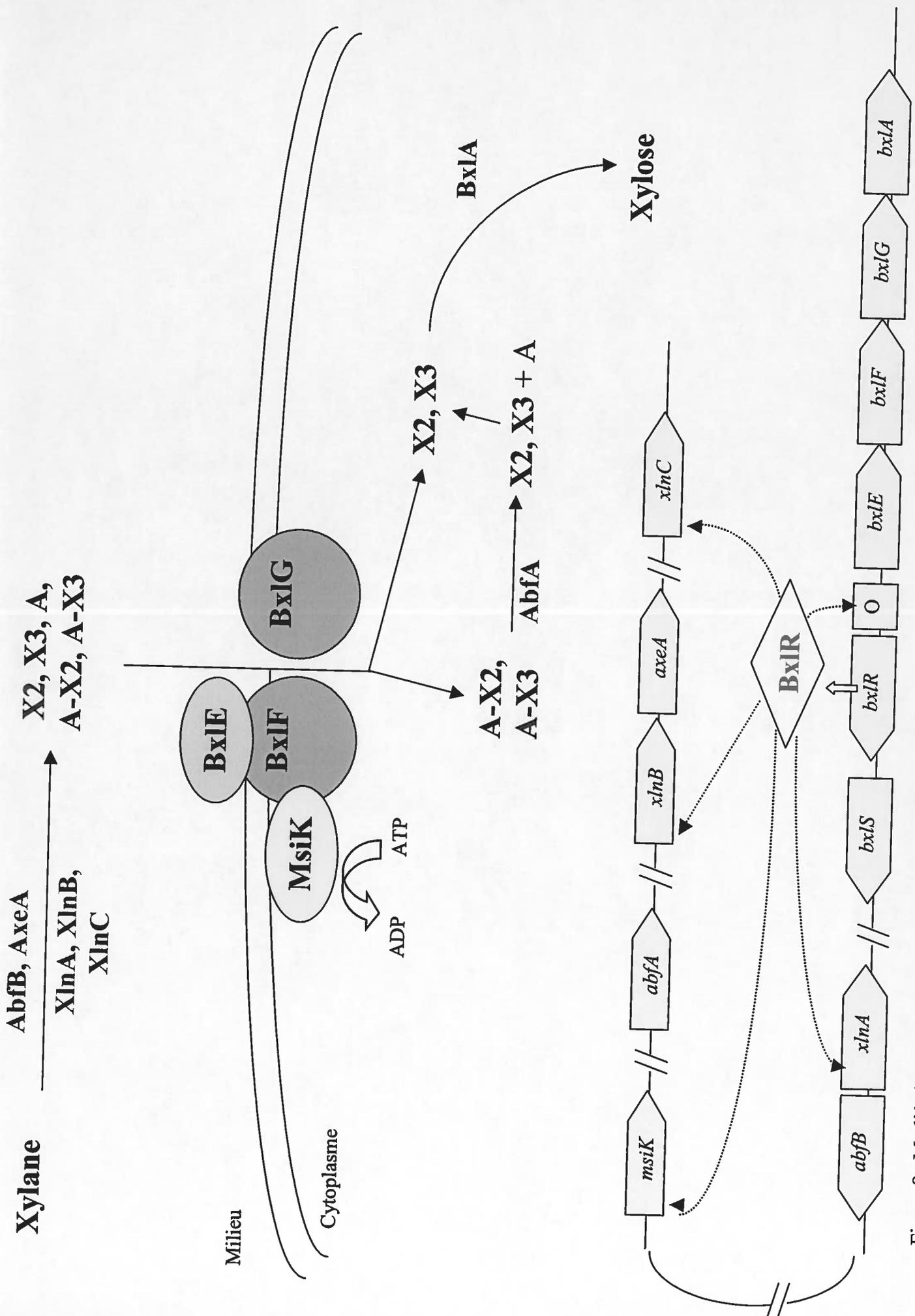


Figure 8. Modèle hypothétique de la dégradation du xylane par *S. lividans*.

2. APPROCHE EXPÉRIMENTALE

2.1 LISTE DES PRODUITS CHIMIQUES UTILISÉS

Acétonitrile	EM Science (HPLC grade)
Acétone	Anachemia (Accusolv)
Acide acétique	BDH
Acide molybdique	Sigma
Acrylamide	Roche
Bacto agar	Difco
Bacto casamino acid	Difco
Bleu de bromophénol	Bio-Rad
Borate de sodium	J.T. Baker
BSA	Bio-Rad
Carbonate d'ammonium	MCB
Carbonate de sodium	BDH
Cellobiose	Sigma
CHAPS	Roche
Chlorure de calcium	Sigma
Chlorure de cobalt	J.T. Baker
Chlorure de cuivre	Fisher
Chlorure de fer	Anachemia
Chlorure de magnésium	BDH
Chlorure de manganèse	J.T. Baker
Chlorure de sodium	BDH
Chlorure de zinc	Fisher
Dithiothréitol	Amersham Pharmacia Biotech
Éther	J.T. Baker
Extrait de levure	Difco
Glucose	BDH
Glycérol	ICN
Glycine	Bio-Rad
Huile minérale	Amersham Pharmacia Biotech
Hydroxyde de sodium	J.T. Baker
Iodoacétamide	Sigma
Isopropanol	Anachemia (Accusolv)
Maltose	Difco
Méthanol	Gouvernement du Québec
Méthionine	Sigma
Nitrate d'argent	BDH
Phosphate de potassium dibasique	BDH
Phosphate de potassium monobasique	Fisher
Phosphate de sodium dibasique	J.T. Baker
Phosphate de sodium monobasique	Fisher
L-proline	Gibco BRL
Résine AG® 501-X8 (D)	Bio-Rad

SDS	Bio-Rad
Saccharose	BDH
Sulfate de cuivre	J.T. Baker
Sulfate de fer	J.T. Baker
Sulfate de magnésium	BDH
Sulfate de manganèse	J.T. Baker
Sulfate de potassium	Anachemia
Sulfate de zinc	J.T. Baker
TCA	J.T. Baker
TES	Sigma
TFA	Sigma
Thiosulfate de sodium	J.T. Baker
Tris	EM Science
TSB	Difco
Tween 80	Aldrich
Tyrosine	Gibco BRL
Urée	USB
Xylane	Sigma
Xylose	Sigma

2.2 SOUCHES BACTÉRIENNES

Les souches bactériennes de *Streptomyces* utilisées lors des travaux sont décrites au tableau suivant. La souche IAF6 dérive de la souche sauvage 1326 et possède le plasmide pIAF6 (Figure 9). La souche M145 provient de la souche sauvage A3(2), est prototrophe et ne contient aucun plasmide.

Tableau 4. Souches bactériennes utilisées.

Souche	Génotype/phénotype	Source/référence
<i>S. lividans</i>		
1326	souche sauvage	Sir D.A. Hopwood (John Innes Center, Angleterre)
IAF6	1326/pIAF6	Ce travail
<i>S. coelicolor</i>		
M145		Prof. M. J. Bibb (John Innes Center, Angleterre)

2.3 PLASMIDES

Les plasmides utilisés au cours de ce projet sont décrits au Tableau 2. Le vecteur à copies multiples pIJ702 (Katz *et al.*, 1983) est utilisé pour surexprimer les gènes *bxIR* et *bxIS* chez la souche sauvage. Il possède le gène de résistance au thiostrepton (*tsr*) et le gène de biosynthèse de la tyrosinase qui permet la production de la mélanine (*mel*). Cette dernière donne une pigmentation noire aux colonies.

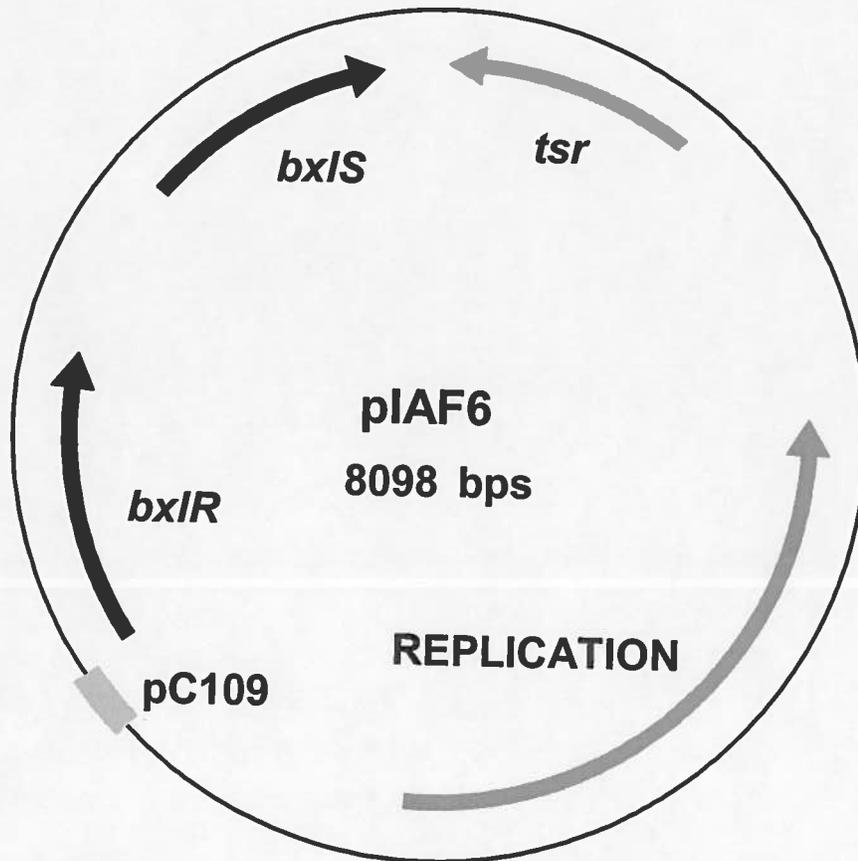


Figure 9. Carte physique du plasmide IAF6 utilisé afin d'obtenir la souche *S. lividans* IAF6. *tsr*, résistance à la thioestrepton; pC109, promoteur C109; *bxlR*, gène de la protéine BxlR; *bxlS*, gène de la protéine BxlS.

L'insertion d'un gène inactive la tyrosinase, ce qui permet de détecter les clones positifs puisque ceux-ci ne synthétisent plus de pigment noir. Le plasmide pIAF6 provient donc du plasmide pIJ702 et possède un insert de 2,7 kb comprenant les gènes *bxIR* et *bxIS*. En amont de *bxIR* se trouve le promoteur muté du gène de la xylanase C de *S. lividans*, le promoteur pC109.

Tableau 5. Description des plasmides.

Vecteur	Description	Source/référence
pIJ702	<i>tsr^r</i> , <i>mel^r</i>	Dr E. Katz
pIAF6	<i>tsr^r</i> , pC109, <i>bxIR</i> , <i>bxIS</i>	Ce travail

2.4 MILIEUX DE CULTURES

2.4.1 Milieu solide pour la conservation des souches

La souche de *S. coelicolor* a été conservée à 4°C sur milieu R5 solide (Hopwood *et al.*, 1985). Ce milieu est composé de 103 g de saccharose, 10 g de glucose, 250 mg de K₂SO₄, 10,12 g de MgCl₂·6H₂O, 100 mg de Bacto Casamino Acid, 5 g d'extrait de levure, 5 g de TES, 22 g de Bacto agar et 2 ml d'une solution d'éléments traces par litre d'eau distillée. La solution d'éléments traces contient 40 mg de ZnCl₂, 200 mg de FeCl₃, 10 mg de CuCl₂, 10 mg de MnCl₂, 10 mg de Na₂B₄O₇·10H₂O et 10 mg de (NH₄)₆Mo₇O₂₄ par litre d'eau distillée.

Suite à la stérilisation (121°C pendant 20 minutes), le milieu a été refroidi et les solutions suivantes ont été ajoutées stérilement: 10 ml de KH₂PO₄ 0,5% (p/v), 4 ml de CaCl₂ 5 M, 1 ml de CuSO₄ 0,5% (p/v), 15 ml de L-proline 20% (p/v), 10 ml de méthionine 1% (p/v) et 400 mg de tyrosine préalablement diluée dans 5 ml de NaOH 1 N. Le milieu a été réparti dans des boîtes de Pétri, qui ont ensuite été séchées pendant cinq

jours à la température de la pièce de façon à obtenir une perte de poids de 15%. Elles ont ensuite été entreposées à 4°C.

2.4.2 Milieu liquide riche pour la production d'inoculum

Le milieu riche employé pour la production d'inoculum (préculture) est le bouillon tryptone de soya sans glucose (TSB). Il a été préparé en réhydratant 27,5 g de TSB par litre d'eau distillée et stérilisé à 121°C pendant 20 minutes.

2.4.3 Milieu liquide minimal pour la croissance cellulaire

Le milieu M14 est composé de 1,4 g de $(\text{NH}_4)_2\text{SO}_4$, 5 g de K_2HPO_4 , 1 g de KH_2PO_4 , 2 ml de Tween 80 et 1 ml d'une solution de sels minéraux par litre d'eau distillée. La solution de sels minéraux de Mandels comprend 140 mg de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 160 mg de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 500 mg de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ et 200 mg de $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ par 100 ml d'eau distillée (Mandels et Reese, 1957). Le pH a été ajusté à 7,4 et le milieu a été réparti dans des Erlenmeyers. La source de carbone a été ajoutée dans chacun des Erlenmeyers avant la stérilisation (xylane) ou après la stérilisation (xylose) à la concentration finale de 1%.

Avant son utilisation, le milieu a été refroidi et les solutions suivantes ont été ajoutées aseptiquement: 0,6% (v/v) de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 5% (p/v) et 1% (v/v) de CaCl_2 3% (p/v).

2.5 CONDITIONS DE CULTURES

2.5.1 Production d'inoculum

Après croissance sur milieu solide R5, le mycélium et les spores des souches de streptomycètes ont été récoltés à l'aide d'un fil de platine bouclé et déposés dans un Erlenmeyer de 250 ml contenant 25 ml de milieu TSB et trois billes de verre stériles. Les billes servaient à fragmenter le mycélium afin de permettre une croissance plus homogène. La culture a été incubée à 34°C avec agitation à 240 rpm pendant 24 heures.

2.5.2 Culture en fermenteur

Une préculture de 25 ml a servi à ensemer 400 ml de milieu nutritif M14 dans un Erlenmeyer de 2 litres et ce dernier a été incubé à 34°C avec agitation à 240 rpm pendant 24 à 48 heures. Par la suite, un fermenteur de 14 litres contenant 7 litres de milieu M14 a été ensemercé avec la deuxième préculture. La température a été réglée à 34°C, l'aération à 7 litres d'air par minute et l'agitation à 400 rpm. La croissance a été poursuivie pendant 48 heures.

2.6 PRÉPARATION DES ÉCHANTILLONS PROTÉIQUES

Le mycélium a été récupéré en filtrant la culture sous vide avec un papier filtre Whatman #1 déposé sur un filtre Büchner. Le mycélium a été récupéré de la surface du filtre à l'aide d'une spatule stérile et resuspendu dans un volume minimal de tampon phosphate de sodium 10 mM à pH 7,4. Tout en maintenant les cellules sur glace afin de minimiser l'activité des protéases, celles-ci ont été lysées aux ultrasons à cinq reprises pendant 1 minute avec un sonicateur XL (Mandel). Afin de vérifier l'efficacité de la lyse

cellulaire, un examen microscopique a été réalisé. Après centrifugation à 27000 x g pendant 30 minutes à 4°C (centrifugeuse Beckman modèle J2-21, rotor JA20) le surnageant, contenant les protéines cytoplasmiques, a été récupéré puis conservé à -20°C.

2.7 DOSAGE DES PROTÉINES

La concentration en protéines des échantillons à analyser a été déterminée par la méthode de Lowry (1951), en utilisant comme protéine standard l'albumine de sérum bovin (BSA). La concentration en protéines des échantillons a été estimée par spectrophotométrie à une densité optique (DO) de 700 nm à l'aide d'une courbe standard de la DO en fonction de la concentration en BSA.

2.8 ÉLECTROPHORÈSE EN DEUX DIMENSIONS

2.8.1 Incorporation des protéines par réhydratation des languettes

La réhydratation du gel IPG des languettes déshydratées (Amersham Pharmacia Biotech) servant à la focalisation isoélectrique a été effectuée selon la méthode décrite dans le manuel du manufacturier. Les protéines ont d'abord été précipitées par deux volumes de TCA10%/acétone pendant au moins une heure à -20°C. Le culot récupéré après centrifugation pendant 5 minutes a été lavé avec de l'acétone froid à deux reprises afin d'éliminer le TCA résiduel. Les protéines présentes dans le culot ont ensuite été solubilisées pendant une heure avec agitation dans le tampon de réhydratation (urée 9 M déionisée avec la résine AG501-X8(D), CHAPS 4% (p/v), tampon pour IPG 0.5% (v/v), DTT 15 mM et quelques grains de bleu de bromophénol). Le tampon contenant les protéines (250 µl pour une languette de 13 cm, 350 µl pour celle de 18 cm) a été déposé en un point central dans le portoir, en évitant les bulles d'air. Ensuite la portion gel de la

languette a été déposée lentement sur la solution afin d'éviter que du tampon ne se retrouve sur le film protecteur. À l'aide de pinces, la languette a été glissée de gauche à droite pour permettre à toute la surface du gel d'être humectée. Chaque languette a été recouverte goutte à goutte d'huile minérale en commençant par une extrémité du portoir jusqu'à ce que l'huile atteigne la moitié de celui-ci. Ensuite, l'autre moitié a été recouverte de la même façon. L'huile prévient l'évaporation et la cristallisation de l'urée. Une fois le couvercle du portoir en place, le gel a été réhydraté pendant 12 à 16 heures, à 20°C. La réhydratation, ainsi que la focalisation isoélectrique ont été réalisées avec l'appareil IPGphor™ (Amersham Pharmacia Biotech). Ce dernier possède une source de courant et un système de réfrigération intégrés.

2.8.2 La première dimension: la focalisation isoélectrique

Cette technique permettait de séparer les protéines selon leur charge dans le gradient de pH immobilisé dans le gel des languettes. Elle a été effectuée selon la méthode décrite dans le manuel d'électrophorèse en 2-D de la compagnie Amersham Pharmacia Biotech. Ainsi, l'extrémité pointue du portoir a été déposée sur l'anode, la température a été réglée à 20°C et le courant à 50 μ A par languette. Les conditions d'électrophorèse pour des quantités préparatives de protéines, soit plus de 200 μ g, sont les suivantes: une première étape avec un voltage maximum de 500 V jusqu'à l'obtention de 500 Vh, une deuxième étape à 1000 V jusqu'à 1000 Vh. Le nombre de volt-heures à atteindre lors de la dernière étape varie selon l'étendue de l'intervalle de pH et la longueur de la languette. Les paramètres de la troisième étape sont décrits au Tableau 6. Chaque étape de la focalisation isoélectrique a été réalisée avec l'option de voltage 'step-n-hold'. Finalement, suite à la focalisation isoélectrique, les languettes ont été conservées à -80°C ou équilibrées immédiatement avant la deuxième dimension.

Tableau 6. Conditions d'électrophorèse utilisées lors de la dernière étape de la focalisation isoélectrique.

Languette		Voltage max. (V)	Vh
longueur	pH		
13 cm	4-7	8000	40000
18 cm	Gradient de pH étroit	8000	64000

2.8.3 Équilibration des languettes

Les languettes ont été équilibrées dans le tampon d'équilibration (urée 6M, glycérol 30% (p/v), SDS 2% (p/v) et Tris-HCl 0,05M pH 8.8) afin de permettre aux protéines de migrer de la première à la seconde dimension. L'équilibration a été réalisée selon le protocole développé par Görg *et al.*, (1988, 1995) et décrit dans le manuel d'électrophorèse 2-D de la compagnie Amersham Pharmacia Biotech. Après la focalisation isoélectrique, les languettes ont été introduites dans un tube de verre contenant 10 ml de tampon I (10 ml de tampon d'équilibration, 100 mg de DTT et 50 μ l de la solution de bleu de bromophénol 0,25% (p/v)) et équilibrées pendant 15 minutes sur un agitateur. Ensuite, elles ont été transférées dans un autre tube contenant 10 ml de tampon II (10ml de tampon d'équilibration, 250 mg d'iodoacétamide et 50 μ l de la solution de bleu de bromophénol 0,25% (p/v)) et équilibrées pendant 15 minutes sur un agitateur.

2.8.4 La deuxième dimension: l'électrophorèse en gel de polyacrylamide en présence de SDS

L'électrophorèse a été effectuée selon la méthode conventionnelle établie par Laemmli (1970) avec l'appareil Protean® II de Bio-Rad. La deuxième dimension permettait de séparer les protéines selon leur masse moléculaire dans un gel de polyacrylamide 12% de 20 cm x 20 cm x 0,15 cm. Une fois la languette équilibrée, celle-ci a été trempée quelques secondes dans le tampon d'électrophorèse 1X (Tris-HCl 25 mM

pH 8,3, glycine 1,4% (p/v), SDS 0,1% (p/v)) afin d'éliminer l'excès de tampon d'équilibration puis placée sur le dessus du gel. La languette a été pressée doucement sur le gel avec une spatule afin d'assurer un meilleur contact. Un standard de poids moléculaire a été déposé sur un morceau de papier filtre, qui a été lui aussi placé sur le dessus du gel, à côté de la languette. Le standard utilisé était le Broad Molecular Weight (Bio-Rad) et il comprenait les protéines suivantes : l'aprotinine (6,50 kDa), le lysozyme (14,4 kDa), l'inhibiteur de trypsine (21,5 kDa), l'anhydrase carbonique (31,0 kDa), l'ovalbumine (45,0 kDa), l'albumine de sérum bovin (66,2 kDa), la phosphorylase B (97,4 kDa), la β -galactosidase (116 kDa) et la myosine (200 kDa). Afin de dénaturer ces protéines avant leur séparation dans le gel de SDS, le standard a été chauffé à 95°C pendant 5 minutes. L'électrophorèse a été réalisée pendant 30 minutes avec un courant maximum de 15 mA/gel et par la suite avec un courant maximum de 30 mA/gel jusqu'à la fin de la migration. Le courant initial faible facilite le transfert des protéines du gel de la languette à celui de la deuxième dimension.

2.9 VISUALISATION DES PROTÉINES PAR COLORATION AU NITRATE D'ARGENT COMPATIBLE AVEC LA DIGESTION TRYPTIQUE

La coloration suivante a été utilisée lorsque les protéines résolues par électrophorèse en 2-D devaient être digérées par la trypsine afin d'être analysées par spectrométrie de masse (Pappin et Perkins, 1999). Après l'électrophorèse, les protéines ont été fixées avec 400 ml d'une solution de méthanol 50% (v/v) et d'acide acétique 5% (v/v) pendant 1 heure et ensuite avec 400 ml d'une solution de méthanol 50% (v/v) pendant 10 minutes. Après avoir rincé le gel à deux reprises avec 400 ml d'eau déionisée pendant 10 minutes, celui-ci a été incubé dans 300 ml d'une solution de $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$ 0,02% (p/v) pendant 1 minute. À la suite de deux lavages de 1 minute avec 400 ml d'eau déionisée, la coloration a été effectuée avec 300 ml d'une solution d' AgNO_3 0,2% (p/v) pendant 20 minutes. Finalement, après deux lavages de 1 minute avec 400 ml d'eau déionisée, le développement a été effectué en ajoutant 300 μl de formaldéhyde 37% à 300 ml d'une solution de Na_2CO_3 2% (p/v). Une fois la coloration suffisamment développée, la réaction a été stoppée en incubant le gel dans une solution d'acide acétique 5% (v/v). Il

était préférable de préparer des solutions fraîches de $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$, d' AgNO_3 et de Na_2CO_3 .

2.10 ANALYSE DES GELS 2-D

Suite à la coloration au nitrate d'argent, les gels ont été digitalisés avec un scanner (Umax, Astra 600S). Les images obtenues avec le scanner ont été converties en fichier TIFF de 8 bits afin d'être compatibles avec le logiciel d'analyse de gels 2-D ImageMaster 2-D Elite v3.01 (Amersham Pharmacia Biotech). Suite à l'acquisition d'un système de capture d'image par caméra CCD (Kodak Image Station IS440CF), les images de gels ont été numérisées avec ce dernier et elles ont été enregistrées directement dans le format adéquat.

Le logiciel a été utilisé afin de détecter automatiquement et de façon manuelle les protéines, de les quantifier et de normaliser leur volume, d'éliminer le bruit de fond et les stries et de faire correspondre les protéines entre les gels. De plus, le point isoélectrique et la masse moléculaire apparents ont été déterminés par une calibration à l'aide de standards utilisés dans une dimension seulement (1-D calibration).

2.11 DIGESTION DES PROTÉINES AVEC LA TRYPSINE

La digestion de protéines en peptides à l'intérieur d'un gel de polyacrylamide a été réalisée selon un protocole transmis par Madame France Dumas de l'Institut de Recherche en Biotechnologie. Suite à la coloration au nitrate d'argent, les protéines séparées par électrophorèse en 2-D ont été excisées du gel à l'aide d'un emporte pièce. Celles-ci étaient conservées à -20°C lorsque la digestion n'était pas effectuée immédiatement après l'électrophorèse. Le morceau de gel a été déposé dans un microtube et recouvert avec 500 μl de NH_4HCO_3 100 mM pendant au moins 1 heure. Afin d'éviter la formation de ponts disulfures entre les fragments peptidiques produits pendant la digestion enzymatique, leurs groupements sulfidriles ont d'abord été réduits puis alkylés.

Ainsi, le gel a été incubé pendant 30 minutes à 60°C dans 150 µl de NH_4HCO_3 100 mM et 10 µl de DTT 45 mM. Une fois les microtubes refroidis à la température de la pièce, 10 µl d'iodoacétamide 100 mM ont été ajoutés et le tout a été incubé 30 minutes à la température de la pièce et à l'abri de la lumière. Par la suite, le gel a été incubé avec agitation pendant une heure dans une solution d'acétonitrile 50%/ NH_4HCO_3 100 mM. Le surnageant a été rejeté et remplacé par 50 µl d'acétonitrile pendant 10 à 15 minutes, de façon à réduire le volume du gel. Afin de déshydrater complètement le morceau de gel, celui-ci a été asséché à l'évaporateur rotatif (SpeedVac, Savant). Ensuite, le gel a été réhydraté avec 20 µl de NH_4HCO_3 25 mM contenant 0,2 µg de trypsine. La trypsine qui a été utilisée est la trypsine modifiée de qualité de séquençage de la compagnie Promega (# de catalogue: V5111) et elle a été reconstituée selon les recommandations du fabricant. Après 10 à 15 minutes (ou jusqu'à ce que le gel ait repris son volume original), 20 µl de NH_4HCO_3 25 mM ont été ajoutés afin de maintenir le gel dans le tampon durant la digestion enzymatique. Celle-ci a été réalisée pendant 16 heures à 37°C.

Après la digestion enzymatique il était nécessaire de combiner la digestion de quatre protéines identiques afin d'obtenir un signal suffisamment intense en spectrométrie de masse. Ainsi, le surnageant de la digestion de chaque protéine a été réservé et les extraits de fragments peptidiques subséquents ont été ajoutés au microtube en utilisant le même embout afin de minimiser les pertes. Deux extractions ont été effectuées en incubant le morceau de gel avec 50 µl d'acétonitrile 60%/TFA 0,1% pendant 20 minutes. Finalement, les extraits combinés ont été évaporés complètement à l'évaporateur rotatif avant leur injection dans le spectromètre de masse.

2.12 ANALYSE DES PEPTIDES PAR SPECTROMÉTRIE DE MASSE

2.12.1 Concentration et purification de l'échantillon de peptides

Avant d'injecter un échantillon dans un spectromètre de masse possédant un mode d'ionisation «electrospray», il est nécessaire d'éliminer les sels, les détergents, les agents chaotropiques et les particules de polyacrylamide présents dans l'échantillon. En effet, ces contaminants inhibent l'ionisation des biomolécules (protéines et peptides) et empêchent par conséquent leur détection par le spectromètre de masse. De plus, la présence de contaminants rend l'analyse des spectres de masse complexe puisque certains ions correspondant à des peptides peuvent être masqués par ces ions interférents.

Les trois solutions nécessaires à l'utilisation des embouts ZipTip™ (Millipore) ont été filtrées sur un filtre de nylon de 0,2 µm résistant aux solvants (Cole-Parmer) et le matériel servant à préparer et à conserver ces solutions a préalablement été rincé à l'eau déionisée, à l'isopropanol, à l'acétone et à l'éther afin d'éliminer toutes traces de contaminants. De plus, les solvants utilisés étaient de qualité supérieure et les microtubes provenaient de la compagnie Bio-Rad (# catalogue: 223-9305).

Les échantillons de peptides ont été purifiés et concentrés à l'aide d'un embout de pipette ZipTip™. Cet embout de 10 µl contient environ 0,6 µl de résine C₁₈ (silice sphérique de 15 µm, porosité de 200Å). Les peptides sont fixés à la résine, les contaminants sont éliminés par lavage et l'échantillon concentré et purifié est élué dans le volume désiré (4 µl).

L'échantillon a d'abord été dissous dans 18 µl d'eau déionisée et 2 µl d'une solution de TFA 20%, de façon à augmenter la concentration en acide de l'échantillon à 2%, ce qui améliore la fixation des peptides à la résine. Après avoir réglé le volume de la pipette à 10 µl et fixé l'embout ZipTip™, la résine a été humectée en aspirant et en rejetant (un cycle) une solution d'acétonitrile 50%/TFA 0,1%. L'opération a été répétée trois fois avec une solution fraîche. Ensuite, l'embout a été équilibré en vue de la fixation

des peptides en réalisant cinq cycles d'aspiration/élimination d'une solution de TFA 0,1%. Une solution fraîche a été utilisée pour chaque cycle. L'échantillon acidifié a été déposé sur le dessus de la résine, à l'intérieur de l'embout ZipTip™ à l'aide d'un embout de chargement à pointe fine et 10 cycles d'aspiration/élimination ont été pratiqués dans un microtube afin de fixer les peptides. Par la suite, 10 cycles supplémentaires ont été effectués avec une solution fraîche de TFA 0,1% entre chaque cycle afin d'éliminer les contaminants. Finalement, 4,5 µl d'une solution d'acétonitrile 80%/TFA 0,1% ont été déposés dans un microtube propre et les peptides ont été élués en effectuant cinq cycles d'aspiration/élimination. Afin d'éliminer les bulles d'air qui peuvent nuire à l'ionisation de l'échantillon, celui-ci a été incubé pendant cinq minutes dans un bain à ultrasons (Branson 1200). L'échantillon de peptides purifié et concentré a été conservé à -20°C lorsque l'analyse par spectrométrie de masse n'était pas effectuée immédiatement.

2.12.2 Injection dans le spectromètre de masse

Tous les échantillons furent analysés par nanoESI MS. Ceux-ci ont été injectés dans un capillaire pour nanoES spray recouvert d'un mélange d'or et de plomb (Micromass) à l'aide d'une seringue de 10 µl (Hamilton). Les capillaires ont été insérés dans la source nanoflow Z *SPRAY*™ d'un spectromètre de masse à triple quadrupole (Quattro II, Micromass). La manipulation de l'instrument, l'acquisition des données et l'analyse des spectres ont été effectuées avec le logiciel Masslynx/Biolynx v3.3 (Micromass). Le voltage du cône, du capillaire, de même que l'énergie de collision appliquée lors de la fragmentation ont été optimisés pour chaque échantillon.

2.12.3 Recherche dans les bases de données

La masse des peptides a été comparée à la masse théorique des peptides de chacune des protéines présentes dans la base de données SWISS-PROT/TrEMBL, mais la recherche a été restreinte aux espèces bactériennes. Les protéines ont été identifiées avec le logiciel de recherche par profil de la masse des peptides, PeptIdent

(expasy.cbr.nrc.ca/tools/peptident.html). Les paramètres suivants ont été utilisés lors de la recherche (Tableau 7): aucune restriction par rapport au point isoélectrique et à la masse moléculaire; la masse chimique moyenne et sans charge (neutre, [M]) des peptides, en tenant compte que les cystéines ont été traitées avec l'iodoacétamide; une marge d'erreur de +/- 0,5 Da et un site de clivage manqué par la trypsine. La moitié des peptides devaient concorder pour que le logiciel accepte la protéine comme un résultat possible.

La masse neutre (m) est calculée de la façon suivante, par exemple pour un ion d'un rapport m/z de 500 et possédant 2 charges (z), donc protoné (H⁺) 2 fois:

$$m + 2H^+ / z = 500$$

$$m + 2H^+ = z \times 500 \text{ et } z = 2$$

$$m = (2 \times 500) - (2H^+) \text{ et H possède une masse de 1 Da}$$

$$m = 1000 - 2$$

$$m = 998 \text{ Da}$$

Quant à la charge d'un peptide (z), elle est déterminée par la différence entre la valeur du rapport m/z du pic de l'ion moléculaire et de son pic isotopique. L'ion moléculaire est aussi appelé l'ion parent et il correspond à la forme ionisée d'un composé neutre, ici le peptide chargé (Lee, p.5). Le terme isotope est utilisé pour décrire les atomes d'un élément avec un nombre variable de neutrons (Lee, p.1). Ainsi, le carbone (composant majoritaire des protéines) est retrouvé de façon naturelle sous deux formes isotopiques, soit ¹²C (6 protons et 6 neutrons) et ¹³C (6 protons et 7 neutrons). Étant donné que ces deux formes n'ont pas le même nombre de neutrons, elles n'ont pas la même masse. Ainsi, un peptide (ion moléculaire, ¹²C) d'une masse de 1000 Da et qui a été chargé une fois apparaîtra à une valeur du rapport m/z de 1000 (1000/1) et le pic isotopique (¹³C) apparaîtra à 1001 (1000 + 1 neutron/1) (Figure 10A). Par contre, le même peptide d'une masse de 1000 mais qui a été chargé deux fois apparaîtra à une valeur du rapport m/z de 500 (1000/2) et son pic isotopique apparaîtra à 500.5 (1000 + 1 neutron/2) (Figure 10B).

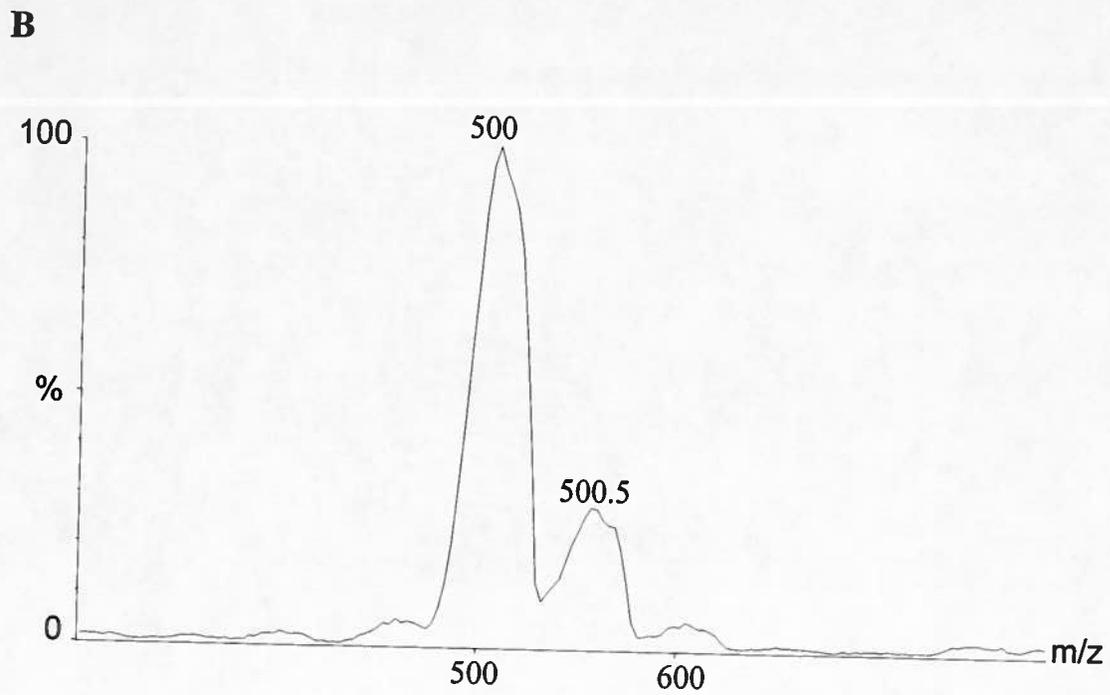
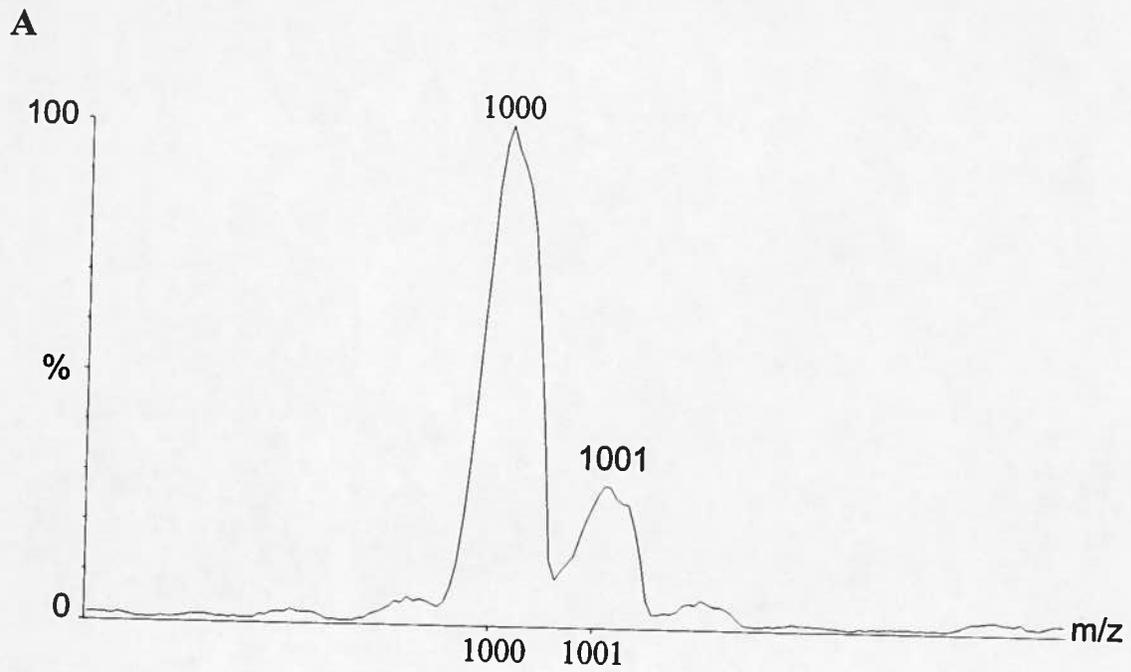


Figure 10. Exemple d'un ion d'une masse de 1000 Da possédant une charge ($+1H^+$) en A et d'un ion d'une masse de 1000 Da et possédant deux charges ($+2H^+$) en B.

Tableau 7. Paramètres utilisés lors de la recherche dans les bases de données avec le logiciel PeptIdent.

Peptide Mass Fingerprinting

Name of the unknown protein: <input type="text"/>		pI: <input type="text"/> within pI range: <input type="text" value="1.00"/>
Database: <input type="text" value="SWISS-PROT and TrEMBL"/>		Mw: <input type="text"/> within Mw range (in percent): <input type="text" value="20"/>
Note: For proteins from TrEMBL, peptides with masses >10000 Da have not been indexed.		
Species to be searched: <input type="text" value="BACTERIA"/>		
Enter a list of peptide masses (separated by spaces or newlines) that correspond to the unknown protein: <input type="text"/>		Or upload a file in one of the supported formats from your computer. The peptide masses will be extracted automatically from this file: <input type="text"/>
<p>All peptide masses are</p> <p><input type="text" value="mh"/> $[M+H]^+$ or</p> <p><input type="text" value="m"/> $[M]$ or</p> <p><input type="text" value="mminus"/> $[M-H]^-$, and</p> <p><input type="text" value="monoisotopic"/> monoisotopic or</p> <p><input type="text" value="average"/> average.</p> <p>The peptide masses are</p> <p>with cysteines treated with:</p> <p><input type="text" value="Iodoacetamide"/></p> <p><input type="text" value="acrylamide"/> with acrylamide adducts on cysteines</p> <p><input type="text" value="oxidize"/> with methionines oxidized.</p> <p>Mass tolerance: ± <input type="text" value="0.5"/> Dalton</p>		<p>Enzyme: <input type="text" value="Trypsin"/></p> <p>Allow for <input type="text" value="1"/> missed cleavage sites (MC).</p> <p>Report only proteins with at least <input type="text"/> peptide hits.</p> <p>Display a maximum of <input type="text" value="20"/> matching proteins.</p> <p><input type="text" value="yes"/> Print information about sequence portion covered by the matching peptides.</p>

3. RÉSULTATS

3.1 MISE AU POINT DES DIFFÉRENTES ÉTAPES DE L'ANALYSE DU PROTÉOME

3.1.1 L'électrophorèse en deux dimensions et détection des protéines

L'optimisation des gels en 2-D a été réalisée en analysant les protéines cytoplasmiques des souches *S. lividans* 1326 et IAF6. Tout d'abord parce que la souche *S. coelicolor* M145 (dont le génome est actuellement séquencé) nous a été offerte beaucoup plus tard au cours du projet et ensuite en raison de l'intérêt pour la souche IAF6. En effet, l'identification des protéines induites et/ou réprimées par la surexpression de BxlR par comparaison des protéomes de la souche sauvage et de la souche IAF6, serait une façon d'approfondir le rôle de cette protéine. Ainsi, les gels en 2-D des protéines cytoplasmiques de la souche sauvage et de la souche IAF6 ont d'abord été comparés (Figure 11). Il a été visuellement possible d'observer des différences entre le patron de protéines de chacune des souches: une protéine présente seulement chez la souche IAF6 (flèche bleu), certaines présentes seulement chez 1326 (flèches rouges) et une dont le niveau d'expression est plus élevé chez IAF6 (flèche jaune). Mais le patron général des protéines a été retrouvé chez les deux gels (cercles blancs).

Il a été possible d'observer sur cette même figure qu'une bonne proportion des protéines ont été séparées et résolues (taches bien définies), mais que les stries horizontales du côté acide (pH 4) ont rendu difficile la visualisation des protéines dans cette portion du gel. Aussi, une forte proportion des protéines ont été retrouvées entre les valeurs de pH de 4 à 5,5 environ. Il semble que les protéines se soient accumulées du côté acide, puisqu'une strie verticale, plus visible dans l'électrophorèse des protéines cytoplasmiques de la souche IAF6, a été observée.

3.1.2 Analyse des gels: estimation du point isoélectrique et du poids moléculaire

Afin d'utiliser ces deux caractéristiques des protéines pour confirmer leur identification, le gel des protéines cytoplasmiques de la souche IAF6 (Figure 11A) a été analysé avec le logiciel ImageMaster 2D Elite. L'image de gel sans le bruit de fond a été choisie pour représenter le gel et un total de 129 protéines ont été détectées et numérotées (Figure 12). Le point isoélectrique ainsi que le poids moléculaire de chacune a été estimé (1-D calibration) (Tableau 8). La protéine SL82 (en bleu), dont le niveau d'expression était plus élevé chez IAF6 (flèche jaune, Figure 11A), a été analysée par spectrométrie de masse.

3.1.3 Analyse des peptides par spectrométrie de masse: profil de la masse des peptides

La protéine SL82, en raison de son abondance comparativement aux autres protéines et aussi parce qu'elle est moins abondante chez la souche 1326, a été la première à être digérée à la trypsine afin que les peptides générés soient analysés par spectrométrie de masse. Quatre morceaux de gels contenant la protéine SL82 ont été digérés et combinés avant d'être injectés dans le spectromètre (Figure 13). Dix ions, suffisamment intenses pour être pris en considération, ont été retenus pour l'analyse par profil de la masse des peptides, soit: 475,09, 489,50, 506,56, 515,10, 562,52, 579,22, 689,36, 697,73, 738,78 et 864,53. Le pic d'un rapport m/z de 679,00 n'a pas été retenu en tant qu'ion pour le profil de la masse des peptides puisqu'il a aussi été retrouvé dans le blanc. L'échantillon servant de blanc consiste en un morceau de gel ne contenant aucune protéine et ayant subi les mêmes traitements qu'un morceau de gel avec protéine. Leur charge a aussi été déterminée et chaque peptide était chargé deux fois. La masse neutre, utilisée pour la recherche dans les bases de données, pour chacun des ions sont les suivantes: 948,18, 977,00, 1011,12, 1028,20, 1123,04, 1156,44, 1376,72, 1393,46, 1475,56 et 1727,06.

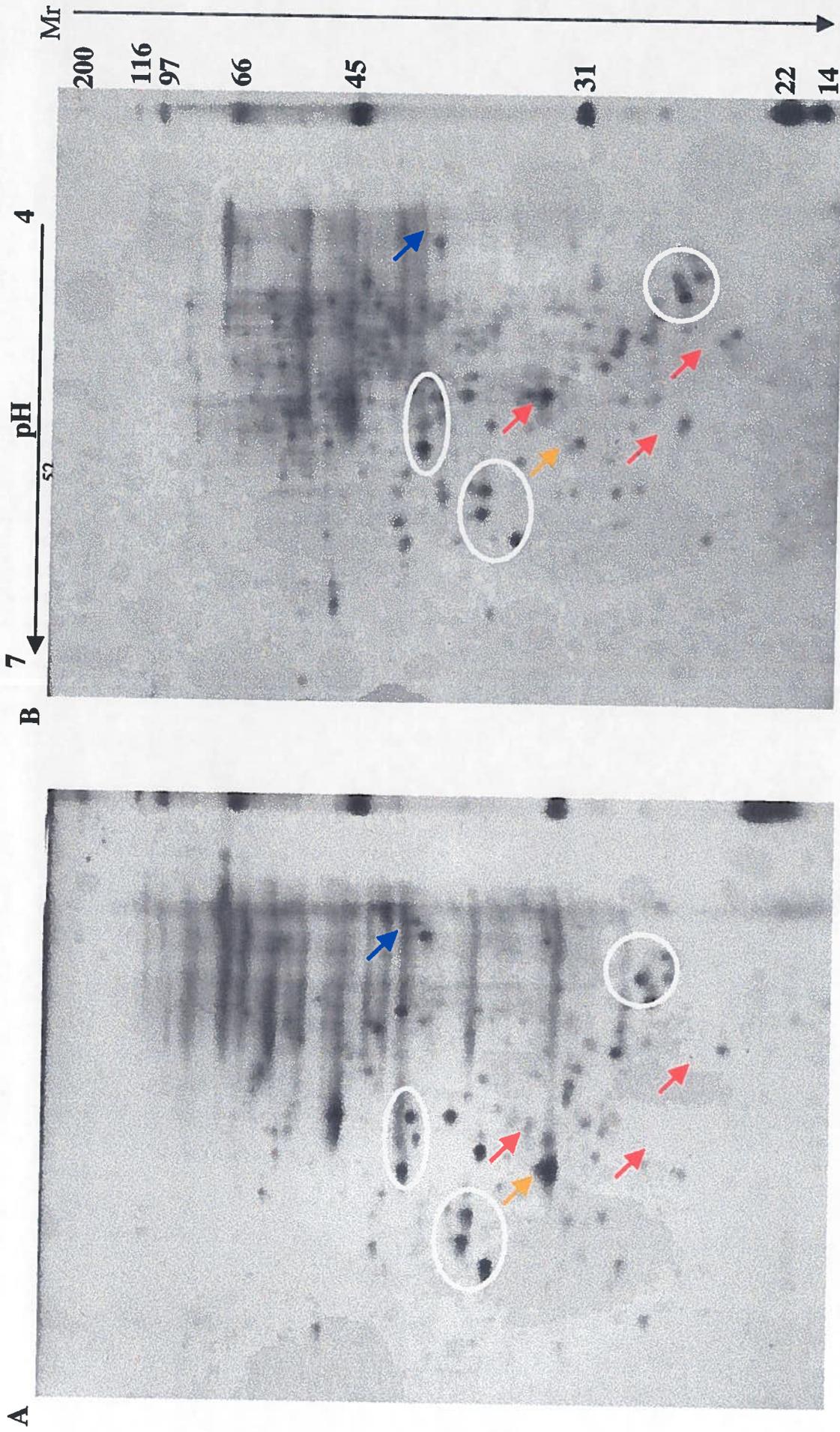


Figure 11. Gel en 2-D de pH 4 à 7 (13 cm) des protéines cytoplasmiques de *S. lividans* IAF6 (A) et de *S. lividans* 1326 (B). La croissance des cellules a été effectuée dans le milieu M14 contenant 1 % de xylose et 200 µg de protéines ont été incorporées dans chacun des gels. Les flèches signalent les différences entre les 2 souches et les cercles, le patron de protéines retrouvé dans les 2 gels.

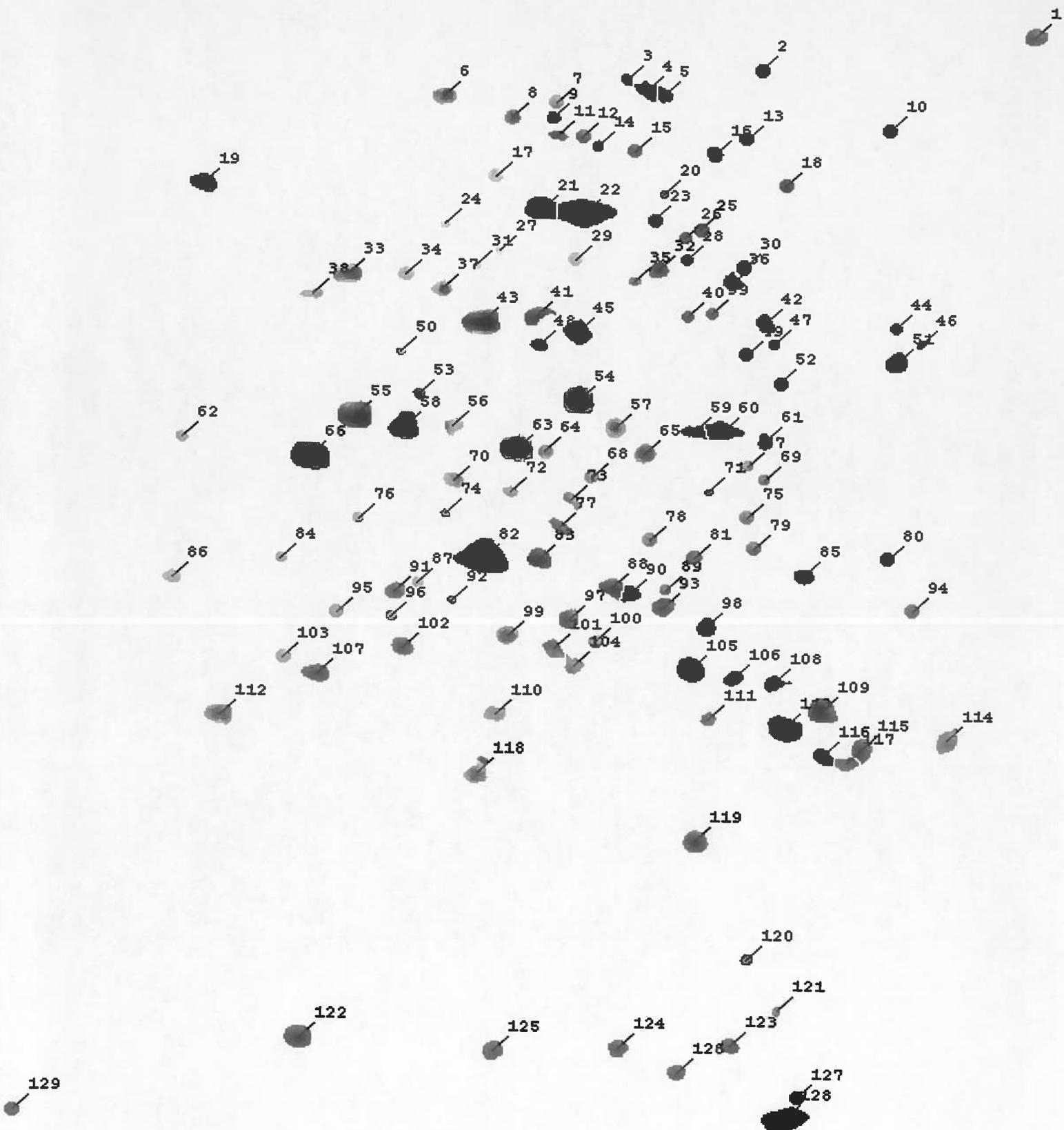


Figure 12. Image numérisée du gel 2-D de pH 4-7 de la souche *S. lividans* IAF6 (Figure 11A) analysée avec le logiciel ImageMaster 2-D Elite.

Tableau 8. Liste des protéines cytoplasmiques séparées par gel 2-D de pH 4-7 de la souche IAF6 détectées par le logiciel ainsi que leur point isoélectrique et leur poids moléculaire apparents.

#	pI	MW	#	pI	MW	#	pI	MW	#	pI	MW
1	3,84	65,275	33	5,95	33,293	65	5,03	27,370	97	5,25	25,423
2	4,69	56,850	34	5,77	33,293	66	6,06	27,245	98	4,83	25,389
3	5,10	54,551	35	5,06	33,107	67	4,71	27,184	99	5,44	25,313
4	5,04	52,709	36	4,76	33,107	68	5,18	26,954	100	5,17	25,282
5	4,99	51,831	37	5,66	32,571	69	4,65	26,954	101	5,29	25,242
6	5,66	51,261	38	6,06	32,233	70	5,61	26,872	102	5,76	25,233
7	5,32	50,429	39	4,83	31,597	71	4,83	26,767	103	6,13	25,168
8	5,45	48,081	40	4,90	31,446	72	5,44	26,717	104	5,23	25,150
9	5,33	48,080	41	5,37	31,371	73	5,25	26,618	105	4,87	25,142
10	4,29	47,106	42	4,66	31,225	74	5,64	26,411	106	4,74	25,100
11	5,31	45,719	43	5,53	31,083	75	4,71	26,411	107	6,02	25,092
12	5,24	45,495	44	4,26	31,083	76	5,90	26,326	108	4,61	25,083
13	4,73	45,495	45	5,24	30,737	77	5,28	26,263	109	4,47	24,965
14	5,19	44,415	46	4,18	30,473	78	5,00	26,126	110	5,47	24,932
15	5,07	43,794	47	4,63	30,410	79	4,68	26,051	111	4,81	24,926
16	4,83	43,591	48	5,36	30,222	80	4,27	25,963	112	6,32	24,913
17	5,50	40,790	49	4,72	29,980	81	4,87	25,946	113	4,58	24,900
18	4,61	40,281	50	5,78	29,921	82	5,52	25,912	114	4,08	24,865
19	6,40	39,626	51	4,25	29,804	83	5,34	25,912	115	4,34	24,832
20	4,98	39,309	52	4,61	29,107	84	6,14	25,879	116	4,46	24,810
21	5,36	37,818	53	5,73	28,630	85	4,53	25,800	117	4,39	24,790
22	5,22	37,539	54	5,23	28,540	86	6,47	25,695	118	5,52	24,748
23	5,01	37,130	55	5,92	28,081	87	5,72	25,680	119	4,85	24,607
24	5,65	36,604	56	5,62	27,889	88	5,12	25,666	120	4,68	24,458
25	4,87	36,477	57	5,12	27,889	89	4,95	25,653	121	4,59	24,421
26	4,91	35,856	58	5,77	27,852	90	5,06	25,624	122	6,06	24,401
27	5,48	34,818	59	4,87	27,852	91	5,79	25,611	123	4,73	24,400
28	4,91	34,389	60	4,79	27,852	92	5,61	25,545	124	5,07	24,398
29	5,25	34,284	61	4,65	27,671	93	4,96	25,520	125	5,46	24,396
30	4,73	33,975	62	6,45	27,568	94	4,19	25,520	126	4,89	24,387
31	5,56	33,874	63	5,42	27,434	95	5,97	25,446	127	4,52	24,377
32	4,99	33,774	64	5,33	27,402	96	5,80	25,423	128	4,56	24,370
									129	6,94	24,369

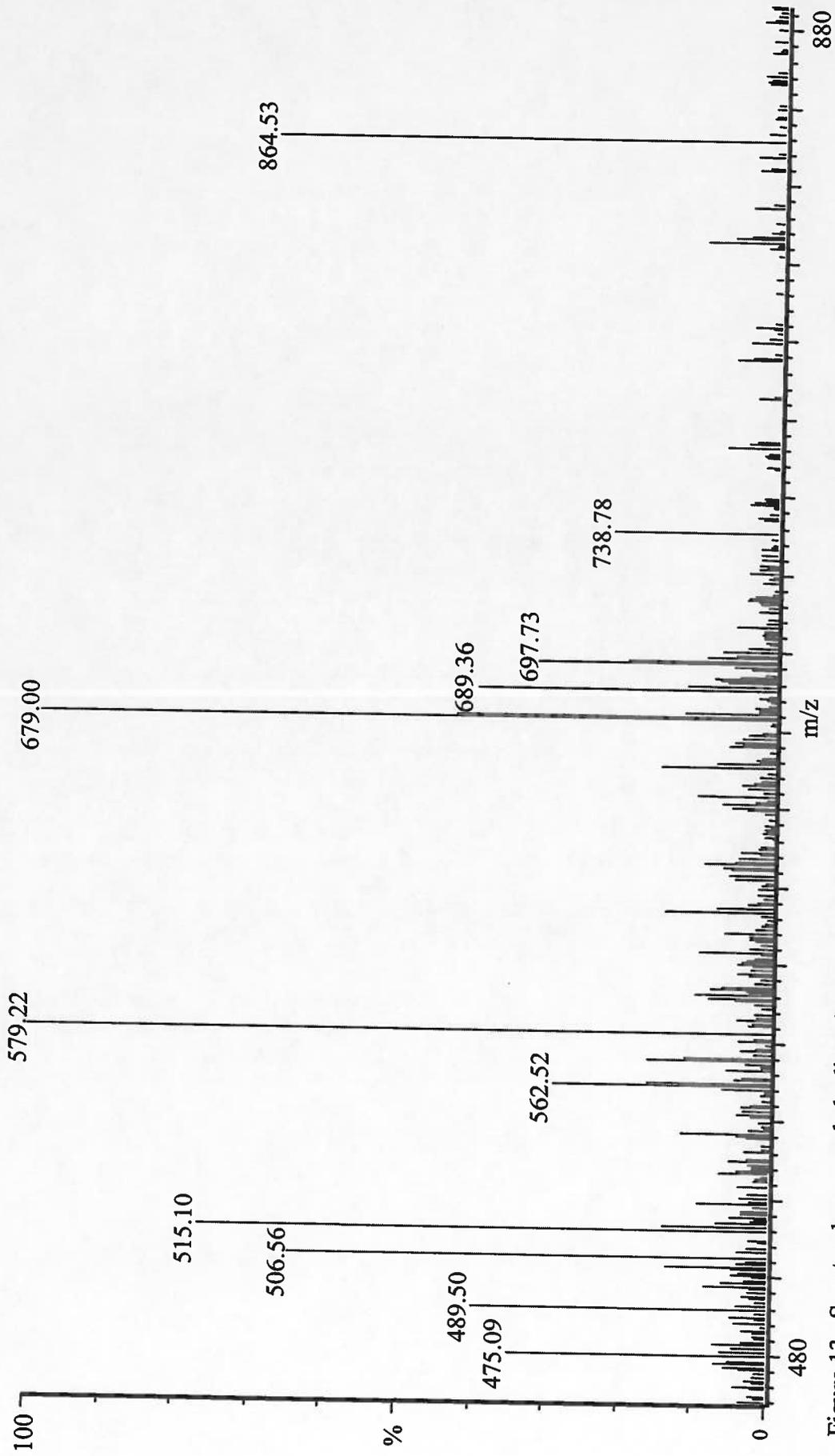


Figure 13. Spectre de masse de la digestion tryptique de la protéine SL82.

3.1.4 Recherche dans les bases de données avec le logiciel PeptIdent

Suite à l'obtention du spectre de masse des peptides issu de la digestion tryptique de la protéine SL82 et à la détermination de la charge et de la masse neutre de chacun des ions retenus pour le profil, la recherche a été effectuée. Le Tableau 9 présente les huit protéines possibles trouvées par le logiciel avec le nom (*ID*) et la description de chacune. Les protéines 4 et 8 ont des valeurs de pI beaucoup plus élevées que la protéine SL82 (pI de 5,52) soit de 9,44 et 9,00 respectivement. Les autres protéines (1, 2, 3, 5, 6 et 7) ont des valeurs de pI similaires à celle de la protéine SL82, soit d'environ 5. Les protéines 2, 3, 5, 6, 7 et 8 ont des valeurs de masse moléculaire plus grandes que la protéine SL82 (*Mr* de 25,9 kDa) soit de 477 kDa, 546 kDa, 573 kDa, 298 kDa, 130 kDa et 216 kDa respectivement. Quant aux protéines 1 et 4, leur masse moléculaire est du même ordre de grandeur que celle de la protéine SL82 (dans les dizaines de kDa), soit de 28,9 kDa et 68,3 kDa respectivement. Cependant, la protéine 4 possède tout de même une masse moléculaire plus élevée que celle de la protéine SL82. Le profil théorique qui possède le plus d'homologie (# peptide matches) avec le profil observé appartient à la protéine 1 (8 peptides sur 10), les autres profils possèdent moins d'homologie avec le profil de la protéine SL82 avec 6 et 5 peptides sur 10. Les peptides obtenus lors de la digestion tryptique de la protéine SL82 couvrent 30,5% de la protéine 1 (% of sequence covered) et couvrent moins de 6,1% des autres protéines.

Dans le Tableau 10, le profil de la masse des peptides de la SL82 est comparé avec le profil théorique de la protéine 1, une méthylase qui confère la résistance au thiostrepton chez *S. azureus*. La plus grande différence de masse (mass (Da)) entre un peptide théorique (matching mass) et un peptide appartenant à la protéine SL82 (user mass) est de 0,3079 Da. Tous les peptides théoriques, sauf celui d'une masse de 1156,44, sont des peptides qui ne contiennent aucun site de clivage manqué (#MC) par la trypsine. Les 2 ions qui ne concordent pas avec la protéine 1 sont 1475,56 (m/z de 738,78) et 1727,06 (m/z de 864,53). De plus, la séquence des peptides de la protéine 1 qui concordent avec les peptides obtenus est fournie

(peptide). La position de ces peptides sur la protéine est indiquée en rouge et en lettres majuscules. De cette façon il a été possible de remarquer que les peptides sont répartis sur toute la protéine.

Tableau 9. Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SL82.

8 matches found.

#	peptide matches	ID	Description	pI	Mw	% of sequence covered
1	8	TSN_ STRAZ	RRNA (ADENOSINE-2'-O-)-METHYLTRANSFERASE (EC 2.1.1.66) (23S RRNA [AM1067] 2'-O-METHYLTRANSFERASE) (23S RRNA METHYLASE) (THIOSTREPTON- RESISTANCE METHYLASE). - <i>Streptomyces azureus</i> .	5.24	28901.04	30.5
2	6	PKSM_ BACSU	PUTATIVE POLYKETIDE SYNTHASE PKSM. - <i>Bacillus subtilis</i> .	5.47	477463.16	1.7
3	5	O31783	POLYKETIDE SYNTHASE OF TYPE I. - <i>Bacillus subtilis</i> .	5.59	545720.42	0.9
4	5	Q9PHY9	PUTATIVE PENICILLIN-BINDING PROTEIN. - <i>Campylobacter jejuni</i> .	9.44	68377.73	6.1
5	5	Q9PPZ5	HYPOTHETICAL MEMBRANE LIPOPROTEIN. - <i>Ureaplasma urealyticum</i> .	6.45	572700.16	1.1
6	5	Q9R9I9	MYCC. - <i>Bacillus subtilis</i> .	5.15	297946.47	2.0
7	5	Q9S6P0	PYRUVATE FERREDOXIN OXIDOREDUCTASE. - <i>Clostridium pasteurianum</i> .	5.92	129711.98	4.1
8	5	HMW2_ MYCPN	CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 2 (CYTADHERENCE ACCESSORY PROTEIN 2). - <i>Mycoplasma pneumoniae</i> .	9.00	215622.05	4.0

Tableau 10. Description détaillée de la comparaison entre le profil de la protéine 1 versus le profil de la protéine SL82.

Score: 0.80, 8 matching peptides: P18644 (TSNR_STRAZ) pI: 5.24, Mw: 28901.04 RRNA (ADENOSINE-2'-O-)-METHYLTRANSFERASE (EC 2.1.1.66) (23S RRNA [AM1067] 2'-O-METHYLTRANSFERASE) (23S RRNA METHYLASE) (THIOSTREPTON- RESISTANCE METHYLASE). - <i>Streptomyces azureus</i> .					
user mass	matching mass	mass (Da)	#MC	position	peptide
948.18	948.087	-0.0931	0	179-186	EEAIAFIR
977	977.169	0.1686	0	213-221	LALLFGSEK
1011.12	1011.233	0.1126	0	126-135	IVGNIGAIIV
1028.2	1028.217	0.0169	0	18-26	IIDVTKPSR
1123.04	1123.348	0.3079	0	187-196	DSGMQLMTLK
1156.44	1156.348	-0.0922	1	115-125	RGDVVVLDGVK
1376.72	1376.616	-0.1042	0	78-89	LIDSSIVNQLFK
1393.46	1393.649	0.1889	0	166-178	GYVFSLPVVLSGR
30.5% of sequence covered:					
mteldtianpsdpavqr IIDVTKPSR snikttliedveplmhsiaagvefievygsdsspfpsellclgrqipvr LIDSSIVNQLFK gerkaktfgiarvprparfgdiasr RGDVVVLDGVK IVGNIGAIIVR tslalgasgiilvdsditsiadrrlqrasr GYVFSLPVVLSGR EEAIAFIR DSGMQLMTLK adgdisvkelgdnpr LALLFGSEK ggpsdlfeeassasvsipm msqteslnvsvglialheridrnl aanr					

3.1.5 Analyse des peptides par spectrométrie de masse: fragmentation et recherche de la séquence en acides aminés

Les ions 506,5; 515,10; 579,22; 679,00; 689,36; 738,78 et 864,53 ont été fragmentés et un spectre de fragmentation typique d'un peptide a été obtenu pour chacun d'eux, sauf pour l'ion d'un rapport m/z de 679,00. Cependant, seule la fragmentation des ions 515,10, 579,22 et 689,36 a permis de générer un spectre où une séquence continue a pu être obtenue (Figures 14, 15 et 16). Le logiciel Biolynx a été utilisé pour retrouver la séquence en acides aminés des peptides de la protéine 1 (Tableau 10) dans les spectres de fragmentation obtenus. Ainsi, pour le spectre de fragmentation de l'ion d'un rapport m/z de 515,10 (1028,20) la séquence recherchée était IIDVTKPSR. Le logiciel, à partir des pics les plus intenses, a trouvé une partie de cette séquence, soit IDVT, en lisant de la droite vers la gauche. La séquence recherchée dans le spectre de fragmentation de l'ion d'un rapport m/z de 579,22 (1156,44) était la suivante: RGDVVVLDGVK. Le logiciel a été en mesure d'identifier une partie de cette séquence, soit VVLDGV, en lisant de la gauche vers la droite. À nouveau, les pics les plus intenses ont été retenus afin d'obtenir cette séquence partielle. Finalement, pour le spectre de fragmentation de l'ion d'un rapport m/z de 689,36 (1376,72), la séquence recherchée était LIDSSIVNQLFK et le logiciel a retrouvé une partie de cette séquence, soit IDSSIVNQ, en lisant de la droite vers la gauche.

Quant à l'ion 679,00 sa fragmentation a été tentée (Figure 17) en raison de sa forte intensité, même si cet ion se retrouvait également dans le blanc. Cependant, il n'a pas fragmenté à la façon d'un peptide avec l'énergie de collision habituelle (45 eV), ni lorsqu'une énergie de collision plus élevée a été appliquée (55 eV). En effet, l'augmentation de l'énergie n'a fait qu'augmenter le bruit de fond. De plus, certains pics sont apparus avec une énergie de 55 eV, mais ils n'étaient pas très larges.

A

	Ile	Ile	Asp	Val	Thr	Lys	Pro	Ser	Arg
B	1029.22	916.07	802.91	687.82	588.68	487.58	359.41	262.29	175.21
	--	0.62	-0.07	-0.55	0.39	0.09	--	--	--

Daughters of 515ES+

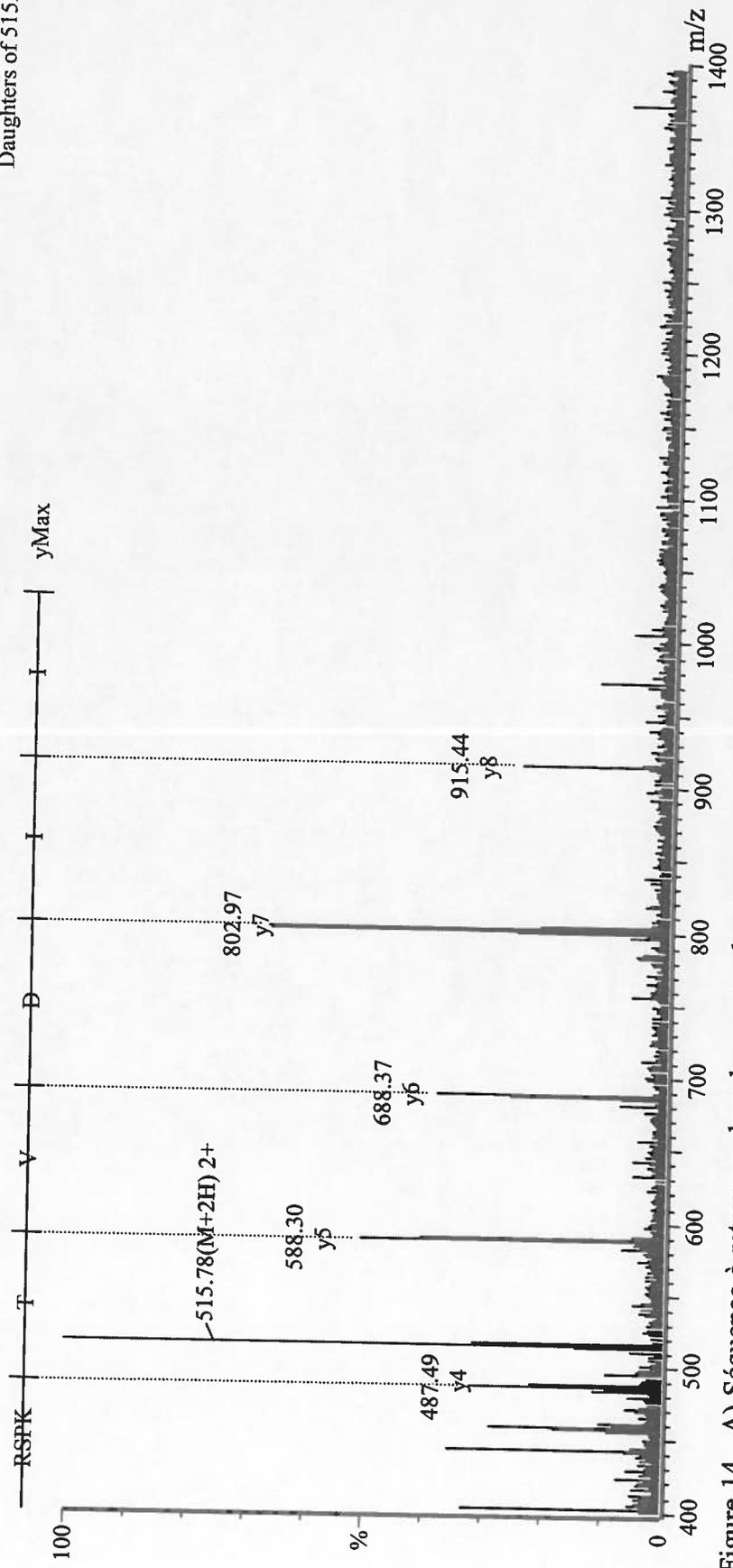


Figure 14. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion parent d'un rapport m/z de 515 obtenu avec une énergie de collision de 35 eV.

A		V		V		L		D		G		V		K		bMax	
b	m/z	m/z	Arg	Gly	Asp	m/z	Val	m/z	Val	m/z	Val	m/z	Leu	Asp	m/z	Val	Lys
---	157.20	---				428.47		527.60		626.73		739.89		854.98	1011.17	1139.34	---
---	---	214.25				-0.31		-0.26		0.17		-0.04		-0.48	-0.58	---	---
	---	---				---		---		---		---		---	---	---	---

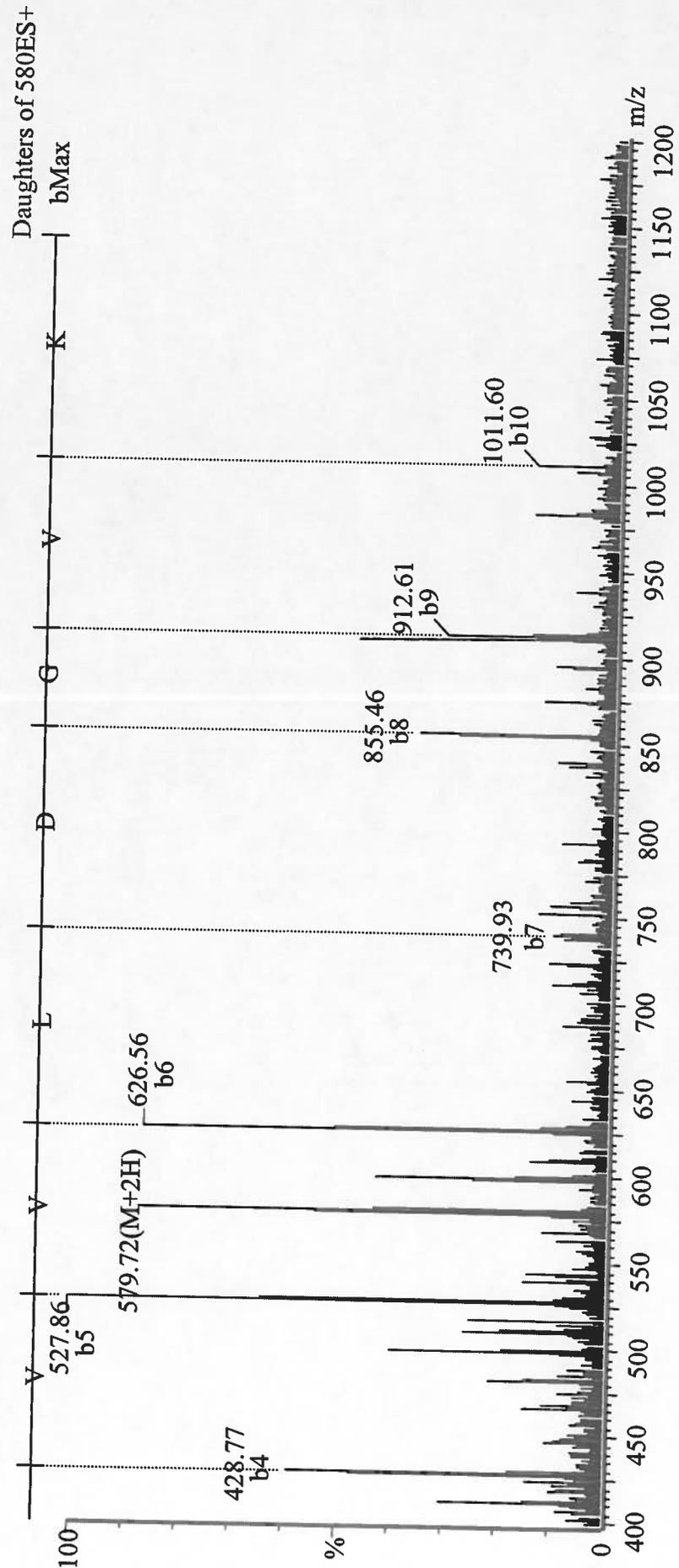


Figure 15. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 580 obtenu avec une énergie de collision de 40 eV.

	Leu	Ile	Asp	Ser	Ser	Ile	Val	Asn	Gln	Leu	Phe	Arg	Lys
y	1377.59	1264.43	1151.27	1036.18	949.11	862.03	748.87	649.74	535.63	407.50	294.34	147.16	-9.02

---	-0.27	-0.60	-0.53	-0.31	0.31	0.57	0.44	0.56	-0.02	--	--	--	--
-----	-------	-------	-------	-------	------	------	------	------	-------	----	----	----	----

B

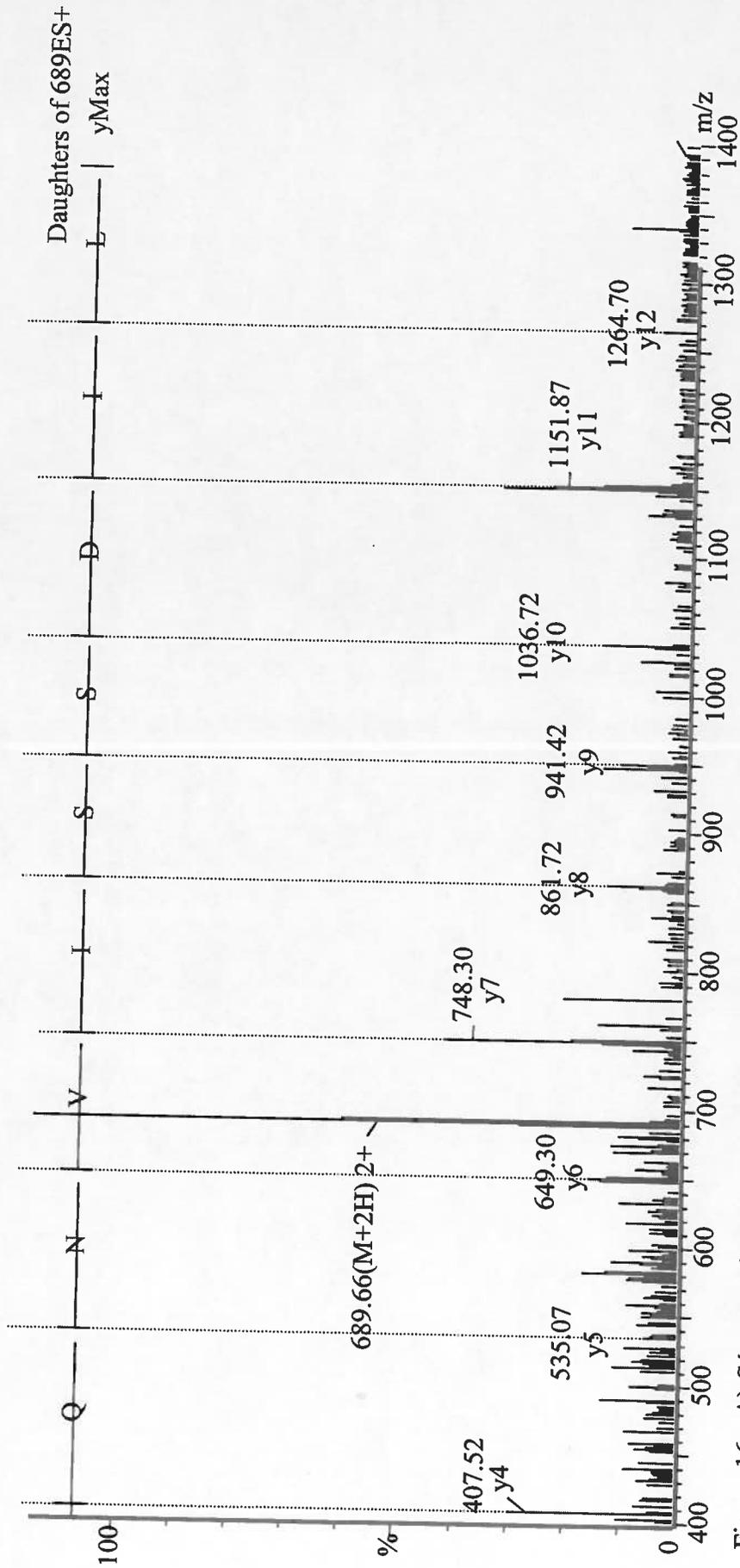


Figure 16. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 690 obtenu avec une énergie de collision de 35 eV.

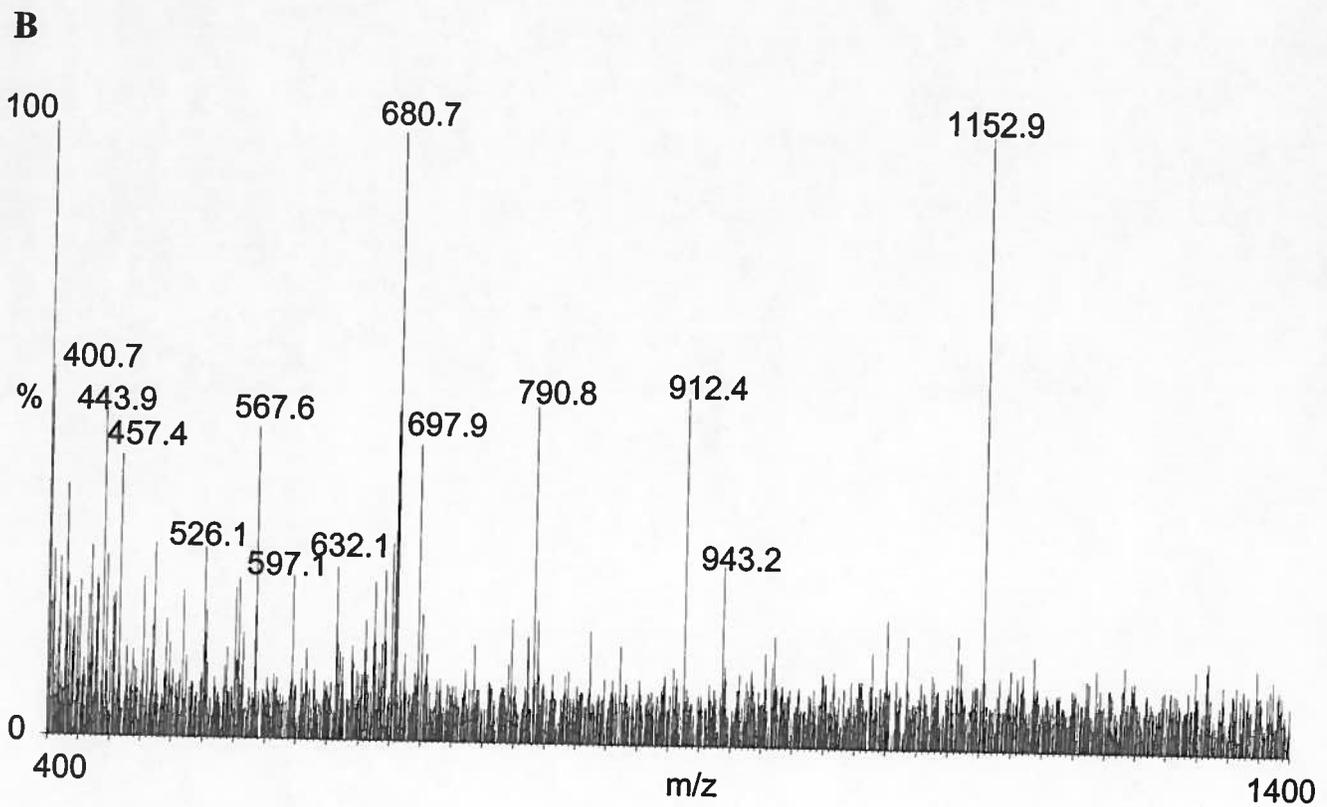
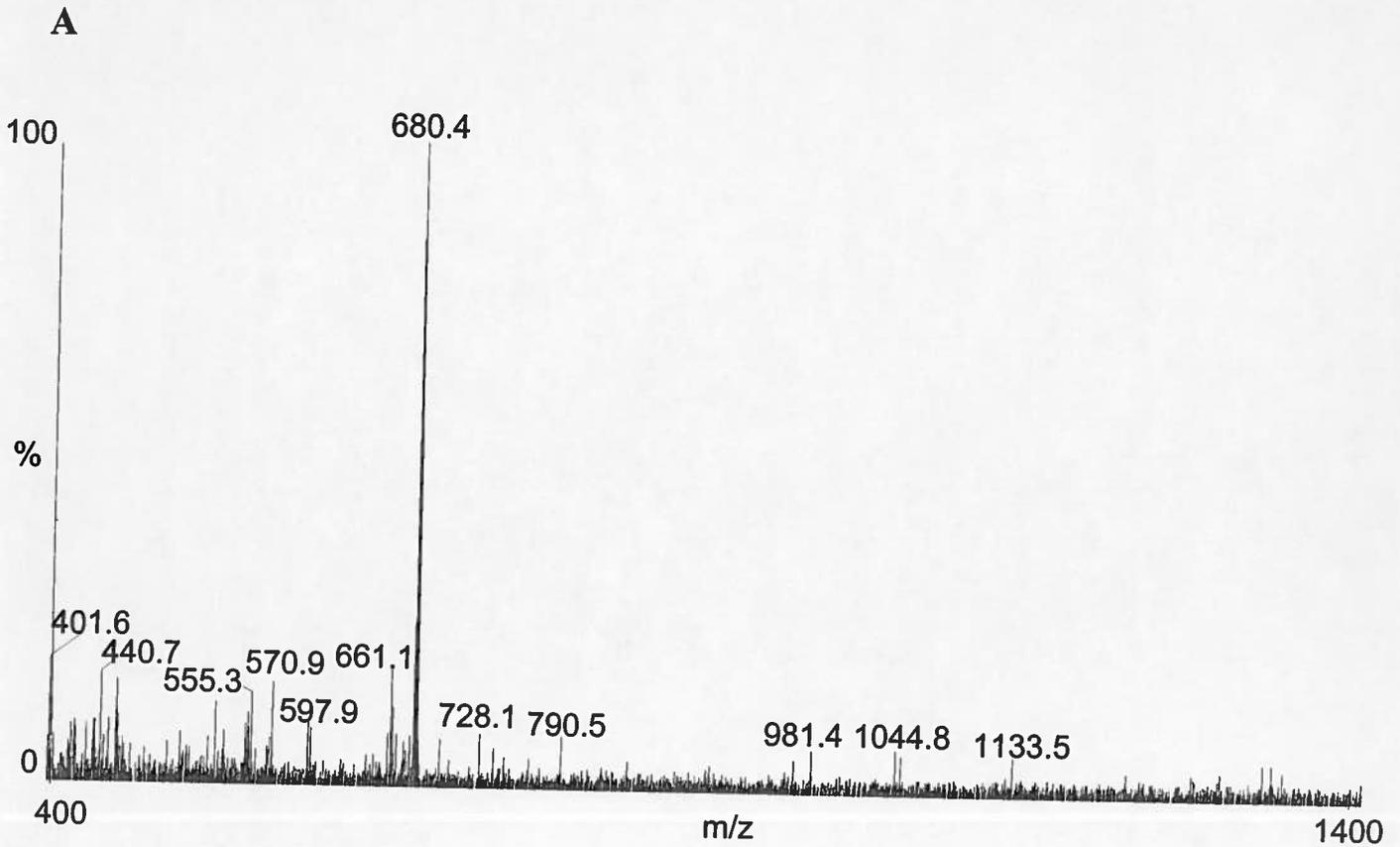


Figure 17. Spectre de fragmentation de l'ion d'une m/z de 679, avec une énergie de collision de 45 eV (A) et une énergie de collision de 55 eV (B).

3.2 ANALYSE DU PROTÉOME DE *S. COELICOLOR* M145

3.2.1 Séparation par électrophorèse en deux dimensions des protéines possédant des pI acides

Une fois les différentes étapes de l'analyse du protéome optimisées, l'étude du protéome de *S. coelicolor* a débuté. Les expériences d'électrophorèse en 2-D réalisées avec *S. lividans* 1326 et IAF6 ont permis de commencer directement avec des languettes de gradient de pH acide pour la séparation des protéines cytoplasmiques. Étant donné que la majorité des protéines se retrouvaient entre les valeurs de pH de 4 à 5,5 environ et qu'il y avait même une accumulation des protéines à pH 4 (section 3.1.1, Figure 11) les languettes de pH étroit et acide ont été utilisées. Ainsi, il a été possible d'obtenir un patron des protéines cytoplasmiques de *S. coelicolor* possédant des valeurs de pI entre 3,5 et 4,5 (Figure 18) et entre 4,0 et 5,0 (Figure 19). Contrairement au gel de pH 4-7 (Figure 11), les protéines sont bien séparées et résolues sur toute la surface du gel et les stries horizontales dans la partie la plus acide des gels sont toujours présentes mais n'occupent qu'une partie minime du gel. Étant donné que l'intervalle de pH des languettes se recouvre (4,0 à 4,5), il a aussi été possible de retrouver les mêmes protéines d'un gel à l'autre pour ces valeurs de pI. À titre de point de repère, les protéines 34, 68, 69 et 73 du gel de pH 3,5-4,5 correspondent respectivement aux protéines 71, 124, 122 et 127 du gel de pH 4,0-5,0.

3.2.2 Analyse des gels: identification des protéines les plus abondantes

Visuellement il a été possible d'observer des protéines en quantité plus importante (surface) et plus intenses (coloration) que d'autres. Néanmoins, les deux gels ont été analysés avec le logiciel ImageMaster 2D Elite afin de déterminer de façon précise les protéines les plus abondantes par rapport à leur surface, cette dernière étant déterminée par le logiciel (Tableau 11 et 12). Ainsi, les protéines les plus abondantes ont été retenues

en tenant compte de leur surface mais aussi en raison de leur intensité sur le gel. Les protéines les plus abondantes pour le gel pH 3,5-4,5 sont indiquées en rouge sur les images de gel (Figure 18 et 19) ainsi que dans les tableaux 11A et 12B. Les protéines 117, 124, 127 et 151 sont quant à elles les protéines les plus abondantes pour le gel de pH 4,0-5,0. Elles sont indiquées en bleu sur l'image du gel (Figure 19) ainsi que dans le Tableau 12A. La protéine indiquée en rose représente une protéine analysée par spectrométrie de masse mais qui n'était pas parmi les plus abondantes. Le pI seulement de ces protéines a été estimé (Tableaux 11B et 12B). En effet comme les languettes étaient de la même longueur que la largeur du gel de polyacrylamide (deuxième dimension) il n'a pas été possible d'insérer le morceau de papier filtre contenant le standard de poids moléculaire. De plus, le logiciel a détecté un total de 87 protéines dans le gel de pH 3,5-4,5 et un total de 154 protéines dans le gel de pH 4,0-5,0.

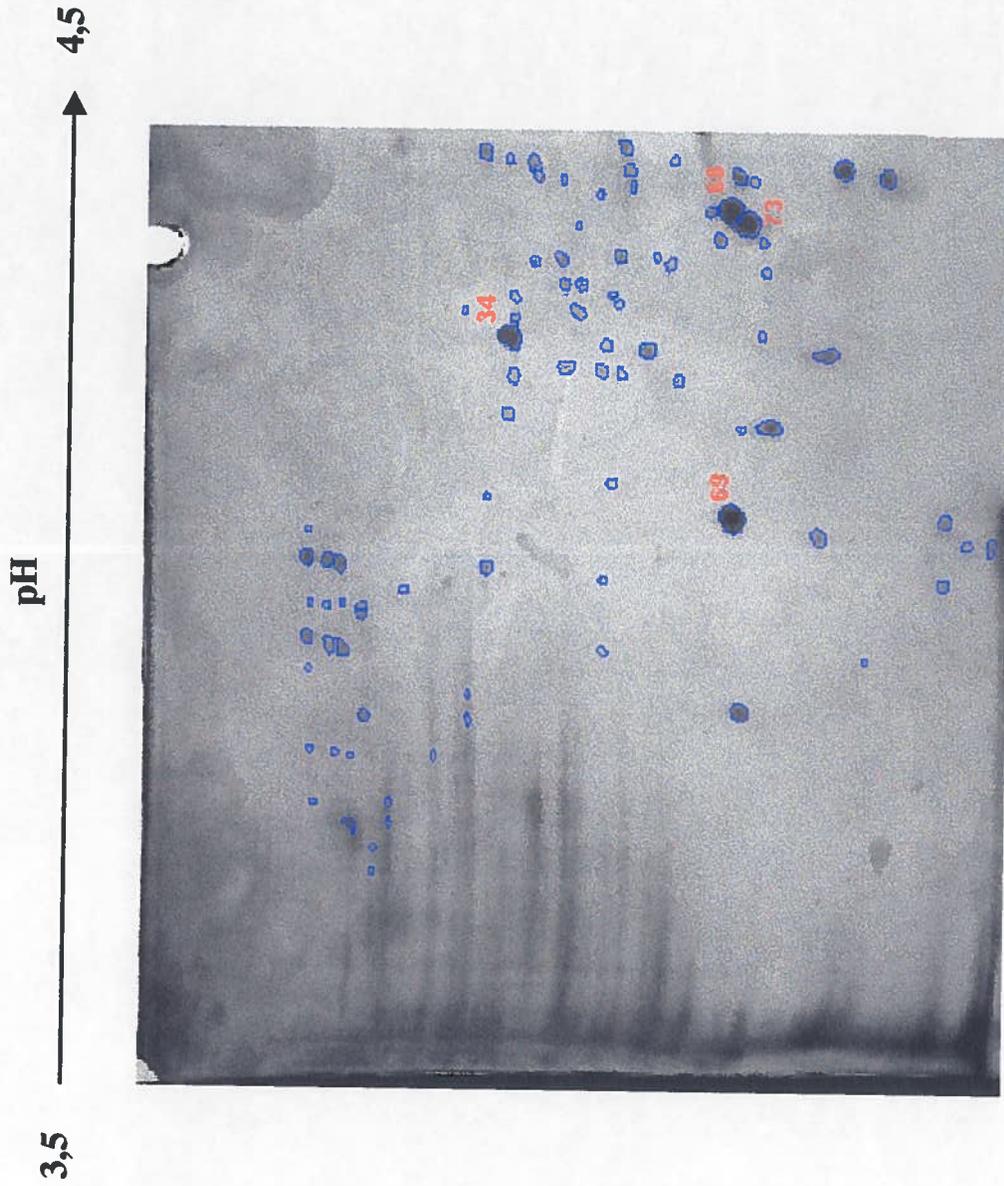


Figure 18. Image numérisée du gel 2-D de pH 3,5-4,5 (18 cm) de la souche *S. coelicolor* M145 analysée avec le logiciel ImageMaster 2-D Elite. La croissance des cellules a été effectuée dans le milieu M14 contenant 1 % de xylane et 500 µg de protéines ont été incorporées dans le gel. Les protéines les plus abondantes sont numérotées en rouge.

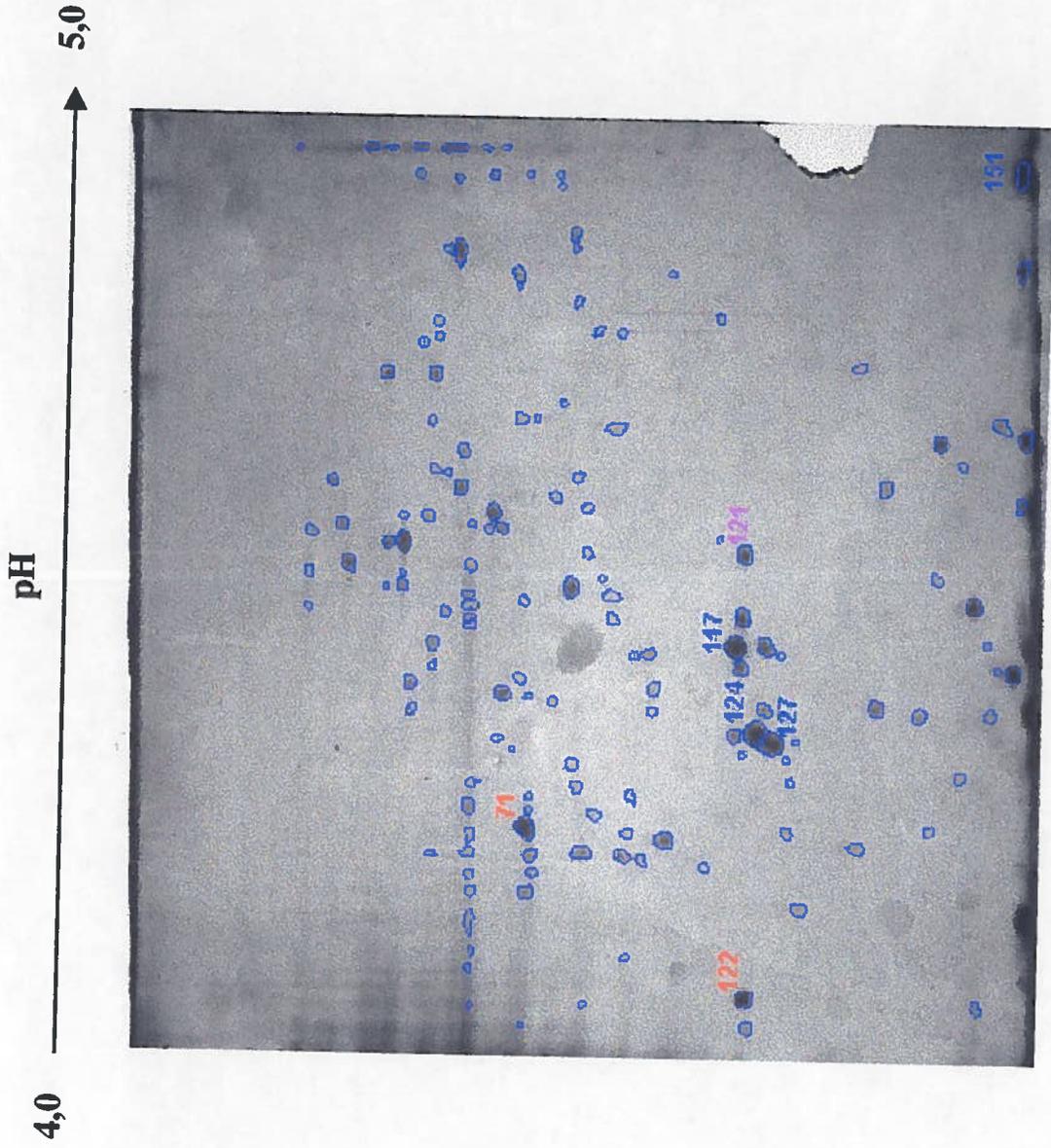


Figure 19. Image numérisée du gel 2-D de pH 4,0-5,0 (18 cm) de la souche *S. coelicolor* M145 analysée avec le logiciel ImageMaster 2-D Elite. La croissance des cellules a été effectuée dans le milieu M14 contenant 1 % de xylane et 500 µg de protéines ont été incorporées dans le gel. Les protéines les plus abondantes sur le gel de pH 3,5-4,5 sont numérotées en rouge, les plus abondantes sur ce gel sont indiquées en rose et la protéine indiquée en rose représente une protéine analysée par spectrométrie de masse.

Tableau 11. Liste des protéines cytoplasmiques séparées par gel 2-D de pH 3,5-4,5 de *S. coelicolor* M145 détectées par le logiciel ainsi que leur surface (A). Les protéines les plus abondantes sont indiquées en rouge et leur pI est indiqué dans le Tableau du bas (B).

A)

#	Surface	#	Surface	#	Surface	#	Surface
1	12 536	24	3 051	47	5 555	70	12 887
2	2 322	25	2 997	48	3 055	71	2 213
3	2 219	26	1 399	49	5 356	72	16 175
4	9 182	27	3 913	50	2 923	73	34 682
5	2 397	28	2 124	51	2 221	74	3 946
6	2 523	29	10 121	52	4 133	75	4 074
7	2 506	30	7 073	53	2 631	76	4 527
8	2 166	31	1 858	54	2 969	77	4 121
9	9 098	32	5 582	55	2 665	78	18 414
10	8 259	33	3 993	56	7 744	79	11 237
11	2 517	34	28 320	57	3 752	80	17 001
12	9 727	35	6 108	58	8 381	81	21 239
13	9 618	36	4 208	60	3 628	82	2 351
14	2 261	37	3 908	61	10 728	83	16 392
15	2 465	38	7 358	62	2 471	84	7 047
16	9 396	39	3 249	63	4 976	85	9 516
17	4 801	40	4 630	64	3 628	86	6 204
18	4 113	41	7 737	65	4 150	87	9 592
19	7 157	42	2 514	66	5 972		
20	2 929	43	5 544	67	7 472		
21	2 915	44	7 182	68	34 002		
22	3 610	45	7 373	69	37 852		
23	3 149	46	1 812				

B)

#	pI
34	4,28
68	4,42
69	4,09
73	4,41

Tableau 12. Liste des protéines cytoplasmiques de *S. coelicolor* M145 séparées par gel 2-D de pH 4,0-5,0 détectées par le logiciel ainsi que leur surface (A). Les protéines les plus abondantes sur le gel de pH 3,5-4,5 sont numérotées en rouge, les plus abondantes sur ce gel sont indiquées en bleu. La protéine indiquée en rose représente une protéine analysée par spectrométrie de masse mais qui n'était pas parmi les plus abondantes. Leur pI est indiqué dans le tableau du bas (B)

A)

#	Surface	#	Surface	#	Surface	#	Surface	#	Surface
1	3 003	32	3 939	63	5 782	94	1 794	125	7 614
2	2 127	33	2 057	64	1 694	95	3 337	126	11 669
3	3 705	34	14 105	65	1 515	96	4 351	127	23 222
4	4 096	35	4 951	66	5 028	97	1 434	128	1 868
5	5 515	36	14 390	67	3 735	98	3 931	129	1 721
6	6 492	37	9 024	68	2 398	99	4 536	130	4 837
7	9 875	38	5 891	69	3 156	100	7 310	131	2 952
8	6 470	39	3 476	70	1 933	101	2 539	132	1 710
9	2 825	40	3 870	71	18 090	102	6 843	133	7 253
10	8 121	41	2 974	72	1 684	103	1 829	134	4 982
11	1 994	42	5 250	73	7 720	104	3 550	135	7 632
12	6 780	43	5 089	74	2 033	105	3 109	136	10 896
13	5 348	44	2 779	75	1 721	106	2 398	137	9 178
14	1 989	45	4 904	76	6 396	107	3 589	138	8 030
15	13 854	46	7 695	77	2 262	108	4 674	139	3 244
16	2 315	47	3 484	78	3 522	109	2 580	140	5 530
17	6 410	48	4 957	79	2 713	110	4 864	141	12 465
18	7 292	49	2 626	80	3 971	111	2 666	142	3 321
19	4 052	50	8 459	81	2 262	112	11 936	143	4 364
20	3 547	51	3 874	82	2 639	113	2 801	144	15 603
21	2 951	52	3 541	83	1 743	114	4 241	145	7 469
22	6 102	53	2 900	84	6 771	115	1 321	146	3 171
23	2 729	54	2 988	85	13 251	116	6 571	147	7 421
24	8 448	55	2 856	86	3 648	117	22 034	148	11 914
25	4 213	56	5 048	87	2 961	118	8 706	149	2 761
26	2 934	57	3 379	88	3 898	119	11 982	150	14 093
27	2 112	58	11 007	89	3 327	120	1 850	151	26 841
28	6 522	59	2 359	90	4 070	121	13 095	152	13 527
29	2 962	60	2 237	91	9 048	122	14 144	153	9 574
30	2 036	61	5 419	92	2 766	123	6 542	154	17 302
31	4 519	62	8 966	93	3 142	124	24 328		

B)

#	pI
71	4,24
117	4,44
121	4,54
122	4,06
124	4,34
151	4,95

3.2.3 Analyse par spectrométrie de masse des protéines les plus abondantes: profil de la masse des peptides

Les protéines SC69 (122) et SC34 (71) ont été découpées du gel de pH 3,5-4,5 et les protéines SC124 (68), SC117 et SC121 ont été découpées du gel de pH 4,0-5,0. La protéine SC151, même si très abondante, n'a pas été analysée par spectrométrie de masse puisqu'elle se trouve à la limite du front de migration de la deuxième dimension.

3.2.3.1 Protéine SC69 (122)

Suite à l'injection des peptides issus de la digestion tryptique de la protéine SC69 (122) dans le spectromètre, un spectre de masse a été obtenu (Figure 20). Cependant, en comparant l'aspect global du spectre de la digestion tryptique de la protéine SL82 (Figure 13) avec le spectre de la digestion tryptique de la protéine SC69 (122), ce dernier ne ressemble pas à un profil de la masse des peptides. En effet, il n'a pratiquement pas été possible d'observer des pics intenses par rapport au bruit de fond. Cependant, l'intensité (hauteur) des pics est relative au pic le plus intense présent sur le spectre. C'est à dire que la valeur de 100% est attribuée au pic correspondant à l'ion le plus abondant dans l'intervalle du rapport m/z choisi et que la valeur des autres pics présents dans le même intervalle est déterminée par rapport à celui-ci. En changeant d'intervalle du rapport m/z , le pic le plus intense change aussi et par le fait même l'intensité relative des autres pics peut augmenter ou diminuer. Ainsi, le spectre complet, où l'ion 679 est le plus intense, a été visualisé en deux parties (Figure 21) afin d'augmenter l'intensité des pics de hautes masses, soit de 700 à 1120 (Figure 21B). Le pic le plus intense dans la région des hautes masses est devenu l'ion 738,68, avec une d'intensité de 100% et les valeurs des autres pics ont augmenté par rapport à leur valeur dans le spectre complet. Comme les peptides sont des biomolécules de hautes masses, il est plus probable que les pics de masse plus élevée soient des peptides que les pics de masse plus faible. C'est pourquoi, dans la région des basses masses, soit de 500 à 680, l'ion 679 a été conservé comme ion le plus intense.

Dans la région des basses masses, six pics ont été identifiés, soient: 503,42; 523,47; 547,41; 603,21 et 639,52 qui sont chargés une fois et 652,13 qui est chargé deux fois. L'ion 679,56, retrouvé dans le blanc, n'a pas été pris en considération, même s'il est très intense. Dans la région des hautes masses, 13 pics ont été identifiés, soient: 692,56; 717,16; 738,68; 863,37; 968,80 et 1089,53 chargés deux fois et 697,42; 749,55; 1000,34; 1029,85; 1047,92; 1061,86 et 1075,81 chargés 1 fois.

Parmi tous ces ions, seuls les ions 652,13; 692,63; 717,16; 863,37; 968,80; 1000,34 et 1089,53 ont été fragmentés et un spectre de fragmentation typique d'un peptide a été obtenu pour chacun d'eux. Ceux-ci ont donc été retenus pour la recherche par profil de la masse des peptides. Même si les autres ions n'ont pas été fragmentés, ce qui n'a pu confirmer qu'ils étaient vraisemblablement des peptides, ils ont tout de même été retenus pour la recherche par profil de la masse des peptides, sauf les ions 503,42; 523,47; 547,41; 603,21 et 639,52. Étant donné que la masse moléculaire neutre de ces derniers est faible (puisque'ils sont chargés une seule fois et que leur rapport m/z est peu élevé), il est peu probable qu'ils soient des peptides. De plus, ils ont été retrouvés dans une région du spectre où le bruit de fond est très élevé. Les ions retenus (en gras) sont donc les suivants, suivi de la masse neutre de chacun entre parenthèses: 652,13 (1302,26), 692,56 (1383,26), 697,42 (696,42), 717,76 (1433,52), 738,68 (1475,36), 749,55 (748,55), 863,37 (1724,74), 968,80 (1935,60), 1000,34 (999,34), 1029,85 (1028,85), 1047,92 (1046,92), 1061,86 (1060,86), 1075,81 (1074,81) et 1089,53 (2177,06). Le Tableau 13 présente le résultat de la recherche dans les bases de donnée à l'aide de ce profil de la masse des peptides.

Deux protéines possibles ont été trouvées par le logiciel PeptIdent. Les protéines 1 et 2 possèdent des valeurs de masse moléculaire plus élevée que celle de la protéine SC69, soit de 799 kDa et 377 kDa respectivement. Elles ont aussi des valeurs de pI plus grandes que celle de la protéine SC69 (pI de 4,09) soit de 5,29 et 7,62. Même si les deux profils théoriques semblent posséder une bonne homologie avec le profil observé avec 10 peptides et 8 peptides sur 14, les peptides obtenus lors de la digestion tryptique de la protéine SC69 couvrent seulement 2,3% de la protéine 1 et 3,3% de la protéine 2. La

protéine 1 appartient néanmoins au genre *Streptomyces* contrairement à la protéine 2 qui provient du genre *Chlamydia*.

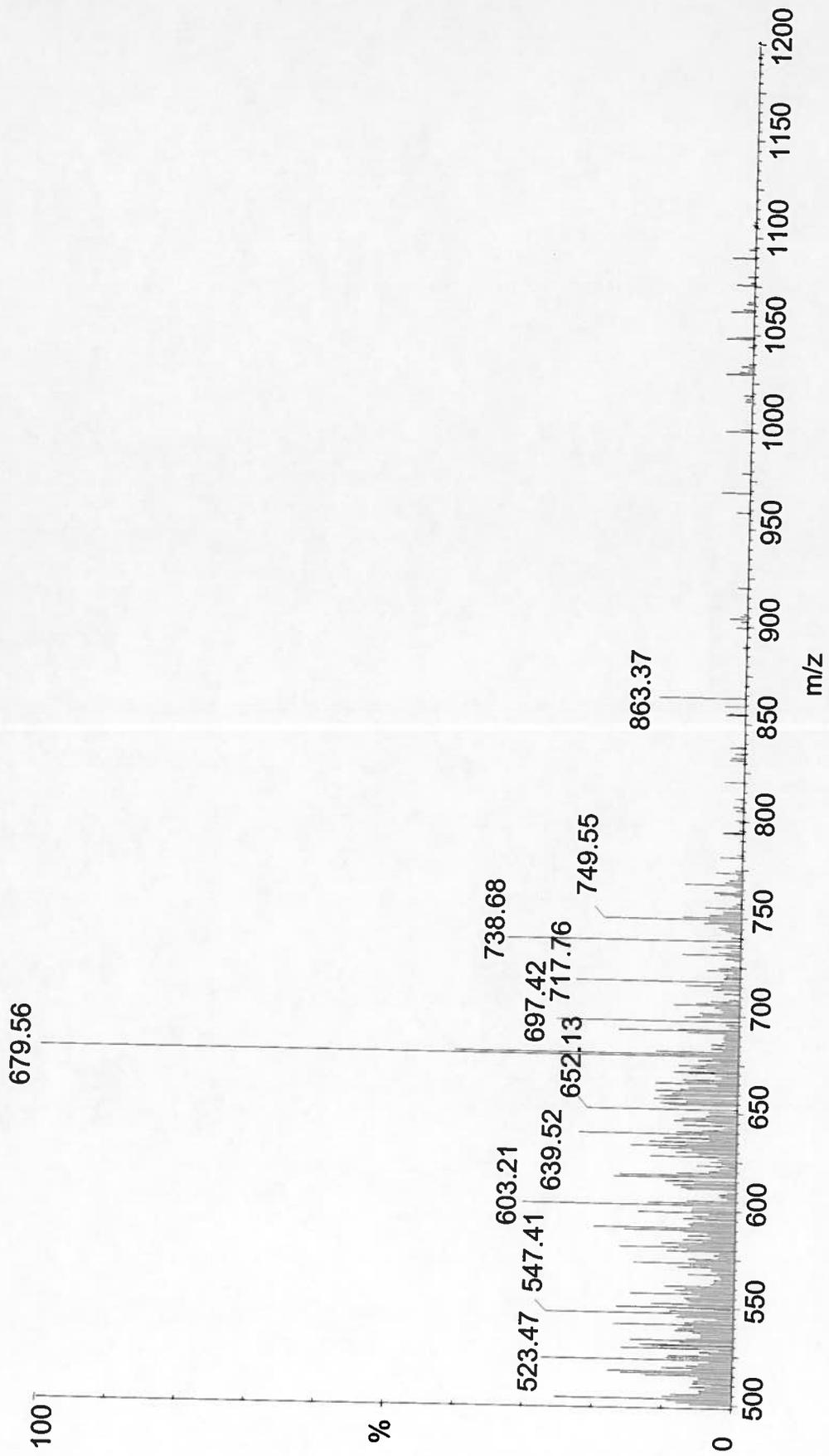


Figure 20. Spectre de masse de la digestion tryptique de la protéine SC69 (122).

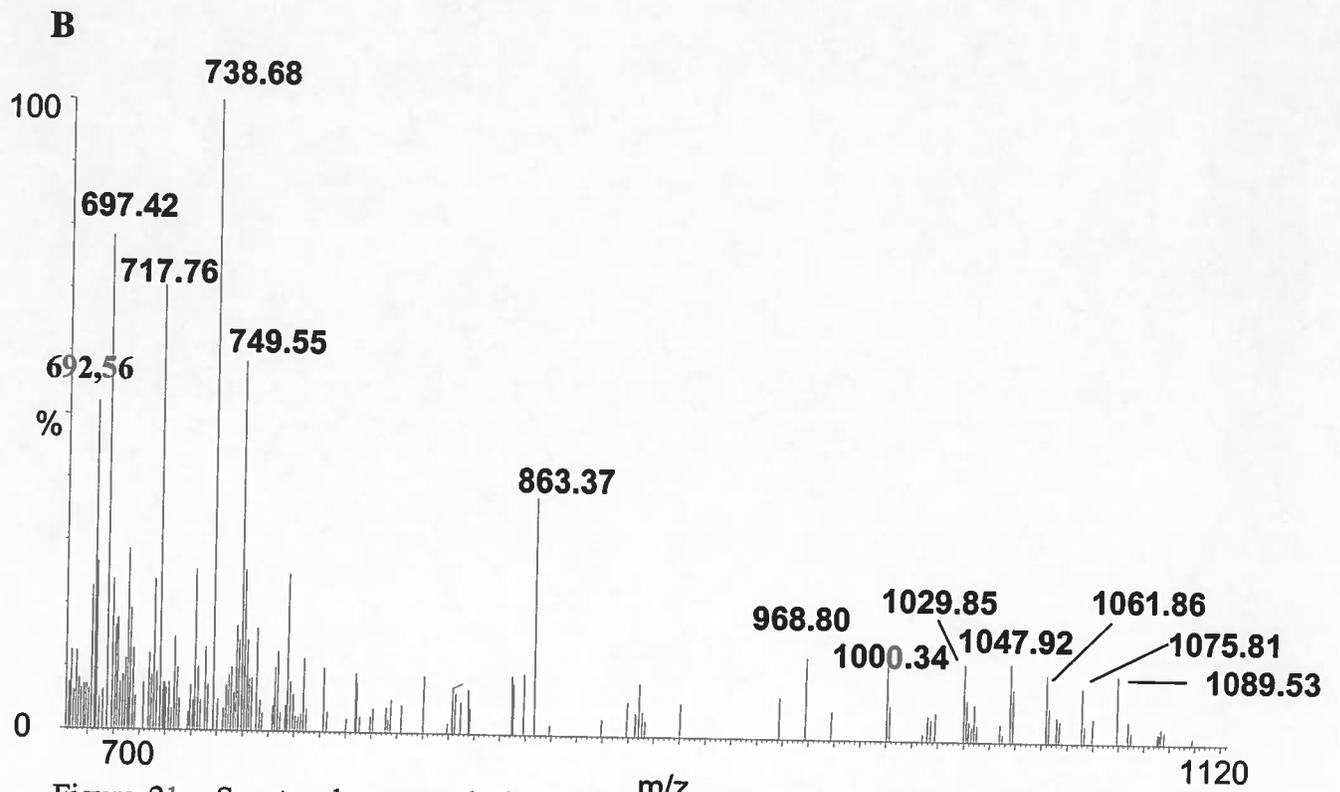
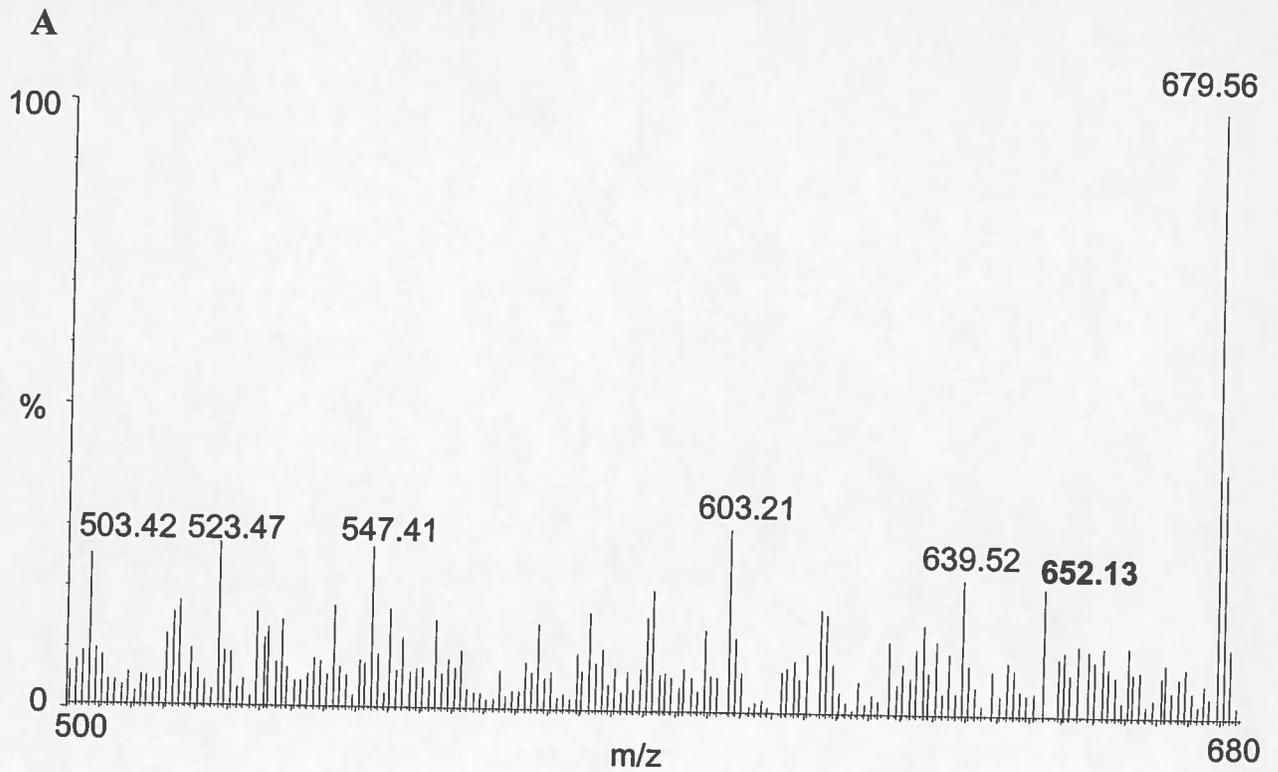


Figure 21. Spectre de masse de la digestion tryptique de la protéine SC69 (122) des valeurs m/z de 500 à 680 (A) et de 700 à 1120 (B).

Tableau 13. Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SC69 (122).

2 matches found.

#	peptide matches	ID	Description	pI	Mw	% of sequence covered
1	<u>10</u>	Q9Z4X6	CDA PEPTIDE SYNTHETASE I. - <i>Streptomyces coelicolor</i> .	5.29	798627.33	2.3
2	<u>8</u>	Q9PKM7	ADHERENCE FACTOR. - <i>Chlamydia muridarum</i> .	7.62	376518.38	3.3

3.2.3.2 Protéine SC34 (71)

À nouveau, le spectre de la digestion tryptique de la protéine SC34 (71) ne ressemble pas à un profil de la masse des peptides (Figure 22) puisqu'il n'a pas été possible d'observer des pics intenses. Néanmoins, afin de mieux évaluer ces pics de faible intensité, le spectre complet a été visualisé en deux parties (Figure 23). Cependant, après comparaison avec le spectre en deux parties de la digestion tryptique de la protéine SC69 (122), peu d'ions différents de ceux présents dans ce spectre ont pu être identifiés. En effet, quatre pics nouveaux (en gras) ont été observés dans le spectre de la digestion tryptique de la protéine SC34 (71), soit: 552,98; 797,71; 911,82 et 1005,98. Ces ions ont tous été fragmentés, mais seulement les ions 911,82 et 1005,98 ont présenté un spectre de fragmentation typique d'un peptide. Ainsi, aucune recherche dans les bases de données à l'aide du profil de la masse des peptides n'a pu être effectuée puisqu'un nombre insuffisant d'ions ont été obtenus.

3.2.3.3 Protéine SC124 (68)

Cette fois-ci, le spectre de la digestion tryptique de la protéine SC124 (68) présente plusieurs pics intenses (en gras, Figure 24), soit les 14 ions suivants, suivi de leur charge et de leur masse neutre entre parenthèses: 459,35 (+1, 458,35), 468,49 (+1, 467,49), 591,37 (+1, 590,37), 607,10 (+2, 1212,20), 709,69 (+3, 2126,07), 738,44 (+2, 1474,88), 762,20 (+2, 1522,40), 784,81 (+2, 1567,62), 790,77 (+3, 2369,31), 794,94 (+2, 1587,88), 900,38 (+2, 1798,76), 1013,05 (+2, 2024,10), 1064,09 (+2, 2126,18) et 1185,74 (+3, 2369,48). Il est à remarquer que les ions 709,69 et 1064,09 correspondent en fait au même peptide d'une masse moléculaire neutre de 2126 Da et que les ions 790,77 et 1185,74 correspondent au même peptide d'une masse moléculaire neutre de 2369 Da. Il a tout simplement été possible de détecter ces peptides chargés deux fois et chargés trois fois lors de l'analyse. Les ions 503, 523 et 547 n'ont pas été retenus puisqu'ils se retrouvent aussi dans les deux derniers spectres (Figures 21 et 23), ni l'ion 679 puisqu'il est retrouvé dans le blanc. De plus, les ions 607,10; 762,20; 790,77; 900,38; 1013,05; 1064,09 et 1185,74 ont été fragmentés et tous ont présentés un spectre de fragmentation

typique d'un peptide. Le résultat de la recherche dans les bases de données obtenu à l'aide du profil de la masse des peptides obtenu est présenté au Tableau 14.

Trois protéines possibles ont été trouvées par le logiciel PeptIdent. Les protéines 2 et 3 ont des valeurs de pI plus élevées que celle de la protéine SC124 (pI de 4,34) soit de 5,29 et 8,83 respectivement. Elles possèdent également des valeurs de masse moléculaire plus grandes que celle de la protéine SC124 soit de 799 kDa et de 117 kDa. De plus les peptides obtenus lors de la digestion tryptique de la protéine SC124 couvrent seulement 1,6% de la protéine 2 et 7,6% de la protéine 3. Par contre, la protéine 1 possède des valeurs de pI et de masse moléculaire très similaire à celles de la protéine SC124 soit de 4,49 pour le pI et de 20,3 kDa pour la masse moléculaire. De plus, 70,7% de cette protéine est couvert par les peptides obtenus lors de la digestion tryptique de la protéine SC124 et elle appartient au genre *Streptomyces*. La protéine 2 appartient aussi à ce genre mais il est à remarquer que cette protéine a aussi été trouvée comme protéine possible lors de la recherche dans les bases de données avec le profil de la masse des peptides de la protéine SC69. Par contre, la protéine 3 appartient au genre *Mycoplasma*.

Dans le Tableau 15, le profil de la masse des peptides de la protéine SC124 est comparé avec le profil théorique de la protéine 1, une protéine hypothétique de 20,3 kDa de *S. coelicolor*. La plus grande différence de masse entre un peptide théorique et un peptide appartenant à la protéine SC124 est de 0,321. Tous les peptides théoriques correspondent à des peptides qui ne contiennent aucun site de clivage manqué par trypsine. Les ions qui ne concordent pas avec la protéine 1 sont 459,35; 468,49; 591,37 et 738,44. La séquence en acides aminés des peptides de la protéine 1 qui concordent avec les peptides obtenus est fournie et leur position sur la protéine est indiquée en rouge et en lettre majuscule. Ainsi, les peptides sont répartis sur toute la protéine 1.

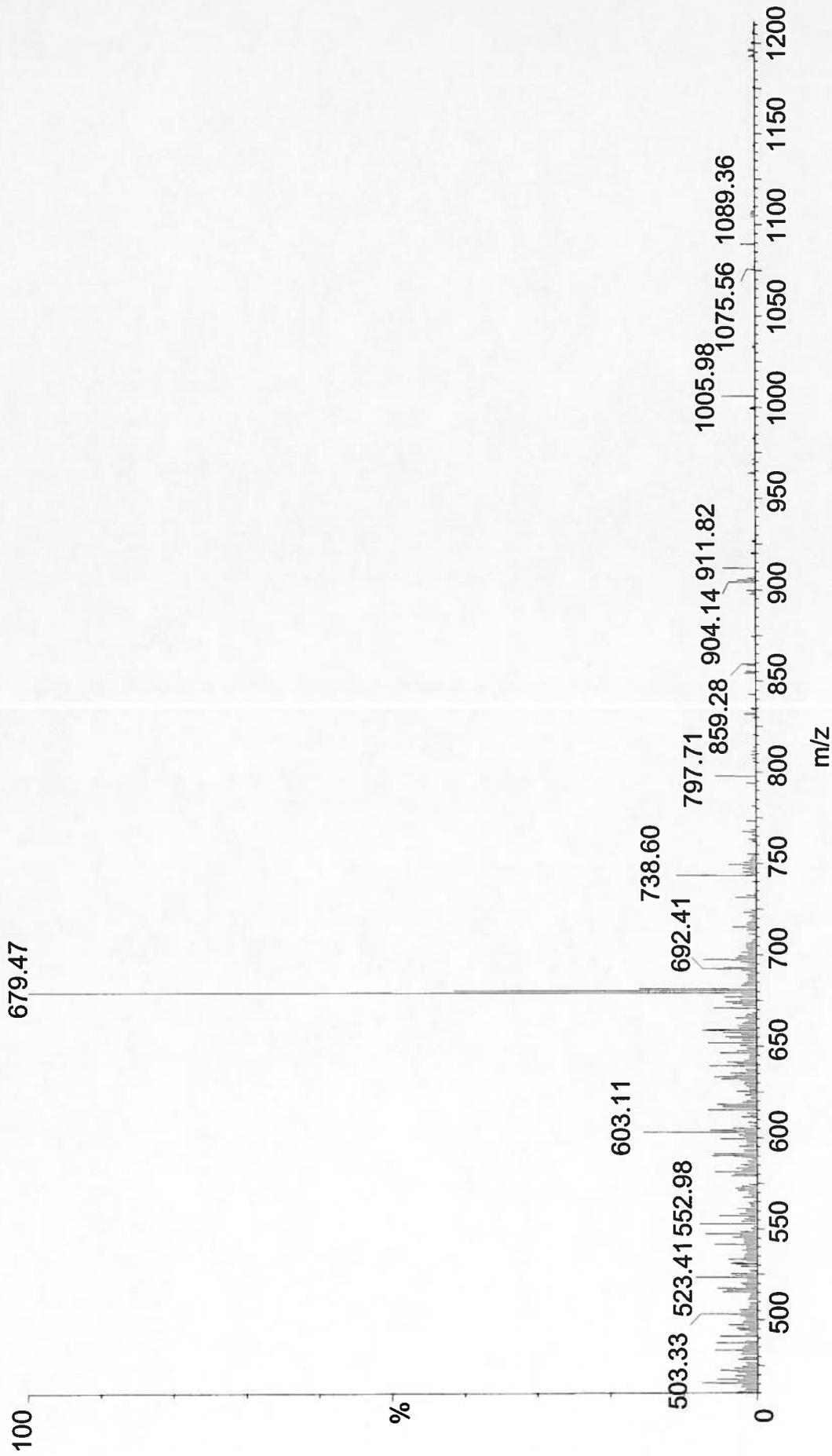


Figure 22. Spectre de masse de la digestion tryptique de la protéine SC34 (71).

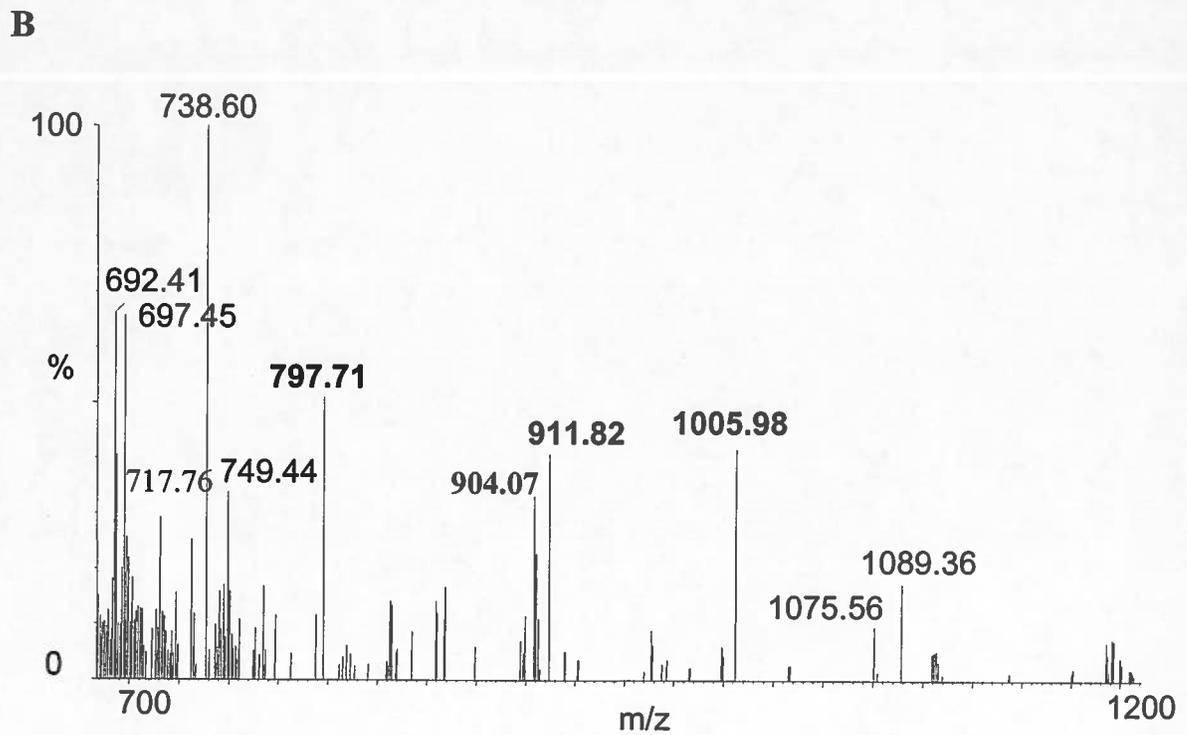
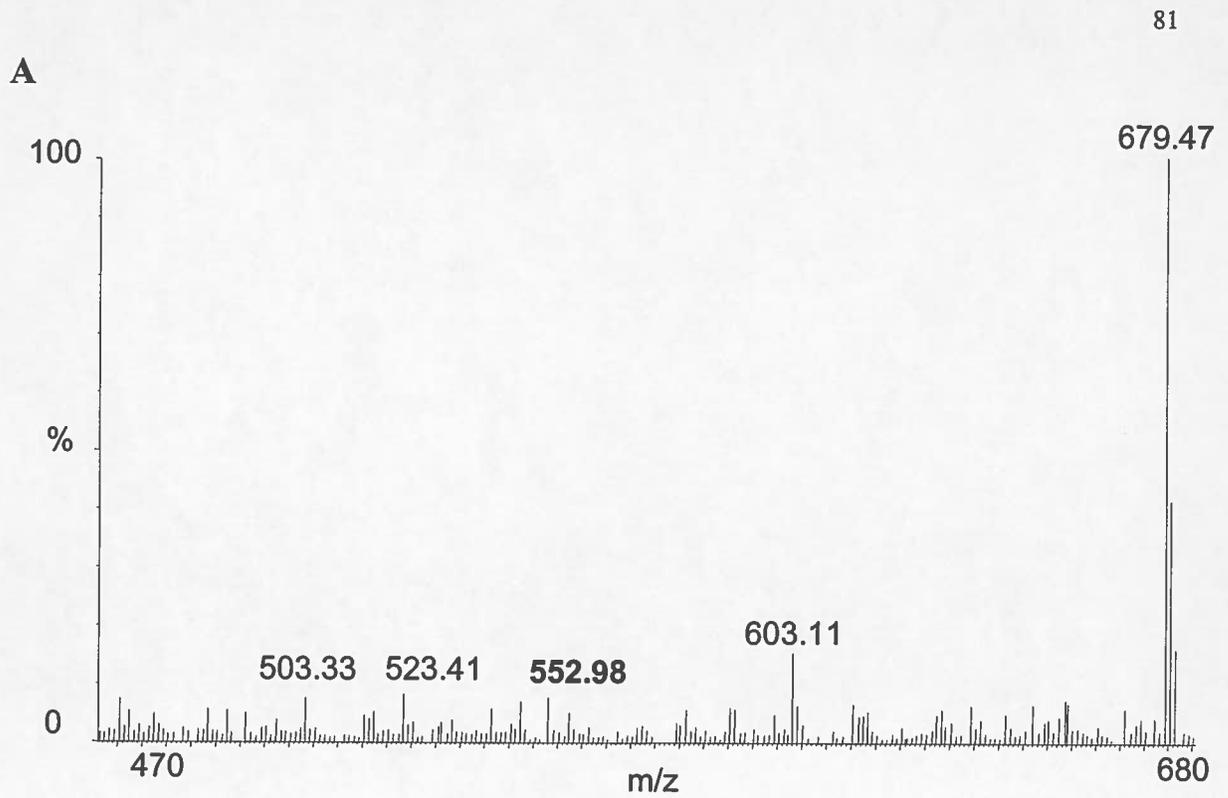


Figure 23. Spectre de masse de la digestion tryptique de la protéine SC34 (71) pour les valeurs des rapports m/z de 470 à 680 (A) et de 700 à 1200 (B).

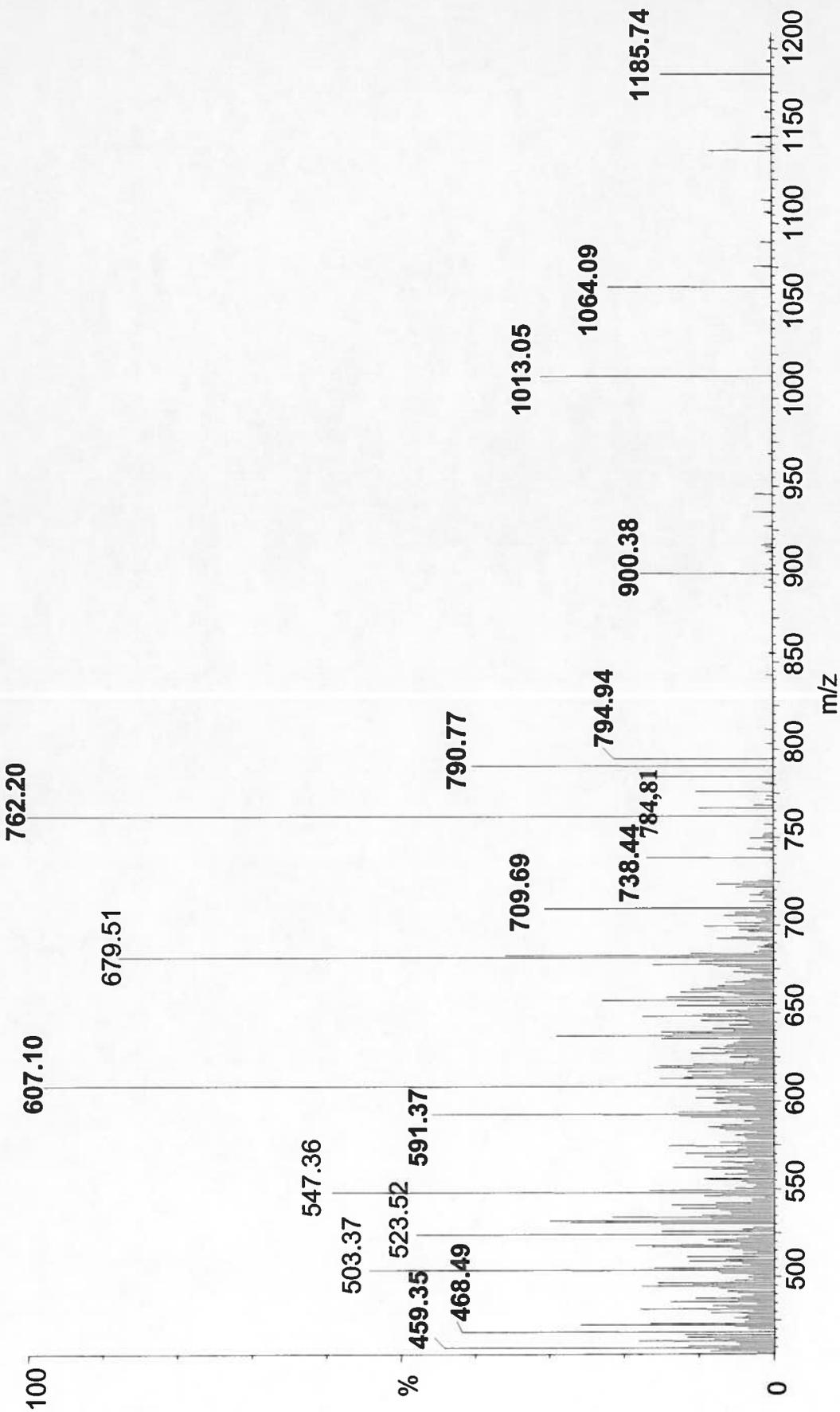


Figure 24. Spectre de masse de la digestion tryptique de la protéine SC124 (68).

Tableau 14. Résultat de la recherche dans les bases de données avec le logiciel Peptide à l'aide du profil de la masse des peptides de la protéine SC124 (68).

3 matches found.

#	peptide matches	ID	Description	pI	Mw	% of sequence covered
1	8	<u>Q9KXW4</u>	HYPOTHETICAL 20.3 KDA PROTEIN. - <i>Streptomyces coelicolor</i> .	4.49	20344.40	70.7
2	8	<u>Q9Z4X6</u>	CDA PEPTIDE SYNTHETASE I. - <i>Streptomyces coelicolor</i> .	5.29	798627.33	1.6
3	7	<u>Q50357</u>	RESTRICTION-MODIFICATION ENZYME SUBUNIT R1. - <i>Mycoplasma pulmonis</i> .	8.83	117357.61	7.2

Tableau 15. Description détaillée de la comparaison entre le profil de la protéine 1 versus le profil de la protéine SC124 (68).

Score: 0.62, 8 matching peptides: Q9KXW4 (Q9KXW4) pI: 4.49, Mw: 20344.40 HYPOTHETICAL 20.3 KDA PROTEIN. - <i>Streptomyces coelicolor</i> .					
user mass	matching mass	mass (Da)	#MC	position	peptide
1212.2	1212.371	0.1712	0	131-142	ILNQAGGAEIAR
2126.07	2126.262	0.1916	0	33-53	TTTGTDFDLASAIAVNTQ GK
1522.4	1522.721	0.321	0	105-117	IVFPVSIYDAENR
1567.62	1567.635	0.0153	0	67-80	QTPDNTIVHTGDNR
1587.82	1587.755	-0.0652	0	54-66	VYSDAHFVFFNNK
1798.76	1799.014	0.2536	0	16-32	EAPGLTEVTVGLGWDVR
2024.1	2024.218	0.1175	0	171-191	AVGQGYASGLTGIAQDFGVNV
2369.48	2369.524	0.0441	0	81-104	TGEGAGDDEAINVNLAGLPADIEK

70.7% of sequence covered:

mavslskggnvsltk **EAPGLTEVTVGLGWDVR TTTGTDFDLASAIAVNTQ GK**
VYSDAHFVFFNNK QTPDNTIVHTGDNR TGEGAGDDEAINVNLAGLPADIEK
IVFPVSIYDAENR sqnfgqvrnayir **ILNQAGGAEIAR** ydlsedaatetamvfgelyrngawkfr
AVGQGYASGLTGIAQDFGVNV

3.2.3.4 Protéine SC117

Le spectre de masse de la digestion tryptique de la protéine SC117 (Figure 25) ressemble aussi à un profil de la masse des peptides sauf qu'il n'a pas été possible d'identifier des pics différents de ceux retrouvés dans le spectre de la digestion tryptique de la protéine SC124 (68). En fait, les deux spectres sont pratiquement identiques, sauf les ions 709,69 et 790,77 qui sont présents mais beaucoup moins abondants dans le spectre de la protéine SC117. Par contre, l'ion 1064,17, qui correspond au même peptide que l'ion 709,69 mais chargé 2 fois, est présent. Ainsi, aucune recherche dans les bases de données n'a été effectuée puisque le même résultat obtenu avec la protéine SC124 (68) est attendu.

3.2.3.5 Protéine SC121

Suite à l'obtention de deux spectres identiques pour les protéines SC124 (68) et SC117 (qui ont pratiquement la même masse moléculaire mais des pI différents, Figure 20), la protéine SC121 (qui possède sensiblement la même masse et qui se retrouve à proximité de ces deux protéines) a aussi été analysée par spectrométrie de masse afin de voir si le même spectre serait à nouveau obtenu. Cependant, moins de peptides intenses ont été observés dans le spectre de la digestion tryptique de la protéine SC121 (Figure 26). Par contre, la plupart étaient différents de ceux retrouvés dans les spectres de masse des protéines SC124 (68) et SC117. Ainsi, les ions suivants ont été retenus pour l'analyse à l'aide du profil de la masse des peptides, suivi de leur charge et de leur masse neutre entre parenthèses: 629,14 (+2, 1256,30), 642,66 (+2, 1283,32), 655,12 (+1,654,12), 657,07 (+1, 656,07), 699,65 (+3, 2095,98), 738,47 (+2, 1474,92), 762,32 (+2, 1522,60), 812,43 (+2, 1622,86), 863,15 (+3, 2586,45) et 1049,59 (+2, 2096,78). Le résultat de la recherche dans les bases de données obtenu à l'aide du profil de la masse des peptides obtenu est présenté au Tableau 16.

Trois protéines possibles ont été trouvées par le logiciel PeptIdent. Les protéines 1 et 3 ont des valeurs de pI plus élevées que celle de la protéine SC121 (pI de 4,54) soit

de 5,59 et 5,49 respectivement. Elles possèdent également des valeurs de masse moléculaire plus grandes que celle de la protéine SC121 soit de 546 kDa et de 522 kDa. De plus les peptides obtenus lors de la digestion tryptique de la protéine SC121 couvrent seulement 1,8% de la protéine 1 et 1,8% de la protéine 3. Par contre, la protéine 2 possède des valeurs de pI et de masse moléculaire très similaires à celles de la protéine SC124 soit de 4,56 pour le pI et de 20,4 kDa pour la masse moléculaire. De plus, 50,8% de cette protéine est couvert par les peptides obtenus lors de la digestion tryptique de la protéine SC121 et elle appartient au genre *Streptomyces*. La protéine 3 appartient aussi à ce genre alors que la protéine 1 appartient au genre *Bacillus*.

Dans le Tableau 17, le profil de la masse des peptides de la protéine SC121 est comparé avec le profil théorique de la protéine 2, une protéine hypothétique conservé de 20,4 kDa de *S. coelicolor*. La plus grande différence de masse entre un peptide théorique et un peptide appartenant à la protéine SC121 est de 0,1675. Tous les peptides théoriques correspondent à des peptides qui ne contiennent aucun site de clivage manqué par trypsine. Les ions qui ne concordent pas avec la protéine 2 sont 655,12; 657,07; 699,65 et 738,47. La séquence en acides aminés des peptides de la protéine 2 qui concordent avec les peptides obtenus est fournie et leur position sur la protéine est indiquée en rouge et en lettre majuscule. Ainsi, les peptides sont répartis sur toute la protéine 2.

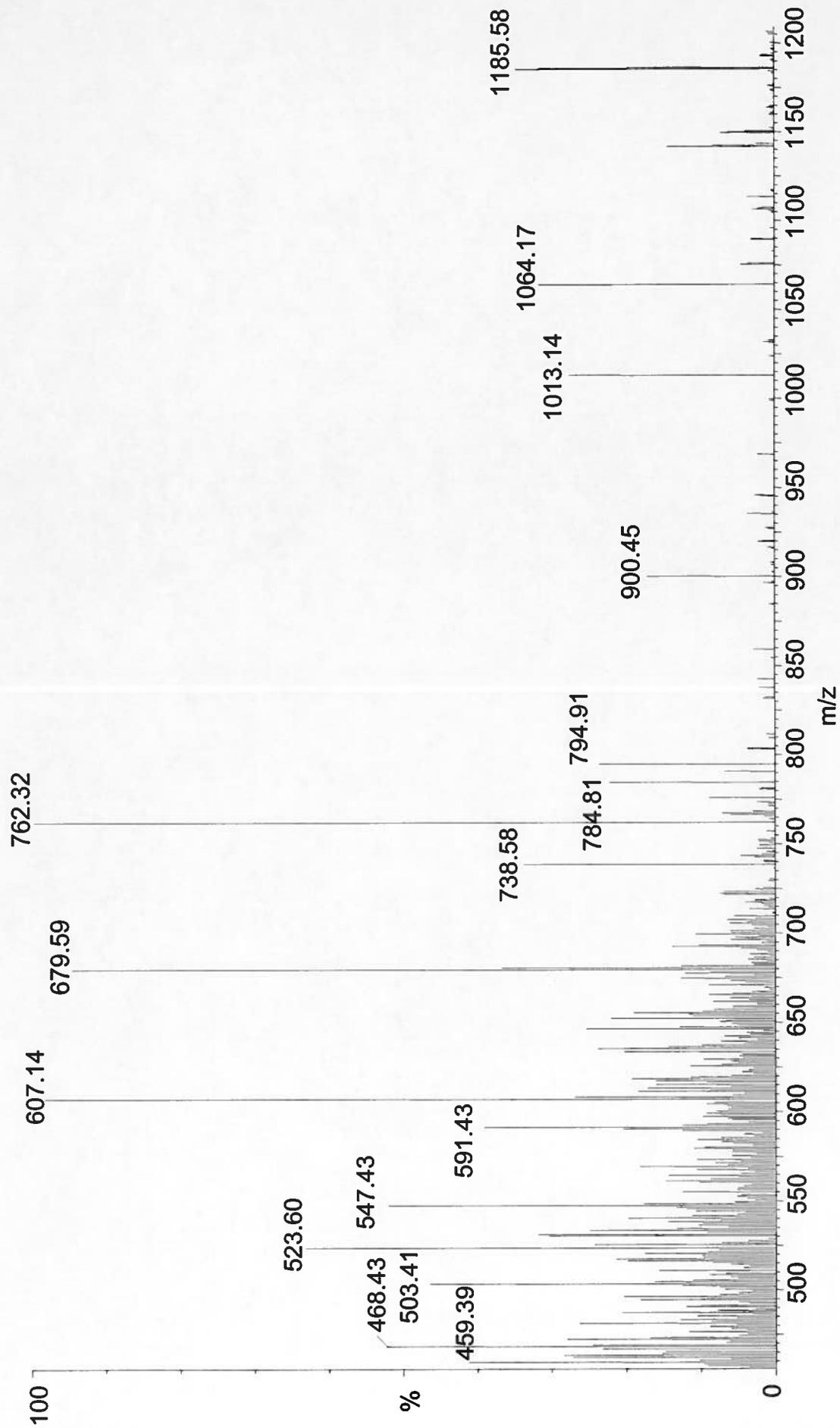


Figure 25. Spectre de masse de la digestion tryptique de la protéine SC117.

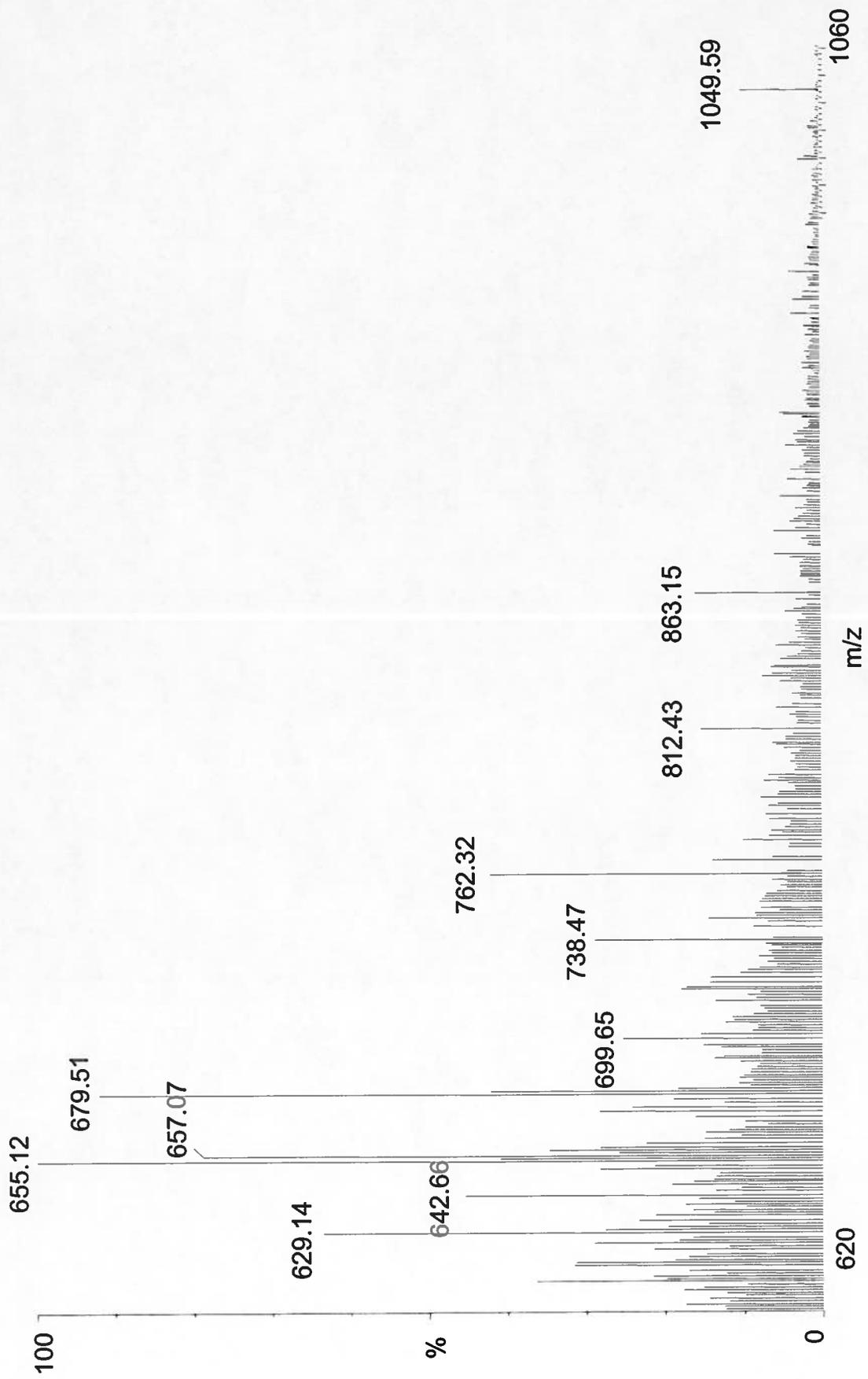


Figure 26. Spectre de masse de la digestion tryptique de la protéine SC121.

Tableau 16. Résultat de la recherche dans les bases de données avec le logiciel PeptIdent à l'aide du profil de la masse des peptides de la protéine SC121.

3 matches found.

#	peptide matches	ID	Description	pI	Mw	% of sequence covered
1	6	O31783	POLYKETIDE SYNTHASE OF TYPE I. - <i>Bacillus subtilis</i> .	5.59	545720.42	1.8
2	6	Q9KY22	CONSERVED HYPOTHETICAL PROTEIN. - <i>Streptomyces coelicolor</i> .	4.56	20387.6	50.8
3	6	O07944	PRISTINAMYCIN I SYNTHASE 3 AND 4. - <i>Streptomyces pristinaespiralis</i> .	5.49	522087.38	1.8

Tableau 17. Description détaillée de la comparaison entre le profil de la protéine 2 versus le profil de la protéine SC121.

Score: 0.60, 6 matching peptides: Q9KY22 (Q9KY22) pI: 4.56, Mw: 20387.60					
CONSERVED HYPOTHETICAL PROTEIN. - <i>Streptomyces coelicolor</i> .					
user mass	matching mass	mass (Da)	#MC	position	peptide
1256.28	1256.381	0.1012	0	131-142	VVNQAGEAEIAR
1283.32	1283.488	0.1675	0	93-104	INLATVPADIEK
1522.64	1522.721	0.081	0	105-117	IVFPVSIYDAENR
1622.86	1622.844	-0.0162	0	54-67	VASDAHFIFFNNLK
2586.45	2586.576	0.1256	0	68-92	SPDGSVEHTGDNITGEGEGDDEQIK
2097.18	2097.263	0.0833	0	33-53	TTGTDFDLASALLNSGGK
50.8% of sequence covered:					
mgvslskggnvsltkeapgltavivglwdir TTGTDFDLASALLNSGGK VASDAHFIFFNNLK SPDGSVEHTGDNITGEGEGDDEQIK INLATVPADIEK IVFPVSIYDAENR qqsfgqvrnafir VVNQAGEAEIAR ydlsedastetam vfgelyrhgaewkfraigqgyasglrgiaqdfgvnv					

3.2.4 Analyse des peptides par spectrométrie de masse: fragmentation et recherche de la séquence en acides aminés

3.2.4.1 Protéine 124 (68)

Seule la fragmentation des ions 607,10 et 762,20 a permis de générer un spectre où une séquence continue a pu être obtenue (Figure 27 et 28). Ainsi, pour le spectre de fragmentation de l'ion d'un rapport m/z de 607,10 (1212,2) la séquence recherchée par le logiciel Biolynx était ILNQAGGAEIAR. Ce dernier, à partir des pics les plus intenses a trouvé une partie de cette séquence, soit LNQAGGA, en lisant de la droite vers la gauche. La séquence recherchée dans le spectre de fragmentation de l'ion d'un rapport m/z de 762,20 (1522,4) était IVFPVSIYDAENR. Le logiciel a été en mesure de trouver la séquence partielle suivante, soit VFPVSIY, en lisant de la droite vers la gauche.

3.2.4.2 Protéine 121

Seule la fragmentation des ions 629,14 et 762,32 a permis de générer un spectre où une séquence continue a pu être obtenue (Figure 29 et 30). Ainsi, pour le spectre de fragmentation de l'ion d'un rapport m/z de 762,32 (1522,64) était IVFPVSIYDAENR. Le logiciel a été en mesure de trouver la séquence partielle suivante, soit VFPVSIY, en lisant de la droite vers la gauche. La séquence recherchée dans le spectre de fragmentation de l'ion d'un rapport m/z de 629,14 (1256,28) la séquence recherchée par le logiciel Biolynx était VVNQAGEAEIAR. Ce dernier, à partir des pics les plus intenses a trouvé une partie de cette séquence soit VNQAGE, en lisant de la droite vers la gauche.

	Ile	Leu	Asn	Gln	Ala	Gly	Gly	Ala	Glu	Ile	Ala	Arg
y	1213.38	1100.22	987.06	872.96	744.83	673.75	616.70	559.64	488.56	359.45	246.29	175.21
	--	0.65	0.41	0.31	0.22	0.57	-0.15	1.04	0.06	--	--	--

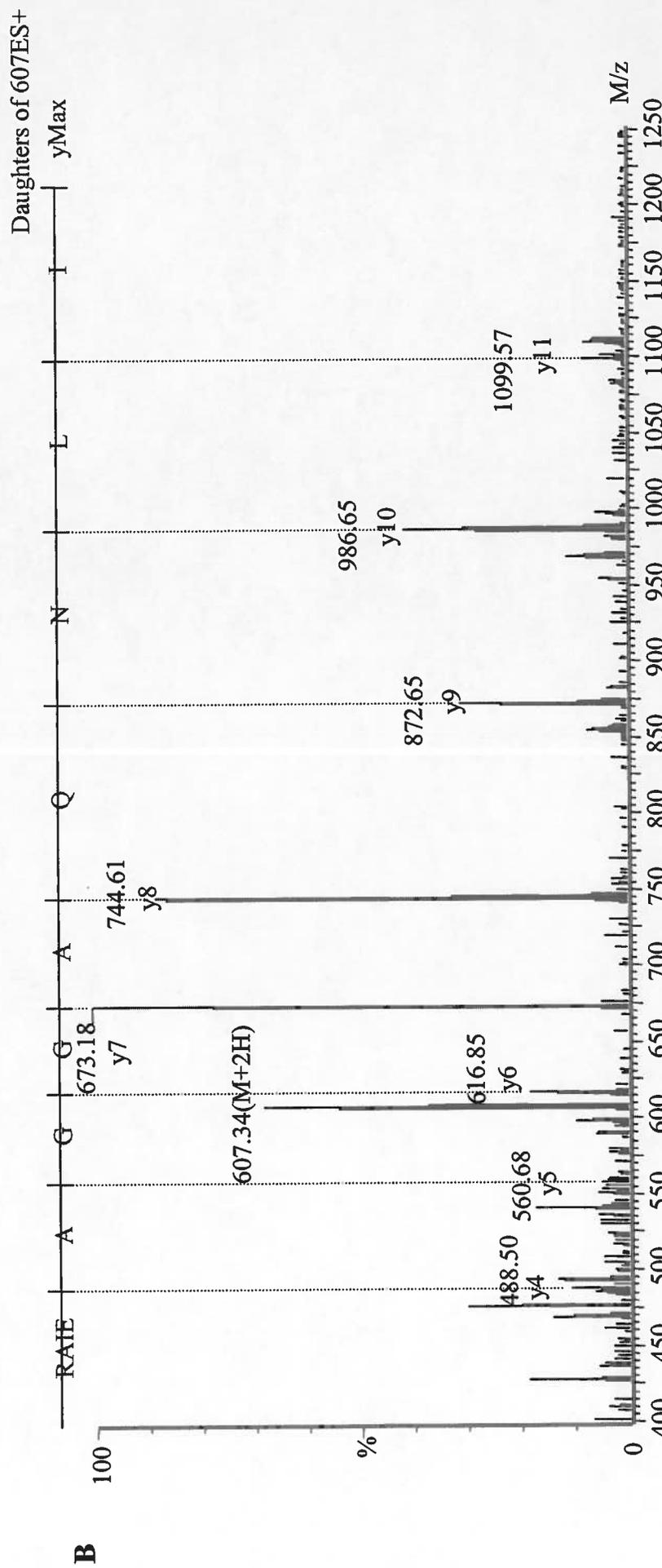
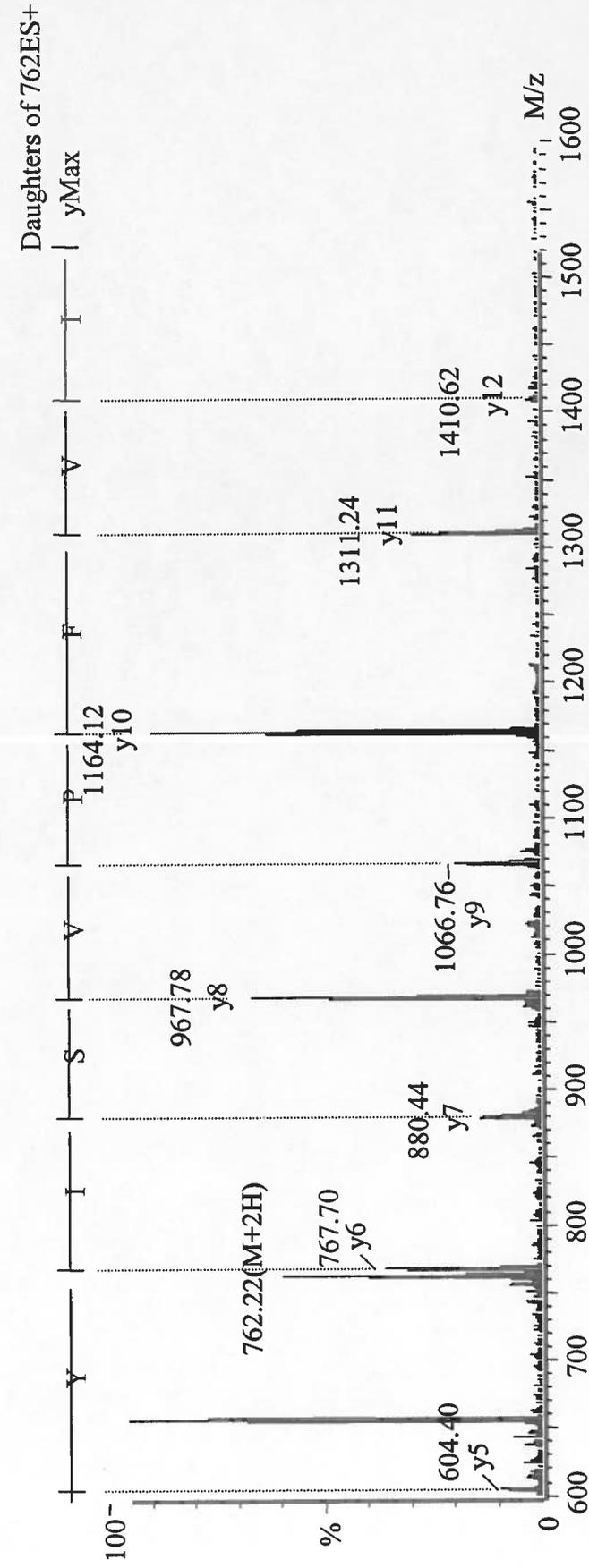


Figure 27. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 607 obtenu avec une énergie de collision de 40 eV.

	Ile	Val	Phe	Pro	Val	Ser	Ile	Tyr	Asp	Ala	Glu	Asn	Arg
y	1523.73	1410.57	1311.44	1164.26	1067.14	968.01	880.93	767.77	604.60	489.51	418.43	289.31	175.21
	---	-0.05	0.20	0.14	0.38	0.23	0.49	0.07	0.20	---	---	---	---

A



B

Figure 28. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 762 obtenu avec une énergie de collision de 40 eV.

	Ile	Val	Phe	Pro	Val	Ser	Ile	Tyr	Asp	Ala	Glu	Asn	Arg
y	1523.73	1410.57	1311.44	1164.26	1067.14	968.01	880.93	767.77	604.60	489.51	418.43	289.31	---
	---	-0.36	0.38	0.47	0.66	0.49	0.57	-0.24	0.06	---	---	---	---

B

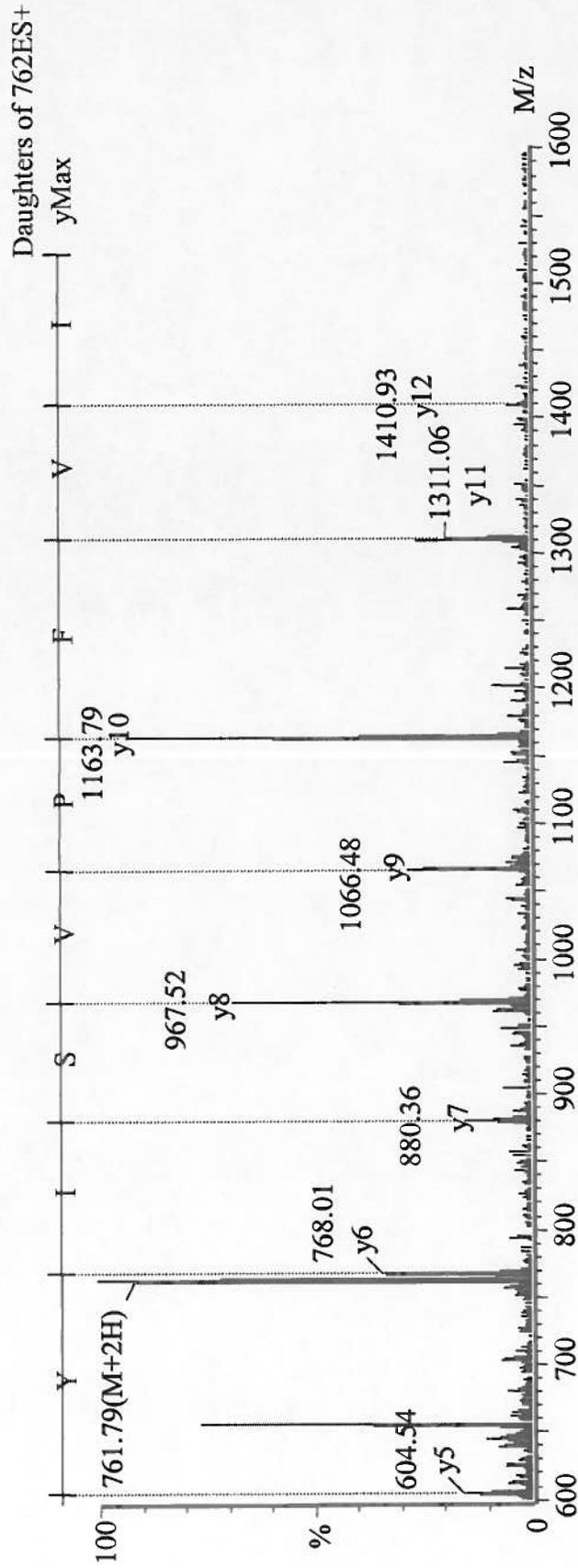


Figure 29. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 762 obtenu avec une énergie de collision de 44 eV.

	Val	Val	Asn	Gln	Ala	Gly	Glu	Ala	Glu	Ile	Ala	Arg
y	1257.39	1158.26	1059.12	945.02	816.89	745.81	688.76	559.64	488.56	359.45	246.29	175.21
	---	0.13	0.47	0.65	0.02	0.32	0.53	0.02	---	---	---	---

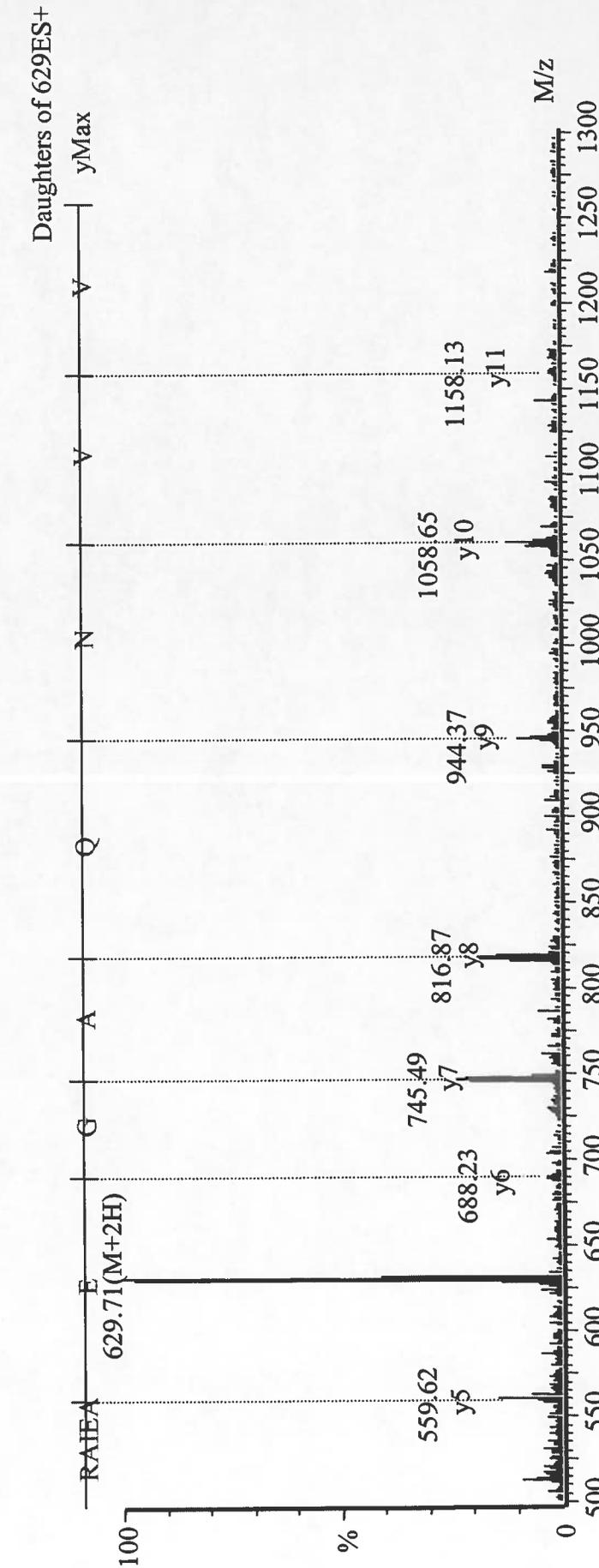


Figure 30. A) Séquence à retrouver dans le spectre de masse, rapport m/z théorique des ions filles et la différence de masse entre le rapport m/z théorique et celui obtenu. B) Spectre de masse de la fragmentation de l'ion d'un rapport m/z de 629 obtenu avec une énergie de collision de 45eV.

4. DISCUSSION

4.1 MISE AU POINT DES ÉTAPES D'UNE ANALYSE PROTÉOMIQUE

4.1.1 L'électrophorèse en deux dimensions et la détection des protéines

Il a été possible de réaliser et d'optimiser l'électrophorèse en 2-D puisque pour différents gradients de pH, soit 4-7, 3,5-4,5 et 4-5, les protéines ont été séparées et résolues (Figures 11, 18 et 19). De plus, les stries horizontales dans la partie acide des gels ont été diminuées par l'utilisation des languettes de 18 cm à gradient de pH étroit (Figures 18 et 19) plutôt que celles de 13 cm (Figure 11). Comme les stries n'occupent maintenant qu'une partie minimale du gel et que les protéines sont résolues sur toute sa surface, l'aspect du gel a été grandement améliorée.

Les paramètres utilisés se sont aussi avérés très reproductibles. Tout d'abord, la comparaison des gels en 2-D des protéines cytoplasmiques des souches *S. lividans* 1326 et IAF6 a permis de constater que le patron général des protéines se retrouvait dans chacun des gels (cercles blancs, Figure 11). Ensuite, il a été possible d'observer que les protéines migrent toujours à leur pI respectif dans des gradients de pH différents. En effet, les protéines 34, 68, 69 et 73 (gel de pH 3,5-4,5, Figure 18) qui correspondent aux protéines 71, 124, 122 et 117 (gel pH 4-5, Figure 19) se retrouvent pratiquement à la même valeur de pI dans les deux gels. Ceci a aussi été vérifié lors de l'estimation du pI par le logiciel ImageMaster (Tableaux 11 et 12).

La reproductibilité permet d'entreprendre des analyses protéomiques comparatives puisqu'il est maintenant possible d'observer les différences réelles entre les gels. En effet, même sans une analyse approfondie avec le logiciel ImageMaster, des différences entre les patrons de protéines cytoplasmiques de la souche *S. lividans* 1326 et IAF6 ont été observées (flèches, Figure 11).

Le seul point à améliorer concernant les gels en 2-D est la capacité de chargement. Même si celle-ci a été augmentée lors de l'utilisation de languettes plus longues (500 ug

pour celles de 18 cm), il est possible, selon le manufacturier, de charger des milligrammes de protéines. Cependant lorsqu'une telle quantité était incorporée, la focalisation isoélectrique n'était jamais complète et des stries horizontales sur toute la largeur du gel étaient obtenues. Le nombre de volt-heures a été augmenté jusqu'à 100 000 Vh, qui est la valeur limite pour l'extrafocalisation (Berkelman et Stenstedt, p.47) mais sans succès. En effet, des périodes de focalisation prolongées peuvent provoquer le déplacement de l'eau et des protéines par électroendosmose et ce phénomène peut être la cause des stries horizontales. Cependant certains auteurs rapportent une focalisation jusqu'à 150-200 kVh sans qu'apparaisse l'extrafocalisation (Morris et Djordjevic, 2001 et Vallorani *et al.*, 2000). Il serait donc avantageux d'incorporer une plus grande quantité de protéines dans chaque languette et d'essayer ces conditions de focalisation isoélectrique. En effet, jusqu'à maintenant la combinaison de la digestion tryptique de protéines provenant de quatre gels identiques était nécessaire afin d'obtenir un signal acceptable en spectrométrie de masse.

4.1.2 Analyse des peptides par spectrométrie de masse: profil de la masse des peptides

Des spectres de masse ont été obtenus après l'injection de la digestion tryptique de protéines inconnues séparées par électrophorèse en 2-D (Figures 13, 24, 25 et 26). Pour chacun de ces spectres, des pics intenses ont été observés. De plus, un nombre suffisant d'ions (10 et plus) a été obtenu pour ces mêmes spectres afin d'effectuer la recherche dans les bases de données à l'aide du profil de la masse des peptides. Cependant, la présence en forte concentration dans tous les spectres de masse de l'ion d'un rapport m/z de 679 est problématique. Celui-ci semble être un contaminant et non un peptide puisque aucune fragmentation typique d'un peptide n'a été observée, même en augmentant l'énergie de collision à 55 eV (Figure 17). En effet, certains pics sont apparus en appliquant une telle énergie mais il est peu probable qu'ils résultent de l'accumulation d'un fragment (ion fille) provenant de la fragmentation du peptide sélectionné (ion parent). La possibilité que cet ion soit en fait un peptide provenant de l'autodigestion de la trypsine est donc éliminée. Étant donné que tous les échantillons de peptides sont

purifiés à l'aide des embouts ZipTip™, il est peu probable que le contaminant provienne des étapes préalables à la purification, soit l'électrophorèse en 2-D, la coloration au nitrate d'argent ou la digestion tryptique. Par contre, le contaminant est peut-être présent dans l'une des trois solutions utilisées avec les ZipTip™, même si celles-ci sont composées d'acétonitrile (à 50% et 80%) et de TFA (0,1%) de qualité supérieure. L'acétonitrile est de qualité "HPLC grade" et le TFA est de qualité "peptide sequencing grade". Il est possible que l'acétonitrile soit suffisamment débarrassée de tous les contaminants pour un tel appareil, mais cette pureté n'est pas suffisante pour un spectromètre de masse étant donné sa sensibilité. L'ion d'un rapport m/z de 679 pourrait donc être présent dans l'acétonitrile. Une qualité supérieure d'acétonitrile ou provenant d'une autre compagnie pourrait être testée de même que le TFA. Il est important d'éliminer ce contaminant, puisque étant présent en très grande quantité, il s'accapare d'une grande partie des charges lors de l'ionisation de l'échantillon. Ainsi, des peptides en plus faible proportion ne sont pas ionisés et par le fait même ne sont pas détectés par le spectromètre de masse.

Finalement, la charge des peptides a pu être déterminée puisqu'une bonne résolution des ions moléculaires avec les ions isotopiques a été obtenue.

4.1.3 Recherche dans les bases de données avec le logiciel PeptIdent

Le logiciel ainsi que les paramètres utilisés lors de la recherche ont permis d'identifier une première protéine de façon certaine, soit la protéine SL82, une méthylase qui confère la résistance au thiostrepton. Les autres protéines possibles identifiées par le logiciel (Tableau 9) ont été éliminées parce que soit les valeurs de pI (protéines 4 et 8), soit les valeurs de masse moléculaire (protéines 2, 3, 5, 6, 7 et 8) ne concordaient pas avec celles de la protéine SL82. Ainsi, la seule protéine dont le pI et la masse moléculaire concordaient était la protéine 1. Afin de confirmer l'identification de la protéine, d'autres critères ont été vérifiés. Premièrement, la protéine 1 est aussi celle dont le profil théorique possédait le plus d'homologie avec le profil observé (8 peptides sur 10) et celle dont le pourcentage de recouvrement des peptides était le plus élevé (30,5%).

C'est pourquoi la description détaillée de la comparaison du profil de la masse des peptides théorique de la protéine 1 avec le profil de la protéine SL82 a été examiné de plus près (Tableau 10). Tout d'abord, il a été possible de remarquer que la plus grande différence de masse entre un peptide théorique et un peptide de la protéine SL82 est très faible, soit de 0,3079 Da. De plus, seulement un peptide théorique correspond à la masse d'un peptide où la trypsine a raté un site de clivage. En effet, si tous les peptides qui concordent contiennent un site de clivage manqué, l'identification peut être mis en doute. Ainsi, tous les critères semblent confirmer que la protéine SL82 est bel et bien la méthylase responsable de la résistance au thiostrepton. Il peut cependant sembler étrange que la protéine ne provienne pas de *S. lividans* mais de *S. azureus*. Mais ceci s'explique par le fait que la souche IAF6 possède le plasmide pIAF6. Ce dernier est dérivé du plasmide pIJ702 et confère la résistance au thiostrepton aux souches de streptomycètes. Or le gène de résistance au thiostrepton utilisé lors de la construction du plasmide pIJ702 à partir de pIJ101 (Katz et Thompson, 1983) provient de *S. azureus*. Il est donc normal de retrouver cette protéine dans l'extrait de protéines cytoplasmiques des souches qui portent ce plasmide et ce en grande quantité puisque pIJ702 est un plasmide à copies multiples. C'est pourquoi la protéine SL82 est si importante en comparaison avec les autres protéines du gel (Figure 11).

Ce premier résultat confirme ainsi que les étapes précédentes particulièrement l'estimation du pI et MW par le logiciel, l'obtention d'un profil de la masse des peptides par spectrométrie de masse et la détermination de la charge des peptides ont été effectuées correctement. De plus, la masse des peptides déterminée par le spectromètre est suffisamment près de la masse réelle (théorique) pour faire une recherche par profil de la masse des peptides.

4.1.4 Analyse des peptides par spectrométrie de masse: fragmentation des peptides et recherche de la séquence en acides aminés

Étant donné que la séquence des peptides de la méthylase conférant la résistance au thiostrepton qui concordait avec les peptides obtenus était fournie, celle-ci a été utilisée pour vérifier que les spectres de fragmentation générés permettraient d'en déduire la même séquence. Chaque fois que le spectre permettait de lire une séquence continue, une partie de la séquence a pu être trouvée (Figures 15, 16 et 17). Étant donné que les pics les plus intenses étaient utilisés pour déduire la séquence et que l'erreur sur la masse attendu du fragment est relativement faible (entre 0,02 et 0,62) la séquence est fort probablement exacte.

4.2 ANALYSE DU PROTÉOME DE *S. COELICOLOR* M145

4.2.1 Profil de la masse des peptides des protéines SC69 (122) et SC34 (71)

Il n'a pas été possible d'obtenir un spectre de masse ressemblant à un profil de la masse des peptides pour ces deux protéines (Figures 21 et 23). Pourtant, leur intensité sur gel était comparable à celle des protéines SC124, SC117 et SC121, pour lesquelles un profil a été obtenu. La recherche dans les bases de données a tout de même été tentée avec la première protéine de *S. coelicolor* à être analysée par spectrométrie de masse, soit SC69 (122) (Tableau 13). Cependant, aucune des protéines trouvées par le logiciel PeptIdent ne semble correspondre à la protéine SC69. Tout d'abord ni les valeurs de pI ni les valeurs de masse moléculaire des deux protéines possibles ne concordent avec celles de la protéine SC69 (122). Ensuite, les peptides obtenus lors de la digestion tryptique de la protéine SC69 (122) couvrent seulement 2,3% de la protéine 1 et 3,3% de la protéine 2. De plus, il ne s'agit pas d'une protéine hydrolysée par des protéases, ce qui expliquerait le faible pourcentage de la protéine couvert par les peptides. Si tel avait été le cas, les peptides couvriraient une même région de la protéine et ne seraient pas distribués partout sur celle-ci.

4.2.2 Profil de la masse des peptides des protéines SC124 (68), SC117 et SC121

4.2.2.1 Protéine SC124 (68)

Le logiciel PeptIdent a identifié cette protéine comme étant une protéine hypothétique de 20,3 kDa appartenant à *S. coelicolor*. Les autres protéines possibles trouvées par le logiciel (Tableau 14) ont été éliminées parce que leur valeur de pI ainsi que les valeurs de masse moléculaire ne concordaient pas avec celles de la protéine SC124 (68). Ainsi, la seule protéine dont les pI et la masse moléculaire concordaient était la protéine hypothétique de 20,3 kDa. De plus, cette protéine est aussi celle qui possédait à la fois le profil théorique ayant le plus d'homologie avec le profil observé (8 peptides sur 12) et le pourcentage de recouvrement des peptides le plus élevé (70,7%). C'est pourquoi la description détaillée de la comparaison du profil de la masse des peptides théorique de la protéine de 20,3 kDa avec le profil de la protéine SC124 (68) a été examinée de plus près (Tableau 15). Tout d'abord, il a été possible de remarquer que la plus grande différence de masse entre un peptide théorique et un peptide de la protéine SC124 est très faible, soit de 0,321 Da. De plus, aucun peptide théorique ne correspond à la masse d'un peptide où la trypsine a manqué un site de clivage. La séquence en acides aminés de deux peptides (m/z de 607,10, LNQAGGA et 762,20, VFPSIY) trouvée dans les spectres de fragmentation (Figures 27 et 28) a aussi confirmé cette identification.

Ainsi, tous les critères semblent confirmer que la protéine SC124 (68) est cette protéine de 20,3 kDa de l'espèce *S. coelicolor*, ayant comme numéro d'identification dans la base de données SWISS-PROT/TrEMBL: Q9KXW4. Cette protéine n'est plus hypothétique mais est effectivement produite par *S. coelicolor*. Étant donné qu'une fonction hypothétique a été attribuée à cette protéine cela signifie que le CLO possède de l'homologie avec une séquence retrouvée chez d'autres organismes mais dont le rôle n'est pas encore complètement défini. Le résultat obtenu lors de la recherche par homologie de séquence a donc été vérifié afin d'obtenir plus d'informations sur la fonction possible de la protéine SC124 (68) (Tableau 18). Cette information est disponible dans la base de

données EMBL, puisque chaque séquence annotée du génome de *S. coelicolor* y est déposée.

Tableau 18. Résultat de la recherche par homologie de séquence de la protéine de 20,3 kDa.

```
CDS complement(8015..8590) /db_xref=SPTREMBL:Q9KXW4 /note=SCD95A.10c,
hypothetical protein, len: 191 aa;
highly similar to SW:TERE_SERMA (EMBL:L38824) Serratia marcescens tellurium
resistance protein TerE, 191 aa; fasta scores: opt: 864 z-score: 1027.3 E(): 0; 65.4% identity
in 191 aa overlap and to TR:CAB62770 (EMBL:AL133424) Streptomyces coelicolor
SCF56.25 tellurium resistance protein TerD (fragment), 185 aa; fasta scores: opt: 950 z-score:
1104.7 E(): 0; 75.0% identity in 184 aa overlap
/transl_table=11 /product=hypothetical protein SCD95A.10c
/gene=SCD95A.10c /protein_id=CAB93037.1
/translation=MAVSLSKGGNVSLTKEAPGLTEVTVGLGWDVRTTTGTDFDL DASA
IAVNTQ GK VYSDAHFVFFN NK QTPDNTIVHTGDNRTGEGAGDDEAINVNLAGLPADI
EKIVFPVSIYDAENRSQNF GQVRNAYIRILNQAGGAEIARYDLSEDAATETAMVFGEL
YRNGAEWKFRAV GQGYASGLTGIAQDFGVNV
```

Cette protéine est donc le produit du gène SCD95A.10c, puisque situé sur le cosmide SCD95A et le numéro de ce dernier dans la base de données EMBL est AL357432. Elle est très similaire à une protéine conférant à la résistance au tellurium, TerE plus précisément, de *S. marcescens* et à TerD de *S. coelicolor*. Le Tableau 19 contient l'information retrouvée dans la base de données EMBL pour la protéine TerE de *S. marcescens*. Chez ce microorganisme, cette protéine a comme rôle l'inhibition de phage, la résistance aux colicines et au tellurite.

Tableau 19. Description de la protéine TerE de *S. marcescens* disponible dans la base de données EMBL.

CDS 6777..7352 /gene="terE" /function="phage inhibition, colicin resistance, tellurite resistance" /codon_start=1 /transl_table=11 /product="putative TerE"
 /protein_id="AAA86851.1" /db_xref="GI:950683"
 /translation="MAVSLVKGGNVSLTKEAPTMNVAMVGLGWDARVTDGQGFDLDAS
 VFAVGEDGKVLSDAHFIFNKTSPDGAVEHQGDNRTGEGDGDDEQVKIDLTKVSA
 DIKKLVFAVTIYDAEARKQNFQGMVSNFMRVYNNNDNGTEIARFDLSEDASTETAMVF
 GELYRHGAEWKFKAVGQGFAGGLAALASQHGUNI"

4.2.2.2 Protéine SC117

Les spectres de masse des protéines SC124 (68) et SC117 sont pratiquement identiques (Figures 24 et 25) et elles ont sensiblement la même masse (même position verticale sur le gel de pH 4,0-5,5) et le même pI. Ces deux protéines sont fort probablement des isoformes de la même protéine. Un isoforme est le produit du même gène (ou d'un variant du gène produit par exemple par épissage alternatif) et il possède donc une constitution moléculaire et une fonction identique ou très similaire.

4.2.2.3 Protéine SC121

Le logiciel PeptIdent a identifié cette protéine comme étant une protéine hypothétique conservée chez *S. coelicolor*. En effet, les autres protéines possibles trouvées par le logiciel (Tableau 16) ont été éliminées parce que leur valeur de pI ainsi que leur valeur de masse moléculaire ne concordaient pas avec celles de la protéine SC121. Ainsi, la seule protéine dont les pI et la masse moléculaire concordaient était la protéine hypothétique conservée. De plus, la protéine est aussi celle qui possédait à la fois le profil théorique ayant le plus d'homologie avec le profil observé (6 peptides sur 10) et le pourcentage de recouvrement des peptides le plus élevé (50,8%). C'est pourquoi la description détaillée de la comparaison du profil de la masse des peptides théorique de la protéine 2 avec le profil de la protéine SC121 a été examiné de plus près (Tableau 17).

Il a été possible de remarquer que la plus grande différence de masse entre un peptide théorique et un peptide de la protéine SC121 est très faible, soit de 0,1675 Da. De plus, aucun peptide théorique ne correspond à la masse d'un peptide où la trypsine a manqué un site de clivage. La séquence en acides aminés de deux peptides (m/z de 629,14, VNQAGE et 762,32, VFPVSIY) trouvée dans les spectres de fragmentation (Figures 29 et 30) a aussi permis de confirmer l'identification.

Ainsi, tous les critères semblent confirmer que la protéine SC121 est cette protéine conservée de *S. coelicolor*, ayant comme numéro d'identification dans la base de données SWISS-PROT/TrEMBL: Q9KY22. Cette protéine n'est plus hypothétique puisque réellement produite par *S. coelicolor*. Le résultat obtenu lors de la recherche par homologie de séquence est indiqué dans le Tableau 20.

Tableau 20. Résultat de la recherche par homologie de séquence de la protéine conservée.

```
CDS complement(25611..26186) /db_xref="SPTREMBL:Q9KY22" /note="SCC8A.26c,
conserved hypothetical protein, len: 191aa; similar to proteins associated with resistance to
tellurium salts eg. SW:Q52357 (TERD_SERMA) tellurium resistance protein TerD from
Serratia marcescens (192 aa) fasta scores; opt: 866, z-score: 1036.9, E(): 0, 65,8% identity in
190 aa overlap. Also strongly similar to others from Streptomyces coelicolor eg.
TR:CAB62770 (EMBL:) and neighbouring CDS SCC8A.26c."
/transl_table=11 /product="conserved hypothetical protein"
/gene="SCC8A.26c" /protein_id="CAB92844.1"
/translation="MGVSLSKGGNVSLTKEAPGLTAVIVGLGWDIRTTTGTDFDLASALLN
SGGKVASDAHFIFNNLKSPDGVEHTGDNITGEGEGDDEQIKINLATVPADIEKIVFP
VSIYDAENRQQSFGQVRNAFIRVVNQAGEAEIARYDLSEDASTETAMVFGELYRHGA
EWKFRAIGQGYASGLRGIAQDFGVNV"
```

Cette protéine est le produit du gène SCC8A.26c, puisque situé sur le cosmide SCC8A et le numéro de ce dernier dans la base de données EMBL est AL356892. Elle

est très similaire à des protéines impliquées dans la résistance au tellurium, TerD plus particulièrement, de *S. marcescens* et à d'autres protéines de *S. coelicolor*. Le Tableau 21 contient l'information retrouvée dans la base de données EMBL pour la protéine TerD de *S. marcescens*. Chez ce microorganisme, cette protéine a comme rôle l'inhibition de phage, la résistance à la colicine et au tellurite.

Tableau 21. Description de la protéine TerD de *S. marcescens* disponible dans la base de données EMBL.

```
CDS 6130..6708 /gene="terD" /function="phage inhibition, colicin resistance, tellurite
resistance" /codon_start=1 /transl_table=11 /product="putative TerD"
/protein_id="AAA86850.1" /db_xref="GI:950682"
/translation="MSVSLSKGGNVSLSKAAPSMKNVLVGLGWDARSTDGQDFDLAS
AFLASNGKVRGDSDFIFYNNLTSSDGSVTHTGDNRTGEGDGDDESLKIKLDAVPSE
VDKIIFVVTIHDAQARRQSFQVSGAFIRLVNDDNQTEVARYDLTEDASTETAMLFGE
LYRHNGEWKFRAVGQGYAGGLASVCAQYGINAS"
```

Chez *S. marcescens* les gènes *terD* et *terE* sont retrouvés sur le plasmide R478 qui fait partie du groupe de plasmide IncH12, propre aux entérobactéries (Whelan, Collieran et Taylor, 1995). Une région de ce plasmide code pour les phénotypes de résistance au tellurite (Te^r), l'inhibition de phages (Phi) et de résistance aux colicines (PacB). Ainsi, sept CLOs ont été identifiés suite au séquençage de cette région, soit *terZ*, *-A*, *-B*, *-C*, *-D*, *-E* et *-F*. Cependant le rôle exact de chacun au niveau des différents phénotypes n'est pas précisément défini, d'où la possibilité de trois fonctions pour TerD et TerE.

En ce qui concerne *S. coelicolor*, il est peu probable que les protéines SC124 et SC121 soient impliquées dans l'inhibition de phages ou dans la résistance aux colicines. En effet, l'inhibition de bactériophages Phi est associée à la résistance au tellurite dans les plasmides IncH (Taylor et Grant, 1976) et ceux-ci ne sont pas présents chez les streptomycètes mais plutôt chez les entérobactéries. De plus, le phénotype Phi inhibe la propagation de phages spécifiques aux entérobactéries, tels que λ , T1, T5, T7 et ϕ 80 (Whelan, Collieran et Taylor, 1995).

Les bactéries produisant des colicines, (bactériocines dirigées contre *E. coli*) synthétisent aussi une protéine d'immunité, qui leur confère une protection (résistance) contre la colicine en question (Espeset *et al.*, 1994). Étant donné que les bactériocines agissent uniquement contre les souches apparentées à la bactérie productrice, *S. coelicolor* ne produit manifestement pas de colicines (puisque très éloigné d'*E. coli*) et donc aucune protéine d'immunité contre celles-ci. Les deux protéines sont vraisemblablement impliquées dans la résistance au tellurite. Les composés de tellurium sont utilisés dans plusieurs procédés industriels mais ils sont relativement rares dans l'environnement. Ce type de résistance n'est pas unique au groupe de plasmide IncH puisque d'autres systèmes chromosomiques et plasmidiques ont été rapportés (Taylor *et al.*, 1994, Walter *et al.*, 1991). La résistance au tellurite est retrouvée tant chez les bactéries Gram-négatives que Gram-positives. En effet, le tellurite a un spectre d'activité très large puisque son mécanisme de toxicité serait due à son pouvoir d'oxydation. Cependant, le mécanisme n'est pas complètement élucidé.

Ainsi, l'étude en profondeur de ces deux protéines pourra confirmer leur implication au niveau de la résistance au tellurite. Par exemple, la comparaison du gel en 2-D des protéines cytoplasmiques de *S. coelicolor* obtenues après croissance en présence d'un sel de tellurite (typiquement le tellurite de potassium) avec le gel en 2-D des protéines cytoplasmiques obtenues sans la présence du sel pourrait être réalisée. En plus de vérifier si la souche est réellement résistante au tellurite, une augmentation de la quantité des deux protéines confirmera leur rôle au niveau de la résistance. Aussi, l'augmentation et/ou l'apparition de protéines permettra d'identifier d'autres protéines impliquées dans la résistance au tellurite. En effet, d'autres CLOs possiblement impliqués dans cette résistance ont été identifiés (Jobling et Ritchie, 1987, Pillage *et al.*, 1994 et Whelan, Collieran et Taylor, 1995). Des expériences de dislocation génétique afin de vérifier si les mutants deviendraient plus sensibles au tellurium pourraient aussi être effectuées.

4.2.3 Comparaison des protéines SC124 (68) et SC121

Étant donné que ces protéines possédaient chacune 191 acides aminés, que leur valeur de pI et de masse moléculaire sont très similaires et qu'elles sont toutes deux impliquées dans la résistance au tellurite, leur séquence ont été comparées afin d'évaluer le degré d'homologie. La comparaison a été effectuée avec le logiciel BLAST (www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) qui permet d'aligner deux séquences. Le résultat est présenté au Tableau 22.

Tableau 22. Résultat de l'alignement de séquences des protéines SC124 (Q9KXW4, Sbjct 1) et SC121 (Q9KY22, Query 1).

Score = 323 bits (827), Expect = 1e-87	
Identities = 156/191 (81%), Positives = 174/191 (90%)	
Query: 1	MGVSLSKGGNVSLTKEAPGLTAVIVGLGWDIRTTTGTDFDLDASALLLNSGGKVASDAHF 60
	M VLSKGGNVSLTKEAPGLT V VGLGWD+RTTGTDFDLDASA+ +N+ GKV SDAHF
Sbjct: 1	MAVSLSKGGNVSLTKEAPGLTEVTVGLGWDVTRTTTGTDFDLDASAIAVNTQGKVVYSDAHF 60
Query: 61	IFFNNLKSPDGSVEHTGDNITGEGEGDDEQIKINLATVPADIEKIVFPVPSIYDAENRQQS 120
	+FFNN ++PD ++ HTGDN TGEG GDDE I +NLA +PADIEKIVFPVPSIYDAENR Q+
Sbjct: 61	VFFNNKQTPDNTIVHTGDNRTGEGAGDDEAINVNLAGLPADIEKIVFPVPSIYDAENRQQS 120
Query: 121	FGQVRNAFIRVNVQAGEAEIARYDLSEDASTETAMVFGELYRHGAEWKFRAIGQGYASGL 180
	FGQVRNA+IR++NQAG AEIARYDLSEDA+TETAMVFGELYR+GAEWKFRA+GQGYASGL
Sbjct: 121	FGQVRNAYIRILNQAGAEIARYDLSEDAATETAMVFGELYRNGAEWKFRVAVGQGYASGL 180
Query: 181	RGIAQDFGVNV 191
	GIAQDFGVNV
Sbjct: 181	TGIAQDFGVNV 191

Ainsi, 81% (156/191) des acides aminés sont parfaitement identiques (identities) et 90% (174/191) sont similaires (même fonction, c'est à dire polaires ou non polaires) (positives). Ces deux protéines sont donc codées par des gènes homologues situés sur des cosmides différents et éloignés l'un de l'autre (représentation graphique de la position des cosmides, www.sanger.ac.uk/Projects/S_coelicolor/Lnewmap.gif) Malgré cette forte homologie entre les deux protéines, elles ont été identifiées comme étant deux protéines différentes grâce au profil de la masse des peptides. Par exemple, si les protéines avaient été identifiées par leur séquence en acides aminés à l'extrémité N terminale, il aurait été difficile de les discriminer l'une de l'autre. En effet, les 21 premiers acides aminés, sauf un, sont parfaitement identiques. La comparaison de la digestion théorique par la trypsine de la protéine hypothétique de 20,3 kDa (Tableau 23) avec celle de la protéine hypothétique conservée (Tableau 24) illustre bien la puissance de la spectrométrie de masse. La digestion tryptique théorique a été effectuée avec le logiciel PeptideMass (expasy.cbr.nrc.ca/tools/peptide-mass.html). En effet, certains peptides ont des séquences et donc des masses identiques (les peptides 8 et 11 de la protéine hypothétique de 20,3 kDa correspondent respectivement aux peptides 6 et 12 de la protéine hypothétique conservée) et plusieurs sont différents par un, deux ou trois acides aminés seulement (peptides 2, 5, 6, 9, 10, 12, 13 et 14 correspondent respectivement aux peptides 2, 4, 5, 8, 11, 13, 14 et 15). Cependant, cette mince différence au niveau de la séquence en acides aminés engendre une différence de masse suffisante pour être discriminée par le spectromètre de masse, qui se reflète alors dans le profil de la masse des peptides.

La forte homologie entre les homologues de TerD et TerE chez *S. coelicolor* n'est pas inattendue puisque le même phénomène est retrouvé chez *S. marcescens* (66,3%) et *E. coli* (66,5%) (Whelan, Colleran et Taylor, 1995 et Jobling et Ritchie, 1987 et 1988). En effet, Jobling et Ritchie démontrèrent que *terD* et *terE* étaient fonctionnellement interchangeable et il fut suggéré que ces deux gènes proviendraient de la duplication d'un gène ancestral. Or, la duplication génétique chez streptomyces est bien connue. Par exemple, la duplication génétique du domaine de fixation au xylane (XBD) de la xylanase

A et de l'arabinofuranosidase B (cosmides 10A5/7H1), du XBD de la xylanase B et de l'acétyl xylane estérase (cosmides C75/C30) et Bx1R (cosmides 1H10/F85), (Redenbach *et al.*, 1996). De plus, l'analyse de la région promotrice en amont de ces deux gènes ne révélait aucune homologie significative. Ceci suggère que les deux gènes ne sont pas soumis à une régulation identique et donc qu'ils ne semblent pas former un régulon.

Tableau 23. Digestion tryptique théorique de la protéine hypothétique de 20,3 kDa.

HYPOTHETICAL 20.3 KDA PROTEIN.

[Theoretical pI: 4.49 / Mw (average mass): 20344.40]

mass	position	#	peptide sequence
2369.524	81-104	1	TGEGAGDDEAINVNLAGLPA DIEK
2281.476	143-162	2	YDLSEDAATETAMVFGELYSR
2126.262	33-53	3	TTTGTDFDLASAIQVNTQG K
2024.217	171-191	4	AVGQGYASGLTGIAQDFGVN V
1799.014	16-32	5	EAPGLTEVTVGLGWDVSR
1587.755	54-66	6	VYSDAHFVFFNNK
1567.635	67-80	7	QTPDNTIVHTGDNR
1522.721	105-117	8	IVFPVSIYDAENR
1212.371	131-142	9	ILNQAGGAEIAR
935.007	118-125	10	SNFGQVR
774.872	8-15	11	GGNVSLTK
734.909	1-7	12	MAVSLSK
703.753	163-168	13	NGAEWK
635.721	126-130	14	NAYIR

Tableau 24. Digestion tryptique théorique de la protéine hypothétique conservée.

CONSERVED HYPOTHETICAL PROTEIN.

[Theoretical pI: 4.56 / Mw (average mass): 20387.60]

mass	position	#	peptide sequence
2586.576	68-92	1	SPDGSVEHTGDNITGEGEGD DEQIK
2297.475	143-162	2	YDLSASTETAMVFGELYR
2097.263	33-53	3	TTGTDFDLASALLNSGG K
1767.058	16-32	4	EAPGLTAVIVGLGWDIR
1622.844	54-67	5	VASDAHFIFNNLK
1522.721	105-117	6	IVFPVSIYDAENR
1283.487	93-104	7	INLATVPADIEK
1256.381	131-142	8	VVNQAGEAEIAR
1092.220	171-181	9	AIGQGYASGLR
1019.122	182-191	10	GIAQDFGVNV
949.034	118-125	11	QQSFGQVR
774.872	8-15	12	GGNVSLTK
726.790	163-168	13	HGAEWK
720.882	1-7	14	MGVSLSK
619.721	126-130	15	NAFIR

5. CONCLUSION

Tout d'abord, le premier objectif du projet a été réalisé, soit la mise au point des différentes étapes d'une analyse protéomique. Ainsi, il est maintenant possible de séparer par électrophorèse en deux dimensions un échantillon complexe de protéines (ici les protéines cytoplasmiques) et de les détecter. L'analyse des gels générés peut aussi être effectuée, permettant en outre de déterminer le point isoélectrique, le poids moléculaire de même que la surface apparente de chaque protéine présente dans le gel. La digestion des protéines par la trypsine à l'intérieur du gel de polyacrylamide est suffisamment optimisée pour permettre l'obtention d'un profil de la masse des peptides de la protéine digérée lors de l'analyse en spectrométrie de masse. Finalement, les données générées (rapport m/z et charge des peptides) permettent l'identification des protéines suite à la recherche dans les bases de données avec le logiciel PeptIdent.

En effet, lors de la mise au point des ces étapes avec les protéines cytoplasmiques des souches *S. lividans* IAF6 et 1326, la méthylase qui confère la résistance au thiostrepton a été identifiée sur le gel en 2-D de la souche IAF6. Suite à l'analyse du protéome de *S. coelicolor* M145, trois protéines (dont deux isoformes de la même protéine) ont été identifiées comme étant des homologues de TerD et TerE de *S. marcescens*. Ces trois protéines sont vraisemblablement impliquées dans la résistance au tellurium. Des expériences complémentaires devront cependant être effectuées afin de confirmer et d'approfondir le rôle de chacune. Ainsi, l'analyse de ces protéines a soulevé la difficulté qui se présente lorsque la fonction d'une protéine au sein d'un organisme tente d'être définie. Il a néanmoins été possible de confirmer que ces protéines autrefois classifiées comme étant hypothétiques sont réellement produites par *S. coelicolor*.

L'analyse du protéome de *S. coelicolor* M145 se poursuit dans nos laboratoires puisqu'il reste encore des milliers de protéines à identifier tant au niveau des protéines cytoplasmiques, membranaires et sécrétées. La tâche sera plus facile puisque les données sur le génome de *S. coelicolor* ne cessent d'augmenter et que le séquençage sera bientôt complété.

Étant donnée les appareils disponibles et la façon dont les techniques furent effectuées au cours de mon projet, l'analyse protéomique fut plutôt fastidieuse. Ainsi, certaines techniques doivent être améliorées et d'autres automatisées afin d'identifier les protéines plus rapidement. Tout d'abord, la quantité de protéines dans les gels doit être augmentée afin que la digestion d'une seule protéine provenant d'un seul gel soit suffisante pour obtenir un signal en spectrométrie de masse. Ainsi, au lieu de digérer quatre protéines identiques, quatre protéines différentes seraient traitées en même temps. De plus, il existe aujourd'hui des appareils qui prélèvent les protéines à partir d'un gel dont l'image a été préalablement numérisée et d'autres qui digèrent la protéine à l'intérieur du gel. Ces derniers sont reliés à des spectromètres de masse et les échantillons sont injectés un à la suite des autres de façon automatisée.

Ainsi, beaucoup de temps était passé à l'obtention de données (injection de l'échantillon, acquisition du spectre de masse et fragmentation des peptides) de même qu'à leur analyse (sélection des peptides pour le profil et détermination de leur charge). Or les spectromètres de masse de type Q-TOF (double quadrupole couplé un time of flight) sont en mesure de générer automatiquement la liste du rapport m/z de tous les peptides détectés avec la charge de chacun. Cette liste peut être utilisée directement par les logiciels utilisés pour la recherche dans les bases de données. De la même façon, de nouveaux logiciels utilisent les données brutes provenant du spectre de fragmentation afin d'identifier une protéine à l'aide de la séquence en acides aminés déduite à partir du spectre. Comme il est très complexe de déduire manuellement une séquence à partir d'un spectre de fragmentation, cette dernière possibilité est très utile dans les cas où un nombre insuffisant de peptides est détecté. En effet, dans de tels cas, il n'est pas possible de faire une recherche par profil de la masses des peptides. Ce problème est d'ailleurs survenu pour les protéines SC69 (122) et SC34 (71).

Ainsi, la technologie relative aux analyses protéomiques permet maintenant l'automatisation de l'acquisition et de l'interprétation des données. Comme de plus en plus de groupes ont entrepris des études de protéomes, les données sur les gènes et les

protéines augmentent sans cesse. Les techniques et les ressources étant donc disponibles, il ne reste qu'à poursuivre ce travail ardu mais nécessaire à la compréhension des processus biologiques.

BIBLIOGRAPHIE

- Baltz, R. H., M. A. McHenney, C. A. Cantwell, S. W. Queener et P. J. Solenberg. 1997. «Applications of transposition mutagenesis in antibiotic producing streptomycetes». Antonie Van Leeuwenhoek, vol. 71. p. 179-87.
- Berkelman, T. et T. Stenstedt. 1998. 2-D Electrophoresis using immobilized pH gradients: principles and methods. Amersham Pharmacia Biotech.
- Bibb, M.J. 1996. «The regulation of antibiotic production in *Streptomyces coelicolor* A3(2)». Microbiology, vol. 142. p. 1335-1344.
- Bibb, M. J. et D. A. Hopwood. 1981. «Genetic studies of the fertility plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2)». 1981. J. Gen. Microbiol., vol. 126. p.427-442.
- Bibb, M. J., J. M. Ward et D. A. Hopwood. 1981. «Transformation of plasmid DNA into *Streptomyces* at high frequency». Nature, vol. 274. p. 398-400.
- Bibb, M. J., J. M. Ward, T. Kieser, S. N. Cohen et D. A. Hopwood. 1981. «Excision of chromosomal DNA sequences from *Streptomyces coelicolor* forms a novel family of plasmids detectable in *Streptomyces lividans*». Mol. Gen. Genet., vol. 184. p. 230-240.
- Bini, L., H. Heid, S. Liberatori, G. Geier, V. Pallini et R. Zwillig. 1997. «Two-dimensional gel electrophoresis of *Caenorhabditis elegans* and identification of protein spots by microsequencing». Electrophoresis, vol. 18. p. 557-562.
- Bjellqvist, B., K. Ek, P. G. Righetti, E. Gianazza, A. Görg, R. Westermeier et W. Postel. 1982. «Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications». J. Biochem. Biophys. Meth., vol. 6. p. 317-339.
- Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scottt, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kaine, M. Borodovsky, H-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese et J. C. Venter. 1996. «Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*». Science, vol. 273. p. 1058-1073.
- Cordwell, S. J., D. J. Basseal et I. Humphery-Smith. 1997. «Proteome analysis of *Spiroplasma melliferum* (A56) and protein characterisation across species boundaries». Electrophoresis, vol. 18. p. 1335-1346.

Dupont, C., N. Daigneault, F. Shareck, R. Morosoli et D. Kluepfel. 1996. Purification of an acetyl xylan esterase produced by *Streptomyces lividans*. Biochem. J., vol. 319. p. 881-886.

Espeset, D., P. Piet, C. Lazdunski et V. Geli. 1994. «Immunity proteins to pore-forming colicins: structure-function relationships». Mol Microbiol., vol. 13. p. 1111-20.

Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J-F. Tomb, B. A. Dougherty, K. F. Bott, P-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison III et J. C. Venter. 1995. «The minimal gene complement of *Mycoplasma genitalium*». Science, vol. 270. p. 397-403.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith et J. C. Venter. 1995. «Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd». Science, vol. 269. p. 496-512.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin et S. G. Oliver. 1996. «Life with 6000 genes». Science, vol. 274. p. 563-567.

Görg, A., W. Postel et S. Günther. 1988. «The current state of two-dimensional electrophoresis with immobilized pH gradients». Electrophoresis, vol. 9. p. 531-546.

Görg, A., G. Boguth, C. Obermaier, A. Posch et W. Weiss. 1995. «Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): The state of the art and the controversy of vertical versus horizontal systems». Electrophoresis, vol. 16. p. 1079-1086.

Gusek, T. W., et J. E. Kinsella. 1992. «Review of the *Streptomyces lividans*/vector pIJ702 system for gene cloning». Crit. Rev. Microbiol., vol. 18. p. 247-260.

Hodges, P. E., W. E. Payne, J. I. Garrels. 1998. «The yeast protein database (YPD): a curated proteome database for *Saccharomyces cerevisiae*». Nucleic acids research, vol. 26. p. 68-72.

Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward et H. Schrempf. 1985. Genetic manipulation of *Streptomyces*. Norwich: John Innes Foundation.

- Jobling, M. G. et D. A. Ritchie. 1987. «Genetic and physical analysis of plasmid genes expressing inducible resistance of tellurite in *Escherichia coli*». Mol. Gen. Genet., vol. 208. p. 288-293.
- Jobling, M. G. et D. A. Ritchie. 1988. «The nucleotide sequence of a plasmid determinant for resistance to tellurium anions». Gene, vol. 66. p. 245-258.
- Katz, E., C. J. Thompson et D. A. Hopwood. 1983. «Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*». J. Gen. Microbiol., vol. 129. p. 2703-2714.
- Kieser, H. M., T. Kieser et D. A. Hopwood. 1992. «A combined genetic and physical map of the *Streptomyces coelicolor* A3(2)». J. Bacteriol., vol. 174. p. 5496-5507.
- Kirby, R. et D. A. Hopwood. 1977. «Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2)». J. Gen. Microbiol., vol. 98. p. 229-252.
- Kluepfel, D., S. Vats-Metha, F. Aumont, F. Shareck et R. Morosoli. 1990. Purification and characterisation of a new xylanase (xylanase B) produced by *Streptomyces lividans* 66. Biochem. J., vol. 267. p. 45-50.
- Kluepfel, D., N. Daigneault, R. Morosoli et F. Shareck. 1992. Purification and characterisation of a new xylanase (xylanase C) produced by *Streptomyces lividans* 66. Appl. Microbiol. Biotechnol., vol. 36. p. 626-631.
- Lam, T. D. T. 1999. Purification et caractérisation de la protéine codée par le gène *bxlA* de *Streptomyces lividans*. Mémoire de maîtrise, INRS-Institut Armand-Frappier, Université du Québec. 138 p.
- Laemmli, U.K. 1970. «Cleavage of structural proteins during the assembly of the head of bacteriophage T4». Nature, vol. 227. p. 680-685.
- Laroque, D. 1998. Analyse moléculaire du locus *bxl* de *Streptomyces lividans*. Mémoire de maîtrise, Institut Armand-Frappier, Université du Québec. 184 p.
- Lee, T. A. 1998. A beginner's guide to mass spectral interpretation. John Wiley & sons, Baffins Lane, Chichester.
- Lin, Y.-S., H. M. Kieser, D. A. Hopwood, C. W. Chen. 1993. «The chromosomal DNA of *Streptomyces lividans* 66 is linear». Mol. Microbiol., vol. 10. p. 923-933.
- Link, A. J., L. G. Hays, E. B. Carmack et J. R. Yates III. 1997. «Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143». Electrophoresis, vol. 18. p. 1314-1334.

- Lowry, O. H., N. J. Rosebrough, A. L. Farr et R. J. Randall. 1951. «Protein measurement with the Folin phenol reagent». J. Biol. Chem., vol. 193. p. 265-275.
- Mandels, M. et E. T. Reese. 1957. «Induction of cellulase in *Trichoderma viridae* as influenced by carbon sources and metals». J. Bacteriol., vol. 73. p. 269-278.
- Manin, C., F. Shareck, R. Morosoli et D. Kluepfel. 1994. Purification and characterisation of an α -L-arabinofuranosidase from *Streptomyces lividans* 66 and DNA sequence of the gene (*abfA*). Biochem. J., vol. 239. p. 587-592.
- McCarthy, A. J. et S. T. Williams. 1992. «Actinomycetes as agents of biodegradation in the environment - a review». Gene, vol. 115. p. 189-192.
- Miyadoh, S. 1993. «Research on antibiotic screening in Japan over the last decade: a producing microorganism approach». Actinomycetologica, vol. 7. p. 100-106.
- Morosoli, R., J. L. Bertrand, F. Mondou, F. Shareck et D. Kluepfel. 1986. «Purification and properties of a xylanase from *Streptomyces lividans*». Biochem J., vol. 239. p. 587-92.
- Morris, A. C. et M. A. Djordjevic. 2001. «Proteome analysis of cultivar-specific interactions between *Rhizobium leguminosarum* biovar *trifolii* and subterranean clover cultivar Woogenellup». Electrophoresis, vol. 22. p. 586-598.
- O'Connor, C. D., M. Farris, R. Fowler et S-Y. Qi. 1997. «The proteome of *Salmonella enterica* serovar *typhimurium*: Current progress on its determination and some applications». Electrophoresis, vol. 18. p. 1483-1490.
- O'Farrell, P. H. 1975. «High resolution two-dimensional electrophoresis of proteins». J. Biol. Chem., vol. 250. p. 4007-4021.
- Pappin, D. J. C. et D. N. Perkins. 1999. «An integrated search engine for protein identification using mass spectrometry». Affiche présentée au congrès de l'ABRF: Bioinformatics and biomolecular technologies: linking genomes, proteomes and biochemistry. (Durham, Caroline du Nord, 19 et 20 mars 1999).
- Penzès, C. 1995. Purification et étude de la protéine MsiK de *Streptomyces lividans*. Mémoire de maîtrise, Institut Armand-Frappier, Université du Québec. 144 p.
- Pillage, C. J., J.-J. Gordon, K. Jury, C. A. Shearman et M. J. Gasson. 1994. «Tellurium resistance carried on the *Lactococcus lactis* sex factor». Affiche présentée au Satellite workshop on plasmid diversity. (Montréal, Québec).
- Redenbach, M., H. M. Kieser, D. Denapaité, A. Eichner, J. Cullum, H. Kinashi et D. A. Hopwood. 1996. «A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome». Mol. Microbiol., vol. 21. p. 77-96.

- Taylor, D. E. et R. B. Grant. 1976. «Inhibition of bacteriophage lambda, T1 and T7 development by R plasmids of the H incompatibility group». Antimicrob. Agents chemother., vol. 10. p. 762-764.
- Taylor, D. E., Y. Hou, R. J. Turner et J. H. Weiner. 1994. «Location of a potassium tellurite resistance operon (*tehA tehB*) within the terminus of *Escherichia coli* K-12». J. Bacteriol., vol. 176. p. 2740-2742.
- The *C. elegans* sequencing consortium. 1998. «Genome sequence of the nematode *C. elegans*: a platform for investigating biology». Science, vol. 282. p. 2012-2018.
- Urquart, B. L., T. E. Atsalos, D. Roach, D. J. Basseal, B. Bjellqvist, W. L. Britton et I. Humphery-Smith. 1997. «Proteomic contigs of *Mycobacterium tuberculosis* and *Mycobacterium bovis* (BCG) using novel immobilized pH gradients». Electrophoresis, vol. 18. p. 1384-1392.
- Vallorani, L., F. Bernardini, C. Sacconi, R. Pierleoni, B. Pieretti, G. Piccoli, M. Buffalini et V. Stocchi. 2000. «Identification of *Tuber borchii* Vittad. mycelium proteins separated by two-dimensional polyacrylamide gel electrophoresis using amino acid analysis and sequence tagging». Electrophoresis, vol. 21. p. 3710-3716.
- VanBogelen, A. R., K. Z. Abshire, B. Moldover, E. R. Olson et F. C. Neidhardt. 1997. «*Escherichia coli* proteome analysis using the gene-protein database». Electrophoresis, vol. 18. p. 1243-1251.
- Vincent, P., F. Shareck, C. Dupont, R. Morosoli et D. Kluepfel. 1997. New α -L-anabinofuranosidase produced by *Streptomyces lividans*: Cloning and DNA sequence of the *abfB* gene and characterisation of the enzyme. Biochem. J., vol. 322. p. 845-852.
- Volff, J. N. et J. Altenbuchner. 1998. «Genetic instability of the *Streptomyces* chromosome». Mol. Microbiol., vol. 27. p. 239-246.
- Walter, E. G., C. M. Thomas, J. P. Ibbotson et D. E. Taylor. 1991. «Transcriptional analysis, translational analysis and sequence of *kilA*-tellurite resistance region of plasmid RK2Te^R». J. Bacteriol., vol. 173. p. 1111-1119.
- Wasinger, V. C., S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams et I. Humphery-Smith. 1995. «Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*». Electrophoresis, vol. 16. p. 1090-1094.
- Whelan, F. K., E. Colleran et D. E. Taylor. 1995. «Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncH12 plasmid R478». Journal of bacteriology, vol. 177. p. 5016-5027.

Wilkins, M. R., C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J-C Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams et D. F. Hochstrasser. 1996. «From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis». Biotechnology, vol. 14. p. 61-65.

Wilkins, M. R., K. L. Williams, R. D. Appel et D. F. Hochstrasser. 1997. Proteome research: new frontiers in functional genomics. Springer, 243 p.

Wong, K. K. Y., L. U. L. Tan et J. N. Saddler. 1988. Multiplicity of β -1,4-xylanase in microorganisms: functions and applications. Microbiol. Rev., vol. 52. p. 305-317.