

## THESIS / THÈSE

### MASTER EN SCIENCES BIOLOGIQUES DES ORGANISMES ET ÉCOLOGIE

#### Etude structurale du système de sécrétion de type IV chez *Brucella suis*

Castado, Cindy

*Award date:*  
2000

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX  
NAMUR**

**Faculté des Sciences**

**ETUDE STRUCTURALE DU SYSTEME DE SECRETION DE TYPE IV CHEZ  
*BRUCELLA SUIS***

**Mémoire présenté pour l'obtention du grade de  
licencié en Sciences biologiques**

**Cindy CASTADO  
Juin 2000**

## ERRATA

- ◆ Les figures 6 et 7 sont inversées. Quand le texte fait référence à la figure 6, c'est la n°7 qu'il faut regarder et *vice versa*.
- ◆ Légende de la figure 29 : « Les résidus soulignés représentent l'alignement qui est conservé par rapport à la figure **28** » (et non la figure 26).
- ◆ La légende de la figure 40 correspond à la structure inférieure gauche. Ne pas tenir compte des autres structures.
- ◆ Point 4.4.2.4, page 50 : Les VirB5 sont évidemment à remplacer par VirB1.
- ◆ Figure 46 : VirB5 est également à remplacer par VirB1.
- ◆ Page 52 : Conclusions et Perspectives, pour le paragraphe concernant VirB5, une référence est nécessaire pour les v-SNARE et les t-SNARE : Weimbs *et al* (1997). « A conserved domain is present in different families of vesicular fusion proteins : A new superfamily ». Proc. Natl. Acad. Sci. **94** : 3046-3051.

## Etude structurale du système de sécrétion de type IV chez *Brucella suis*

CASTADO Cindy

### Résumé

Les *Brucella* sont des pathogènes intracellulaires qui infectent les humains et autres mammifères. Il existe peu d'informations à propos des mécanismes de virulence de cette bactérie. Récemment, un système de sécrétion de type IV nommé VirB a été identifié chez *Brucella suis* et semble être essentiel pour sa survie dans les macrophages. La région *virB* code pour douze protéines : VirB1 à 11 et l'orf 12. Des homologues de ces protéines ont été trouvés et ils remplissent différentes fonctions dans d'autres bactéries comme la conjugaison, l'infection d'hôtes, le transfert de macromolécules et d'ADN.

Nous avons suivi une approche bioinformatiques afin de prédire, sur base de la séquence primaire, des éléments structuraux et fonctionnels pour les douze protéines encodées par la région *virB* sur base de leur séquence primaire. Cette analyse est basée sur la similarité du système de sécrétion de type IV de *Brucella suis* avec d'autres systèmes homologues de bactéries dont celui de l'organisme modèle, *Agrobacterium tumefaciens*, qui induit des tumeurs par sécrétion de T-DNA dans la cellule végétale.

Nous avons collecté les informations qui pouvaient être intéressantes pour une meilleure compréhension de la relation entre la structure de ces douze protéines et leurs fonctions et des éventuelles interactions entre elles. Nous avons effectué des recherches dans les banques de données pour trouver les séquences homologues des douze protéines (avec Blast N.R., EST, PDB), des alignements de séquences, des recherches de motifs, des prédictions de structures secondaires et des reconnaissances de repliement protéique.

Au terme de ce travail, des prédictions de fonctions pour les protéines VirB1, VirB4 et VirB11 ont pu être renforcées et des arguments pour la résolution d'ambiguïté de localisation cellulaire pour les protéines VirB2, VirB5 et VirB9 ont été avancés. Certains composants du système de sécrétion de type IV nous ont semblé avoir des rôles plus importants dont la protéine VirB5. C'est ce composant qui serait impliqué dans la déviation du trafic intracellulaire de *Brucella suis*.

En premier lieu, je remercie le Professeur Eric Depiereux de m'avoir acceptée dans son équipe de biologie moléculaire structurale.

Je dis un grand merci et accorde toute ma gratitude à Katalin sans qui ce travail n'aurait pas été le même, pour son soutien dans les moments les plus difficiles et pour son perpétuel optimisme avec ses mémorants.

Je remercie, tout particulièrement, les membres du jury, M. Raes, I. Housen, RM Delrue et C. Vinals qui ont accepté de lire ce mémoire.

Merci à toute l'équipe URBM pour leur joyeuse humeur et la bonne ambiance qu'ils font régner dans le laboratoire.

Pour les passionnantes discussions et les encouragements, je voudrais remercier le Professeur Jean-Jacques Letesson.

Un tout grand merci aussi à Xavier, Cri, Anne et RoseMay.

Je tiens à dire merci à toutes les personnes de mon année grâce à qui ces quatre années se sont déroulées dans la joie et la bonne humeur.

Enfin, je tiens à associer toute ma famille ainsi que Cédric pour le grand réconfort moral qu'ils ont pu m'apporter tout au long de ce travail.

Un merci particulier à Cédric et Ben pour leur soutien lors de la dernière nuit.

## Table des Matières

<b>1. Introduction.....</b>	<b>3</b>
1.1. La bioinformatique.....	3
1.1.1. Généralités .....	3
1.1.2. Recherche d'homologues .....	3
1.1.3. Les banques de données .....	5
1.1.3.1. Les banques de données de séquences nucléotidiques .....	5
1.1.3.2. Les banques de données de séquences protéiques .....	5
1.1.3.2.1. Les banques de données de séquences primaires.....	5
1.1.3.2.2. Les banques composées.....	6
1.1.3.2.3. Les banques de structures tertiaires .....	6
1.1.3.2.4. Les banques de classification de structures .....	7
1.1.4. Alignements .....	8
1.1.5. Caractérisation de motifs .....	8
1.1.6. Localisation cellulaire .....	9
1.1.7. Prédiction de structures secondaires .....	9
1.1.8. Prédiction de structure tertiaire .....	10
1.1.8.1. Modélisation par homologie .....	10
1.1.8.2. Reconnaissance de « fold » .....	11
1.1.8.2.1. Threading.....	12
1.1.8.2.2. Pseudo-threading.....	12
1.2. Entre théorie et pratique .....	12
1.3. Le Contexte biologique .....	13
1.3.1. Les bactéries.....	13
1.3.1.1. Gram positive, Gram négative .....	13
1.3.1.2. Systèmes de transport .....	13
1.3.1.2.1. Transport « Out –In » .....	14
1.3.1.2.2. Transport « In – Out » .....	15
1.3.2. Le système de sécrétion de type IV chez <i>Agrobacterium tumefaciens</i> .....	21
1.3.2.1. VirB1 .....	22
1.3.2.2. VirB2 .....	22
1.3.2.3. VirB3 .....	22
1.3.2.4. VirB4 .....	23
1.3.2.5. VirB5 .....	23
1.3.2.6. VirB6 .....	23
1.3.2.7. VirB7 .....	24
1.3.2.8. VirB8 .....	24
1.3.2.9. VirB9 .....	24
1.3.2.10. VirB10 .....	25
1.3.2.11. VirB11 .....	25
1.3.2.12. VirD4 .....	26
1.3.3. L'organisme d'intérêt : <i>Brucella suis</i> .....	26
1.3.3.1. <i>Brucella</i> et brucellose .....	26
1.3.3.2. La pathogénie .....	26
1.3.3.3. La survie intracellulaire de <i>Brucella</i> .....	27
1.3.3.4. Le système de sécrétion de type IV de <i>Brucella suis</i> .....	27
<b>2. Objectifs .....</b>	<b>28</b>

<b>3. Matériel et méthodes.....</b>	<b>29</b>
3.1. Blast et ses améliorations .....	29
3.2. Les méthodes d'alignement.....	30
3.2.1. Align .....	30
3.2.2. Match-Box .....	30
3.2.3. ClustalW .....	31
3.3. Blocks.....	31
3.4. ScanProsite .....	32
3.5. ProDom .....	32
3.6. Prédiction de structures secondaires .....	32
3.6.1. PHD .....	32
3.6.2. PREDATOR .....	33
3.6.3. Prof .....	34
3.6.4. Jpred .....	34
3.6.5. Psi-Pred.....	35
3.7. PSORT .....	35
3.8. TOPS .....	36
3.9. Prédiction de structure tertiaire.....	37
3.9.1. 3D PSSM .....	37
3.9.2. 123D .....	38
3.9.3. Ucla-Doe.....	38
<b>4. Résultats et discussion .....</b>	<b>40</b>
4.1. Données préliminaires.....	40
4.2. « Débroussaillage » du complexe .....	41
4.3. Synthèse .....	45
4.4. Analyse plus détaillée de deux composants du système .....	46
4.4.1. VirB5 .....	46
4.4.1.1. <i>Analyse des alignements</i> .....	46
4.4.1.2. <i>Le « coiled coil »</i> .....	47
4.4.2. VirB1 .....	48
4.4.2.1. <i>Prédiction de fonction</i> .....	48
4.4.2.2. <i>TOPS</i> .....	49
4.4.2.3. <i>Consensus de structures secondaires</i> .....	49
4.4.2.4. <i>Alignement séquence/structure</i> .....	50
<b>5. Conclusion et perspectives .....</b>	<b>51</b>
<b>6. Bibliographie.....</b>	<b>53</b>

# 1. Introduction

## 1.1. La bioinformatique

### 1.1.1. Généralités

Depuis ces cinq dernières années, de grands projets de séquençage ont vu le jour et plusieurs ont déjà abouti grâce à un développement rapide des techniques de séquençage d'ADN (Boguski 1998). Des génomes entiers ont été mis à la disposition du monde scientifique, accroissant un déficit séquence/structure qui s'était déjà un peu installé auparavant (Fig.1). Voici quelques chiffres indiquant ce déficit : au début de l'an 2000, plus de 450000 séquences protéiques étaient connues alors que moins de 12000 structures protéiques tridimensionnelles étaient disponibles dans les banques de données. Ce déficit est compréhensible étant donné que l'information de séquence est beaucoup plus rapide à obtenir qu'une structure 3D protéique.

Actuellement, la bioinformatique a pour but de combler ce déficit et aussi d'annoter de nouvelles séquences. En effet, avant de connaître la structure tridimensionnelle d'une protéine, la bioinformatique peut aider à en savoir plus, par exemple sur ses fonctions ou sa localisation. Les annotations des séquences doivent s'automatiser afin d'obtenir un maximum de données sur la séquence protéique en un minimum de temps.

De plus, il faut pouvoir stocker et rendre disponible toute cette information (banques de données et programmes utilisés pour les traiter) aux chercheurs du monde entier. C'est à cette étape de la bioinformatique qu'intervient « Internet » (Ouellette 1998).

### 1.1.2. Recherche d'homologues

Deux séquences sont homologues quand elles proviennent d'un ancêtre commun. Homologue n'est pas un synonyme de similarité mais vu que l'homologie n'est pas quantifiable, on passe par un degré de similarité. Des séquences orthologues partagent la même fonction dans des espèces différentes. Par contre, on parlera de paralogues si deux séquences d'un même organisme ont une fonction similaire (Parry-Smith 1999).

La première possibilité pour obtenir des informations sur une séquence est de la comparer avec ses orthologues. Ceux-ci ne sont pas forcément identifiables car on ne connaît pas encore tous les génomes, on réalise donc une recherche d'homologues dans les différentes banques de données et on espère que le meilleur



résultat trouvé pour une espèce sera le vrai orthologue (Sternberg 1996). En effet, si on trouve des homologues, différentes analyses pourront être effectuées afin de déterminer si l'homologie de structure primaire se traduit par une fonction semblable.

Plusieurs méthodes ont été mises au point pour faire ces recherches d'homologie dont BLAST (Basic Local Alignment Search Tool) (Parry-Smith 1999).

Quand on soumet une séquence à BLAST, le programme considère les trois premiers résidus comme un mot. Ensuite, il génère tous les mots de trois lettres qui seraient similaires au premier mot, la similarité devant être supérieure à un seuil au choix de l'utilisateur. L'exemple suivant montre le cas d'une similarité dont le seuil T vaut "13" pour un mot de trois résidus YRN.

**Soit une séquence:** EGDNCVFDGMIYRNGETFQP

Tous les mots de 3 lettres pouvant être similaires au premier.	}	YRN	18	T = 13
		YKN	15	
		YQN	14	
		FRN	14	
		WRN	13	
		<u>YNN</u>	<u>13</u>	
		YRR	12	
		YAN	12	

Ces différents mots ne sont plus considérés comme des acides aminés quand BLAST fait l'analyse contre toutes les séquences des banques de données mais comme des lettres qui pourront être présentes sur les séquences de la banque. Pour effectuer cette comparaison, BLAST utilise la méthode du « suffix tree » (Fig.2). Quand un mot similaire est trouvé sur une séquence de la banque, cela s'appelle un HSP (« High-scoring Segment Pair »). Le programme recommence cette étape pour les résidus 2 à 4, etc... On obtient ainsi une collection d'HSP. Ensuite, le programme essaie d'étendre la similarité entre les résidus des deux séquences de chaque côté du HSP. Au moment où la similarité devient trop faible, le programme arrête l'extension, on obtient alors un MSP (« Maximum-scoring Segment Pair ») qu'il garde en mémoire. Quand le programme a parcouru toute la séquence, il compare les scores des différents MSP qui sont quantifiés par une E-Value et classe ceux-ci par ordre croissant.

Deux améliorations ont été apportées à BLAST : BLAST2 (ou « gapped BLAST ») et PSI-BLAST (« Position-Specific Iterated Blast ») (Altschul, Madden *et al.* 1997). La première permet l'utilisation d'indel (insertion – délétion) pour grouper deux HSP ne se chevauchant pas et situés à une certaine distance l'un de l'autre. Cela permet aussi d'augmenter la vitesse de recherche. La deuxième amélioration est une méthode itérative, capable de trouver des similarités faibles mais biologiquement significatives entre les séquences. Cette méthode est basée sur

la création d'une matrice de score « position-spécifique » générée par les alignements significatifs de l'itération  $i$  qui sera utilisée pour l'itération  $i+1$ .

BLAST utilise différents algorithmes selon la nature de la séquence d'intérêt et de la banque de séquences utilisée (tabl. 1 et 2).

Pour pouvoir faire une recherche optimale d'homologues, il est indispensable d'établir, de tenir à jour et de disséminer des banques de données afin de classer et de rendre aisément disponible toute cette information. Différentes banques de données existent, selon la nature de l'information qui y est stockée (cf. Matériel et Méthodes) (Parry-Smith 1999).

### 1.1.3. Les banques de données

Il est possible d'effectuer une recherche par mot-clés dans toutes les banques de données grâce au système SRS (Sequence retrieval system). Ce programme a été développé par l'Institut de Bioinformatique Européen (EBI).

ENTREZ est un autre système de recherche qui intègre toute l'information des banques de données du NCBI, incluant les séquences nucléotidiques, protéiques, les structures macromoléculaires, les génomes entiers et un « MEDLINE » (banque de publications dans le domaine biomédical) au travers de PubMed.

Ces systèmes sont indispensables pour la recherche d'informations scientifiques (Lewitter 1998).

#### 1.1.3.1. Les banques de données de séquences nucléotidiques

Les principales banques de séquences nucléotidiques sont GenBank (USA), EMBL (Europe) et DDBJ (Japon), elles collaborent et s'échangent fréquemment leurs données (Parry-Smith 1999). D'autres banques existent, notamment les banques de génomes complets. Ces banques comprennent les séquences provenant de la littérature scientifique et des projets de séquençage.

GenBank est une banque de données, créée en 1988, au Centre National pour l'Information Biotechnologique (NCBI) (Fig.3) (Benson, Boguski *et al.* 1999). EMBL (European Molecular Biology Library) a été créée par l'Institut européen de Bioinformatique (EBI). DDBJ (DNA Data Bank of Japan) créée en 1998, est entretenue par l'Institut National de Génétique (NIG) (Parry-Smith 1999).

#### 1.1.3.2. Les banques de données de séquences protéiques

##### 1.1.3.2.1. Les banques de données de séquences primaires

La première banque à avoir été créée est PIR (Protein Information Resource) au NBRF (National Biomedical Research Foundation) par Margaret Dayhoff en 1960. Depuis 1988, cette banque est sans cesse renouvelée grâce à un consortium

entre PIR au NBRF, JIPID (Japan International Protein Information Database) au Japon et MIPS (Martinsried Institute for Protein Sequences) (Parry-Smith 1999).

En 1986, Swiss-Prot est produite par le Département de Biochimie Médicale à l'université de Genève et par EMBL (Hamm and Cameron 1986). Cette banque se distingue par la richesse des informations qui accompagnent la séquence : description de la fonction de la protéine, structure de ses domaines, ses modifications post-traductionnelles,... (Fig.4). De plus, ce site fournit un certain nombre de liens susceptibles de donner plus d'information sur chaque protéine. Le stockage des séquences protéiques, dont les analyses sont insuffisantes pour une description claire de la protéine, se fait dans une banque appelée Tr-Embl (Translated EMBL) (Bairoch and Apweiler 1996) créée en 1996 en supplément à Swiss-Prot (Parry-Smith 1999). Le passage d'une banque à l'autre se fait après plusieurs annotations automatiques de la séquence vérifiées par des experts : caractéristiques de séquence (peptide signal, hélices transmembranaires, régions « coiled-coil »), la fonction de la protéine, prédictions de structures secondaires, recherche de similarité avec d'autres séquences,... (Bairoch and Apweiler 2000).

Selon les différentes analyses faites sur la séquence primaire, on peut regrouper les différents résultats dans des banques que l'on nomme banques secondaires (Parry-Smith 1999).

#### 1.1.3.2.2. Les banques composées

Pour faire une recherche efficace d'homologues d'une protéine chaque banque, cela demande un certain temps. De plus, il risque d'y avoir une certaine redondance.

Pour faciliter ce genre de recherche, des banques ont été composées au départ de plusieurs autres banques de données. Un exemple de banque composée est la banque de séquences non-redondantes (NRDB). Elle comprend l'information contenue dans quatre banques différentes : GenBank, Protein DataBank (PDB), Swiss-Prot (SP) et PIR. De plus, NRDB possède deux programmes pour renouveler SP et GenBank toutes les semaines pour la première et tous les jours pour la seconde (Parry-Smith 1999).

#### 1.1.3.2.3. Les banques de structures tertiaires

Avec l'utilisation de la diffraction aux rayons X par des cristaux de protéines et de la résonance magnétique nucléaire (RMN), de nombreuses structures 3D ont pu être déterminées. Elles ont été recensées dans la PDB (Ouellette 1998). Actuellement, cette banque compte environ 12000 entrées mais cela ne veut pas dire 12000 protéines différentes. En effet, de nombreuses structures de protéines mutantes sont présentes dans la banque en plus de la structure de la protéine sauvage. PDB accepte les structures protéiques issues de modélisation mais leur nombre est restreint par rapport au nombre total d'entrées dans PDB.

Dans cette banque, il existe deux degrés de fiabilité, la **résolution** (R) et le **facteur B**. Plus la résolution est faible, meilleure est la structure. Le facteur B représente le pourcentage du spectre de diffraction qui n'est pas expliqué par la

structure. Plus ce facteur est faible, mieux la structure proposée correspond au spectre.

Avant d'entrer dans PDB, les protéines sont dans une banque en attente, la « pending list ». Les chercheurs y placent les noms et descriptions des protéines qui seront bientôt dans la PDB. Cela permet aux autres scientifiques de tester différentes méthodes de modélisation et de pouvoir ensuite évaluer leur modèle et donc leur méthode.

#### 1.1.3.2.4. Les banques de classification de structures

Beaucoup de protéines partagent des similarités de structures reflétant, dans certains cas, une origine évolutive commune. La nature de l'information présentée par un schéma de classification de structure dépend de la méthode utilisée pour identifier et évaluer la similarité structurale.

Deux banques principales de classification existent, SCOP et CATH.

La banque de données SCOP (Structural Classification of Proteins) est construite avec une hiérarchisation qui reflète la relation entre la structure de la protéine et son origine. Les trois niveaux de classification principaux sont le repliement de la protéine, la super-famille et la famille (Parry-Smith 1999) (Murzin, Brenner *et al.* 1995). Les protéines sont classées comme ayant le même repliement (« fold ») si elles possèdent des structures secondaires identiques dans le même enchaînement (topologie identique), que ces protéines aient, ou non, la même origine évolutive. Les protéines sont classées en super-famille quand, malgré une faible identité de séquence, leurs caractéristiques structurales et fonctionnelles suggèrent un ancêtre commun. Les protéines d'une même famille possèdent une relation entre elles d'un point de vue évolutif. Les protéines sont classées dans une même famille lorsqu'elles ont des identités de séquences  $\geq 30\%$ . Notons que cette mesure dépend du programme d'alignement utilisé (Parry-Smith 1999).

Comme son nom l'indique, la banque de données CATH (Class, Architecture, Topology, Homology) possède quatre niveaux de hiérarchisation (Orengo, Michie *et al.* 1997) (Parry-Smith 1999). La classe est dérivée du contenu des structures secondaires. Cinq classes structurales sont reconnues pour les structures :

- majoritairement composées d'hélices  $\alpha$  (all  $\alpha$ ),
- majoritairement composées de brins  $\beta$  (all  $\beta$ ),
- composées de structures avec hélices  $\alpha$  et brins  $\beta$  alternés ( $\alpha/\beta$ ),
- composées de successions distinctes  $\alpha$  et  $\beta$  ( $\alpha+\beta$ )
- composées de peu de structures secondaires classiques (ex. : le collagène) (Parry-Smith 1999).

L'architecture est le niveau qui décrit l'arrangement et la succession des structures secondaires mais ignore leurs liens. La topologie donne une description qui englobe les liens entre les structures secondaires. L'homologie groupe des domaines qui partagent une identité de séquence  $\geq 35\%$  et dont on suspecte qu'ils partagent un ancêtre commun (Parry-Smith 1999).

#### 1.1.4. Alignements

Si des homologues sont trouvés, on peut passer à une étape suivante afin d'obtenir des informations sur la séquence d'intérêt au départ de ses homologues. On va générer des alignements pairés et/ou multiples entre la séquence d'intérêt et ses homologues (Ouellette 1998).

Un alignement peut être fait entre 2 séquences (on parlera alors d'alignement pairé) afin d'établir les correspondances entre les différents résidus. Par contre, lorsque la séquence d'intérêt possède plusieurs homologues, on effectuera des alignements multiples. Une similarité de séquence est plus significative si elle est partagée par plusieurs séquences.

L'alignement pairé est moins complexe que l'alignement multiple et demande moins de temps de calcul. C'est la méthode pairée qui est utilisée dans différentes banques de séquences pour retrouver les homologues d'une séquence d'intérêt excepté pour le programme Blast.

Lorsque des fragments de séquence sont conservés chez différents homologues protéiques, cela peut apporter des informations quant à la structure et la fonction de la séquence d'intérêt grâce à l'identification de motifs (Ouellette 1998).

Deux types d'alignements pairés et multiples existent : les alignements globaux et locaux. Les alignements globaux essaient d'ajuster les séquences dans leur entièreté (Schulder 1999). Si les séquences sont de tailles inégales, la plus petite sera remplie d'« indels » pour s'aligner sur toute la longueur de la grande séquence. Un exemple de logiciel d'alignement global est ClustalW (Thompson, Higgins *et al.* 1994).

Les alignements locaux ne doivent pas obligatoirement se faire sur la totalité des séquences. Les « indels » sont totalement absents des parties de séquences alignées. Ce type d'alignement convient mieux pour une recherche d'homologie de domaines, de conservation de site actif. Un exemple de programme d'alignement local est Match-Box (Depiereux and Feytmans 1992).

#### 1.1.5. Caractérisation de motifs

Un autre moyen de trouver des indications d'une fonction protéique est l'identification de motifs sur la séquence primaire. Cette analyse nécessite une banque de motifs et des outils informatiques qui exécutent efficacement la recherche (Sternberg 1996).

Une protéine inconnue peut posséder des caractéristiques de séquence particulières (motifs) qui la rattachent à une famille donnée de protéines ou qui sont structurellement significatives, pour exemple, le motif de liaison à l'ATP et au GTP (ATP/GTP-binding site motif A) dont le consensus est le suivant: [AG]X<sub>(4)</sub>GK[ST].

### 1.1.6. Localisation cellulaire

Certains programmes peuvent prédire la localisation sub-cellulaire de protéines grâce à la détection de signaux sur la séquence protéique. Pour les différentes analyses, un programme de référence existe: PSORT. Ce système peut décrire 83 % des sites de localisation de protéines d'une banque de données contenant 106 protéines (Nakai and Kanehisa 1991). PSORT regroupe plusieurs méthodes qui interviennent à différents moments de la prédiction de localisation.

La version de PSORT pour les bactéries Gram négatives disponible sur Internet est décrite ci-dessous (Fig.5).

Le programme débute par la détection d'une séquence signal permettant l'exportation à travers la membrane interne. A partir de cette séquence signal, le programme recherche la présence d'un consensus de clivage pour la « signal peptidase II » (SPII) (Fig.6), spécifique des lipoprotéines. Quand ce consensus est présent, PSORT analyse la portion N-Terminale de la protéine mature; quand elle possède une charge négative, elle sera localisée en membrane interne, sinon, la localisation se fait en membrane externe. Si ce motif est absent, un motif de clivage classique (« signal peptidase I » (SPI)) est recherché. Ce motif permettra la localisation de la protéine dans la membrane interne et le périplasme ou dans la membrane externe. La distinction entre les deux localisations se fait par une analyse de la composition des acides aminés (Fig.7). Signalons qu'un peptide signal peut ne pas avoir de site de clivage reconnu par SPI ou SPII. La protéine restera alors ancrée en membrane interne *via* cette portion hydrophobe.

Ensuite, le logiciel détectera les segments transmembranaires (TMS) qui sont nécessaires à l'insertion des protéines en membrane interne. Un TMS est suffisant à condition qu'il soit assez hydrophobe.

Enfin, une séquence qui ne possède aucune caractéristique « significative » sera par défaut localisée dans le cytoplasme.

### 1.1.7. Prédiction de structures secondaires

Les techniques de prédiction de structures secondaires sont les méthodes les plus utilisées généralement pour obtenir des informations structurales à partir de séquences nouvellement découvertes.

Elles sont impliquées dans une grande variété de problèmes liés aux protéines (Sternberg 1996). Par exemple :

- 1) l'assignement de structures secondaires peut aider à confirmer une relation structurale et fonctionnelle entre deux protéines.
- 2) la prédiction de structures secondaires est importante dans l'établissement d'alignements pendant la construction d'un modèle et dans la vérification de la validité du modèle cristallographique.

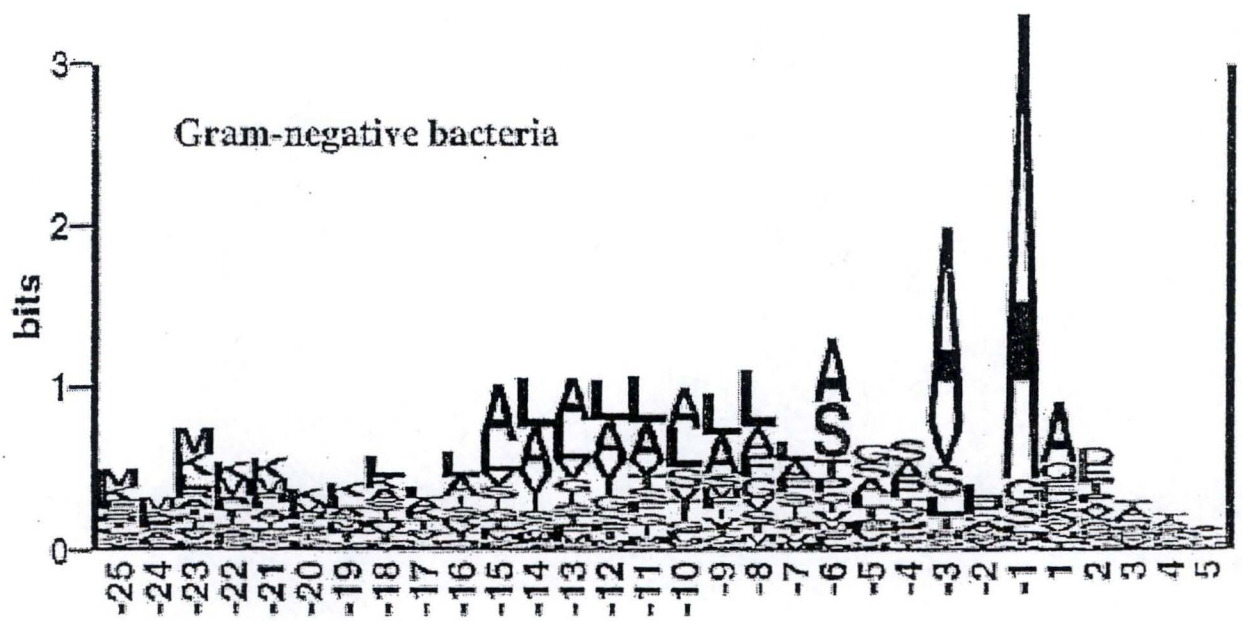


Fig.7 : Répartition des acides aminés dans la séquence des peptides signaux. Le site de clivage se trouve en position 0. Plus la probabilité de trouver un certain résidu à une position déterminée est grande, plus la lettre représentant cet acide aminé est de grande taille.

3) la prédiction de structures secondaires est aussi la première étape pour générer un modèle 3D d'une protéine [Sternberg, 1996 #114].

De nombreuses techniques de prédiction de structures secondaires existent mais la plus renommée actuellement est PHD (Rost and Sander 1993). PHD est une méthode qui utilise des réseaux neuronaux constitués de différentes unités. Celles-ci intègrent leurs propres données et les transforment en réponses, qui sont transmises aux autres unités connectées en parallèle. Ces systèmes sont, en premier lieu, entraînés sur plusieurs protéines tests et peuvent ensuite être utilisés pour des protéines inconnues (Rost, Sander *et al.* 1994). PHD exécute plusieurs analyses sur la séquence soumise : les prédictions de structures secondaires (PHDsec) et d'accessibilité au solvant ainsi que la recherche d'hélices transmembranaires.

On peut prédire différentes structures secondaires car elles ont chacune leurs caractéristiques propres. Par exemple, l'hélice  $\alpha$  qui a une préférence pour des résidus non chargés ou de petit volume comme l'alanine, la leucine ou la phénylalanine. La proline est souvent absente d'une hélice car elle perturberait sa conformation. D'autres résidus, comme l'arginine, la sérine et le glutamate qui sont polaires, ont tendance à se repousser et donc perturber aussi la conformation de l'hélice (Rawn 1990) (Creighton 1993).

#### 1.1.8. Prédiction de structure tertiaire

Le but des techniques de prédiction de structure tertiaire est de passer de la séquence protéique d'un polypeptide à sa structure tridimensionnelle. Pour cela, différentes méthodes peuvent être employées : soit on utilise un programme pour chaque étape de modélisation, soit on utilise un logiciel qui effectue toutes les étapes tout seul.

L'idéal serait des programmes qui exécutent toutes les opérations nécessaires à la réalisation du modèle 3D d'une protéine inconnue au départ de sa seule séquence. Il en existe déjà mais leurs résultats ne sont pas dignes de confiance, et ils nécessitent un temps de calcul considérable. De ce fait, ce type de modélisation (*ab initio*) n'est applicable que pour de petites protéines.

##### 1.1.8.1. Modélisation par homologie

Pour connaître la structure tertiaire d'une protéine, on peut utiliser la modélisation par homologie (« homology modeling »). Cela requiert deux conditions. La première est élémentaire: la séquence primaire de la structure d'intérêt doit être connue. La seconde est que cette protéine ait un homologue dont la structure 3D est connue et qui servira de protéine de référence (« template »). En effet, deux séquences similaires ont une forte probabilité d'avoir une structure tertiaire identique (Doolittle 1981). La faisabilité de l'alignement dépend du taux d'identité entre les deux séquences :



- 80 %, l'alignement est possible et facile
- 50 à 80 %, l'alignement est possible
- 25 à 50 %, l'alignement est possible mais difficile
- 25 %, la question de faisabilité de l'alignement se pose

Une bonne modélisation par homologie dépendra donc du taux de similarité entre la séquence d'intérêt et la séquence de référence. C'est pour cela que les techniques de comparaison de séquences sont très importantes. La modélisation par homologie s'effectue en plusieurs étapes (Vinals 1996):

- 1) Analyse de la composition en acides aminés de la séquence d'intérêt et clivage de cette protéine en domaines distincts.
- 2) Recherche d'un maximum de protéines homologues. Plus le nombre de protéines homologues intégrées dans l'alignement est grand, plus les régions conservées (SCR) qui sortiront de l'alignement seront fiables.
- 3) Alignement des séquences des protéines homologues avec la séquence d'intérêt.
- 4) Etablissement des correspondances entre la séquence cible et la (les) structure(s) similaire(s) connue(s) au départ de l'alignement.
- 5) Construction d'une partie du squelette sur base des SCRs d'une part, des VR (région variable) d'autre part.
- 6) Positionnement des chaînes latérales.
- 7) Mécanique et dynamique moléculaire pour affiner le modèle. L'affinement demande un important temps de calcul (contrairement aux autres étapes).

Toutes ces étapes peuvent être effectuées par différents programmes mais d'autres logiciels automatisés existent et effectuent la totalité des étapes. SWISS-MODEL en est un exemple. Actuellement, il donne peu de résultats car il est très sélectif quand les conditions d'homologie entre la séquence d'intérêt et le « template » ne sont pas optimales. Par exemple, il ne donne aucun résultat quand l'homologie est inférieure à 25 %. Ces programmes restent des boîtes noires pour un utilisateur moyen : les paramètres sont difficilement modifiables, on ne peut contrôler aucune étape.

#### 1.1.8.2. Reconnaissance de « fold »

Quand une protéine inconnue ne possède pas d'homologue dont la structure 3D est connue, il est possible de trouver un repliement pour cette protéine parmi les structures tridimensionnelles connues. En effet, même en l'absence de similarité de séquence ou de fonction, des protéines peuvent adopter des « folds » similaires (Orengo 1994).

Le but des méthodes de reconnaissance de « fold » est d'essayer de trouver des repliements protéiques qui sont compatibles avec une séquence particulière. Dans ces techniques, on se pose la question à l'envers: plutôt que de prédire comment une

protéine se replie, ces méthodes prédisent si une conformation connue convient à la séquence d'intérêt. Les techniques de reconnaissance de « fold » peuvent être classées en deux catégories : le threading et le pseudo-threading.

#### 1.1.8.2.1. Threading

Ces méthodes sont basées sur des algorithmes qui « enfilent » (to thread) la séquence de la protéine d'intérêt sur la structure 3D connue d'une autre protéine et qui cherchent un alignement correspondant à un modèle structural énergétiquement favorable. Ces techniques sont basées sur un calcul d'énergie (potentiel de force moyenne). Quelques logiciels tel que ProFIT (Sippl 1993) et THREADER (Miller, Jones *et al.* 1996) effectuent ces prédictions.

#### 1.1.8.2.2. Pseudo-threading

Ces méthodes n'utilisent pas de potentiels empiriques. Elles détectent *et alignent* des motifs similaires de structure secondaire et d'accessibilité entre une séquence de structure inconnue et un « fold » connu. Elles sont basées sur le principe des réseaux neuronaux. Quelques logiciels effectuent le pseudo-threading : TOPITS (Rost 1995) (Rost, Schneider *et al.* 1997), 123D (Alexandrov, Nussinov *et al.* 1996), 3D PSSM (Fischer, Barret *et al.* 1999) et le site de l'Ucla-Doe.

Une fois le repliement prédit par threading ou pseudo-threading, on peut l'argumenter à l'aide d'un consensus de prédictions de structures secondaires. Il sera réalisé à l'aide de différents logiciels de prédictions. Si le consensus de structure secondaire de la séquence d'intérêt possède, par exemple, des hélices  $\alpha$  et que le « fold » prédit ne possède que des brins  $\beta$ , ce ne sera pas correct.

La dernière étape est l'alignement de la séquence d'intérêt à la structure tertiaire.

## 1.2. *Entre théorie et pratique*

Dans un futur proche, tous les biologistes moléculaires et cellulaires devront pouvoir se servir d'outils informatiques afin d'optimiser leur travail en laboratoire. En effet, actuellement la bioinformatique pour elle-même n'est plus le seul objectif, elle peut aussi être très utile dans l'étude d'un contexte biologique.

Les analyses théoriques, effectuées grâce à un ordinateur, ont plusieurs avantages : quand les chercheurs possèdent une protéine pour laquelle ils se posent des questions, ils peuvent avoir un premier contact avec cette protéine d'intérêt de façon relativement immédiate et éventuellement cibler les différentes expériences à pratiquer en laboratoire. Cela restreint le temps passé à chercher quelle direction prendre pour étudier cette protéine.

Certaines unités de recherche ont regroupés les équipes de théorie et de recherche en laboratoire afin qu'il puisse plus facilement y avoir un échange entre ces deux groupes.

Dans ce travail, nous avons utilisé des outils bioinformatiques pour étudier une question biologique dont le contexte décrit ci-dessous.

### **1.3. Le Contexte biologique**

#### 1.3.1. Les bactéries

##### 1.3.1.1. Gram positive, Gram négative

Historiquement, les bactéries ont été classées en deux catégories par l'utilisation d'une méthode de coloration développée par C. Gram : Gram positif et Gram négatif. Cette différence de coloration reflète une différence de structure de l'enveloppe. En effet, les Gram positives, comme *Bacillus subtilis*, ont une seule membrane appelée la membrane cytoplasmique. Elle est entourée d'une paroi épaisse de peptidoglycane et d'acide teichoïque. Les Gram négatives, comme *Escherichia coli*, ont deux membranes : la membrane interne, qui est similaire à la membrane des bactéries Gram positives, et la membrane externe. L'espace situé entre les deux membranes est nommé périplasme et contient une fine couche de peptidoglycane (Fig.8).

L'ultrastructure d'une bactérie peut sembler simple, comparée à celle des eucaryotes, mais il n'en reste pas moins qu'elle possède des compartiments différents. Ainsi, les bactéries Gram positives possèdent trois sites distincts d'adressage de protéines (le cytoplasme, la membrane, l'extérieur) et les bactéries Gram négatives en ont cinq (le cytoplasme, la membrane interne, le périplasme, la membrane externe et l'extérieur).

##### 1.3.1.2. Systèmes de transport

Les membranes sont constituées essentiellement d'une double couche de phospholipides formant une barrière hydrophobe qui ne laisse passer que l'eau et de petites molécules hydrophobes. Il existe des transports, spécifiques ou non, pour différentes molécules, protéiques ou non, pour qu'elles puissent traverser la membrane (Fig.9). Ces transports doivent pouvoir se faire dans les deux sens. En effet, la bactérie est un organisme qui, comme tel, doit interagir avec son milieu. S'il lui faut des molécules qu'elle ne peut synthétiser, elle doit les incorporer (importation) et comme la bactérie doit pouvoir rendre son environnement favorable à son développement, elle doit pouvoir exporter des substances qui le modifieront (John E. Cronan 1987).

Les prochains développements seront basés sur les bactéries Gram négatives.

### 1.3.1.2.1. Transport « Out –In »

Certaines substances peuvent traverser par diffusion simple ou facilitée. La diffusion simple d'une protéine est son passage au travers d'une membrane grâce à un gradient de concentration. Pour la diffusion facilitée, la molécule traverse la membrane à l'aide de transporteurs. Pour les deux transports, cela se fait grâce à un gradient de concentration de la molécule transportée.

Exemple : La diffusion facilitée du glycérol chez *Escherichia coli*. La protéine qui facilite l'entrée du glycérol est codée par le gène *glpF*, elle forme un pore, assez spécifique, dans la membrane cytoplasmique qui permet l'influx et l'afflux de glycérol. De ce fait, les concentrations interne et externe à la cellule sont vite équilibrées. L'entrée de glycérol ( $6 \cdot 10^5$  molécules/sec/pore), est couplée avec sa conversion en sn-glycerol 3-phosphate par une enzyme appelée : glycérol kinase. Cette nouvelle molécule ne peut repasser par le pore, elle est donc emprisonnée à l'intérieur de la cellule. L'entrée est indépendante de la température et ce transport ne nécessite pas d'énergie métabolique. (John E. Cronan 1987).

Cependant la plupart des molécules nécessitent un transport actif. On retrouve dans les bactéries trois types de transports actifs.

Le gradient chimi-osmotique est un transport qui ne nécessite qu'une protéine située intégralement en membrane. Le substrat est transporté par symport ou antiport avec un ion. L'énergie requise pour concentrer le substrat à l'intérieur de la cellule est obtenue soit directement (symport de proton) soit indirectement (symport de sodium) au départ du gradient chemiosmotique (John E. Cronan 1987).

Exemple : Le transport du lactose par *Escherichia coli*. Le lactose est transporté par la perméase au lactose codée par le gène *lacY*. C'est une protéine très hydrophobe, située intégralement en membrane. La prédiction de structure secondaire basée sur la séquence ADN de *lacY* trouve 8 à 12 hélices  $\alpha$  hydrophobes qui traversent la membrane. Cela peut former un canal qui traverse la membrane où les régions hydrophobes vont interagir avec les phospholipides membranaires et où l'intérieur du canal sera hydrophile. D'après des études sur la perméase, on sait que ce transport ne nécessite que cette protéine. La perméase lie un proton en premier lieu, cela l'active afin qu'elle puisse lier une molécule de lactose. Ensuite, le complexe change de conformation et libère lactose et proton à l'intérieur de la cellule. La libération du proton est l'étape limitante de ce transport qui est beaucoup plus lent que le contreflux de lactose. Ce contreflux est l'échange d'une molécule de lactose interne contre une externe, cela ne demande pas d'énergie et n'apporte pas de gradient net de lactose entre les deux milieux. Ce flux de lactose implique la rapide libération du lactose sans libération du proton. Pour que le proton lié soit libéré, il faut une concentration du lactose à l'intérieur de la cellule. Donc, la perméase transforme un gradient chimi-osmotique ( $[H^+]_{ext} > [H^+]_{int}$ ) en un gradient de lactose ( $[lactose]_{ext} > [lactose]_{int}$ ) (John E. Cronan 1987).

Le transport dépendant d'une liaison à une protéine demande une haute affinité entre le substrat et la protéine qui le lie pour le transporter. L'énergie nécessaire pour ce transport est obtenue par phosphorylation du substrat mais le mécanisme

précis est encore inconnu. Environ 40% des substrats d'*Escherichia coli* qui entrent dans la cellule utilisent ce transport (John E. Cronan 1987).

Exemple : La perméase à histidine d'*Escherichia coli*. Quatre protéines sont requises pour ce transport : HisJ qui est une protéine soluble du périplasma liant l'histidine, HisQ et HisM qui sont des protéines situées intégralement en membrane et HisP qui est une protéine dimérique associée à la membrane. HisJ a, à son extrémité N-Terminale, un peptide signal qui lui permet de passer la membrane interne et d'être ainsi libérée dans le périplasma. Elle possède deux domaines, l'un lie l'histidine et l'autre interagit avec HisP. HisQ et M ont des séquences très similaires. Elles sont très hydrophobes et les prédictions de structures secondaires déterminent trois hélices  $\alpha$  composées d'environ 20 résidus et qui traversent la membrane. Leur rôle est de former un canal par lequel le substrat est transloqué. HisP interagit avec le périplasma grâce à HisJ, il est amphipatique. HisQ, HisM ou HisP n'ont pas de peptide signal comme les protéines exportées (John E. Cronan 1987). Ce transport requiert de l'énergie qui est fournie par HisP qui hydrolyse l'ATP (Nikaido and Ames 1999).

Aucun des deux transports précédents ne modifie le substrat pendant le transfert. Ce n'est pas le cas du dernier mécanisme de transport : la translocation de groupement. Ce type de transfert couple le transport du substrat avec sa modification chimique, ce qui fournit l'énergie au système et cela empêche le substrat de ressortir de la cellule. Ce système nécessite plusieurs composants membranaires (John E. Cronan 1987).

Exemple : Le transport du mannitol dépendant du phosphoénol pyruvate (PEP) chez *Escherichia coli*. Ce transport requiert de l'énergie, qui provient du PEP, et trois protéines : l'enzyme I (*ptsI*), HPr (*ptsH*) et l'enzyme II (*II<sup>mtl</sup>* dont le gène est *mtlA*). Les deux premières protéines, situées en périphérie de la membrane, initient le transfert du groupement phosphoryl du PEP vers la protéine de transport liée à la membrane. En effet, elles sont en contact avec les composants de la machinerie de transport du mannitol. Le groupement phosphoryl est transféré, en premier lieu, sur l'enzyme I et ensuite sur HPr. Quand ce dernier possède le groupement, il lie le domaine cytoplasmique de l'enzyme II qui est spécifique au mannitol et qui peut catalyser un échange de groupement. Cette protéine, qui traverse la membrane et forme un canal grâce à 7 régions très hydrophobes, va lier le mannitol qui se fera phosphoryler pendant sa translocation. Ce qui est relâché dans le cytoplasme est le mannitol 1-phosphate. (John E. Cronan 1987).

Ces trois transports diffèrent par la forme d'énergie requise, par le nombre et la localisation des protéines nécessaires au transport et par la forme du substrat transporté.

#### 1.3.1.2.2. Transport « In – Out »

Ce type de transport peut, évidemment, se faire pour différentes molécules : sucres, lipides et protéines.

Ici, nous allons développer le cas particulier du transport protéique. Les polypeptides peuvent être envoyés vers différentes localisations selon leur fonction

(Driessen 1999). Les différents systèmes de transport de ces protéines sont répartis en deux grandes classes selon le nombre de membranes traversées.

On parlera d'**exportation** quand les protéines ne traversent que la membrane interne (du moins en partie) et de **sécrétion** quand les protéines passent au travers des deux membranes et que leur transport au niveau de la deuxième membrane s'effectue de façon spécifique.

#### 1.3.1.2.2.1. L'exportation

##### 1.3.1.2.2.1.1. Les différents signaux

Les protéines peuvent avoir deux types de signaux d'adressage:

##### - Le peptide signal « classique »

Ce signal est situé à l'extrémité amino-terminale (N-Terminale) et est clivé pendant le processus de l'exportation par des peptidases. Les peptides signaux doivent avoir des caractéristiques communes afin de pouvoir être reconnus par les signal-peptidases. La longueur de ces signaux varie entre 18 et 30 acides aminés et sont divisés en trois : une région N-terminale composée de résidus chargés positivement, un noyau central de résidus hydrophobes qui forment une hélice  $\alpha$  et une extrémité C hydrophile à la fin de laquelle on trouve le site de clivage (Fig.10). Une mutation de résidus hydrophobes en résidus non-hydrophobes dans la deuxième région du peptide signal entraîne une accumulation de précurseurs dans le cytoplasme de la cellule. La dernière région n'est pas essentielle à l'exportation. Une mutation dans cette partie ralentit ou supprime la maturation (clivage du peptide signal) mais n'affecte pas l'exportation du précurseur (Oliver 1987).

Une variante de ce peptide signal existe, c'est celui de type « lipoprotéine » où le consensus du site de clivage est différent. En effet, la présence d'une cystéine est indispensable pour ces protéines qui resteront ancrées en membrane après le clivage de leur peptide signal par la signal-peptidase II (Driessen 1999).

##### - Des segments transmembranaires

Pour les protéines de membrane interne, en plus du peptide signal qui n'est pas forcément présent, des signaux supplémentaires sont utiles pour un bon adressage. Ce sont des segments hydrophobes qui sont internes à la séquence protéique et qui se replieront en hélices (Beckwith 1996).

##### 1.3.1.2.2.1.2. Les différents voies d'exportation

Trois mécanismes d'exportation sont à distinguer (Fig.11).

##### 1.3.1.2.2.1.2.1. Sec-dépendant

Ce type de transport est le système d'exportation le plus étudié. C'est par lui que transitent la plupart des protéines devant être exportées.

Le système Sec-dépendant utilise plusieurs partenaires, dont six ont été identifiés. SecB est une protéine chaperonne cytoplasmique qui forme un tétramère. SecA est un dimère situé à la périphérie de la membrane interne. Elle contient un domaine pour lier et hydrolyser l'ATP, et des domaines d'interaction avec l'ARN et

d'autres protéines. Ce système se compose aussi de protéines complètement localisées en membrane interne : SecD, E, F et Y qui forment la translocase. SecB lie la protéine d'intérêt et l'amène près du complexe SecA qui a un mécanisme très spécifique pour lier le précurseur (c'est la région du 267 au 340ème résidu de SecA qui reconnaît le peptide signal de la protéine). La liaison de l'ATP résulte de la libération du précurseur, et donc du passage de celui-ci vers la translocase. Tout cela induit un changement de conformation de SecA qui sera alors capable de lier à nouveau un précurseur. L'ATP est hydrolysé et SecA peut alors insérer le précurseur en membrane. On ne sait pas si l'interaction SecA / SecB se fait grâce au précurseur *via* la reconnaissance du peptide signal par SecA, ou si SecA a un site d'interaction avec SecB. SecY est une protéine qui traverse dix fois la membrane interne et ses deux extrémités sont situées dans le cytosol. C'est pour cela qu'on suppose qu'elle forme, soit avec elle-même, soit par interaction avec d'autres protéines, un pore par lequel les précurseurs transloquent. Il semble qu'elle constitue le corps de l'appareil d'exportation. Le peptide signal sera clivé, chez *Escherichia coli*, par les peptidases LspA (pour la signal peptidase de type II) ou par Lep (signal peptidase de type I). La première a pour cible les lipoprotéines et la seconde, toutes les autres protéines exportées possédant une peptide signal classique. Le peptide signal, une fois clivé, est dégradé par des protéases (Beckwith 1996).

Exemple : MalE chez *Escherichia coli*. Cette protéine traverse la membrane interne grâce au système Sec-dépendant. Son rôle est de lier le maltose dans le périplasme.

#### 1.3.1.2.2.1.2.2. SRP

Ce système utilise une ribonucléoprotéine, SRP (signal recognition protein), qui a une forte affinité pour les régions hydrophobes des protéines. SRP lie la protéine pendant qu'elle est traduite. Ensuite, il la conduit vers un homologue de SecA, FtsY, qui se liera à une translocase (probablement la même que pour le système Sec-dépendant). Le récepteur FtsY est associé à la membrane interne. Il possède deux domaines : une région fortement chargée, le domaine A et une région C-terminal, le domaine NG. Ce dernier domaine est divisé en deux segments : le N-Terminal et une GTPase. L'association à la membrane se fait *via* le domaine A (Driessen 1999).

Exemple : MtlA chez *Escherichia coli*. Le produit de ce gène est un composant de la perméase au mannitol (PTS : phosphoenolpyruvate-dépendant sugar phosphotransferase system). Cette protéine se localise en membrane interne avec un domaine périplasmique (Konstanze Beck 2000).

En analysant des protéines qui traversaient la membrane interne en l'absence de composants du système Sec et en l'absence de translocase, on a pu remarquer qu'elles possédaient un motif conservé en N-Terminal, RRXFXK, essentiel pour l'efficacité de cette translocation. On nomme ce système de transport TAT (twin arginin transfert) route en raison des deux arginines (R) qui se suivent. Les composants du système TAT sont encodés par l'opéron *tatABCD* et par *tatE* (Cristobal, de Gier *et al.* 1999).

Les molécules qui utilisent ce système sont des protéines impliquées dans les réactions rédox du périplasma. Le mécanisme de ce transport est encore très mal compris étant donné que c'est une voie découverte très récemment. La protéine doit avoir sa conformation active pour pouvoir se faire transporter, ce qui est différent des autres types de transport. De plus, elle semble devoir être liée à son (ses) cofacteur(s) avant l'exportation (Driessen 1999).

Certaines protéines n'ont apparemment pas besoin de mécanisme particulier pour traverser la membrane interne. Elles possèdent des régions suffisamment hydrophobes que pour se diriger et traverser la membrane interne seule.

Ce qui se passe en réalité est loin d'être parfaitement défini. Quand une protéine naissante émerge du ribosome, un grand nombre de chaperonnes, de facteurs d'adressage et d'enzymes de repliement conformationnel attendent de pouvoir se lier à cette nouvelle molécule. A ce moment, les différents facteurs d'adressage entrent en compétition afin de lier le précurseur.

La première chaperonne à entourer le précurseur est le facteur « trigger ». Il a la capacité de maintenir la conformation du précurseur qui lui permettra de transloquer (Hesterkamp and Bukau 1996). Un autre candidat pour lier le précurseur est DnaJ. Cette chaperonne lie le segment N-Terminal du précurseur. Elle est accompagnée de DnaK et de GrpE (en présence d'ATP). Ce complexe peut séparer le précurseur du ribosome (Kudlicki, Odom *et al.* 1996). Quand la protéine naissante semble posséder une séquence signal hydrophobe, cette équipe de chaperonnes est en compétition avec le SRP. En contraste avec les autres chaperonnes, le seul facteur, qui peut dissocier SRP du précurseur, se trouve en membrane cytoplasmique. Cela permet à SRP de mettre hors concours les autres compétiteurs et cela stabilise son interaction avec la séquence signal jusqu'à ce qu'il arrive à destination. Quand la séquence n'est pas assez hydrophobe, SRP ne peut interagir avec le précurseur. Cela permet à d'autres chaperonnes d'intervenir. SecB est l'une d'entre elles. Elle peut se lier à la protéine naissante à condition que cette dernière ne soit pas trop courte. La stabilité de cette interaction dépend du type de précurseur. Comme SRP, SecB ne libère son substrat qu'après interaction avec SecA qui est situé en membrane interne (Driessen 1999).



#### 1.3.1.2.2.2. La sécrétion

Les protéines sécrétées sont nombreuses et diverses d'un point de vue fonctionnel. Or, il n'existe que quatre types de transport pour faire passer ces protéines du cytoplasme bactérien vers l'espace extracellulaire ou vers une autre cellule (Hueck 1998).

##### 1.3.1.2.2.2.1. Type I

La machinerie de sécrétion de type I est de type ABC transporteur (ATP binding cassette).

Tous les membres de la famille de protéine utilisant la machinerie de sécrétion de type I partagent une répétition de neufs résidus caractéristiques : GGBGBBXLX ou B représente un aspartate (D) ou une asparagine (N) (Holland 1996). Les protéines qui utilisent ce transport ne possèdent pas de peptide signal en N-Terminal mais elles ont un signal de sécrétion en C-Terminal qui ne sera pas clivé. Ce signal est encore inconnu. Pour ce transport, il n'y a pas d'intermédiaire périplasmique (Lory 1998). En effet, des études topologiques et de fractionnement cellulaire ont montré que les composants des membranes interne et externe interagissent entre eux pour former un canal transmembranaire par lequel les protéines peuvent être sécrétées (Reeves 1993). Ce transport requiert trois protéines : une ATPase, située en membrane interne et qui confère l'énergie au système, une protéine de membrane externe et une protéine de fusion. Cette dernière, ancrée dans la membrane interne, s'étend dans le périplasme et a une interaction avec la membrane externe (Hueck 1998).

Exemple :  $\alpha$ -hemolysin d'*Echerichia coli*. Cette protéine, de 107Kd, est une toxine extracellulaire codée par *hlyA*. Elle se lie aux membranes des eucaryotes et provoque leur lyse. Elle se situe intégralement en membrane externe de l'hôte mais elle n'interagit qu'avec une monocouche de phospholipides grâce à ses dix hélices  $\alpha$  amphiphatiques (Ana Soloaga 1999). Les gènes, codant pour son transport, sont nommés : *hlyB*, *hlyD* et *tolC*. Les deux premiers gènes sont transcrits et traduits en protéines de membrane interne et TolC est une protéine de membrane externe possédant une structure de type porine. L'ensemble forme un complexe qui traverse les deux enveloppes des bactéries. Le domaine C-terminal de l'hémolysine est reconnu par HlyB. HlyD possède un grand domaine périplasmique et son domaine C-terminal est impliqué dans l'interaction avec la membrane externe (Fig.12) (Holland 1996).

##### 1.3.1.2.2.2.2. Type II

C'est le système de sécrétion le mieux connu (Reeves 1993). C'est le seul transport qui s'effectue en deux étapes. Les protéines sont, en premier lieu, exportées dans le périplasme grâce au système Sec-dépendant puis, en second lieu, sécrétées au travers de la membrane externe. Tous ces transports se font à l'aide soit d'un domaine « helper » soit grâce à des protéines « helper ». Cela peut se faire de trois manières différentes :

Soit la protéine possède un tonneau  $\beta$  à l'extrémité carboxy-terminale (C-Terminal), qui lui sert de domaine « helper » et il ne lui faut rien d'autre. Les protéines qui utilisent ce type de transport sont des auto-transporteurs.

Exemple : IgA protéase de *Neisseria gonorrhoeae*. Cette protéine est synthétisée sous la forme d'un précurseur de 169kD qui possède cinq domaines fonctionnels différents : le peptide signal en N-terminal, le domaine protéasique (106kD), le petit  $\gamma$ -domaine (30 résidus), l' $\alpha$ -domaine (12kD) et le  $\beta$ -domaine en C-terminal (45kD) (Thomas Klauser 1993). Après le clivage du peptide signal, la protéine se retrouve dans le périplasma. Elle est, rapidement, incorporée à la membrane externe *via* son domaine C-terminal, le  $\beta$ -domaine. Celui-ci va former un pore dans la membrane externe qui permettra la translocation des domaines protéasiques,  $\alpha$  et  $\gamma$  vers l'extérieur. Ensuite, l'enzyme se replie dans sa conformation active et est scindée du  $\beta$ -domaine qui sera ôté de la membrane externe par clivage autoprotéolytique. Le précurseur extracellulaire subit ensuite deux clivages qui libéreront les domaines  $\alpha$  et  $\gamma$  (Wandersman 1996) (Fig.13).

Ou bien, le transport de la protéine au travers de la membrane externe requiert une protéine « helper ».

Exemple : La sécrétion de l'hémolysine de *Serratia marcescens*. ShlA est une grande protéine. Elle possède un peptide signal et passe la membrane interne *via* le système Sec-dépendant. Le passage de la membrane externe se fait à l'aide de ShlB. L'expression de ces deux protéines est suffisante pour la sécrétion de l'hémolysine. En l'absence de ShlB, ShlA s'accumule dans le périplasma (Wandersman 1996).

Enfin, il peut y avoir formation d'un canal entre le périplasma et l'extérieur grâce à 12 à 14 protéines helper (Driessen 1999).

Exemple : La sécrétion de pullulanase par *Klebsiella oxytoca*. La pullulanase est une enzyme catabolique qui hydrolyse le lien  $\alpha(1-6)$  des polymères de glucose. Le gène qui code pour cette enzyme est nommé *pulA*. En plus du système Sec, un set complémentaire de 14 facteurs de sécrétion, les 'pul', est nécessaire pour sa sécrétion. Les gènes sont organisés en un opéron. Il contient les gènes dont les protéines constituent le complexe de transport et les gènes qui codent pour la pullulanase. Au moins 7 de ces protéines sont localisées en membrane interne alors que PulS et PulD sont situées en membrane externe (Hueck 1998).

#### 1.3.1.2.2.2.3. Type III

L'appareillage de sécrétion de type III est responsable de la libération de protéines cytoplasmiques de la bactérie directement dans le cytosol de l'hôte (Cornelis, Biot *et al.* 1989) (Cornelis, Boland *et al.* 1998).

La machinerie est composée d'au moins 17 polypeptides situés en membrane interne et externe. De plus, ce transport possède une ATPase cytoplasmique associée à la membrane interne qui fournit l'énergie au système (Hueck 1998). Les protéines utilisant ce transport ne possèdent pas de peptide signal classique en N-Terminal (Lory 1998). Deux signaux seraient présents, l'un serait les 15 premiers codons de l'ARN messager et l'autre serait le site de liaison à la chaperonne.

Exemple : Le système Yop de *Yersinia spp* (Guy R. Cornelis, 1998). Quand la bactérie entre en interaction avec un macrophage, l'appareil de sécrétion Ysc est installé et un stock de protéines Yop est synthétisé. Onze protéines, qui agissent à différents moments de l'infection, sont sécrétées par le transport de type III. La sécrétion se fait de façon co-translationnelle (Schneewind 1999). Les gènes, dont les protéines sont impliquées dans la machinerie de sécrétion, s'organisent en trois opérons. Les protéines Yop sécrétées ont des effets antiphagocytaires pour la cellule cible (Hueck 1998). Les protéines YopB et D ont une action de 'translocation'. En plus de ces polypeptides sécrétés, de petites protéines cytoplasmiques ont été identifiées comme guidant les YopD, E, H vers la machinerie de sécrétion. Elles portent, respectivement, les noms de SycD, E et H. Elles constituent des chaperonnes individuelles (Fig.14)(Wandersman 1996).

#### 1.3.1.2.2.4. Type IV

Ce transport utilise plusieurs composants qui formeront un complexe traversant les deux membranes d'une bactérie.

Il sert à la sécrétion et parfois à l'injection de complexe protéine/ADN, de toxines et autres molécules à l'intérieur de la cellule cible. Depuis ces dix dernières années, plusieurs données ont montré que ce transfert et la conjugaison bactérienne sont analogues, sinon équivalents. Les protéines du transport de type IV ont des homologues dans le système de conjugaison d'*Escherichia coli* (Tab.3). Il existe aussi des similarités entre le transport d'ADN et le transport de protéines comme la sécrétion de la toxine de pertussis par *Bordetella pertussis*, ou le système d'injection d'ADN-T chez *Agrobacterium tumefaciens* et cela, jusque dans l'organisation de l'opéron.

Un modèle schématique du système de transport de type IV a été représenté en se servant d'un consensus de localisation des différents composants chez les différents homologues (Fig.15) (A. Covacci 1999) (Christie 1997).

Le transport d'*Agrobacterium tumefaciens* consiste en un assemblage de sous-unités qui traverse les deux membranes des bactéries Gram négatives. Ce transport est expliqué en détail dans le point suivant.

### 1.3.2. Le système de sécrétion de type IV chez *Agrobacterium tumefaciens*

*Agrobacterium tumefaciens* est une bactérie extracellulaire. Elle transforme les cellules de l'hôte en cellules tumorales provoquant, ainsi, la galle du collet chez les plantes.

La virulence de cette bactérie est due au plasmide Ti. Les gènes contenus dans ce plasmide codent pour la virulence d'*Agrobacterium tumefaciens*. Ces gènes induisent la formation de la tumeur et sont entourés de deux séquences en répétitions directes de 23 bases. Cette région qui porte le nom d'ADN-T est injectée via le transport de type IV et induit la formation de la tumeur. L'ADN-T contient

des gènes pour la synthèse d'hormones de croissance végétale et de protéines synthétases d'opines, source de carbone et d'azote d'*Agrobacterium tumefaciens* (Prescott 1995).

Chez *Agrobacterium tumefaciens*, le transport de type IV se compose de douze protéines appelées Vir (pour virulence). Leur transcription est activée par un système à deux composantes : VirA et VirG (Thorstenson and Zambryski 1994). VirA est une protéine impliquée dans la perception des phénols, comme l'acetosyringone (AS), générés par la blessure d'une plante. VirG est un activateur de transcription du régulon *vir* qui est induit en présence d'AS (Karine E. Finberg 1995). Le transfert vers la plante infectée requiert dix protéines de l'opéron *virB* et une protéine supplémentaire nommée VirD4 (Christian Baron 1997).

#### 1.3.2.1. *VirB1*

VirB1 se situe, de manière prédominante, en membrane interne et est composée de 239 acides aminés. On peut séparer en deux les activités de VirB1. La première activité est celle de la région N-Terminale, qui est homologue au lysosyme (Fullner 1998), et qui dégrade le peptidoglycane. Ceci permet de bien situer toute la machinerie de sécrétion de type IV qui se place de manière stable, uniquement aux endroits où la couche de peptidoglycane est dégradée. La deuxième fonction est active après le clivage de la protéine par une protéase. En effet, VirB1\*, qui correspond à 73 résidus de la région C-Terminale de VirB1, a une activité différente de celle de la région N-Terminale de VirB1, étant donné que cette protéine se fait sécréter (fig.16). On ne connaît pas encore sa fonction hors de la bactérie (Christian Baron 1997) mais on pourrait supposer qu'elle facilite le contact cellulaire entre la bactérie et la plante hôte (Fullner 1998).

#### 1.3.2.2. *VirB2*

VirB2 se situe dans les deux membranes des bactéries avec une plus grande association à la membrane interne. Elle est composée de 105 résidus et est le constituant majeur du pilus d'*Agrobacterium*. VirB2 partage 48% de similarité avec TraA, qui est la sous-unité structurale du pilus F d'*Escherichia coli*. Comme TraA, VirB2 est maturée d'un précurseur de 12.3kDa en une protéine de 7.2kDa (Erh-Min Lai 1998). Le clivage se fait entre l'alanine et la glutamine dans le motif : Pro-Ala-Ala-Ala-Glu-Ser (résidus 44 à 49). Il se fait *via* une peptidase spécifique (A. Louise Jones 1996). De plus, la protéine VirB2 mature est un peptide cyclique tout comme son homologue à TrbC qui est la sous-unité du pilus de conjugaison IncP chez *Escherichia coli*. Cette cyclisation confère au pilus une résistance à diverses dégradations comme par exemple, la dégradation d'Edman, la digestion par les protéases,... (Ralf Eisenbrant 1999).

#### 1.3.2.3. *VirB3*

VirB3 est située en membrane interne dans les souches mutées au niveau de VirB4 et se situe dans les deux membranes pour les souches sauvages. Elle est

constituée de 116 acides aminés et promeut la formation du pilus. VirB3 est homologue à TraL du plasmide F de *Escherichia coli* qui est essentiel pour la formation de pilus de conjugaison (John R Zupan 1998). Le rôle de cette protéine est prédit mais non vérifié : elle s'assemblerait, en membrane externe, en une structure qui permettrait le passage du complexe ADN / protéine (Tu Anh Dang 1999). Ce qui est certain à propos de VirB3, c'est qu'elle est essentielle pour la virulence (Heike Schmidt-Eisenlohr 1999).

#### 1.3.2.4. *VirB4*

VirB4 est étroitement associée à la face cytoplasmique de la membrane interne tout en étant située dans le cytoplasme. Elle se compose de 789 acides aminés et forme un dimère. Elle possède un motif « Walker A nucleotide-binding » qui lie l'ATP. Ce motif est localisé entre les résidus 432 et 440. Elle l'hydrolyse et fournit de l'énergie au système de sécrétion pour le processus d'assemblage du complexe (Christie 1997). De plus, cette protéine est essentielle à la stabilisation de VirB3 en membrane externe. VirB4 possède plusieurs régions hydrophobes de longueur suffisante pour traverser la membrane qui sont interrompues par plusieurs résidus chargés. Cela est donc incompatible avec une traversée de la membrane. Le motif « Walker » de VirB4 est essentiel pour la virulence. Quand on le supprime, la bactérie n'est plus capable infecter la plante vu que le complexe ne peut plus s'assembler (Barbara R. Berger 1993). VirB4 possède en fait trois domaines cytoplasmiques et deux domaines périplasmiques, le premier est en région N-Terminale (résidus 14 à 131) et le second se situe juste après le motif « Walker » (Tu Anh T. Dang 1997). Le domaine de dimérisation est situé à l'extrémité N-Terminale (Tu Anh Dang 1999).

#### 1.3.2.5. *VirB5*

C'est une protéine sécrétée et qui serait étroitement associée au pilus. Elle est constituée de 238 acides aminés. D'après des études faites sur son homologue du système de conjugaison TraC, VirB5 est assignée comme étant un composant mineur de l'extrémité du pilus et interviendrait dans l'adhésion avec la cellule végétale. La quantité de VirB5 est fortement réduite quand les autres Vir sont absentes. Cela indique que la stabilisation de la liaison extracellulaire de VirB5 avec VirB2 est assurée par des autres composants de la machinerie du transport de type IV. (Heike Schmidt-Eisenlohr 1999).

#### 1.3.2.6. *VirB6*

C'est une protéine étroitement associée à la membrane interne de la bactérie de 295 résidus. Il n'existe, actuellement, aucune donnée de fonction ou d'une interaction entre elle et une autre protéine VirB. Elle possède six domaines transmembranaires (John R Zupan 1998).

### 1.3.2.7. *VirB7*

Cette protéine, de 55 acides aminés, se situe dans le périplasma et est ancrée en membrane externe. Elle est essentielle pour l'assemblage fonctionnel de la machinerie du transport de type IV. Elle possède un peptide signal caractéristique des lipoprotéines. Une interaction a été démontrée entre *VirB7* et *VirB9*. Il y a deux façons d'expliquer l'association entre ces deux protéines. La première est une association covalente entre la cystéine-24 de *VirB7* et la cystéine-262 de *VirB9* formant un pont disulfure. La deuxième association se fait grâce au segment du 173<sup>ème</sup> au 275<sup>ème</sup> résidu de *VirB9* et grâce à l'extrémité N-Terminale de *VirB7*. En effet, chez certaines espèces possédant ce complexe, le pont disulfure n'est pas utile à la formation du dimère (Das, Anderson *et al* 1997). Il n'en reste pas moins que cette interaction a un rôle prépondérant dans la stabilisation de tout le complexe de type IV. Une étude récente montre que l'homologue de *VirB7* (*PtlI*) dans le système *ptl* de *Bordetella pertussis* forme aussi un lien disulfure avec l'homologue de *VirB9* (*PtlF*) du même système. L'hétérodimère *VirB7/B9* stabilise *VirB10* en l'absence de la plupart des autres protéines *VirB* (Christie 1997). En fait, *VirB7* a pour rôle la stabilisation de *VirB9* et l'hétérodimère *B7/B9* établit des interactions avec les autres protéines *VirB* afin de stabiliser l'entièreté du complexe. En effet, quand on mute *VirB7*, certaines protéines du complexe sont absentes (*VirB4*, *VirB10*), ou ne sont plus bien représentées (*VirB8*, *VirB9*) et d'autres sont présentes mais en moindre quantité (*VirB5*, *virB11*) (David Fernandez 1996).

### 1.3.2.8. *VirB8*

Cette protéine, située de façon prédominante en membrane interne, la traverse grâce à une hélice  $\alpha$  transmembranaire (hydrophobe) composée de 230 acides aminés et localisée dans le premier tiers de la séquence. *VirB8* contribue à la stabilisation du dimère *VirB7/VirB9*. C'est donc un facteur essentiel de la virulence d'*Agrobacterium tumefaciens* et ce, même quand elle n'est présente qu'en petite quantité. Une hypothèse de son mode d'action est qu'elle faciliterait l'alignement correct des deux cystéines, impliquées dans le pont disulfure, pendant le repliement des deux *VirB* (David Fernandez 1996). En fait, *VirB8* interagit avec *VirB9*, *VirB10* et avec elle-même. La zone d'interaction avec elle-même se situe sur 113 résidus à l'extrémité N-Terminale. Pour l'interaction avec *VirB9*, ce sont les résidus 145 et 172 qui sont impliqués (Anath Das 2000). *VirB9* a un homologue dans le système de conjugaison d'*Escherichia coli* (*TraE*) et chez *Bordetella pertussis* pour la sécrétion de la toxine de pertussis (*PtlE*) (Thorstenson and Zambryski 1994).

### 1.3.2.9. *VirB9*

*VirB9* est ancrée en membrane interne mais est située de manière prédominante dans le périplasma. Elle est associée à *VirB7* (Christian Baron 1997). Elle est constituée de 293 résidus et forme un hétérodimère avec *VirB7* qui stabilise les autres composants du système de transport. La synthèse de cette protéine n'est

pas essentielle pour la stabilisation de VirB7 (même si l'inverse est vrai), mais en absence de VirB9, les autres protéines Vir ne sont pas stabilisées (David Fernandez 1996). VirB9 interagit, grâce aux résidus 17-122 situés en région N-Terminale, avec VirB7, VirB10 et avec une autre VirB9. De plus, il y a une autre zone d'interaction, avec VirB7 et avec lui-même, qui est située à l'extrémité C-Terminale de la protéine. VirB9 interagit donc avec lui-même grâce à deux domaines différents : l'un est impliqué lors de la formation d'un dimère, l'autre d'un oligomère (Anath Das 2000).

#### 1.3.2.10. *VirB10*

C'est une protéine qui possède un large domaine périplasmique (David Fernandez 1996) et qui forme des complexes de 120, 160 et 225 kDa situés en membrane interne. Ils seront composés de VirB10 et d'autres protéines (Christine E. Beaupré 1997). VirB10 est composée de 377 acides aminés. Cette protéine traverse la membrane interne grâce à un segment hydrophobe situé entre les résidus 22 et 55. Cette partie de VirB10 est nécessaire pour la virulence de la bactérie ainsi que pour l'oligomérisation de VirB10 (Karine E. Finberg 1995). Des interactions avec les protéines VirB8, B9 et B10 ont été démontrées en région N-Terminale de cette protéine. L'interaction entre VirB10 et VirB9 est nécessaire pour l'oligomérisation de VirB10. Celle-ci dépend donc aussi de VirB7, étant donné que cette protéine sert à la stabilisation de VirB9 (Anath Das 2000)

#### 1.3.2.11. *VirB11*

C'est une protéine de 343 acides aminés qui se situe exclusivement en face cytoplasmique de la membrane interne. VirB11 possède un motif Walker qui lie l'ATP. Quand elle l'hydrolyse, cela fournit de l'énergie au système de sécrétion pour le processus d'exportation du substrat (Tu Anh Dang 1999) (Christie 1997) (Tu Anh T. Dang 1997). L'association de VirB11 à la membrane interne se fait grâce à plusieurs domaines de la protéine qui sont contenus aux deux extrémités. Le motif « Walker » est essentiel pour la fonction de VirB11. Ses homologues TraG (*Escherichia coli*) et PtlH (*Bordetella pertussis*) ont aussi des motifs liant l'ATP qui sont nécessaires pour leur fonction. Il existe, en fait, trois régions critiques pour le bon fonctionnement de VirB11 : près de l'extrémité N-Terminale, la région juste en amont du motif « Walker » et ce motif. Quand on introduit une mutation dans le motif « Walker », cela a un effet sur la localisation subcellulaire de VirB11. On peut donc penser que la liaison ou l'hydrolyse de l'ATP intervient dans la liaison de cette protéine avec la membrane, comme c'est le cas pour SecA (*cfr. Infra* : exportation Sec-dépendante) (Rashkova, Spudich *et al.* 1997) d'autant plus qu'elle possède aussi une activité d'autophosphorylation (Stephens, Roush *et al.* 1995). VirB11 possède un domaine d'interaction à son extrémité C-Terminale. Il s'associe à lui-même et forme aussi des complexes avec VirB9 et VirB10 pendant la biogenèse de l'appareillage de transport. Il est donc très probable que VirB11 soit

fonctionnel sous la forme d'un multimère et que ce domaine d'interaction soit important pour son oligomérisation (Xue-Rong Zhou 1997).

#### 1.3.2.12. *VirD4*

Cette protéine se localise dans la membrane interne. Elle possède un motif « Walker » qui lie l'ATP. Elle est essentielle pour la virulence de la bactérie (Tu Anh T. Dang 1997). L'opéron *virD* d'*Agrobacterium tumefaciens* est important pour la fonction « Dtr » (DNA transfer and replication) du système de transfert de l'ADN-T, tout comme son homologue TraG (*Escherichia coli*) est essentiel dans le transfert d'ADN lors de la conjugaison (Hamilton, Lee *et al.* 2000). Le locus *virD* est composé de quatre gènes *virD1* à *virD4*. Ce dernier code pour un composant du système de transport de type IV. Cette protéine possède à son extrémité N-Terminale un peptide signal pour son transport vers la membrane interne de la bactérie et il l'ancre dans celle-ci. Sa portion C-Terminale est située dans le périplasma (Okamoto, Toyoda-Yamamoto *et al.* 1991).

### 1.3.3. L'organisme d'intérêt : *Brucella suis*

#### 1.3.3.1. *Brucella* et brucellose

Les *Brucella* sont de petits coccobacilles Gram négatifs pathogènes, asporulés et aérobies stricts. Ce sont des bactéries intracellulaires facultatives. Le genre *Brucella* est constitué de trois espèces principales: *Brucella abortus*, *Brucella melitensis*, et *Brucella suis*. Les deux dernières espèces sont les plus virulentes pour l'homme (Alton 1988). Il est le plus souvent infecté par la consommation d'aliments contaminés ou par contact avec des animaux infectés chez qui elles provoquent la brucellose encore appelée fièvre ondulante. La maladie se caractérise par différents symptômes d'ordre général (fièvre, sueurs, anorexie et dépression) mais peut s'accompagner de complications plus localisées (P. Berche 1988).

#### 1.3.3.2. La pathogénie

Les mécanismes d'adhésion aux muqueuses et la pénétration de l'épithélium sont encore inconnus. Le passage de cette première barrière provoque une réaction inflammatoire aiguë dans la sous-muqueuse avec infiltration de neutrophiles et de monocytes. Après avoir franchi cette deuxième défense, l'infection s'étend par voie lymphatique aux ganglions locaux. Ensuite, la bactérie se propage par le sang dans les macrophages, sans doute sous forme intracellulaire. Chez les bovins, les étapes intracellulaires sont cruciales pour l'établissement de l'infiltration qu'il s'agisse de survie et/ou de multiplication en cellule phagocytaire ou en cellule non phagocytaire (Enright 1990).



### 1.3.3.3. La survie intracellulaire de *Brucella*

Afin de comprendre comment *Brucella* est capable de se multiplier à l'intérieur des cellules de l'hôte, son trafic intracellulaire a été étudié *ex vivo* dans une lignée de cellules épithéliales, les cellules HeLa. Après environ dix minutes d'infection de ces cellules par *Brucella abortus*, les phagosomes contenant les bactéries virulentes interagissent avec des endosomes précoces. Ensuite, *Brucella* évite les endosomes tardifs et se localise dans des compartiments ayant les propriétés d'autophagosomes. Finalement, la bactérie est localisée dans une vacuole de réplication que l'on sait associée au réticulum endoplasmique, étant donné qu'elle possède des marqueurs de cet organelle. Cette vacuole est le siège d'une réplication massive de *Brucella* (Fig.17) (Pizarro-Cerda, Meresse *et al.* 1998).

*Brucella* est donc capable d'échapper à la voie de phagocytose classique qui la conduit à la dégradation dans un phagolysosome. L'identification de facteurs bactériens, qui permettent cet établissement dans une niche de réplication spécifique, fait l'objet de plusieurs travaux de recherche notamment en laboratoire. Plusieurs mutants transpositionnels ont été créés afin d'étudier ces différents facteurs.

### 1.3.3.4. Le système de sécrétion de type IV de *Brucella suis*

L'implication du système de sécrétion de type IV dans la virulence de la bactérie a été mise en évidence grâce à un mutant transpositionnel de *Brucella suis* 1330 qui était incapable de se multiplier en cellule HeLa. Une région de 11,8 kb, nommée *virB*, code pour ce système de transport de type IV (D.O'Callaghan 1999). Cette région est essentielle pour la survie intracellulaire de *Brucella*, mais ne l'est pas pour son entrée dans l'hôte (Delrue, communication personnelle).

La séquence d'ADN comprend 12 ORF dont onze protéines, encodées par ces ORF, sont homologues aux composants du système de sécrétion de type IV d'*Agrobacterium tumefaciens*. La douzième protéine, l'ORF 12, n'a pas encore été identifiée dans d'autres systèmes de sécrétion de type IV. Une région intergénique de 439bp sépare *virB1* de *virB2* et les couples *virB5/B6* et *virB6/B7* sont séparés par une région de 183 et 163bp respectivement (D.O'Callaghan 1999). Ces 12 ORF forment un opéron, étant donné qu'elles sont co-transcrites en un même ARN messager (O'Callaghan, communication personnelle).

On peut remarquer une similarité, jusque dans l'organisation de l'opéron, entre *Brucella suis*, *Agrobacterium tumefaciens* et *Bordetella pertussis* (Fig.18)

## 2. Objectifs

Notre travail s'inscrit dans le cadre de l'étude de *Brucella* sp. Actuellement, le mode d'infection de ces bactéries est encore très mal connu. Un transport de type IV est présent chez *Brucella suis* (D.O'Callaghan 1999) et il est, très probablement, impliqué dans la virulence de cette bactérie.

Ce travail se décompose en deux étapes.

La première partie du mémoire consiste à obtenir, sur la totalité des protéines de ce système de transport, un maximum d'information théorique comme, par exemple, les prédictions de structures secondaires, la localisation... en vue de prédire une fonction probable au sein du complexe.

Pour ce faire, une recherche de tous les homologues du transport de type IV chez d'autres bactéries Gram négatives va être effectuée et nous allons comparer les différents composants entre eux. De plus, une étude comparative plus poussée entre *Brucella suis* et *Agrobacterium tumefaciens* va être réalisée. Cette bactérie a été choisie comme organisme de référence car c'est chez elle que le complexe de type IV a été le mieux étudié et le nombre de données de littérature à ce sujet est élevé.

Les différentes analyses seront effectuées au départ des programmes disponibles sur Internet.

Dans une deuxième étape, nous allons analyser de façon plus approfondie les composants du complexe de sécrétion *a priori* les plus intéressants pour l'interaction des protéines du complexe entre elles et leur interaction avec les protéines de l'hôte. Des modèles tridimensionnels seront envisagés afin d'obtenir le plus d'information possible sur la relation structure – fonction. Cette analyse se fait en vue d'étudier ces interactions.

### 3. Matériel et méthodes

Les différents programmes utilisés dans ce travail sont disponibles sur Internet. Leurs adresses Internet sont représentées dans le tableau 4.

Pour les programmes de visualisation (InsightII : Biosym/MSI, San Diego) de protéines et de calculs de threading, nous avons utilisé d'autres stations de travail, les Silicon Graphics Octane duo fonctionnant sous les système d'exploitation IRIX 6.5.

#### 3.1. Blast et ses améliorations

Comme décrit dans l'introduction, Blast est un programme qui effectue des recherches d'homologues dans différentes banques de données.

Selon la nature de la séquence soumise et la nature du contenu de la banque dans laquelle on veut trouver des homologues, il existe différents algorithmes :

	Banque de séquences	Séquence d'intérêt
Blastp	Protéiques	Protéique
Blastn	Nucléotidiques	Nucléotidique
Blastx	Nucléotidiques (traduites dans les six phases)	Protéique
Tblastn	Protéiques	Nucléotidique (traduite dans les six phases)
Tblastx	Nucléotidiques (traduites dans les six phases)	Nucléotidique (traduite dans les six phases)

Les différentes banques accessibles *via* Blast sont représentées en tableaux 1 et 2 de l'introduction.

Les résultats sont quantifiés par un score pour chaque séquence provenant de la banque. Ce score est normalisé en une **E-Value** (E).

$$E = Y \times Z \times K \times e^{-\lambda S}$$

Y : longueur de la séquence

Z : taille de la banque

K et  $\lambda$  : paramètres de Karlin *et altschul*

S : score du MSP considéré

La valeur E représente le nombre de séquence qui auraient obtenus le même score par hasard, pour une séquence de même longueur et une banque de même taille. Plus la E-Value est petite, plus l'homologie trouvée est significative. Une valeur

seuil d'E souvent utilisée est 0,01. Cependant, il se peut que certains homologues pouvant apporter une information importante soient perdus dans le bruit de fond d'une recherche trop générale suite à une E-value trop faible. Dans ce cas, pour faire une recherche spécifique des homologues dont la structure tridimensionnelle est connue, il vaut mieux effectuer une recherche dans PDB (Fig.19) (Altschul, Gish *et al.* 1990).

La banque de Blast utilisée par défaut est la NR (non redondante) car elle regroupe l'information contenue dans les autres banques.

En ce qui concerne notre travail, nous avons utilisé la version « Gapped Blast ». Les recherches d'homologues ont été effectuées dans la banque NR, la banque PDB et la banque d'ESTs humaines (expressed sequences tags). Les différents paramètres ont été utilisés par défaut.

## **3.2. Les méthodes d'alignement**

### **3.2.1. Align**

Ce programme effectue des alignements pairés globaux et c'est un logiciel qui possède peu de paramètres modifiables.

Deux versions d'ALIGN seront utilisées dans ce travail : une version disponible sur Internet (Myers and Miller 1988) et une version « routine » fonctionnant sur Silicon Graphics (station de travail Unix).

Les résultats de la version Internet contiennent un alignement caractérisé par « : » pour une identité entre les deux séquences et « . » pour une similarité, un pourcentage d'identité entre les deux séquences, deux valeurs de pénalité données à l'insertion d'indels (le premier chiffre pour débiter un nouvel indel et le deuxième chiffre pour étendre un indel existant) et un score d'alignement global.

Nous avons utilisé ALIGN avec les paramètres par défaut. Notre analyse portait sur une vision globale de l'alignement et sur le pourcentage d'identité.

L'algorithme tournant sur Unix et utilisant ALIGN nous servira à établir une matrice d'identité pour différentes séquences prises 2 à 2. Les pourcentages d'identité sont calculés grâce aux alignements effectués par la matrice Blossum 50.

### **3.2.2. Match-Box**

C'est un programme d'alignement multiple local (Fig.20) (Depiereux and Feytmans 1992). Il comprend deux logiciels bien distincts : EXPLORE *et* ALIGN. Le premier détermine si les similarités observées se démarquent du hasard. Le deuxième logiciel est divisé en trois : le scanning, le matching et le screening. Le scanning calcule le meilleur ensemble de paramètres qui sera utilisé dans l'étape de matching. Celle-ci effectue toutes les comparaisons pairées et génère une banque de

régions prédites comme structurallement conservées (pSCR). Le screening génère l'alignement local au départ des pSCR.

L'alignement donné par Match-Box est quantifié par des coefficients de confiance qui ne sont calculés qu'au niveau des boîtes conservées pour toutes les séquences. Ce coefficient s'échelonne de 1 à 9, 1 étant le meilleur résultat (Tab.5).

Les résultats sont envoyés par E-mail. L'utilisateur reçoit quatre messages. Le premier est envoyé pour signaler que le travail a bien débuté. Le deuxième message est l'EXPLORE. On peut y voir des graphiques et des tables montrant la similarité globale cumulée pour toutes les séquences. Le message suivant est ALIGN, l'alignement y est représenté avec les boîtes de résidus conservés dans toutes les séquences. Sous ces boîtes, les coefficients de confiance sont notés. Les acides aminés inscrits en majuscules ne sont pas alignés. Le quatrième message envoie les différentes séquences soumises au programme en format FASTA *et alignées les unes aux autres*.

Nous avons utilisé tous les paramètres de Match-Box par défaut excepté le format de la fiche de résultat où l'on demandait le format paysage.

### 3.2.3. ClustalW

Ce programme effectue des alignements multiples globaux (Thompson, Higgins *et al.* 1994). ClustalW possède beaucoup de paramètres modifiables au gré de l'utilisateur, ce qui le rend très « malléable ». L'idée de base de ce programme est que deux séquences similaires sont liées d'un point de vue évolutif. Le logiciel se base donc sur un arbre phylogénétique dérivé d'une matrice de distances entre les paires de séquences alignées. Les deux séquences les plus similaires sont alignées en premier, ensuite elles forment un groupe qui sera aligné à la troisième séquence la plus similaire à ce groupe, etc...

Les résultats fournis par ClustalW sont donnés dans deux formats différents, l'alignement des séquences en format ClustalW et l'alignement FASTA (séquences en format FASTA où des indels sont introduits pour aligner des séquences).

Nous avons utilisé ce programme avec tous les paramètres par défaut.

### 3.3. Blocks

Block recherche dans la séquence soumise des blocs de séquences conservés et significatifs d'une famille de protéines. Chaque bloc est un alignement multiple local sans indels fait au départ de groupes de protéines proches d'un point de vue fonctionnel et/ou évolutif et donc partageant une similarité de séquence (Henikoff and Henikoff 1994).

Ce programme possède différents paramètres dont, par exemple, un seuil modifiable pour la **E-Value**. Cette valeur peut être utilisée pour quantifier le degré

de similarité globale entre la séquence d'intérêt et le bloc. Par exemple, avec une E-Value de  $10^{-3}$ , on s'attend à trouver un même score de similarité dans la même banque une fois toutes les 1000 recherches.

Les résultats peuvent être envoyés par E-mail. Ils fournissent les noms des différentes familles de protéines dont certains blocs ont été retrouvés sur la séquence soumise. Le nombre de blocs retrouvés sur la séquence d'intérêt est noté ainsi que le nombre total de blocs qui caractérise la famille. Les alignements de la séquence d'intérêt avec les différents blocs sont également fournis.

Nous avons soumis des séquences protéiques à ce programme en utilisant tous les paramètres par défaut.

### **3.4. ScanProsite**

ScanProsite est un programme qui permet de comparer une séquence protéique à toute la banque de données PROSITE (Hofmann, Bucher *et al.* 1999). Cette banque est composée de motifs caractéristiques d'un point de vue fonctionnel. L'utilisateur peut demander au logiciel d'utiliser un filtre, c'est à dire ne pas tenir compte des petits motifs qui sont trop fréquents dans les séquences en raison de leur peu de spécificité, par exemple le site d'Asn-glycosylation : N-{P}-[ST]-{P} où P peut être n'importe quel acide aminé excepté la proline.

Les résultats contiennent la séquence soumise et les différents motifs trouvés.

Nous avons soumis toutes les séquences de ce travail à ce programme en utilisant le filtre.

### **3.5. ProDom**

ProDom est une banque de données qui contient des familles de domaines protéiques générées au départ de la banque Swiss-Prot par comparaison automatique de toutes les séquences (Corpet, Gouzy *et al.* 1999).

## **3.6. Prédiction de structures secondaires**

### **3.6.1. PHD**

Ce programme, basé sur des systèmes neuronaux, est le premier à atteindre une efficacité supérieure à 70 % après avoir été validé sur 100 protéines uniques (Rost, Sander *et al.* 1994).

Ce programme effectue différentes prédictions : les structures secondaires (PHDsec), l'accessibilité au solvant (PHDacc), l'éventuelle présence d'hélices transmembranaires (PHDhtm), les régions contenant des boucles, des motifs

fonctionnels annotés par des experts (PROSITE), d'éventuels domaines (ProDom), la présence de cystéine liées en pont disulfure (CYSPRED), et d'autres... (Fig.21). L'utilisateur peut limiter les analyses à effectuer. On peut soumettre à PHD soit une séquence, soit une liste de séquences homologues dont il fera l'alignement pour effectuer la prédiction de structure secondaire de la première séquence de la liste avec une plus grande confiance.

Les résultats sont envoyés par E-mail. On peut y trouver l'alignement effectué par le logiciel si on lui a fourni une liste de séquences, deux tableaux représentant les pourcentages en acides aminés et en structures secondaires (H « helix » pour hélice, E « extended » pour brin et L « loop » pour boucle) de la structure soumise et la prédiction de structures secondaires longeant toute la séquence. Cette prédiction est quantifiée par un coefficient de confiance s'échelonnant de 0 à 9 et plus celui-ci est élevé, plus la prédiction est confiante. Sous la séquence, on trouve également les accessibilités au solvant (b « buried » pour enfoui et e « exposed » pour exposé). Si PHD a détecté des hélices transmembranaires, il donne une deuxième fois la séquence et y note les hélices transmembranaires (T). Dans ce cas, un message annonce que les premières prédictions (H, E, L) sont moins confiantes.

Nous avons surtout utilisé PHD pour les prédictions de structures secondaires et d'hélices transmembranaires.

### 3.6.2. PREDATOR

La méthode de prédiction est basée sur la reconnaissance de résidus potentiellement liés par ponts hydrogène sur la séquence cible. Par exemple, pour une hélice  $\alpha$ , un résidu  $i$  se lie par pont hydrogène avec le résidu  $i+4$  (Frishman and Argos 1996). Ce logiciel de prédiction a subi un changement en 1997, on peut maintenant lui soumettre un ensemble de séquences homologues pour trouver les structures secondaires de la première séquence de la liste. PREDATOR effectue des alignements locaux pairés et ensuite un alignement multiple. Cela permet d'augmenter le degré de confiance des prédictions de structure. L'efficacité de prédiction de ce programme tourne autour de 75 % (Frishman and Argos 1997) alors qu'elle était de 68 % pour une seule séquence soumise (Frishman and Argos 1996).

PREDATOR présente différents paramètres, comme par exemple le format de la fiche de résultats.

Les résultats sont envoyés par E-mail au format demandé par l'utilisateur. Le format court permet de visualiser la séquence (code 1 lettre) et les prédictions de structures secondaires : H pour hélice, E pour brin et \_ (underscore) pour boucle. Le format long est présenté en colonnes avec les acides aminés en code 1 et 3 lettres, les prédictions de structures secondaires (H pour hélice, E pour brin et C pour boucle) et une estimation de la confiance de la prédiction.

### 3.6.3. Prof

L'ancien nom de ce logiciel est DSC (Discrimination of protein Secondary structure Class). Ce programme est basé sur la décomposition de la prédiction de structures secondaires en ces concepts basiques :

- la conformation préférencielle du résidu pour une structure particulière (potentiel moyen de GOR, « Garnier Osguthorpe Robson »), pour exemple : l'alanine est favorable à la formation d'hélices  $\alpha$
- le profil moyen d'hydrophobicité adopté par une hélice  $\alpha$  ou un brin  $\beta$ , un certain profil d'hydrophobicité est associé avec un type de structure secondaire; cela se remarque plus pour les hélices  $\alpha$
- les effets bordants des séquences sont importants car plus les résidus sont proches de la fin d'une chaîne protéique, moins ils ont de contraintes structurales ; la partie C-terminale d'une protéine est généralement plus mobile que le reste de la protéine
- l'existence d'insertions et de délétions dans un alignement de séquences homologues, leur tolérance varie avec la classe de structure secondaire, ils témoignent de la présence de boucles; les délétions interrompent plus facilement une hélice  $\alpha$  que les insertions alors que pour les brins  $\beta$ , les deux ont une action identique
- l'existence de parties conservées lors d'un alignement d'homologues qui permet une application plus fiable de la conformation préférencielle d'un résidu (King and Sternberg 1996)

Toutes ces considérations sont appliquées à chaque résidu lors des quatre itérations du programme.

La fiche de résultats se présente sous un format de colonnes. On peut y voir la séquence soumise, les prédictions pour chaque résidus de chacune des étapes. La dernière colonne représente les probabilités pour les trois structures (H : hélice, E : brin, C : boucle) et le programme choisi la meilleure.

### 3.6.4. Jpred

Jpred est un serveur qui effectue un consensus de prédiction de structures secondaires au départ de six algorithmes différents : DSC, PHD, NNSSP (Salamov and Solovyev 1995), PREDATOR, ZPRED (Zvelebil, Barton *et al.* 1987) et MULPRED. L'efficacité de ce consensus est d'environ 72,9 % (Cuff, Clamp *et al.* 1998).



Ce serveur permet la soumission d'une séquence seule ou d'un alignement multiple de séquences homologues. Si la séquence soumise est seule, un procédé automatique créera un alignement multiple de séquences *via* Psi-Blast.

Les résultats sont envoyés par E-mail. Ils se présentent sous la forme d'une adresse Internet accompagnée de la séquence soumise avec les prédictions de structures secondaires et d'un coefficient de confiance. Sur l'adresse Internet mentionnée dans le mail, on peut trouver les différentes étapes de la prédiction de la protéine soumise.

### 3.6.5. Psi-Pred

PSIpred est constitué de deux réseaux neuronaux. Son efficacité est d'environ 77 % (Jones 1999).

Sur la page web de ce site, on peut demander différentes analyses : PSIpred, une prédiction de topologie (MEMSAT), une reconnaissance de « fold » (GenTHREADER).

Les résultats sont envoyés par E-mail. Le message comprend trouve la séquence soumise avec les prédictions de structures secondaires et un coefficient de confiance s'échelonnant de 0 à 9. Plus le chiffre est proche de 9 et plus la confiance est élevée. Sur cette fiche de résultats, on trouve aussi une adresse Internet où il est possible de voir la topologie de la séquence (Fig.22).

## 3.7. PSORT

Comme décrit dans l'introduction, PSORT est un système expert qui permet la prédiction de la localisation cellulaire de protéines. L'efficacité de la prédiction pour les différentes localisations cellulaires est représentée en figure 23.

La première analyse, la reconnaissance d'un peptide signal se fait *via* la méthode **McGeoch** (McGeoch 1985) qui est applicable pour tous les peptides signaux. Dans cette analyse, deux régions sont définies : la région N-terminale qui est souvent chargée (CR) et la suite, non chargée et hydrophobe (UR). Ensuite, quatre paramètres sont examinés : la longueur et la charge nette de la CR, la longueur et le degré d'hydrophobicité de l'UR. La méthode **von Heijne** (von Heijne 1989) permet de différencier les peptides signaux des lipoprotéines par rapport aux peptides signaux classiques. Pour ce faire, la méthode compare le motif consensus de clivage du peptide signal lipoprotéique à la partie C-terminale de la CR. Cette méthode est aussi utilisée afin de prédire si le peptide signal est clivé ou non (von Heijne 1986). Elle est basée sur une comparaison du site de clivage de la signal peptidase I à la partie C-terminale de l'UR. Ensuite, la méthode de **Klein et al** (Klein, Kanehisa *et al.* 1985) est utilisée dans la reconnaissance de segments transmembranaires hydrophobes afin de prédire une localisation en membrane

interne. Elle se base sur une longueur de 17 résidus hydrophobes pour un segment transmembranaire.

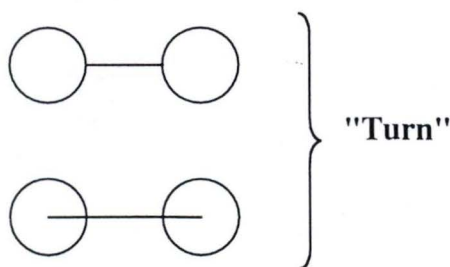
On soumet à PSORT une séquence protéique dont on définit la provenance (bactérie G+, G-, levure, animal, végétal). Aucun autre paramètre n'est accessible.

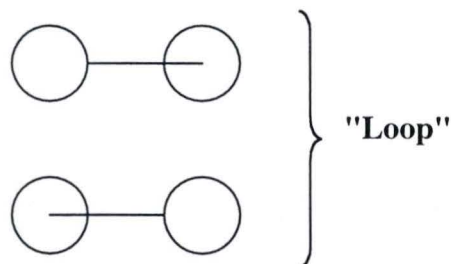
La fiche de résultats de PSORT contient : la séquence soumise, le consensus de clivage du peptide signal s'il s'agit d'une lipoprotéine, le score McGeoch, le score von Heijne. Plus le score McGeoch est positif, plus il est probable qu'il y ait une séquence signal. De même, un score positif pour von Heijne est significatif d'un peptide signal clivé. S'il s'agit d'une protéine qui possède une peptide signal clivé. S'il s'agit d'une protéine qui possède une peptide signal clivé, le logiciel donne la position du premier résidu de la protéine mature. En plus de ces scores, PSORT donne la position des segments transmembranaires existant ainsi qu'une confiance qui est fonction de l'hydrophobicité du segment qu'il prédit comme étant une hélice transmembranaire (htm). Plus le score est négatif, plus le segment protéique est hydrophobe et donc, plus la chance que ce soit une htm est grande. Pour finir, il donne un facteur de confiance de présence de la protéine soumise dans les différents sites de localisation grâce à l'analyse des différents résultats obtenus.

Toutes les séquences soumises à PSORT lors de ce travail provenaient de bactéries Gram négative.

### 3.8. TOPS

TOPS est un logiciel qui fournit des diagrammes 2D symbolisant la structure de protéines se trouvant dans la PDB (Westhead, Slidel *et al.* 1999). On obtient ainsi la topologie de cette protéine. Les triangles représentent des brins  $\beta$  et les cercles des hélices. Ces symboles imitent les structures secondaires qui seraient perpendiculaires au plan du diagramme. Les structures secondaires ont une direction (de  $N_i$  à  $C_{i+1}$ ) et une orientation. Celle-ci est représentée par les connections entre les structures. Si le lien entre deux symboles se fait sous (ou au-dessus) ces symboles, cela veut dire que la connections est un simple « turn ». Si le lien se fait sous le premier symbole et arrive par dessus le deuxième symbole, ce sera un long « loop ».





La base de l'algorithme repose sur l'assignation des structures secondaires exécutée par DSSP (Kabsch and Sander 1983).

TOPS ne permet pas la soumission de protéine n'étant pas dans le format PDB. Nous avons donc utilisé ce programme afin de visualiser la topologie des homologues 3D de nos protéines, retrouvés par Blast PDB ou par les logiciels de reconnaissance de « fold ».

### 3.9. Prédiction de structure tertiaire

#### 3.9.1.3D PSSM

Ce serveur prédit la structure tridimensionnelle (Kelley, MacCallum *et al.* 1999) (Kelley, MacCallum *et al.* 2000). Il « enfile » la séquence soumise sur toutes les structures des protéines contenues dans sa banque de données. Un score (E-value) quantifie la compatibilité entre la séquence soumise et chaque « threading ». Différents éléments entrent en jeu pour calculer ce score : 1D-PSSM (les profils des séquences sont construites sur base d'homologues relativement proches), 3D-PSSM (le profil est plus général et est basé sur des homologues lointains), la correspondance des éléments de structures secondaires et les caractéristiques d'accessibilité au solvant des résidus de la protéine soumise.

Sur la page web du site, il faut donner l'adresse E-mail, une courte description de la protéine et la séquence protéique. Celle-ci est limitée à 800 résidus car pour une protéine plus longue, le temps de calcul serait trop long. De plus, il est rare qu'une protéine de plus de 800 résidus ne soit constituée que d'un seul domaine, alors on peut soumettre au programme les domaines de façon séparée.

Les résultats sont sur une page web dont l'adresse, spécifique aux résultats de la protéine soumise, est envoyée par E-mail. Ils ne restent disponibles à cette adresse que 5 jours mais on peut les télécharger. Ces résultats, représentés sous la forme d'un tableau, contiennent les 20 protéines dont les structures connues sont les plus probables pour la protéine soumise, et sont classés en fonction de leur « E-value ». De plus, une description de type SCOP pour la protéine homologue de structure connue est donnée. Les alignements des séquences de ces protéines avec la protéine d'intérêt sont aussi donnés ainsi qu'une prédiction de structures

secondaires pour la séquence soumise, effectuée par Psi-Pred. Deux « E-values » sont visibles dans le tableau :

- la PSSM E-value représente le score du résultat retrouvé par rapport à la séquence d'intérêt, la case contenant ce score est colorée selon un pourcentage de certitude du résultat et le code couleur se situe au-dessus du tableau
- la SAWTED E\_value mesure la similarité entre la séquence soumise et l'homologue trouvé sur base du texte de description des deux protéines.

### 3.9.2.123D

Ce programme effectue des recherches de grands ensembles d'éléments de structures secondaires protéiques qui seraient compatibles avec la protéine soumise et l'homologue trouvé mais ces structures ne doivent pas forcément être dans le même ordre sur les deux séquences primaires. Ce qui est recherché, c'est un arrangement tridimensionnel similaire des structures secondaires pour les deux protéines. 123D distinguent trois type de similarité. Le premier type est la similarité qu'il existe entre deux protéines de même famille. Le deuxième tient compte des ressemblances qu'il peut exister entre deux protéines de familles et fonctions différentes mais qui partagent la même topologie. Enfin, le troisième type de similarité se retrouve pour deux protéines évolutivement différentes chez qui les connections entre les structures secondaires sont différentes.

Sur la page Web de ce site, on peut y trouver différents paramètres modifiables : la matrice que l'on veut utiliser, la pénalité que l'on veut donner à l'introduction d' « indel » et à leur extension, le type d'alignement utilisé et son affichage.

L'adresse Internet contenant les résultats pour la protéine soumise est envoyée par E-mail. Sur cette page Web, un rappel des paramètres choisis est présenté. Ensuite, un tableau vous donne les différentes protéines ntrouvées comme homologues à la protéine soumise. Ces homologues sont quantifiées par un score Z, le nombre de résidus alignés et le pourcentage d'identité que les deux séquences partagent. Au départ de cette page, un lien direct est fait si l'utilisateur veut plus d'information à propos de l'homologue trouvé.

### 3.9.3.Ucla-Doe

Ce programme permet une reconnaissance de « fold » d'une séquence protéique. Il va rechercher, dans une banque de protéines de structures connues, des homologues à la séquence soumise. Le logiciel d'UCLA-Doe utilise, pour ce faire, plusieurs méthodes (GONP, GONPM, SEQPPRF, SEQPMPRF, PRFSEQ) et en fait un consensus.

La page d'accueil de ce site ne présente aucun paramètre permettant de modifier une quelconque étape de la prédiction. On y trouve uniquement un endroit où mettre son adresse E-mail, une fenêtre pour le nom de la protéine et une fenêtre pour « coller » la séquence.

Les résultats sont envoyés par E-mail et ils se partagent en deux : une adresse Internet et les résultats des différentes méthodes ainsi que le consensus final. Chaque résultat est quantifié par un score Z. Plus il est élevé, plus le résultat trouvé est bon.

L'adresse Internet renvoie à une page où on a accès à différents résultats de programmes tournant sur le site de l'Ucla-Doe : Psi-Blast, alignement multiple des résultats retrouvés par Psi-Blast, PHD pour la protéine d'intérêt, nombre et localisation d'éventuelles hélices transmembranaires sur la séquence soumise, les tableaux qui résument les résultats des différentes méthodes de reconnaissance de « fold » et un tableau résumé de ces méthodes et du consensus.

Nous avons soumis toutes les protéines du complexe étudié dans ce travail à ce site en vue de comparer les prédictions qu'il donne et les prédictions de 3D-PSSM.

## 4. Résultats et discussion

### 4.1. Données préliminaires

Une étude comparative entre différents appareils de sécrétion de type IV a déjà été réalisée, aboutissant à un schéma consensus de l'appareillage de sécrétion de type IV (Fig.15 dans l'introduction) (A. Covacci 1999).

Les données de localisation cellulaire, de motifs et les fonctions « présumées » des composants du complexe de sécrétion de type IV d'*Agrobacterium tumefaciens*, d'*Escherichia coli* et de *Bordetella pertussis* sont résumées dans le tableau 6. Le tableau 7 représente les données connues sur les composants du complexe de sécrétion de type IV de *Brucella suis* ainsi qu'un pourcentage d'identité et de similarité avec les homologues de ces protéines pour les espèces citées ci-dessus.

Les sites de clivage pour les six protéines VirB de *B. suis* possédant un peptide signal sont donnés en figure 24 (D.O'Callaghan 1999).

<i>B. suis</i> VirB		Homologies with other type IV systems <sup>a</sup>						Propertes of homologues in Pti and VirB
ORF	MW <sup>b</sup>	Pti	VirB C58	Tra pKM101				
1	25.2 (23)	-	-	22	46	28.5	56	OM; transglycosylase
2	11 (8.2)	45	76	18	49	19.5	48	Pilus formation
3	13 (12)	37	66	21.5	53	23	48	IM?
4	94.5	51	81	30	65	29	65	Cytoplasmic? ATPase
5	27 (24.7)	-	-	19	52.5	29	66	Pilus component?
6	36	14.5	48	21	51	20	59	IM
7	5.7 (4.3)	27	58	27	52.5	24.5	56	Lipoprotein; complex with VirB9
8	25.5	34.5	67	24.5	56	31.5	67.5	IM
9	32 (30)	42	67	26.5	58	27	51	Periplasmic; complex with VirB7
10	41.5	31	59	32.5	65.5	30	68	Periplasmic
11	44.3	39	64	29.5	59.5	25.5	55	IM, ATPase
12	19 (17.5)	-	-	-	-	-	-	-

Tab.7 : a. Pourcentage d'identité et de similarité calculé avec Clustal.

b. Poids moléculaire prédit en kDa. Entre parenthèses, le poids moléculaire prédit de la protéine mature, après clivage du peptide signal (Callaghan et al, 1999).



Fig.24 : Sites de clivage des peptides signaux prédits pour six protéines du complexe de type IV de *Brucella suis*. Ces prédictions ont été effectuées par le programme PSORT. Les sites de clivage sont représentés par une flèche. Les motifs de sites de clivage lipoprotéiques sont soulignés (Callaghan et al, 1999).

Ci-dessous se trouve un complément de données préliminaires pour différentes protéines du système de sécrétion de type IV (D.O'Callaghan 1999).

**VirB1** d'*A.tumefaciens* possède deux domaines de fonctionnalité différente : VirB1 (la protéine entière) et sa partie C-terminale (VirB1\*) qui se fait exporter. VirB1 possède une activité enzymatique, elle dégrade le peptidoglycane, probablement pour permettre un positionnement correct de tout l'appareillage de sécrétion. VirB1 de *B. suis* contient les motifs AAIVQVES (résidus 21-28), GIAQIN (69-74) et YYAG (128-131) qui sont semblables aux blocs A, B et C associés au site actif de transglycosylase d'*Escherichia coli*.

**VirB4** de *B. suis* possède également un motif de liaison au nucléoside triphosphate dont la séquence est GQSGAGKT (résidus 458-465). Ce motif est conservé dans tous ses homologues.

**VirB7** d'*A. tumefaciens* interagirait avec **VirB9** grâce à un pont disulfure et c'est cela qui stabiliserait l'appareillage de sécrétion. Chez *B. suis*, si l'interaction VirB7 / B9 existe, ce n'est pas grâce à un pont disulfure car VirB9 ne possède qu'un seul résidu cystéine et il se situe dans le peptide signal de celui-ci, donc le pont disulfure n'est pas conservé.

**VirB10** d'*A.t. tumefaciens* est nécessaire pour la virulence de la bactérie mais on ne connaît pas sa fonction exacte.

**VirB11** d'*A. tumefaciens* est essentielle au fonctionnement du complexe de sécrétion. Son homologue chez *B. suis* possède également un site de liaison au nucléoside triphosphate: GPTGSGKT (résidus 169-178).

L'**orf12** de *B.suis* ne possède pas d'homologue connu dans un autre système de sécrétion de type IV. La seule donnée à son sujet est qu'il possède un consensus de lipoprotéine.

#### **4.2. « Débroussaillage » du complexe**

Le premier objectif de ce travail était le « débroussaillage » structural du système de sécrétion de type IV de *Brucella suis*. Nous devons donc étudier les localisations et fonctions de douze protéines qui interagissent sans doute entre elles étant donné qu'elles forment un complexe.

Nous avons utilisé, pour ce faire, des outils informatiques qui permettent d'obtenir un certain nombre d'information sur base uniquement de la séquence. Toutes les séquences protéiques de ce complexe ont été soumises à plusieurs programmes. Ensuite, nous avons analysé les différents résultats obtenus et nous les avons comparé avec la littérature. Ceux qui nous semblaient les plus intéressants, par rapport aux données de la littérature, ont fait l'objet de discussions plus poussées.

La première étape d'un « débroussaillage » est la recherche classique d'homologue (Blast NR).

Lorsque nous avons eu fini toutes ces recherches d'homologues *via* Blast N.R., nous avons repris les séquences de *Brucella suis*, d'*Agrobacterium tumefaciens*, de *Bordetella pertussis*, de *Bartonella henselae*, d'*Helicobacter pylori*, de *Rickettsia prowazekii*, de *Rhizobium etli*, et de *Legionella pneumophila* afin de faire tourner une routine sur toutes ces séquences. Cette démarche utilise le programme ALIGN et donne tous les pourcentages d'identité pour toutes les séquences prises deux à deux. Cela nous a permis de constater que les homologues d'une protéine VirB ne sont homologues qu'à cette protéine et non aux autres composants du système de sécrétion de type IV.

Après cette première démarche, une recherche d'homologues plus poussée sur une banque de structures tridimensionnelles connues et de séquences eucaryotes a été effectuée. Nous avons choisi de sélectionner la première banque dans l'espoir d'obtenir une structure 3D de référence pour la protéine soumise. La sélection de la deuxième banque s'est faite pour savoir si une des protéines du complexe pouvait avoir une homologie avec une protéine de l'hôte et donc pourrait interagir avec celui-ci. Ensuite, nous avons effectué des prédictions de localisation cellulaire (PSORT) étant donné que la localisation peut argumenter une fonction potentielle. En effet des fonctions spécifiques sont le plus souvent localisées dans des compartiments cellulaires bien déterminés. Des détections de motifs conservés dans des familles protéiques connues ainsi que des prédictions de structures secondaires et tertiaires, ont aussi été réalisées.

Bien que toutes ces étapes répondent à des questions différentes, nous espérons avec les données récoltées, préciser des caractéristiques structurales et fonctionnelles pour chacune des protéines du complexe.

Dans le tableau n°8, les différents résultats obtenus pour chaque programme sont représentés.

Les lignes représentent les résultats de chaque protéine et les colonnes représentent les différents programmes. Pour le logiciel Blast, une subdivision a été réalisée afin de présenter les résultats obtenus en sélectionnant différentes banques de protéines.

Ci-dessous se trouvent les analyses des différents résultats obtenus pour une partie des protéines. Les composants du complexe n'ayant pas de résultats concluants ou dont les résultats n'apportaient rien de nouveau par rapport à la littérature n'ont pas fait l'objet d'une discussion plus approfondie.

#### **- VirB1 est apparenté aux lysozymes.**

Grâce aux résultats obtenus pour la protéine VirB1, nous pouvons renforcer son homologie avec les transglycosylases de plusieurs façons. Blockmaker identifie sur la séquence de VirB1 les deux blocs caractérisant la famille des transglycosylases. De plus les sites de reconnaissance de repliement protéique (Ucla-Doe et 3D-PSSM) donnent des transglycosylases comme homologues de structure tridimensionnelle connue. Une autre famille d'homologues de structure 3D connue fournie par Blast PDB et 3D-PSSM est celle des hydrolases dont les



motif		Fold recognition	
Prosite	Block	Ucla-doe	3D-PSSM
Résultats non concluants	"Prokaryotic transglycosidase protein"	"Lytic transglycosylase", isomerase, Transferase et Racemase : <b>alpha et beta</b> Oxydo-reductase et répresseur de la toxine diphtérique : tout alpha.	Répresseur de la toxine diphtérique, "DNA-binding", "lambda repressor-like" : tout alpha <b>1</b> hydrolase : <b>alpha et beta</b> . <b>2</b> lysozymes : <b>alpha et beta</b> .
Résultats non concluants	Aminotransferase Permease.	Facteur de transcription, biosynthèse antibiotique, "tyrosil transfer" : tout hélices alpha.	Saposine, "globin-like" : tout alpha Resolvase : alpha et beta
Motif RGD (16-18)	"Flagellar biosynthetic"	Facteur de transcription, "calcium binding", DNA binding : tout hélices alpha	Oxydoreductase, chloroperoxidase, "tetracyclin repressor" : tout alpha
Site de liaison à l'ATP/GTP(458-465) Motif RGD (325-327)	ABC transporteur "Flagellar motor protein motA." "3'5' cyclic nucleotide phosphodiesterase"	<b>3</b> oxydo-reductases, <b>1</b> déshydrogénase, <b>1</b> lyase : <b>alpha et beta</b>	N'accepte pas la séquence car est trop longue.
Résultats non concluants	Apolipoprotéine	<b>2</b> lipoprotéines, biosynthèse d'antibiotique, delta-endotoxine : tout <b>hélices alpha</b>	<b>4</b> "4 helical up-and-down bund apolipoprotein" : tout <b>alpha</b> .
Résultats non concluants	"Prolipoprotéine Diacylglycerol transferase"	Hydrolase, oxydo-reductase, transferase : <b>alpha et beta</b> .	<b>2</b> Terpenoid synthases, <b>2</b> ferritin-like : tout alpha.
Résultats non concluants	Pas de résultats	<b>2</b> Facteurs de croissance, <b>1</b> protéine de coagulation sguine, initiateur de transcription : small prot. (knottins) "Protein Antifreeze" : <b>alpha et beta</b>	"Antifreeze protein Type III" : tout beta. "Multihem cytochrome" : small prot.
Résultats non concluants	FIQ	Creatinase, méthyl-transferase, dioxygenase : <b>alpha et beta</b> .	Domain phosphatidine , domain ATPase, S-adenosyl-L-methionine : <b>alpha et beta</b> .
Résultats non concluants	Peptidase	<b>2</b> oxydoreductases, <b>1</b> glycosidase, heparin et integrin binding : tout beta Serine proteinase : alpha et beta	<b>2</b> "viral coat" et protéine de capsid : tout beta <b>*1</b> N-term nucleophile amino proteasome sudunits" : alpha et beta
Résultats non concluants	Répétition de proline	Oxydoreductase, coat prot. : tout beta Myeloperoxydase, photosynthetic reaction center : tout alpha.	<b>2</b> virus coat protein : tout alpha et tout beta Leu aminopeptidase : alpha et beta.
Site de liaison à l'ATP/GTP (182-189) "Sigma-54 interaction domain ATP-binding region A signature" (178-191)	"ABC transporter" "Bacterial type II secreton system"	Isomerase, transferase, signal recognition particle receptor : <b>alpha et beta</b> Peroxydase : tout alpha.	Class I aldolase, CheB methyltransferase domain, Zn dpdt exopeptidase : <b>alpha et beta</b> .
Résultats non concluants	"OmpA like domain protein"	Ligase, hydrolase, disulfide oxydo-reductase, DNA repair : <b>alpha et beta (a/b/a)</b>	Immunoglobulin-like : tout alpha et tout beta. Deoxyuridine 5'-triphosphaye nucleotidohydrolase, "nucleic acid binding protein", "common fold of diphteria" : tout beta.

lysozymes font parties. Tous ces résultats corroborent la probable activité de dégradation du peptidoglycane de VirB1.

**- VirB1\* est un domaine distinct du reste de la protéine VirB1.**

Nous pouvons argumenter, grâce à la différence de succession de structures secondaires prédite par PHD pour VirB1 et VirB1\*, que VirB1\* est un domaine distinct du reste de la protéine VirB1. En effet, VirB1 est constitué d'hélices  $\alpha$  et de brins  $\beta$  alors que VirB1\* ne serait composé que de brins  $\beta$ . D'après des données de littérature, VirB1\* serait soit sécrétée, soit en membrane externe. Après avoir analysé les résultats obtenus par PHD, nous opterions pour une localisation en membrane externe étant donné la présence de brins  $\beta$  dans VirB1\*. En effet, la présence de brins  $\beta$  permet à une protéine de s'insérer en membrane externe tout comme des hélices transmembranaires le permettent en membrane interne.

**- Le pilus, composé de VirB2, s'ancre en membrane interne.**

Pour VirB2, les résultats obtenus peuvent nous aider à résoudre une ambiguïté de littérature. En effet, deux localisations différentes de cette protéine, et donc deux ancrages de pilus différents, sont présentés dans deux articles (A. Covacci 1999) (Das and Xie 2000). Ces localisations sont la membrane interne (Fig.15 introduction) et la membrane externe (Fig.25). Le programme PHD prédit des hélices transmembranaires, ce qui est caractéristique d'une localisation en membrane interne et ce positionnement est corroboré par PSORT.

Nous optons donc pour un ancrage du pilus en membrane interne (Fig.15).

**- VirB2 ressemblerait à la protéine pVIII du phage M13.**

Une autre analyse a été effectuée sur VirB2, nous l'avons comparée à la protéine pVIII du phage M13. En effet, l'assemblage du système de sécrétion de type IV pourrait être assez semblable au mécanisme d'élongation de ce phage. Le composant du pilus du système de sécrétion pourrait donc être comparé à la protéine pVIII (Fig.26) (Kay, Winter *et al.* 1996). Après l'analyse d'alignements, il s'est avéré que les séquences de VirB2 de *Brucella suis* et de la protéine pVIII du phage M13 partagent 20 % d'identité, ce qui est loin d'être négligable étant donnée que la protéine pVIII fait la moitié de VirB2 et donc que ces 20 % d'identité s'étale sur 73 acides aminés.

**- Les localisations respectives de VirB3 et VirB4 sont ambiguës : leur localisation pourrait évoluer au fur et à mesure de la formation du complexe.**

VirB3 semble être localisée en membrane interne (PSORT). Cette localisation est renforcée par la présence d'une hélice transmembranaire (PHD). Cela contredit partiellement les données de la littérature de cette protéine. En effet, en présence de VirB4 non muté, VirB3 devrait se localiser à la fois en membrane interne de la bactérie et en membrane externe où elle devrait s'assembler en un pore qui permettrait le passage de complexes ADN/protéines. Par contre, dans une souche bactérienne où la protéine VirB4 est mutée, VirB3 ne se situe qu'en membrane interne.

D'après nos résultats, VirB3 peut se localiser en membrane interne étant donné que PHD prédit une hélice transmembranaire mais en ce qui concerne sa

localisation en membrane externe, il est probable qu'elle ne puisse s'expliquer qu'avec l'interaction VirB3/VirB4.

VirB4 est situé dans le cytoplasme de la bactérie et c'est corroboré par la topologie de cette protéine qui est une alternance d'hélices  $\alpha$  et de brins  $\beta$ . Cette localisation pourrait être partielle car le logiciel Prosite identifie sur la séquence de VirB4 le motif RGD (entre les résidus 325 et 327) et ce motif a, comme rôle potentiel, l'adhésion aux cellules hôtes. Donc ce motif doit être dirigé vers l'extérieur. Le problème est qu'il n'y a qu'une centaine de résidus qui sépare ce motif RGD du motif de liaison à l'ATP retrouvé par le programme MOTIF (résidus 458 à 465) et ce deuxième motif, lui, doit obligatoirement se situer dans le cytoplasme. Plusieurs hypothèses peuvent être proposées en ce qui concerne ces deux motifs :

- les deux motifs sont fonctionnels mais à des moments différents
- lorsque le peptidoglycane est digéré, les deux membranes se rapprochent suffisamment pour que 100 résidus soient suffisants pour traverser l'espace périplasmique et donc les deux motifs peuvent être fonctionnels en même temps
- le motif RGD est retrouvé par hasard, un « mot » de trois lettres successives sur une séquence n'est peut-être pas très spécifique

L'effet de VirB4 sur la localisation en membrane externe de VirB3 pourrait aussi être indirect. VirB4 pourrait agir sur une protéine insérée en membrane interne mais qui posséderait un grand domaine périplasmique pouvant s'étendre jusqu'en membrane externe et agir sur VirB3. Cela pourrait aussi expliquer la présence de VirB3 en membrane externe uniquement quand VirB4 est non muté.

**- VirB5 pourrait mimer l'action d'une protéine associée aux microtubules et donc être la protéine qui dévie la trafic intracellulaire de *Brucella*.**

VirB5 est composée d'une suite d'hélices  $\alpha$ . La famille protéique retrouvée par Block et par différentes méthodes de pseudo-threading (Ucla-Doe et 3D-PSSM) est celle des apolipoprotéines. Un autre résultat intéressant pour VirB5 est l'homologie retrouvée avec la MAP1A (Microtubule Associated Protein 1A) par Blast avec la banque d'ESTs humaines. VirB5 pourrait mimer l'action de cette protéine et agir sur les microtubules de la cellule hôte afin de dévier le trafic de la vacuole qui contient *Brucella*, qui se fait grâce à ces microtubules. Un résultat pourrait être déroutant avec les données de la littérature, c'est la localisation de VirB5 prédite par PSORT. Il la situe dans le périplasma alors que VirB5 d'*A.tumefaciens* est exportée. Ceci est dû à la version de PSORT disponible sur Internet qui ne prédit pas encore les localisations extracellulaires. Toutefois PSORT prédit un peptide signal ce qui signifie que la protéine est exportée. De l'exportation à la sécrétion, il n'y a plus qu'un seul pas à franchir.

**- VirB9 serait située en membrane externe ou serait sécrétée.**

PSORT trouve une hélice transmembranaire pour la protéine VirB9 et donc la localise en membrane interne alors que PHD ne trouve pas d'hélice transmembranaire. La méthode de prédiction de structures secondaires de PSORT étant plus ancienne que celle de PHD, nous tiendrons compte de ce dernier logiciel.

De plus, cette version de PSORT n'est pas très confiante dans ses prédictions de localisation en membrane interne. Les données de la littérature localisent VirB9 en membrane interne mais d'autres articles localisent les homologues de cette protéine en membrane externe ou les prédisent comme protéines sécrétées (Tab.6). Afin d'argumenter l'hypothèse d'une localisation autre qu'en membrane interne pour VirB9, nous avons envoyé cette protéine à un programme de prédictions spécifiques des hélices transmembranaires (TMHMM). Ce logiciel ne trouve aucune hélice transmembranaire, cela nous renforce dans l'idée que VirB9 de *Brucella suis* ne se situerait pas en membrane interne.

**- L'orf12 serait composée de deux domaines dont l'un aurait un « fold » de porine et l'autre un « fold » d'hydrolase.**

L'orf12 possède deux types de successions de structures secondaires différentes, la moitié C-terminale de sa séquence est composée d'une série de brins  $\beta$  alors que la partie N-terminale possède des brins  $\beta$  et des hélices  $\alpha$ . PSORT la localise en membrane externe et reconnaît le consensus lipoprotéique. Le logiciel Blockmaker identifie, sur la séquence de l'orf12, les blocs caractéristiques des protéines « OmpA-like » qui sont des porines. Blast PDB et 3D-PSSM trouvent, comme repliement protéique possible, la famille « Immunoglobulin-like ». C'est compatible avec les prédictions de structures secondaires de PHD pour la première moitié de la séquence étant donné que les immunoglobulines et les porines possèdent des structures composées uniquement de brins  $\beta$ . Par contre, ces programmes retrouvent aussi, comme homologues de structures 3D connues, des hydrolases qui sont composées d'une alternance d'hélices  $\alpha$  et de brins  $\beta$ . Ces prédictions sont compatibles avec les prédictions de structures secondaires de la seconde partie de la protéine.

### **4.3. Synthèse**

VirB1 a une activité enzymatique, il possède des homologies avec les lysozymes qui dégradent le peptidoglycane. Il se pourrait donc que la fonction supposée dans la littérature de VirB1 soit correcte. Il dégraderait le peptidoglycane afin de permettre un bon positionnement du complexe de sécrétion.

L'ancrage du pilus, composé essentiellement de VirB2, se ferait en membrane interne.

VirB3 est probablement localisée en membrane interne mais sa localisation en membrane externe reste incertaine. Nous ne pouvons donc pas prendre position à propos de son rôle supposé dans la littérature qui serait l'assemblage en un pore permettant le passage de complexes ADN/protéines, étant donné que sa localisation semble dépendre de VirB4

VirB4 et VirB11 fournissent l'énergie au système de sécrétion. En effet, les motifs de liaison à l'ATP/GTP sont identifiés sur leurs séquences protéiques.

VirB5 pourrait être la protéine qui dévie la vacuole contenant *Brucella* du trajet de la voie de phagocytose classique.

La localisation de la protéine VirB9 est ambiguë. La littérature la situe en membrane interne mais elle ne posséderait pas d'hélices transmembranaires pour s'y insérer. Rappelons qu'elle pourraient être localisée en membrane externe comme son homologue, la protéine PtlF de *Bordetella pertussis*.

Après avoir effectué ce « débroussaillage » du système de sécrétion de type IV de *Brucella suis*, nous allons approfondir l'analyse de deux de ses composants, ce qui était le deuxième objectif de ce travail.

#### **4.4. Analyse plus détaillée de deux composants du système**

Le choix des deux protéines à analyser plus en profondeur s'est porté sur VirB1 et VirB5.

La sélection de VirB5 s'est fait grâce à son homologie avec la MAP1A. En effet, il est très intéressant qu'une protéine sécrétée soit similaire à une protéine de la cellule hôte. VirB1 a été choisie car elle semble homologue aux lysozymes ce qui corrobore l'action présumée de cette protéine, qui est la dégradation du peptidoglycane.

##### **4.4.1. VirB5**

###### **4.4.1.1. Analyse des alignements**

Un homologue de VirB5 est la protéine humaine (Soares\_NhHMPu\_S1) que nous noterons S1 (trouvé Blast avec la banque de données ESTs humaines).

Dans les résultats obtenus par Blast ESTs humaines, VirB5 a comme homologue une protéine humaine (Soares\_NhHMPu\_S1) que nous appellerons S1 pour plus de facilité et qui est similaire à la MAP1A. S1 partage 31 % d'identité avec la protéine VirB5 et 55 % de similarité sur une longueur de 95 résidus soit 40 % de la séquence de VirB5. L'alignement de ces deux segments protéiques est donc relativement bon (Fig.27). Cependant comme la banque de référence est la banque d'ESTs, la séquence protéique de S1 est une séquence déduite d'une séquence nucléotidique, dans laquelle un codon stop est traduit par un acide aminé hypothétique représenté par la lettre Z (Fig.28). Cela n'a évidemment aucun sens biologique. Nous avons donc sélectionné la partie de la séquence commençant après le dernier Z, que nous avons appelé S1 « réduite » (106 résidus), afin de refaire un alignement plus significatif. Les deux alignements étaient identiques à partir du résidu 183 de VirB5 et ce jusqu'à la fin de la séquence (Fig.29).

Nous avons aligné S1 à la séquence de la MAP1A afin de voir à quel point les deux séquences étaient similaires. En fait pour la séquence de S1 dont l'alignement avec VirB5 est conservé dans les deux alignements précédemment effectués, c'est une identité parfaite entre la MAP1A et S1 « réduite » excepté les résidus X de la MAP1A qui représentent un résidu dont on connaît la présence mais que l'on n'a pas encore identifié (Fig.30).

Nous avons effectué les alignements multiples ClustalW et Match-Box pour les protéines VirB5, S1 et MAP1A « réduite » (résidus 1341 à 1434) (Fig.31). Les deux alignements sont identiques à partir du résidu 163 de la protéine VirB5.

De ces différents alignements, nous pouvons dire qu'il existe une similarité entre ces trois séquences protéiques. Les analyses suivantes servent à qualifier cette homologie.

Une étude plus approfondie de la MAP1A a été réalisée. C'est le motif [K/R]-[K/R]-[D/E] qui est responsable de la liaison aux microtubules et ils sont au nombre de huit localisés dans la partie N-terminale de la protéine (du résidu 415 au résidu 541) (John K. Fink 1996). Or, VirB5 s'aligne à la MAP1A entre les résidus 1341 et 1434. D'après les données de la littérature, VirB5 ne mimerait donc probablement pas la fonction de liaison aux microtubules de la MAP1A. Cette protéine est divisée en plusieurs blocs. VirB5 s'aligne à la MAP1A entre deux blocs, dans une région moins conservée chez les différentes espèces contenant une MAP1A (John K. Fink 1996). Ceci est un argument de plus qui corrobore le fait que la région où s'aligne VirB5, et donc la protéine VirB5 elle-même, ne sont pas impliquées dans la liaison aux microtubules.

Nous avons analysé la région de la MAP1A qui nous intéressait (du résidu 1342 à 1434) et nous avons découvert que cette partie de séquence, répertoriée sous le nom PD000002 dans ProDom (Fig.32), avait beaucoup d'homologues dans diverses protéines qui n'ont aucune fonction en commun. Leur seul lien est un motif : le « coiled-coil » (Fig.33). La similarité partagée par la MAP1A et VirB5 est probablement structurale.

Un argument qui renforce cette idée est l'homologie retrouvée pour VirB5 par 3D-PSSM avec l'apolipoprotéine E4. Cette protéine possède également un domaine « coiled coil » (Fig.34).

#### 4.4.1.2. Le « coiled coil »

La structure « coiled coil » est représentée en figure 35. C'est un ensemble d'hélices  $\alpha$  qui s'assemblent en une superhélice. Le nombre de ces hélices peut varier entre 2 et 5. Les « coiled coils » se distinguent par un motif répété de résidus hydrophobes et hydrophiles, le « heptad repeat » (Lupas 1996).

En ce qui concerne la MAP1A « réduite », le programme PHD prédit trois hélices  $\alpha$  dans la partie de séquence qui s'aligne à VirB5. Cela confirmerait la présence d'un « coiled coil » (Fig.36). Pour VirB5, les prédictions de structures secondaires sont aussi en accords avec la présence d'un « coiled coil » (Fig.37). En effet, PHD prédit trois hélices  $\alpha$  dans le segment protéique de VirB5 qui s'aligne

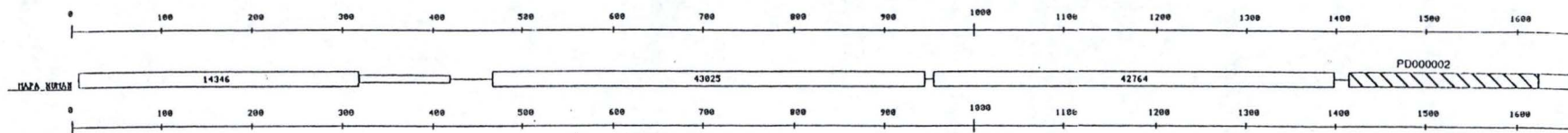


Fig.32 : Représentation des premiers blocs de la MAPIA. Le bloc hachuré (PD000002) correspond à un motif « coiled coil ».

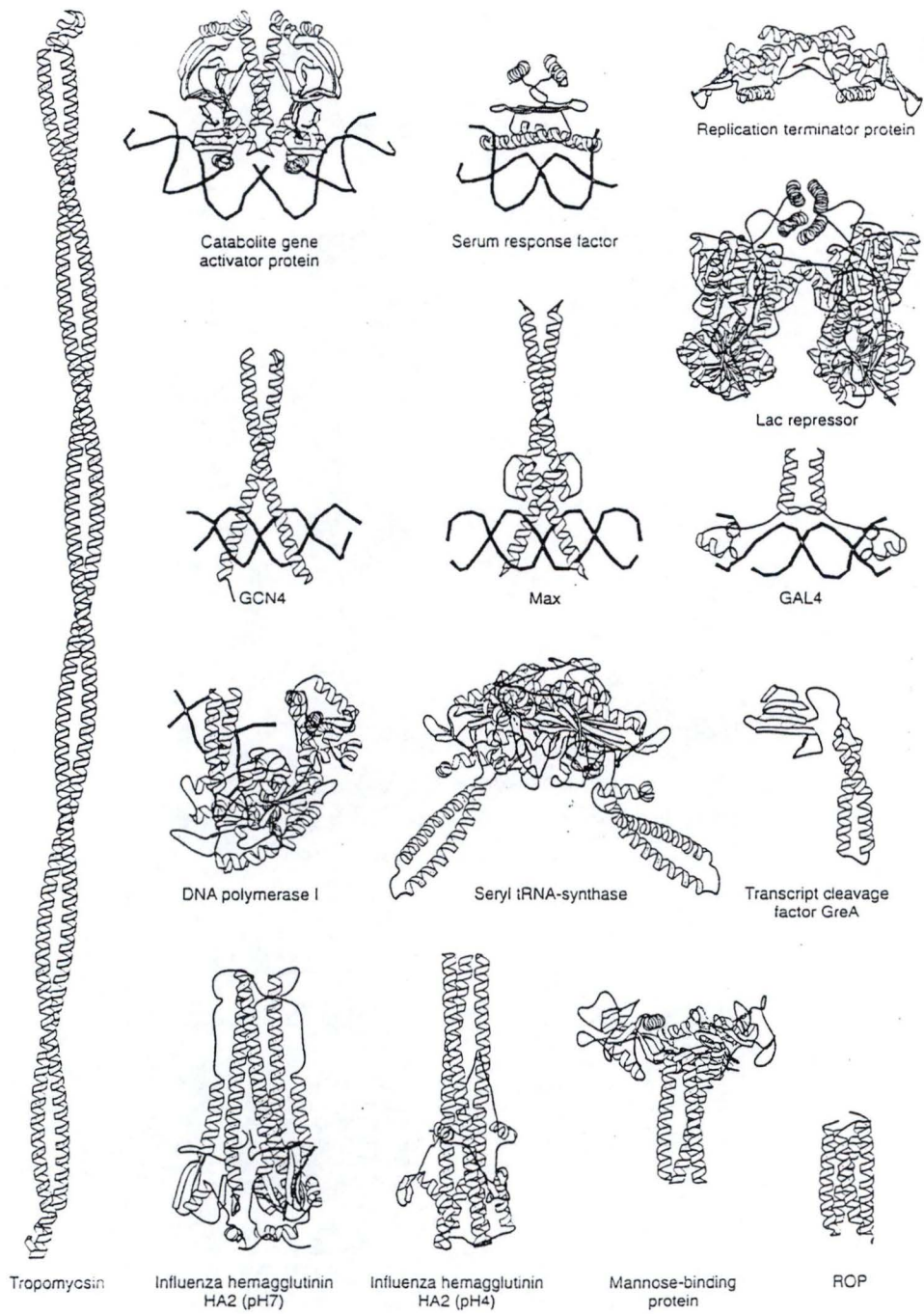


Fig.35 : Exemple d'organisation de « coiled coils » pour des protéines de fonctions différentes (Lupas, 1996).



avec la MAP1A (Fig.38). De plus, il y a une bonne correspondance entre les localisations des ces hélices  $\alpha$  quand on aligne ces deux protéines.

Les structures « coiled coil » ont plusieurs fonctions dans la littérature, généralement la dimérisation de facteurs de transcription, la liaison à l'ADN, le clivage de la polymérase à ARN d'une protéine transcrite, etc... Parmi ces fonctions nous n'avons aucun argument pour en favoriser une par rapport à une autre (Lupas 1996). Par contre, une autre fonction du « coiled coil » décrite dans la littérature pourrait nous intéresser. En effet, ce motif est souvent retrouvé comme étant étroitement associé à la face cytoplasmique des organelles des voies de sécrétion. Plusieurs rôles ont été proposés pour les protéines possédant ce motif dont l'un d'entre eux est impliqué dans l'entrée de vésicules ou d'organelles d'un donneur dans un organelle accepteur avant leur fusion avec celui-ci comme, par exemple, les protéines v-SNARE (qui possèdent également le motif « coiled coil ») qui sont reconnues par des récepteur t-SNARE présents notamment sur le réticulum endoblastique (Weimbs, Low *et al.* 1997) (Bascom, Srinivasan *et al.* 1999).

VirB5 étant déjà, *à priori*, le candidat préférentiel dans la déviation du trafic intracellulaire de la vacuole contenant *Brucella*, ces différentes informations sur le motif « coiled coil » nous renforcent dans l'idée que cette protéine joue, en effet, un rôle important dans la survie intracellulaire de la bactérie.

#### 4.4.2. VirB1

##### 4.4.2.1. Prédiction de fonction

La fonction prédite de VirB1 est la dégradation du peptidoglycane afin de permettre un bon positionnement du complexe de sécrétion de type IV.

Entre les deux membranes des bactéries Gram négatives se trouve le « mur cellulaire » qui est constitué de peptidoglycane. Il se compose de chaînes linéaires d'oligosaccharides qui sont inter-connectées entre elles par de petits peptides (Fig.39) (Stryer 1997). Le peptidoglycane est composé d'oligosaccharides dont le N-acétylglucosamine (NAG) (Fig.40) (Lodish, Baltimore *et al.* 1995).

Le logiciel 3D-PSSM, ainsi que 123D, retrouve, entre autres, pour VirB1 deux protéines « lysozyme-like. Dans ces protéines, en plus de l'homologie structurale il existe une homologie fonctionnelle. En effet, les lysozymes dégradent aussi le peptidoglycane.

Cet argument, qui corrobore la fonction de dégradation du peptidoglycane par VirB1, est renforcé par le résultat retrouvé par Blockmaker. Ce programme identifie sur la séquence de VirB1 les deux blocs caractéristiques de la famille des transglycosylases qui sont des protéines « lysozymes-like » (Fig.41).

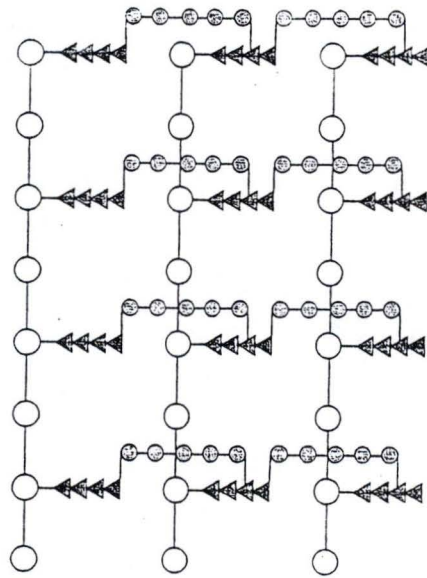


Fig.39 : Diagramme schématique du peptidoglycane. Les sucres sont représentés par les ronds blancs, les tétrapeptides par des triangles et les ponts de pentaglycine par les cercles gris (Stryer, 1997).

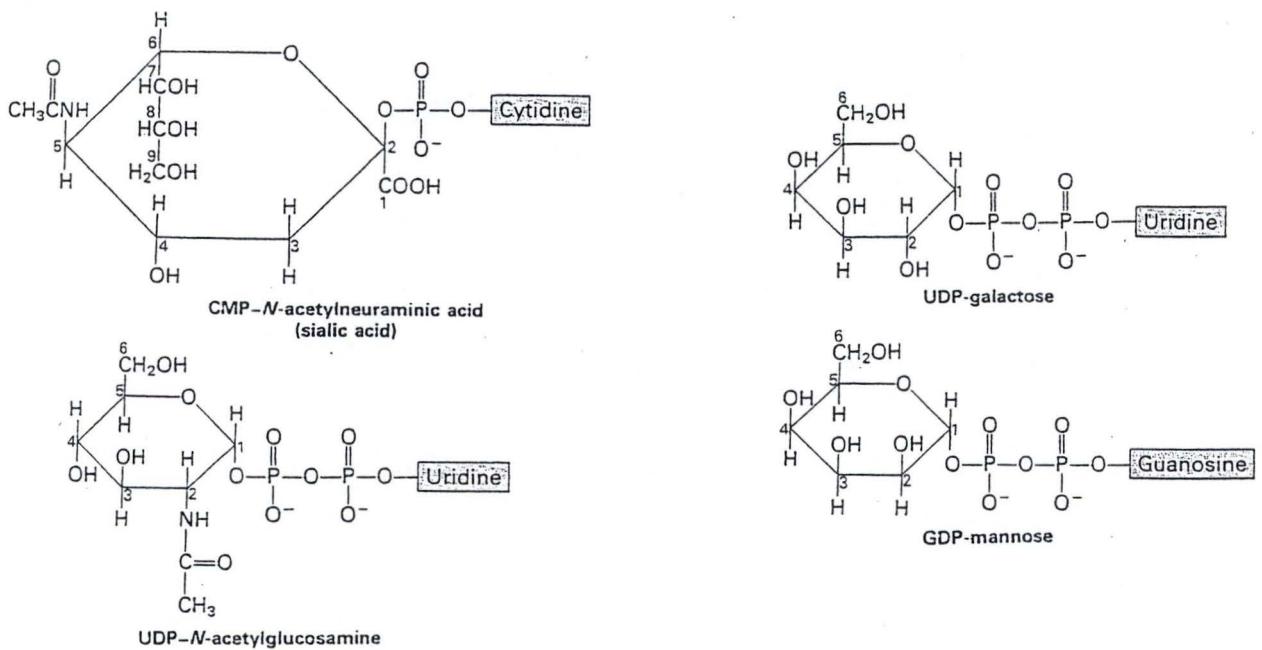


Fig.40 : Structure de l'UDP-N-acetylglucosamine (NAG) (Lodish *et al*, 1995)

#### 4.4.2.2. TOPS

Le programme TOPS nous a permis de mettre en évidence la topologie des lysozymes. Leur topologie est une succession de deux hélices  $\alpha$ , suivies de trois brins  $\beta$  et cela se termine par cinq hélices  $\alpha$  (Fig.42).

Il faut maintenant comparer la topologie des lysozymes aux prédictions de structures secondaires de VirB1.

#### 4.4.2.3. Consensus de structures secondaires

Nous avons effectué un consensus de prédiction de structures secondaires grâce à différentes méthodes : PHD, PsiPred, Prof, JPRED et PREDATOR. Pour chaque prédiction, des scores ont été attribués selon la méthode et selon la structure prédite (en général, il y a « surprédiction » des hélices  $\alpha$  au dépend des brins  $\beta$ ). Cette répartition des scores avait déjà été mise au point, préalablement, dans l'équipe de recherche.

Programme	Hélice $\alpha$		Brin $\beta$	
	confiance	score	confiance	score
PHD	8-9	3	>5	-3
	5-6-7	2	3-4	-2
	<5	1	<3	-1
PsiPred	8-9	3	>5	-3
	5-6-7	2	3-4	-2
	<5	1	<3	-1
	Entre les extrémités	Extrémités	Entre les extrémités	Extrémités
Prof	2	1	-2	-1
PREDATOR JPRED	1		-1	

Après avoir effectué les différentes prédictions, on a sommé les scores donnés pour chaque prédiction de chaque résidu. Les prédictions d'hélices  $\alpha$  étaient confiantes quand le score total était supérieur à 7 et acceptables quand il était à 6. Pour les brins  $\beta$ , les prédictions étaient confiantes quand le score était inférieur à -5 et acceptables quand il était de -4.

Le consensus de structures secondaires, pour VirB1, est présenté en figure 43. On peut y voir cette succession de structures : deux hélices  $\alpha$ , deux brins  $\beta$ , une hélice  $\alpha$ , un brin  $\beta$ , deux hélices  $\alpha$  et trois brins  $\beta$ . C'est différent de la topologie des structures secondaires des lysozymes mais en trois dimensions, il serait possible que VirB1 adopte la même architecture que le lysozyme, excepté ce qui correspond à VirB1\*, si on accepte le fait de permuter deux liens entre deux structures secondaires (Fig.44).



#### 4.4.2.4. *Alignement séquence/structure*

Un alignement a été réalisé par ClustalW entre la protéine VirB1 et le lysozyme 1at6 (Fig.45). Les deux séquences partagent un certain nombre de résidus identiques.

Nous avons observé le site actif du lysozyme (code PDB de la structure :1at6) complexé au NAG. Dans cette structure tridimensionnelle, nous avons mis en évidence les résidus identiques à la protéine VirB5 (Fig.46). Trois des seize résidus identiques à VirB5, sont impliqués dans le site actif du lysozyme. On pourrait donc proposer une mutation ponctuelle

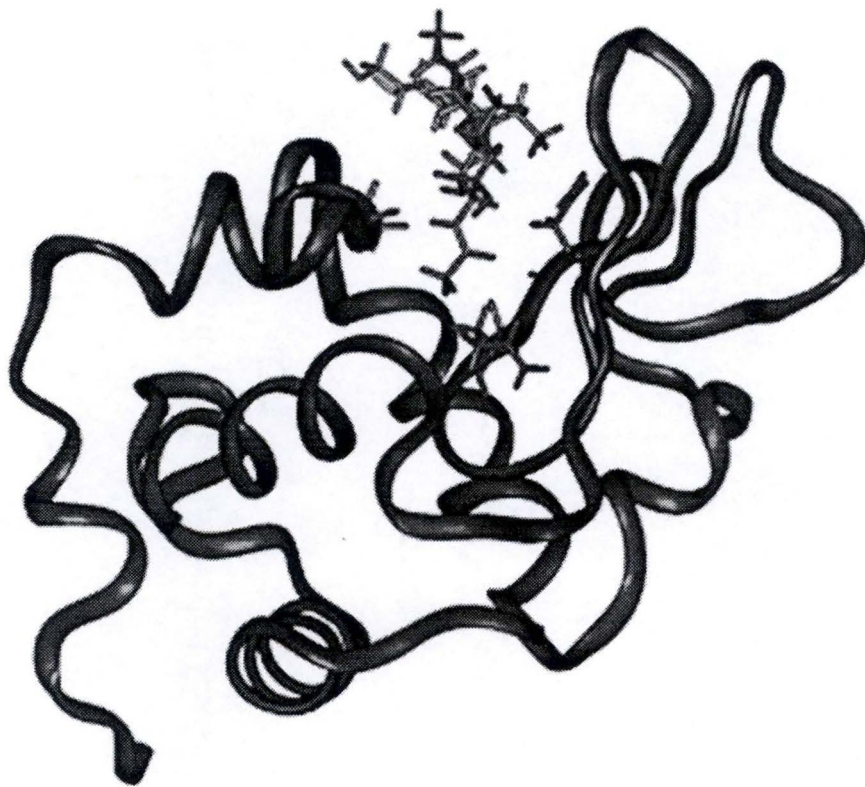


Fig.46 : Structure tridimensionnelle du lysozyme 1at6. Les résidus colorés en clairs sont identiques à la protéine VirB5. De gauche à droite, ce sont alanine 107, glutamine 57 et asparagine 59. Le NAG est représenté en clair dans le site actif du lysozyme.

## 5. Conclusion et perspectives

Dans ce travail, nous avons utilisé divers outils de la bioinformatique pour obtenir un maximum de données sur les différents composants du système de sécrétion de type IV de *Brucella suis*.

Le premier objectif de ce mémoire était le « débroussaillage » structural du système de sécrétion de type IV. Les outils bioinformatiques permettent des recherches d'homologues, des prédictions de localisations cellulaires et de fonctions. Toutes ces analyses se font par des programmes différents mais le but est commun : apporter des éléments qui permettent de poser des hypothèses concernant la structure et la fonction d'une protéine.

Des localisations cellulaires pour toutes les protéines VirB et pour l'orf12 ont été prédites. Les prédictions de localisation de six protéines corroborent les données de la littérature : VirB4, VirB7, VirB8, VirB10, VirB11 en membrane interne et l'orf12 en membrane externe.

De plus, il existe dans la littérature, des hypothèses contradictoires pour la localisation de la protéine VirB2. VirB2 se situerait soit en membrane interne soit en membrane externe. La présence d'hélices transmembranaires détectées grâce à diverses méthodes, nous permettent de prendre position pour une localisation en membrane interne.

Par contre, les données de la littérature pour la localisation des protéines VirB3 et VirB9 sont contredites partiellement pour VirB3 et totalement pour VirB9 par nos différentes prédictions. En effet, VirB3 devrait se situer à la fois dans la membrane interne et dans la membrane externe mais cette dernière localisation ne se détecte pas uniquement sur base de la séquence primaire. Il se pourrait que cette localisation ne s'explique que par une interaction avec la protéine VirB4. VirB9 est localisé en membrane interne dans la littérature mais, d'après nos prédictions, il ne posséderait pas d'hélice transmembranaire pour s'y insérer.

Nous avons pu donner des arguments pour des fonctions prédites dans la littérature de trois protéines VirB du complexe. L'activité de dégradation du peptidoglycane de VirB1 a pu être renforcée à l'aide d'un « fold » similaire à celui du lysozyme. Cette activité devrait permettre un bon positionnement de tout le système de sécrétion. Les protéines VirB4 et VirB11 fourniraient l'énergie au complexe grâce à la liaison et à l'hydrolyse d'ATP.

Un motif RGD est présent dans VirB4 aux résidus 325 à 327. Cette protéine étant cytoplasmique, nous doutons que ce motif puisse jouer un rôle tel qu'une interaction avec les intégrines d'une cellule de l'hôte. Pour le vérifier, nous proposerions une mutagenèse dirigée contre l'un de ces résidus afin de voir si VirB4 est fonctionnel.

La deuxième étape de ce travail était l'analyse plus approfondie de deux composants du système de sécrétion. Notre choix s'était porté sur les protéines VirB1 et VirB5.

Les alignements séquence / structure de VirB1 avec des lysozymes nous ont permis d'observer seize résidus identiques (pour un alignement de 129 résidus) pour les deux séquences dont trois seraient impliqués dans le site actif du lysozyme. On pourrait donc proposer ces résidus de VirB1 pour une mutation ponctuelle afin de tester l'activité enzymatique de cette protéine.

La protéine VirB5 semble particulièrement intéressante de par son éventuel rôle dans le détournement de la voie classique de la phagocytose. Nous proposons comme hypothèse que cette déviation se fait *via* le motif « coiled coil » situé à l'extrémité C-terminale de VirB5 qui mimerait le « coiled coil » des protéines v-SNARE dont la fonction est la reconnaissance des récepteurs t-SNARE et la fusion de la vésicule contenant ces v-SNARE à l'organite présentant les t-SNARE.

Cette fonction doit être confirmée expérimentalement :

- Par co-immunoprécipitation de VirB5, on savoir si VirB5 précipité est associé à des récepteurs du genre t-SNARE.
- Par fusion du motif « coiled coil » à une protéine fluorescente (exemple : GFP) qui n'est pas sensée interagir avec des vacuoles associées au réticulum endoplasmique, qui seraient elles aussi marquées (*via* une molécule fluorescente associée à un marqueur spécifique de la vacuole de réplication) on pourrait tenter d'observer une colocalisation des deux marqueurs fluorescents. OU, on pourrait compare une colocalisation d'une souche de *Brucella* où VirB5 est sauvage et une souche dont le motif « coiled coil » est muté.

Le suivi idéal à ce mémoire serait une vérification expérimentale de toutes les hypothèses produites après l'analyse des différents résultats obtenus pour ces douze protéines.



## 6. Bibliographie

- A. Covacci, J. L. T., G. Del Giudice, J. Parsonnet, R. Rappuoli (1999). "Helicobacter pylori Virulence and Genetic Geography." Science **284**.
- A. Louise Jones, E.-M. L., Ken Shirasu, and Clarence I. Kado (1996). "VirB2 Is a Processed Pilin-Like Protein Encoded by the Agrobacterium tumefaciens Ti Plasmid." Journal of Bacteriology **178**(19): 5706-5711.
- Alexandrov, N. N., R. Nussinov, *et al.* (1996). "Fast protein fold recognition via sequence to structure alignment and contact capacity potentials." Pac Symp Biocomput: 53-72.
- Alton, G. G., L.M.Jones, R.D.Angus, and J.M.Verger (1988). Techniques for the brucellosis laboratory. Paris, INRA.
- Altschul, S. F., W. Gish, *et al.* (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.
- Altschul, S. F., T. L. Madden, *et al.* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.
- Ana Soloaga, M. P. V., Luis M. Garcia-Segura, Helena Ostolaza, Robert Brasseur and Félix M. Goni (1999). "Insertion of Escherichia coli  $\alpha$ -haemolysin in lipid bilayer as a non-transmembrane integral protein : prediction and experiment." Molecular Microbiology **31**(4): 1013-1024.
- Anath Das, Y.-H. X. (2000). "The Agrobacterium T-DNA Transport Pore VirB8, VirB9, and VirB10 Interact with One Another." Journal of Bacteriology **182**: 758-763.
- Bairoch, A. and R. Apweiler (1996). "The SWISS-PROT protein sequence data bank and its new supplement TREMBL." Nucleic Acids Res **24**(1): 21-5.
- Bairoch, A. and R. Apweiler (2000). "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000." Nucleic Acids Res **28**(1): 45-8.
- Barbara R. Berger, P. J. C. (1993). "The Agrobacterium tumefaciens virB4 Gene Product Is an Essential Virulence Protein Requiring an Intact Nucleoside Triphosphate-Binding Domain." Journal of Bacteriology **175**: 1723-1734.
- Bascom, R. A., S. Srinivasan, *et al.* (1999). "Identification and characterization of golgin-84, a novel Golgi integral membrane protein with a cytoplasmic coiled-coil domain [published erratum appears in J Biol Chem 1999 Apr 30;274(18):12950]." J Biol Chem **274**(5): 2953-62.
- Beckwith, C. K. M. a. J. (1996). Export of Proteins to the Cell Envelope in Escherichia coli. Escherichia coli and Salmonella. F. C. Neidhardt. Washington, D;C;, ASM PRESS. **1**: 967-977.

Benson, D. A., M. S. Boguski, *et al.* (1999). "GenBank." Nucleic Acids Res **27**(1): 12-7.

Boguski, M. S. (1998). "Bioinformatics - a new era." trends guide to bioinformatics: 1-3.

Christian Baron, M. L., Sharleen Zhou, Patricia C. Zambrysky (1997). "VirB1, a Component of the T-Complex Transfer Machinery of *Agrobacterium tumefaciens*, Is Processed to a C-Terminal Secreted Product, VirB1\*." Journal of Bacteriology **179**: 1203-1210.

Christian Baron, Y. R. t., Patricia C. Zambrysky (1997). "The Lipoprotein VirB7 Interacts with VirB9 in the Membranes Of *Agrobacterium tumefaciens*." Journal of bacteriology **179**: 1211-1218.

Christie, P. J. (1997). "*Agrobacterium tumefaciens* T-Complex Transport Apparatus : a Paradigm for a New Family of Multifunctional Transporters in Eubacteria." Journal of Bacteriology **179**: 3085-3094.

Christine E. Beaupré, J. B., Elizabeth M. Dale, Andrew N. Binns (1997). "Interactions between VirB9 and VirB10 Membrane Proteins Involved in Movement of DNA from *Agrobacterium tumefaciens* into Plant Cells." Journal of Bacteriology **179**: 78-89.

Cornelis, G. R., T. Biot, *et al.* (1989). "The *Yersinia yop regulon*." Mol Microbiol **3**(10): 1455-9.

Cornelis, G. R., A. Boland, *et al.* (1998). "The Virulence Plasmid of *Yersinia*, an Antihost Genome." Microbiology and Molecular Biology Reviws **62**(4): 1315-1352.

Corpet, F., J. Gouzy, *et al.* (1999). "Recent improvement of the ProDom database of protein domain families." Nucleic Acids Research **27**(1): 263-267.

Creighton, T. E. (1993). Proteins Structures and Molecular Properties. New York.

Cristobal, S., J. W. de Gier, *et al.* (1999). "Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*." Embo J **18**(11): 2982-90.

Cuff, J. A., M. E. Clamp, *et al.* (1998). "JPred: a consensus secondary structure prediction server." Bioinformatics **14**(10): 892-3.

D.O'Callaghan, C. C., A. Allardet-Servent, M.L. Boschiroli, G. Bourg, V. Foulongne, P. Frutos, Y. Kulakov, M. Ramuz (1999). "A homologue of the *Agrobacterium tumefaciens* VirB and *Bordetella pertussis* Ptl type IV secretion system is essential for intracellular survival of *Brucella suis*." Molecular Microbiology **33**: 1210-1220.

Das, A., L. B. Anderson, *et al.* (1997). "Delineation of the interaction domains of *Agrobacterium tumefaciens* VirB7 and VirB9 by use of the yeast two-hybrid assay." J Bacteriol **179**(11): 3404-9.

Das, A. and Y. H. Xie (2000). "The *Agrobacterium* T-DNA transport pore proteins VirB8, VirB9, and VirB10 interact with one another." J Bacteriol **182**(3): 758-63.

David Fernandez, G. M. S., Xue-Rong Zhou, Peter J. Christie (1996). "The Agrobacterium tumefaciens VirB7 Lipoprotein Is Required for Stabilization of VirB Proteins during Assembly of the T-Complex Transport Apparatus." Journal of bacteriology **178**: 3168-3176.

Depiereux, E. and E. Feytmans (1992). "MATCH-BOX : a fundamentally new algorithm for the simultaneous alignment of several protein sequences." CABIOS **8**(5): 501-509.

Doolittle, R. F. (1981). "Similar amino acid sequences: chance or common ancestry?" Science **214**(4517): 149-59.

Driessen, P. F. a. A. J. m. (1999). "Protein Targeting to the Bacterial Cytoplasmic Membrane." Microbiology and Molecular Biology Reviews **63**(1): 161-173.

Enright, F. M. (1990). The pathogenesis and pathobiology of Brucella infection in domestic animals. Animal brucellosis. C. Press, Nielsen and J. R. Duncan: 301-320.

Erh-Min Lai, C. I. K. (1998). "Processed VirB2 Is the Major Subunit of the Promiscuous Pilus of Agrobacterium tumefaciens." Journal of bacteriology **180**: 2711-2717.

Fischer, D., C. Barret, *et al.* (1999). "CAFASP-1: critical assessment of fully automated structure prediction methods." Proteins Suppl(3): 209-17.

Frishman, D. and P. Argos (1996). "Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence." Protein Eng **9**(2): 133-42.

Frishman, D. and P. Argos (1997). "Seventy-five percent accuracy in protein secondary structure prediction." Proteins **27**(3): 329-35.

Fullner, K. J. (1998). "Role of Agrobacterium virB genes in transfer of T complexes and RSF1010." J Bacteriol **180**(2): 430-4.

Hamilton, C. M., H. Lee, *et al.* (2000). "TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58." J Bacteriol **182**(6): 1541-8.

Hamm, G. H. and G. N. Cameron (1986). "The EMBL data library." Nucleic Acids Res **14**(1): 5-9.

Heike Schmidt-Eisenlohr, N. D., Christina Angerer, Gerhard Wanner, Patricia C. Zambrysky, Christian Baron (1999). "Vir Proteins Stabilize VirB5 and Mediate Its Association with the T Pilus of Agrobacterium tumefaciens." Journal of Bacteriology **181**: 7485-7492.

Henikoff, S. and J. G. Henikoff (1994). "Protein family classification based on searching a database of blocks." Genomics **19**(1): 97-107.

Hesterkamp, T. and B. Bukau (1996). "The Escherichia coli trigger factor." FEBS Lett **389**(1): 32-4.

Hofmann, K., P. Bucher, *et al.* (1999). "The PROSITE database, its status in 1999." Nucleic Acids Res **27**(1): 215-9.

- Holland, C. C. a. I. B. (1996). "Random and Directed Mutagenesis To Elicitate the Functional Importance of Helix II and F-989 in the C-Terminal Secretion Signal of *Escherichia coli* Hemolysin." Journal of Bacteriology **178**(4): 1232-1236.
- Hueck, C. J. (1998). "Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants." Microbiology and Molecular Biology Reviews **62**(2): 379-433.
- John E. Cronan, J., Robert B. Gennis, and Stanley R. Maloy (1987). Cytoplasmic membrane. Escherichia coli, Salmonella typhimurium. F. C. Neidhardt. Washington DC 20006, American Society for Microbiology. **1**: 31-55.
- John K. Fink, S. M. J., Constance Esposito and Jodi Wilkowski (1996). "Human Microtubule-Associated Protein 1a (MAP1A) Gene : Genomic Organization, cDNA Sequence, and Developmental- and Tissue-Specific Expression." Genomics **35**: 577-585.
- John R Zupan, D. W., Patricia Zambryski (1998). "Assembly of the VirB transport complex for DNA transfer from *Agrobacterium tumefaciens* to plant cells." Current Opinion in Microbiology **1**: 649-655.
- Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.
- Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-637.
- Karine E. Finberg, T. R. M., Sarah P. Young, Jessica B. Maken, Shannon M. Heitritter, Andrew N. Binns, Lois M. Banta (1995). "Interactions of VirB9, -10, -11 with the Membrane Fraction of *Agrobacterium tumefaciens* : Solubility Studies Provide Evidence for Tight Associations." Journal of Bacteriology **177**: 4881-4889.
- Kay, B. K., J. Winter, *et al.* (1996). Phage Display of Peptides and Proteins. San Diego, ACADEMIC PRESS, Inc.
- Kelley, L., R. MacCallum, *et al.* (2000). "Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM." J. Mol. Biol **299**(2): 501-522.
- Kelley, L. A., R. Maccallum, *et al.* (1999). RECOMB 99, Proceedings of the Third Annual Conference on Computational Molecular Biology. New York, The Association for Computing Machinery.
- King, R. D. and M. J. Sternberg (1996). "Identification and application of the concepts important for accurate and reliable protein secondary structure prediction." Protein Sci **5**(11): 2298-310.
- Klein, P., M. Kanehisa, *et al.* (1985). "The detection and classification of membrane-spanning proteins." Biochim Biophys Acta **815**(3): 468-76.

Konstanze Beck, L.-F. W., Josef Brunner and Matthias Muller (2000). "Discrimination between SRP- and SecA/SecB- dependant substrates involves selective recognition of nascent chains by SRP and trigger factor." The EMBO Journal **19**(1): 134-143.

Kudlicki, W., O. W. Odom, *et al.* (1996). "Binding of an N-terminal rhodanese peptide to DnaJ and to ribosomes." J Biol Chem **271**(49): 31160-5.

Lewitter, F. (1998). "Text-based database searching." trends guide to bioinformatics: 3-5.

Lodish, A., D. Baltimore, *et al.* (1995). Molecular Cell Biology. New York, Scientific American Books, Inc.

Lory, S. (1998). "Secretion of proteins and assembly of bacterial surface organelles : shared pathways of extracellular protein targeting." Current Opinion in Microbiology: 27-35.

Lupas, A. (1996). "Coiled coils: new structures and new functions." Trends Biochem Sci **21**(10): 375-82.

McGeoch, D. J. (1985). "On the predictive recognition of signal peptide sequences." Virus Res **3**(3): 271-86.

Miller, R. T., D. T. Jones, *et al.* (1996). "Protein fold recognition by sequence threading: tools and assessment techniques." Faseb J **10**(1): 171-8.

Murzin, A. G., S. E. Brenner, *et al.* (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." J Mol Biol **247**(4): 536-40.

Myers, E. W. and W. Miller (1988). "Optimal alignments in linear space." Comput Appl Biosci **4**(1): 11-7.

Nakai, K. and M. Kanehisa (1991). "Expert system for predicting protein localization sites in gram-negative bacteria." Proteins **11**(2): 95-110.

Nikaido, K. and G. F. Ames (1999). "One intact ATP-binding subunit is sufficient to support ATP hydrolysis and translocation in an ABC transporter, the histidine permease." J Biol Chem **274**(38): 26727-35.

Okamoto, S., A. Toyoda-Yamamoto, *et al.* (1991). "Localization and orientation of the VirD4 protein of *Agrobacterium tumefaciens* in the cell membrane." Mol Gen Genet **228**(1-2): 24-32.

Oliver, D. B. (1987). Periplasm and Protein Secretion. Escherichia coli and Salmonella typhimurium. F. C. Neidhardt. Washington, DC 20006, American Society for Microbiology. **1**: 56-69.

Orengo, C. (1994). "Classification of protein folds." Curr. Op. Struct. Biol. **4**: 429-440.

Orengo, C. A., A. D. Michie, *et al.* (1997). "CATH--a hierarchic classification of protein domain structures." Structure **5**(8): 1093-108.

Ouellette, A. D. B. a. B. F. F. (1998). Bioinformatics, A Practical Guide to the Analysis of Genes and Proteins. New York, A John Wiley & sons, inc., Publication.

P. Berche, J.-L. G. e. M. S. (1988). Bactéries des infections humaines. Bactériologie. Médecine-Science/Fammarion. Paris.

Parry-Smith, T. K. A. a. D. J. (1999). Introduction to bioinformatics. England, Addison Wesley Longman Limited.

Pizarro-Cerda, J., S. Meresse, *et al.* (1998). "Brucella abortus transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes." Infect Immun **66**(12): 5711-24.

Prescott, H., Klein (1995). Microbiologie. Bruxelles.

Ralf Eisenbrant, M. K., Erh-Min Lai, Rudi Lurz, Clarence I. Kados, and Erich Lanka (1999). "Conjugative Pili of IncP Plasmids and the Ti Plasmid T Pilus Are Composed of Cyclic Subunits." The Journal of Biological Chemistry **274**(32): 22548-22555.

Rashkova, S., G. M. Spudich, *et al.* (1997). "Characterization of membrane and protein interaction determinants of the Agrobacterium tumefaciens VirB11 ATPase." J Bacteriol **179**(3): 583-91.

Rawn, J. D. (1990). Traité de Biochimie. Bruxelles.

Reeves, G. P. C. S. a. P. J. (1993). "Membrane traffic wardens and protein secretion in Gram-negative bacteria." Trends in biochemistry Science **18**: 7-12.

Rost, B. (1995). "TOPITS: threading one-dimensional predictions into three-dimensional structures." Ismb **3**: 314-21.

Rost, B. and C. Sander (1993). "Prediction of protein secondary structure at better than 70% accuracy." J Mol Biol **232**(2): 584-99.

Rost, B., C. Sander, *et al.* (1994). "PHD--an automatic mail server for protein secondary structure prediction." Comput Appl Biosci **10**(1): 53-60.

Rost, B., R. Schneider, *et al.* (1997). "Protein fold recognition by prediction-based threading." J Mol Biol **270**(3): 471-80.

Salamov, A. A. and V. V. Solovyev (1995). "Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments." J Mol Biol **247**(1): 11-5.

Schneewind, D. M. A. a. O. (1999). "Type III machines of Gram-negative pathogens : injecting virulence factors into host cells and more." Current Opinion in Microbiology **18**(2): 18-24.

Schulder, G. D. (1999). Sequence Alignment and Database Searching. Bioinformatics A Practical Guide to the Analysis of Genes and Proteins. A. D. B. a. B. F. F. Ouellette. New York: 145-171.

- Sippl, M. J. (1993). "Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures." J Comput Aided Mol Des **7**(4): 473-501.
- Stephens, K. M., C. Roush, *et al.* (1995). "Agrobacterium tumefaciens VirB11 protein requires a consensus nucleotide-binding site for function in virulence." J Bacteriol **177**(1): 27-36.
- Sternberg, M. J. E. (1996). Protein Structure Prediction : A Practical Approach. New York, Oxford University Press.
- Stryer, L. (1997). Biochemistry. New York, W. H. Freeman and Company.
- Thomas Klauser, J. P. a. T. F. M. (1993). "The Secretion Pathway of IgA Protease-type Proteins in Gram-negative Bacteria." BioEssays **15**(12): 799-805.
- Thompson, J. D., D. G. Higgins, *et al.* (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.
- Thorstenson, Y. R. and P. C. Zambryski (1994). "The essential virulence protein VirB8 localizes to the inner membrane of Agrobacterium tumefaciens." J Bacteriol **176**(6): 1711-7.
- Tu Anh Dang, X.-R. Z., Brenda Graf and Peter J. Christie (1999). "Dimerization of the Agrobacterium tumefaciens VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter." Molecular Microbiology **32**: 1239-1253.
- Tu Anh T. Dang, P. J. C. (1997). "The VirB4 ATPase of Agrobacterium tumefaciens Is a Cytoplasmic Membrane Protein Exposed at the Periplasmic Surface." Journal of Bacteriology **179**: 453-462.
- Vinals, C. (1996). Modélisation de structures tridimensionnelles de protéines : application à l'étude de la stéréospécificité des lactate déshydrogénases. Unité de recherche en biologie moléculaire. Namur, Facultés Universitaires Notre-Dame de la Paix.
- von Heijne, G. (1986). "A new method for predicting signal sequence cleavage sites." Nucleic Acids Res **14**(11): 4683-90.
- von Heijne, G. (1989). "The structure of signal peptides from bacterial lipoproteins." Protein Eng **2**(7): 531-4.
- Wandersman, C. (1996). Secretion across the Bacterial Outer Membrane. Escherichia coli and Salmonella. F. C. Neidhardt. ashington, D.C., ASM PRESS. **1**: 955-966.
- Weimbs, T., S. H. Low, *et al.* (1997). "A conserved domain is present in different families of vesicular fusion proteins: a new superfamily." Proc Natl Acad Sci U S A **94**(7): 3046-51.
- Westhead, D. R., T. W. Slidel, *et al.* (1999). "Protein structural topology: Automated analysis and diagrammatic representation." Protein Sci **8**(4): 897-904.

Xue-Rong Zhou, P. J. C. (1997). "Suppression of Mutant Phenotypes of *Agrobacterium tumefaciens* VirB11 ATPase by Overproduction of VirB Proteins." Journal of Bacteriology **179**: 5835-5842.

Zvelebil, M. J., G. J. Barton, *et al.* (1987). "Prediction of protein secondary structure and active sites using the alignment of homologous sequences." J Mol Biol **195**(4): 957-61.