

## THESIS / THÈSE

### MASTER EN SCIENCES BIOLOGIQUES

#### Caractérisation structurale d'un régulateur appartenant au système de transduction du signal à deux composantes chez *Brucella* sp.

De Fays, Katalin

*Award date:*  
1996

*Awarding institution:*  
Universite de Namur

[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

**FACULTE DES SCIENCES**

-----

**Caractérisation structurale d'un régulateur appartenant au système de  
transduction du signal à deux composantes chez  
*Brucella* sp.**

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

**Katalin de FAYS  
1996**

**Caractérisation structurale d'un régulateur appartenant au système de  
transduction du signal à deux composantes chez  
*Brucella* sp.**

de FAYS Katalin

Résumé

Pour faire face aux variations environnementales, les bactéries ont élaboré des mécanismes d'adaptation, dont les plus largement répandus sont les systèmes de transduction du signal à deux composantes. Ces systèmes utilisent des réactions de phosphorylation comme moyen de communication entre deux partenaires: un senseur, détectant un paramètre environnemental et un régulateur médiant une réponse adaptative, fréquemment une modification de l'expression génique.

Nous avons cloné chez *Brucella abortus* un membre des systèmes à deux composantes (BaCtrA), homologue à OmpR d'*Eschérichia coli* et à CtrA de *Caulobacter crescentus* (CcCtrA). CcCtrA est un gène essentiel encodant un régulateur de réponse qui agit directement au niveau des promoteurs de plusieurs gènes impliqués dans le développement du flagelle ainsi que dans de nombreux évènements liés au cycle cellulaire. BaCtrA présente 80% d'identité avec CcCtrA et possède dans son domaine N-terminal tous les résidus qui identifient les membres de la superfamille des régulateurs. Un modèle tridimensionnel de BaCtrA N-term a été élaboré sur base de l'homologie avec CheY, puis évalué grâce à un potentiel d'énergie extrêmement précis (niveau atomique), mis au point au laboratoire. Un modèle topologique de BaCtrA C-term a été prédit à partir des données obtenues par prédiction de structure secondaire et par threading. Ces modèles structuraux vont nous servir d'outils d'interprétation et de prédiction pour des expériences futures.

Mémoire de licence en Sciences Biologiques

Decembre 1996

**Promoteur:** E. Feytmans

Je tiens tout d'abord à remercier le Professeur Ernest Feytmans de m'avoir ouvert, en m'acceptant dans son laboratoire, les portes du domaine ésotérique qu'est le domaine de la prédiction de structure tridimensionnelle de protéines.

Je voudrais aussi remercier le Professeur Jean-Jacques Letesson pour tous les conseils et les encouragements reçus tout au long de ce travail, arrivant toujours au moment propice. Merci pour "Carpe diem".

Je remercie les membres du jury, X. DeBolle, J-P. Matheise, D. Delforge et P. Van Cutsem, d'avoir la patience de lire ce travail.

Merci Damien de m'avoir guidée à travers toutes les péripéties rencontrées et de m'avoir aidée à les surmonter. Merci pour ton enthousiasme et ton optimisme...

Merci aussi à mes "tuteurs de secours", Pil et Carla, pour l'aide et la patience dont ils ont fait preuve à mon égard, et surtout à Jean-Yves, sans qui ce travail ne serai pas ce qu'il est.

Merci à Anne, Flore, Deborah et Xavier.

Merci à tous les membres du laboratoire et de l'URBM, pour l'ambiance et la bonne humeur apportée quotidiennement.

Merci aux autres mémorants Cynthia, David, Rose-May, Calou, Bernard, Olivier, Sébastien et Laurence, pour le réconfort apporté ces derniers jours.

Enfin, je remercie mes parents et mes petites soeurs d'avoir supporté mon mauvais caractère lors des périodes stressantes de ces cinq dernières années.

# Table des Matières

<b><u>INTRODUCTION.</u></b>	<b>1</b>
I. Régulation des gènes d'adaptation.	1
A. Les bactéries face à leur environnement.	1
B. Type de régulation.	2
1. Amplification et réarrangements géniques.	2
2. Régulation transcriptionnelle	3
3. Régulation post-transcriptionnelle	4
II. Système de transduction du signal à deux composantes.	5
A. Caractéristiques générales	5
1. Les senseurs	6
2. Les régulateurs	7
3. Mécanismes de communication	9
B. Conservation structurale et fonctionnelle des régulateurs	10
1. Historique et liens de parenté entre les différentes familles.	10
2. Structure de CheY	12
a) <i>Fold</i> général et éléments de structure secondaire	12
b) Structure du site actif	13
3. Résidus conservés	13
a) Conservation du noyau hydrophobe et de la structure tertiaire.	13
b) Conservation de la structure secondaire	14
4. Contrôle de l'activité de l'effecteur par le domaine receveur.	16
III. <i>Brucella</i> .	19
<b><u>OBJECTIFS DU TRAVAIL.</u></b>	<b>20</b>
<b><u>MATERIEL et METHODES.</u></b>	<b>21</b>
I. Techniques de génie génétique:	21
A. Matériel	21
1. Souches et plasmides	21
a) <i>E.coli</i> XL1-blue	21
b) pUC19	21
2. Solutions et Tampons	22
a) Solutions	22
b) Tampons	23
3. Milieux de culture	25
B. Méthodes	25
1. Cellules:	25
Transformation par électroporation	25
2. ADN:	26
a) MINIPréparation d'ADN plasmidique	26
b) Extraction au phénol	27
c) Précipitation de l'ADN	27
d) Restriction d'ADN purifié	28
e) Déphosphorylation	28
f) Ligation	29
g) Electrophorèse en gel d'agarose	29

h) Récupération d'ADN sur gel d'agarose par papiers	30
i) Amplification par PCR	31
j) Marquage d'une sonde à la fluorescéine	33
k) Southern blot	34
II. Modélisation de structures tridimensionnelles de protéines	39
A. Matériel	39
B. Méthodes	39
1. Recherche dans les banques de données	40
2. Alignement de séquences	40
3. Attribution de domaines	41
4. Modélisation par homologie	41
5. Modélisation de protéines non-homologues à une structure 3-D connue	
a) Prédiction de structure secondaire	42
b) Reconnaissances de <i>fold</i>	43
c) Prédiction <i>ab initio</i>	45
6. Conclusion	46

## **RESULTATS et DISCUSSION** **47**

### Partie 1: Clonage d'une paire senseur-régulateur chez *B. melitensis* 16M.

I. Amplification PCR de fragments de gènes homologues à OmpB:	47
A. Tentatives d'amplification d'un domaine homologue au domaine effecteur de OmpR.	47
B. Tentatives d'amplification de domaines homologues aux domaines receveur de OmpR et transmetteur de EnvZ.	49
II. Southern blot sur l'ADN génomique de <i>B. melitensis</i> 16M restreint par différentes enzymes de restriction	50
III. Criblage d'une sous-banque génomique de <i>B. melitensis</i> 16M	51
A. Création de la sous-banque	51
1. Récupération des fragments aux alentours de 4Kb sur gel d'agarose	51
2. Préparation du plasmide	51
3. Ligation	52
4. Transformation	52
B. criblage de la sous-banque par Southern blot	53
IV. Criblage de la banque génomique de <i>B. abortus</i> :	53
V. Discussion	55

### Partie 2: Prédiction de la structure 3D du régulateur cloné

I. Recherche d'homologues dans les banques de séquences	58
II. Alignement de séquences et délimitation des domaines de la protéine.	59
III. Modélisation du domaine amino-terminal de BaCtrA	60
A. Recherches complémentaires sur la conformation du domaine amino-terminal de BaCtrA	60
1. Prédiction de structure secondaire	60
2. Reconnaissance de folds par comparaison de structures secondaires et de patterns d'accessibilité	61
3. Reconnaissance de folds par Threading	62
B. Modélisation par homologie du domaine "CheYlike"	62
IV. Tentatives de modélisation du domaine carboxy-terminal "OmpRlike"	64

1. Prédiction de structures secondaires	64
2. Reconnaissance de folds par comparaison de structures secondaires et de patterns d'accessibilité	65
3. Reconnaissance de folds par Threading	65
4. Consensus	66
V. Discussion	67
<b><u>CONCLUSION et PERSPECTIVES</u></b>	<b>69</b>
<b>BIBLIOGRAPHIE</b>	<b>71</b>

# Introduction

## I. Régulation de facteurs d'adaptation:

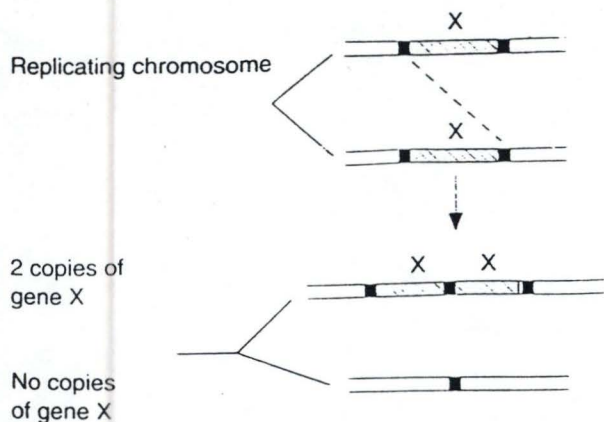
### A. Les bactéries face à leur environnement:

Les bactéries vivent dans des environnements souvent inhospitaliers, dont les caractéristiques peuvent varier considérablement et d'une manière soudaine (température, pH, osmolarité, disponibilité en nutriments, présence de substances toxiques, ...). Prenons le cas d'une bactérie pathogène ingérée par un homme ou un animal avec de l'eau contaminée: dans l'eau, la température était inférieure à 37°C, le pH neutre, la pression osmotique et le taux de nutriments faibles. Brusquement, la bactérie se trouve face à une température et une force osmotique plus élevées, à une multitude de nutriments, mais accessibles sous une forme différente de ceux qu'elle utilisait dans l'eau, et exposée transitoirement à un pH acide dans l'estomac puis à un pH plus basique dans l'intestin.

En plus de la réponse à ces modifications, les bactéries pathogènes doivent aussi coordonner un réseau complexe de facteurs de virulence, dont l'expression doit être précisément contrôlée pour permettre la colonisation de l'hôte. Un facteur de virulence est un produit ou un mécanisme qui contribue à la virulence ou à la pathogénie, comme par exemple les facteurs d'adhérence, toxines, capsules et autres propriétés complexes telles que la motilité et l'invasion cellulaire (Salyers et Whitt, 1994). De plus, la bactérie va devoir élaborer des stratégies pour échapper au système immunitaire de l'hôte. La survie des bactéries dépend donc de la rapidité et de l'originalité de leurs "réflexes" pour contrer cette pression de l'hôte. Pour ce faire, elles doivent constamment percevoir les variations environnementales et ajuster leur structure, leur physiologie, leur métabolisme, leur cycle cellulaire, ..., comme illustré ci-dessous:

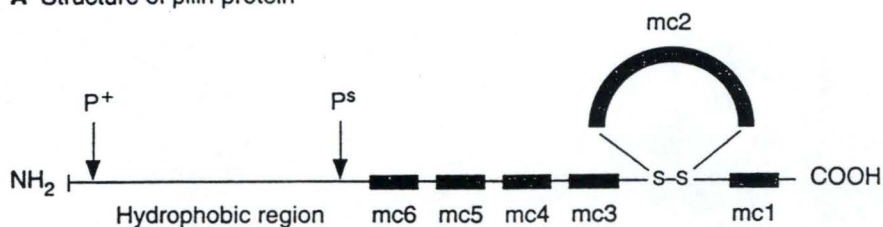
- se déplacer vers des conditions favorables et fuir les conditions nuisibles,
- produire très rapidement la machinerie nécessaire à l'utilisation d'un métabolite opportun,





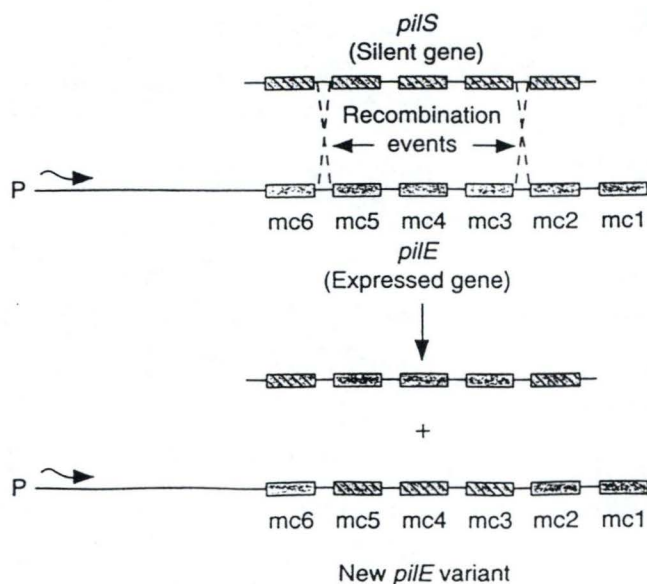
**Figure 1:** Amplification génique due à la recombinaison homologue entre deux copies d'un chromosome répliqué (Salyers et Whitt, 1994).

**A Structure of pilin protein**



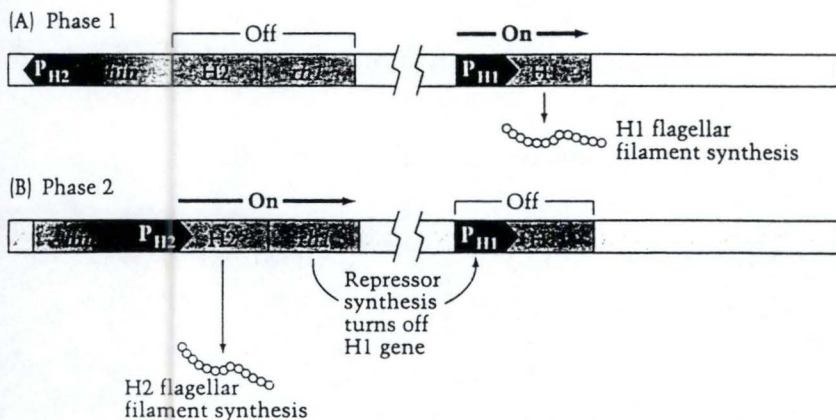
—, Segment highly conserved in different copies of the gene; ■, Variable segment

**B Antigenic variation in PiliE**



**Figure 2:**

- A, Structure d'une protéine de piline chez *Neisseria gonorrhoeae*;
- B, Variation antigénique au niveau des cassettes encodant les différents segments d'une protéine de piline générant ainsi un nouveau type protéique (Salyers et Whitt, 1994).



**Figure 3:** Variation de phase permettant le passage d'un type de flagelle à l'autre chez *Salmonella typhimurium*. (Neidhardt et al., 1990).

- supporter des conditions stressantes exceptionnelles comme la présence d'antibiotiques, les radiations, ou encore les grandes variations de température,
- réguler l'expression de gènes encodant des facteurs de virulence suite à l'invasion de l'hôte, pour les bactéries pathogènes.

## B. Types de régulation:

Nous pouvons rassembler les mécanismes de régulation en trois catégories:

- amplification et réarrangements géniques,
- régulation transcriptionnelle,
- régulation post-transcriptionnelle.

### 1. Amplification et réarrangements géniques:

L'amplification génique est un processus par lequel une copie unique d'un gène est remplacée par des copies multiples de ce même gène. Les mécanismes d'amplification génique sont, par exemple, la recombinaison entre deux copies d'un chromosome répliqué (Fig.1) ou la transposition. Tous ces phénomènes conduisent à la multiplication du nombre de copies d'un gène dans le génome.

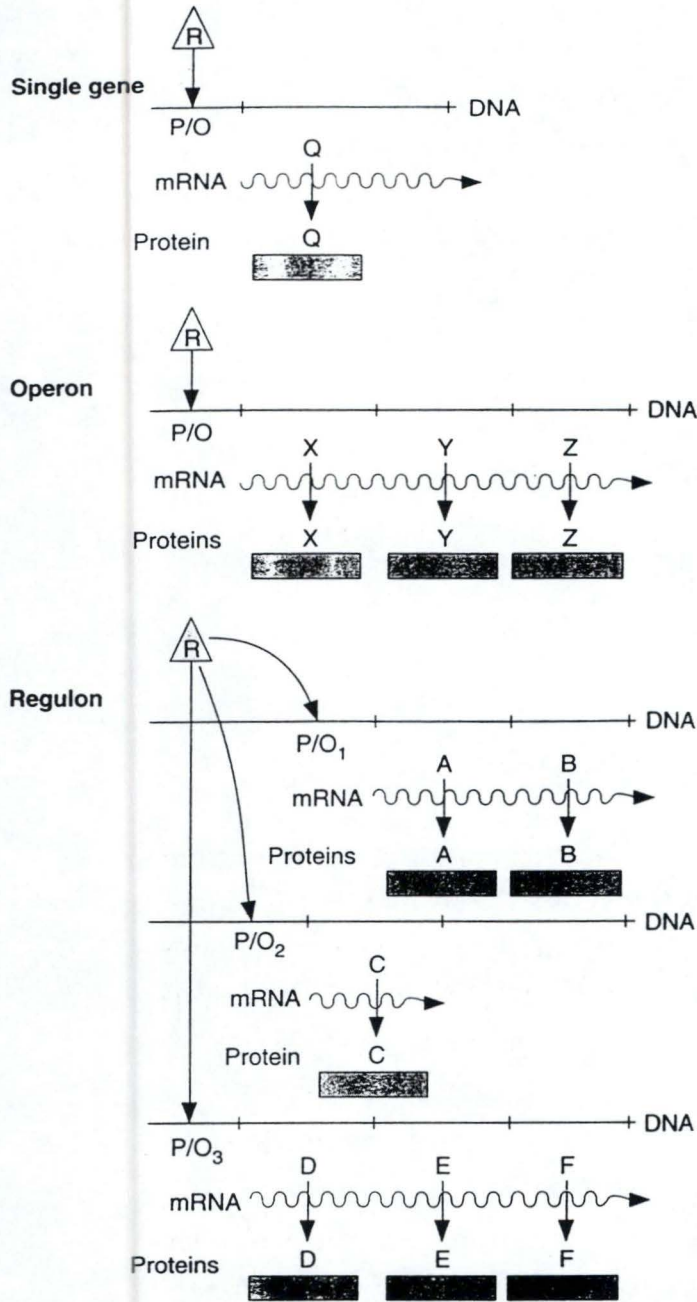
Les réarrangements géniques impliquent des changements dans la structure d'un gène et donc altèrent sa séquence ou son expression. Un exemple de réarrangement est la variation antigénique, impliquant une recombinaison homologue entre une version active d'un gène exprimée jusqu'ici et une version silencieuse de ce gène. Celui-ci encode généralement une protéine de surface reconnue par le système immunitaire de l'hôte. A chaque cycle cellulaire, une protéine de surface différente peut être exprimée suite à la recombinaison, ce qui permet à une bactérie d'échapper à la réponse immunitaire de l'hôte.

ex.: les gènes encodant la piline de *Neisseria gonorrhoeae* , (cfr. fig.2)

Un autre exemple est la variation de phase ("phase variation"): la présence d'éléments d'insertion permet le passage d'un phénotype à un autre.

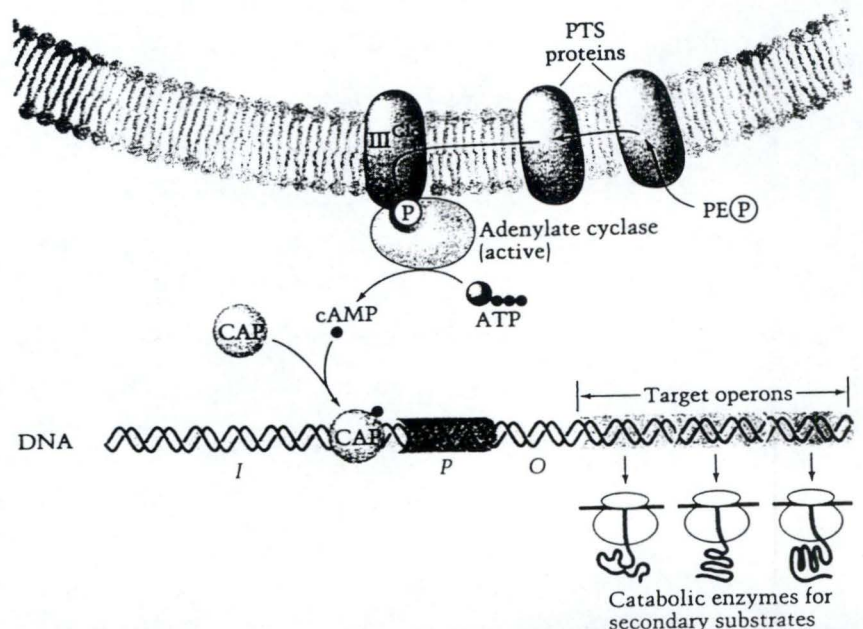
ex: le changement de type de flagelles chez *Salmonella typhimurium* (Fig.3):

Les 2 gènes encodant les flagellines, H1 et H2, sont situés à différents endroits du chromosome. Le promoteur PH2, pour le gène H2, est à l'intérieur



**Figure 4:** Organisation des gènes d'adaptation en opéron et régulon; R est le régulateur et P/O correspond aux promoteur et opérateur (Salysers et Whitt, 1994).

**Figure 5:** Activation de la transcription des gènes impliqués dans le catabolisme de sucres autres que le glucose par le complexe CAP:AMPc (Neidhardt et al., 1990).



d'un élément d'insertion, qui est entouré de deux séquences à répétition inverse et qui contient le gène *hin*, encodant une DNA invertase. L'opéron H2 contient aussi un gène, *rhl*, encodant un répresseur agissant sur le promoteur PH<sub>1</sub> du gène de la flagelline H1. En phase 1, le seul gène actif est celui de la flagelline H1, car PH<sub>2</sub> n'est pas orienté vers le gène H2. En phase 2, PH<sub>2</sub> est orienté vers H2, conduisant à la synthèse de H2 et du répresseur Rh1, qui éteint l'expression du gène H1.

## 2. Régulation transcriptionnelle:

Les gènes d'adaptation peuvent être organisés en (Salyers et Whitt, 1994):

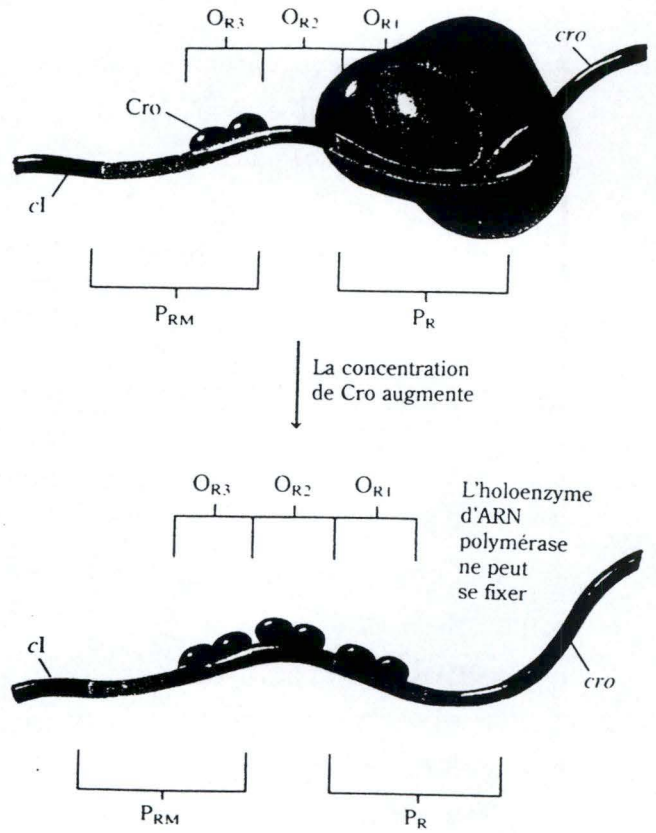
- **opéron**, ensemble de gènes transcrits simultanément à partir d'un seul promoteur;
- **régulon**, ensemble de gènes localisés à différents endroits dans le génome, mais dont les promoteurs répondent à une même protéine régulatrice (fig.4);
- **stimulon**, ensemble de gènes régulés par une même protéine mais pas de la même manière.

Les activateurs transcriptionnels sont des protéines qui se lient à une région opératrice d'un gène, qui interagissent avec l'ARN-polymérase directement ou indirectement, et qui stimulent la transcription de ce gène, soit en favorisant la liaison de la polymérase au promoteur, soit en augmentant le taux de transcription dans certaines régions du génome (formation de complexes ouverts,...).

Ex: CRP (cAMP receptor protein, ou CAP, catabolite activator protein), qui évalue la concentration en glucose dans le milieu extracellulaire. A basse concentration en glucose, la concentration en AMP<sub>c</sub> s'élève. CRP fixe l'AMP<sub>c</sub> et le complexe CRP: AMP<sub>c</sub> stimule la transcription des gènes impliqués dans le métabolisme des sucres autres que le glucose (fig.5).

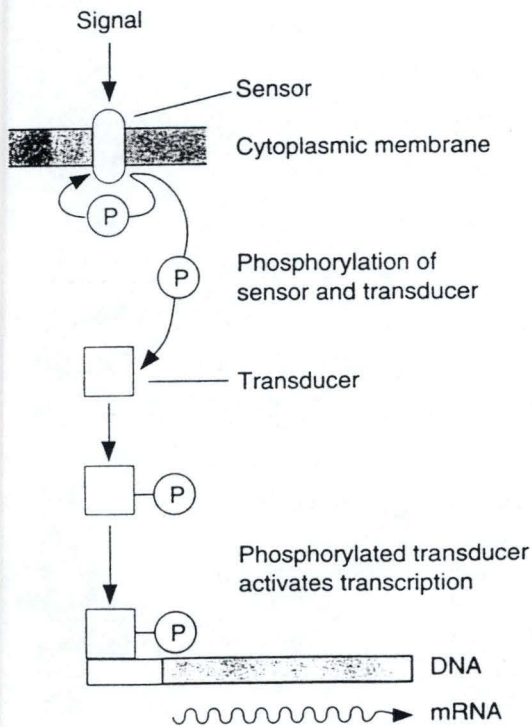
Les répresseurs transcriptionnels se lient aussi à des régions opératrices et soit empêchent la liaison de l'ARN-polymérase quand la région opératrice se superpose au promoteur, soit empêchent la polymérase d'initier la transcription, quand la région opératrice se situe en aval du promoteur.

Ex: Cro (contrôle du répresseur et d'autres fonctions), qui réprime la transcription des gènes précoces immédiats et des gènes précoces du bactériophage  $\lambda$  (Fig. 6).



**Figure 6:** Influence de Cro sur l'opérateur  $O_R$ .

Quand Cro est présent il s'attache à  $O_{R3}$ , le site opérateur envers lequel il présente l'affinité la plus grande. Sa liaison à  $O_{R3}$  prévient la transcription du gène  $cI$  au départ de  $P_{RM}$ . Dès qu'une concentration élevée de Cro est atteinte, Cro s'attache d'une manière non-coopérative à  $O_{R1}$  et  $O_{R2}$ . Cette liaison empêche la transcription du gène Cro au départ de  $P_R$ . Il s'ensuit que la production de Cro est aussi auto-ajustée (Rawl, 1990)



**Figure 7:** Représentation schématique d'un système simple de transduction du signal à deux composants (Salyers et Whitt, 1994)

Le mécanisme le plus largement répandu chez les bactéries est probablement le "**Système de transduction du signal à deux composants**" (Stock *et al.*, 1989). Comme le nom l'indique, ce système implique deux protéines (fig. 7):

- une protéine histidine-kinase ou senseur, qui perçoit les variations environnementales telles l'osmolarité, la disponibilité en glucose, azote, phosphate,... Elle est capable de s'autophosphoryler au niveau d'une histidine et de transmettre son phosphate à la deuxième composante, le régulateur.

- un régulateur de réponse, qui suite à l'activation par le senseur, est capable d'effectuer une réponse adaptative, la plupart du temps une régulation de la transcription.

Ce système sera détaillé par la suite.

### 3.Régulation post-transcriptionnelle:

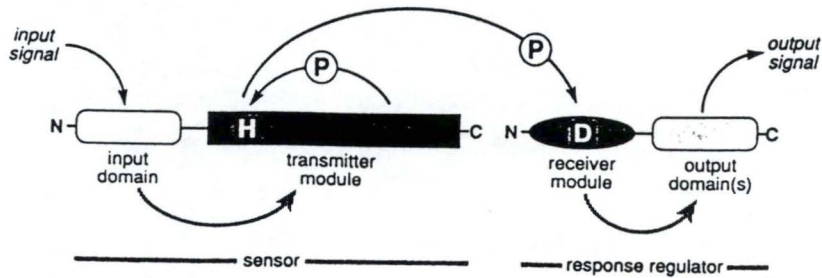
Celle-ci est importante, car elle contrôle le taux de facteurs d'adaptation **actifs**.

Les protéines provenant d'un même transcrit n'ont pas nécessairement le même taux de traduction. Celui-ci peut être modulé par l'affinité des ribosomes pour leur RBS (site de liaison au ribosome) ou encore par le taux de recyclage des ARN messagers.

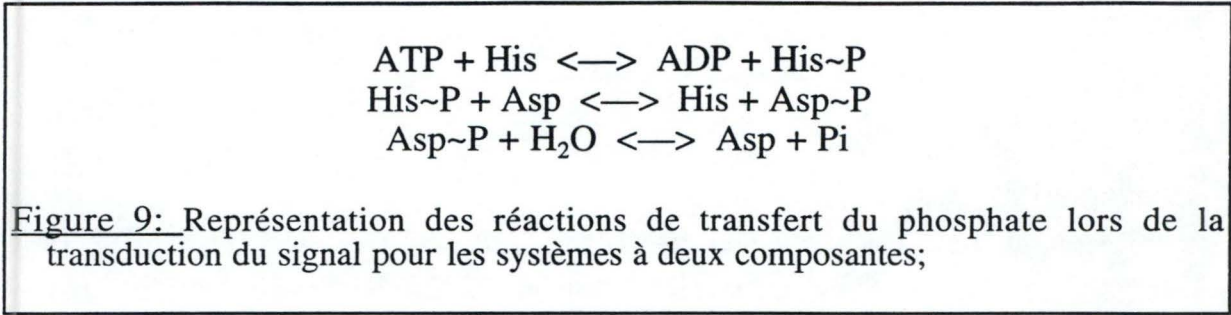
Souvent, le produit de l'ARN messager n'est pas une protéine active et il doit subir une maturation supplémentaire:

- un clivage protéolytique,
- une modification covalente (méthylation, phosphorylation,...),
- une liaison non-covalente (activation allostérique, ...)

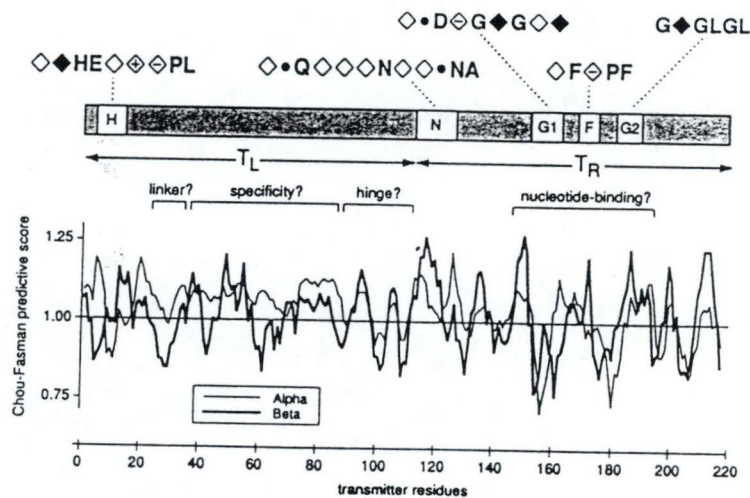
Pour conclure, il n'est pas rare de voir plusieurs de ces stratégies contrôler un ensemble de gènes d'adaptation. En effet, une combinaison de ces mécanismes permet un meilleur contrôle de la multitude des facteurs produits et une réponse plus fine face à une combinaison complexe de signaux.



**Figure 8:** Paradigme des modules de communication des systèmes à deux composantes. L'information se transmet via un contrôle exercé sur un domaine par un autre (flèches hachurées), et par des réactions de phosphorylation (flèches désignées par un P) impliquant des résidus histidine et aspartate. (Parkinson et Kofoid, 1992)



**Figure 9:** Représentation des réactions de transfert du phosphate lors de la transduction du signal pour les systèmes à deux composantes;



**Figure 10:** Caractéristiques structurales des domaines transmetteurs. Les lettres indiquent les acides aminés présents à cette position pour au moins 70% des séquences alignées des transmetteurs. Les losanges indiquent les positions pour lesquelles au moins 50% de acides aminés appartiennent à la même famille chimique: blancs, nonpolaires (I, L, M, V); noirs, polaires (A, G, P, S, T); +, basiques (H, K, R); -, acides (D, E, N, Q) (Parkinson et Kofoid, 1992).

## II. Systèmes de transduction du signal à deux composantes:

### A. Caractéristiques générales:

Le nom "Système à deux composantes" dérive de l'observation initiale que plusieurs systèmes bactériens de transduction du signal (aussi bien chez les organismes Gram-positifs que Gram-négatifs) partagent deux motifs conservés (Nixon *et al.*, 1986). Ces motifs sont appelés "domaine transmetteur" et "domaine receveur". Le domaine transmetteur en combinaison avec une variété de domaines de réception du stimulus forme le senseur et le domaine receveur en combinaison avec une variété de domaines effecteurs de la réponse forme le régulateur. Tous ces domaines peuvent être arrangés en différentes configurations afin de former de multiples circuits de transduction du signal (Fig.8).

Les systèmes à deux composantes sont constitués d'ATP-ases transmettant un signal en utilisant l'énergie libérée lors de l'hydrolyse de l'ATP pour répondre à une variation environnementale (Stock *et al.*, 1989). C'est, jusqu'à présent, le seul mécanisme de communication démontré dans ces systèmes (Parkinson, 1993).

Trois réactions de transfert impliquant deux intermédiaires phosphorylés ont été observées (fig.9):

1° Le groupement  $\gamma$ -phosphoryl de l'ATP est transféré sur un résidu histidine du senseur;

2° le groupement phosphoryl est ensuite transféré de l'histidine du senseur à un aspartate du régulateur,

3° enfin, le groupement phosphoryl est libéré dans le milieu grâce à une activité phosphatase associée.

Toutes ces réactions sont réversibles, mais le transfert du groupement phosphoryl est unidirectionnel.

Il y a donc deux familles de domaines protéiques essentiels fonctionnant par paires:

- les senseurs;
- les régulateurs de réponse.



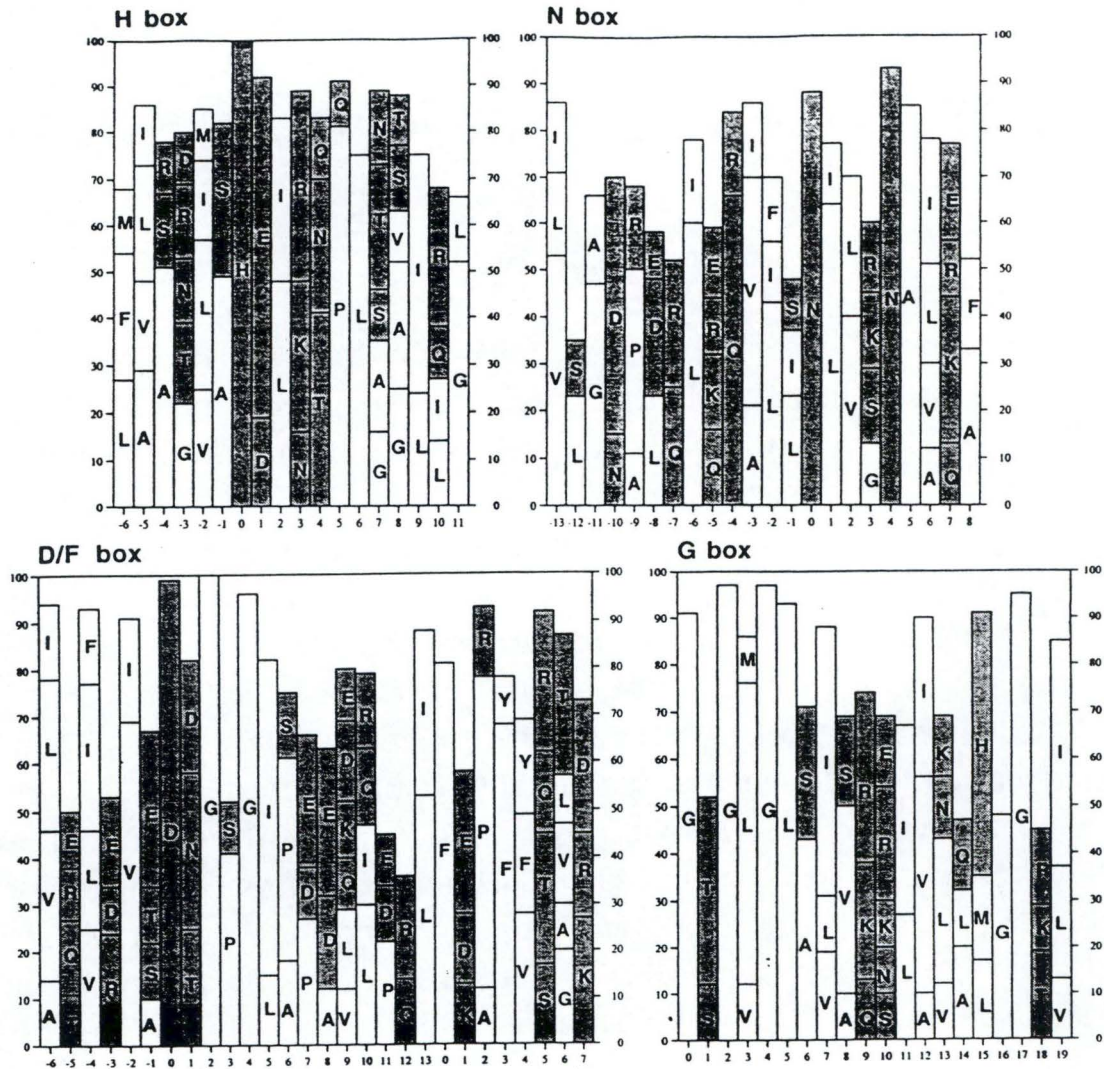


Figure 11: Histogramme des préférences en acides aminés pour les boîtes conservées parmi les domaines transmetteurs. Les fréquences en acides aminés proviennent d'une analyse de 68 domaines transmetteurs à partir des banques de séquences GenEMBL et SwissProt. Les acides aminés polaires sont en gris et les non-polaires sont en blanc (Stock *et al.*, 1995)

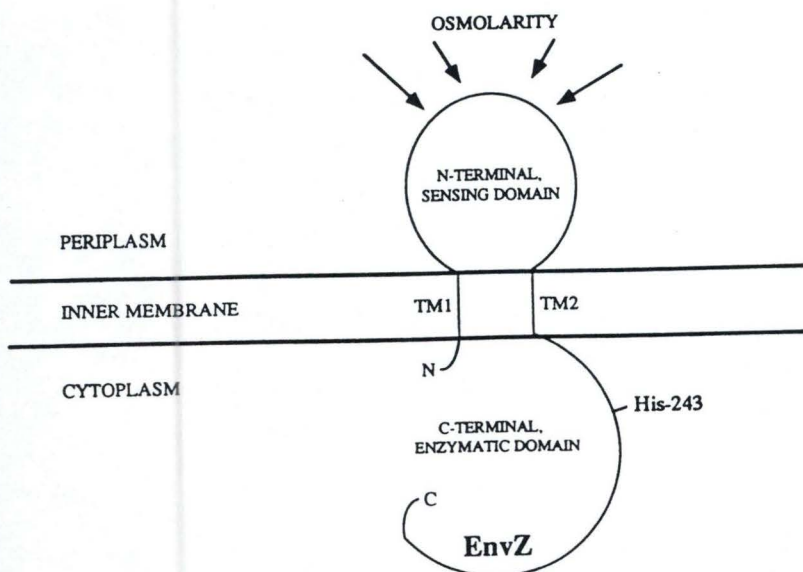


Figure 12: Représentation de l'osmosenseur EnvZ, localisé en membrane interne. Il est constitué d'un domaine de perception du signal se situant dans l'espace périplasmique, de deux segments transmembranaires (TM1 et TM2) et d'un domaine transmetteur cytoplasmique. La position relative de l'histidine catalytique est indiquée (Pratt et Silhavy, 1995)

## 1. Les senseurs:

La caractéristique de cette superfamille est la conservation d'une séquence carboxy-terminale s'étendant sur environ 240 résidus: le domaine transmetteur, capable de lier l'ATP et de catalyser son autophosphorylation au niveau de la chaîne latérale d'une histidine. Cette dernière propriété a valu aux senseurs le nom "d'Histidine-kinase".

On ne connaît pas encore grand-chose de la structure secondaire ou tertiaire des senseurs. Cependant, jusqu'à présent, presque 100 senseurs ont été séquencés, et on retrouve en moyenne 25% de similarité entre les domaines histidine-kinases, comparés deux à deux.

A partir de l'analyse de la séquence, Parkinson et Kofoid (1992) concluent que le domaine histidine-kinase conservé est divisé en deux sous-domaines ( $T_L$  et  $T_R$ ; fig. 10 et 11):

- le premier,  $T_L$ , contient la boîte H conservée qui correspond au site actif où l'histidine est phosphorylée.

Ce sous-domaine est assez variable et pourrait contenir les déterminants impliqués dans la reconnaissance spécifique des régulateurs cibles.

- le deuxième,  $T_R$ , contient les boîtes conservées N, G1, G2 et F.

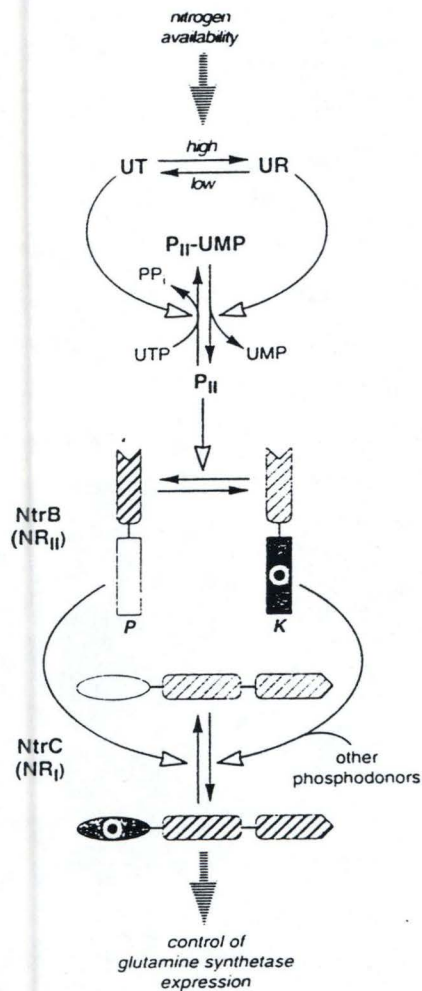
G1 et G2 sont des boîtes riches en glycine et seraient impliquées dans la liaison à l'ATP; la fonction des deux autres boîtes est encore inconnue.

Les sous-domaines  $T_L$  et  $T_R$  auraient des configurations tridimensionnelles distinctes.

La majorité des protéines contenant un domaine transmetteur sont localisées en membrane cytoplasmique, celui-ci se trouvant dans le cytoplasme. La plupart ont aussi deux segments transmembranaires, parfois plus, flanquant le domaine de réception du stimulus qui est déployé de l'autre côté de la membrane cytoplasmique.(ex: EnvZ dont le domaine de réception du stimulus se trouve dans l'espace périplasmique chez *E. coli*, fig.12).

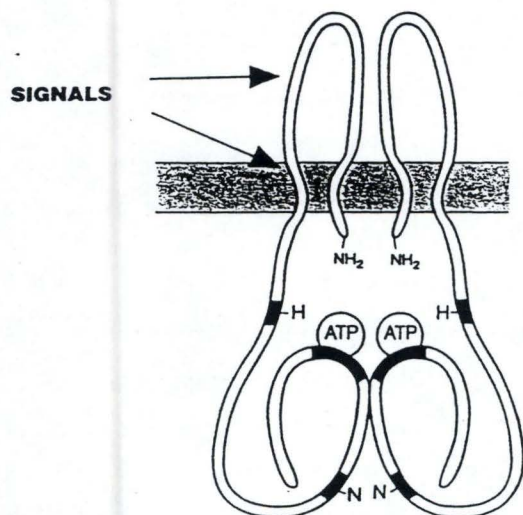
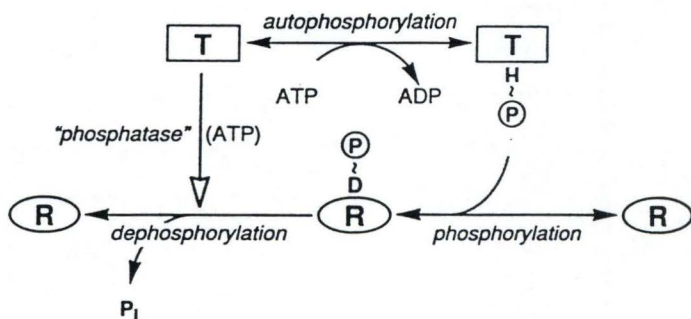
Les domaines de réception du stimulus diffèrent beaucoup au niveau de leur structure primaire, ce qui reflète la variété des stimuli chimiques et physiques qu'ils détectent. Pour certains, on a démontré une fonction de liaison à un ligand, mais les autres sont peu caractérisés en raison de l'ignorance de la nature exacte des stimuli qu'ils perçoivent.

Le signal se propage d'un domaine à l'autre, à travers la membrane cytoplasmique, probablement via un changement conformationnel induit par le stimulus.



**Figure 13:** Voie de régulation de l'expression de la glutamine synthetase en réponse à une variation de la disponibilité en azote, par le couple NtrB-NtrC (Parkinson, 1993). K et P représentent respectivement les activités autokinase et phosphatase de NtrB.

**Figure 14:** Diagramme représentant les 2 voies de modification du régulateur par le senseur (Parkinson, 1993)



**Figure 15:** Représentation d'un dimère formé suite à l'activité d'autophosphorylation des domaines transmetteurs (Chang et Meyerowitz, 1994).

Quelques senseurs sont solubles et cytoplasmiques. Leur domaine amino-terminal pourrait néanmoins avoir ce même rôle de perception du signal.

Ex: NtrB(*E. coli* et *Salmonella typhimurium*) qui répond à une variation de disponibilité en azote à l'extérieur de la cellule, en percevant une modification covalente de P<sub>II</sub> médiée par une variation du taux de deux enzymes, UT et UR (urydyltransferase et urydyl removing) suite à la variation de la quantité en azote (Fig.13).

Le transmetteur régule l'état de phosphorylation de son partenaire régulateur de deux manières (Fig.14):

- en formant par une activité autokinase (Ninfa et Bennet, 1991) un intermédiaire riche en énergie pour un transfert ultérieur sur l'aspartate du régulateur (Sanders *et al.*, 1989; 1992).

Cependant, la phosphorylation n'est pas un mécanisme intramoléculaire. Dans tous les cas étudiés, la formation d'un dimère dont les deux monomères se phosphorylent mutuellement est requise, fig.15 (Yang et Inouye 1991; Ninfa *et al.*, 1993; Swanson *et al.*, 1993)

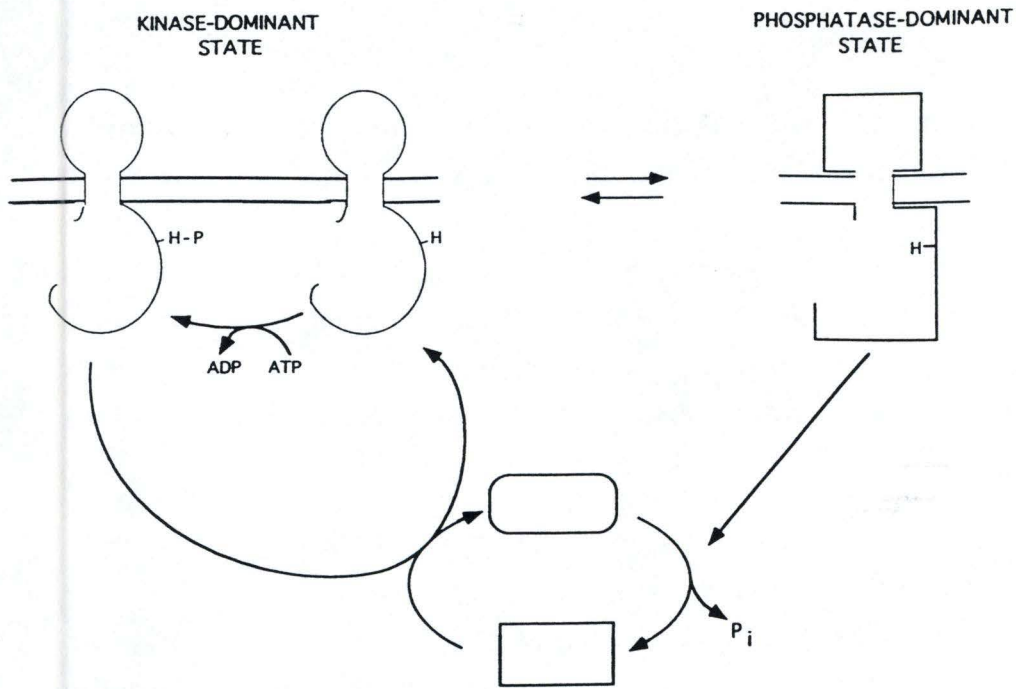
- en induisant directement ou indirectement la déphosphorylation du régulateur (Aiba *et al.*, 1989; Igo *et al.*, 1989; Ninfa et Magasanik, 1986). Le stimulus modulerait l'équilibre entre ces deux activités pour réguler le flux de groupement phosphoryl disponible pour le régulateur (Fig.16).

## 2. Les régulateurs:

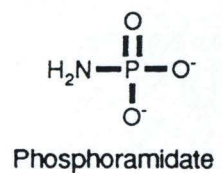
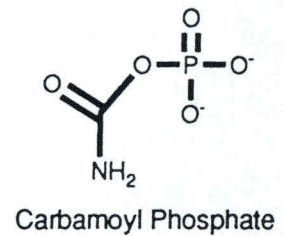
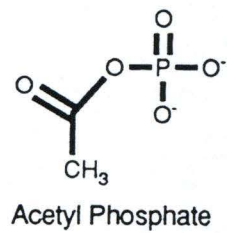
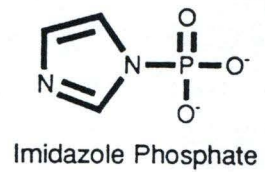
Les régulateurs sont caractérisés par un domaine amino-terminal conservé d'environ 125 résidus (Stock *et al.*, 1990; Volz, 1993) recevant le groupement phosphoryl du senseur. Ce domaine est habituellement attaché à un domaine effecteur via un lien flexible.

Contrairement aux senseurs, leur structure 3D est bien connue au moins au niveau du domaine amino-terminal. En effet, la structure de CheY (fig.20), un régulateur composé seulement de ce domaine receveur et intervenant dans le chémotactisme, a été déterminée par diffraction aux rayons X chez *Salmonella typhimurium* (Stock *et al.*, 1989) et chez *E.coli* (Volz *et al.*, 1990), et par RMN chez *E.coli* (Bruix *et al.*, 1993; Moy *et al.*, 1994).

CheY est une structure  $\alpha/\beta$ , composée de 5 brins  $\beta$  arrangés en feuillet parallèle, entourés par 5 hélices  $\alpha$ . Les brins  $\beta$  s'alignent pour former un noyau interne hydrophobe, autour duquel s'enroulent les hélices  $\alpha$ . Les



**Figure 16:** Représentation de l'équilibre entre les activités d'autophosphorylation et d'induction de la déphosphorylation du régulateur pour les senseurs (Pratt et Silhavy, 1995)



**Figure 17:** Représentation plane de petits donneurs de phosphate pouvant céder leur groupement phosphoryl aux régulateurs à la place des senseurs (Stock *et al.*, 1995).

résidus clés pour la phosphorylation sont localisés dans une poche acide à l'extrémité carboxy-terminale du feuillet  $\beta$  (fig.20).

Au niveau de cette poche (fig. 21), on retrouve:

- une paire d'aspartates proches de l'extrémité amino-terminale,
- une lysine proche de l'extrémité carboxy-terminale,
- et un aspartate central, site de phosphorylation.

Ces caractéristiques ont été retrouvées dans plus de 90 protéines contenant un domaine receveur, suggérant que tous les modules receveurs pourraient être structurés comme CheY (Stock *et al.*, 1989).

La conservation fonctionnelle et structurale des régulateurs est détaillée au point B.

Les régulateurs sont toujours cytoplasmiques.

Les domaines receveurs et effecteurs sont souvent reliés par un lien flexible, le "Q linker" (Wooton et Drummond, 1989), suggérant que des connections souples peuvent être importantes pour que le receveur soit capable de contrôler le domaine effecteur associé.

Le domaine effecteur est, dans la plupart des cas, un domaine de liaison à l'ADN, régulant la transcription d'un ou de plusieurs gènes. Dans d'autres cas le domaine effecteur a une autre fonction comme, par exemple le domaine effecteur de CheB qui a une activité estérase, déméthylant les chémorécepteurs membranaires.

Parfois, le régulateur n'est composé que du seul domaine receveur: CheY ou SpoOF (intervenant dans l'initiation de la sporulation de *Bacillus subtilis* ).

Enfin, il existe des senseurs, appelés kinase-hybrides, possédant un domaine receveur localisé à l'extrémité carboxy-terminale du domaine transmetteur.

Ex: VirA (*Agrobacterium tumefaciens*), une kinase-hybride permettant de percevoir les composants phénoliques exsudés par les plantes et de produire une réponse d'invasion par le transfert de leur plasmide Ti (tumor- inducing plasmid) ou ArcB (*E.coli* ), une kinase-hybride contrôlant l'expression de gènes en conditions de faible disponibilité en oxygène (aerobic respiration control).

Le transfert du phosphate depuis l'histidine du senseur jusqu'à l'aspartate du régulateur est catalysé par le régulateur lui-même. En effet, de petits donneurs de phosphate, comme l'acétylphosphate ou l'imidazolphosphate (fig. 17), peuvent agir à la place de l'histidine kinase phosphorylée pour donner le groupement phosphoryl à l'aspartate du régulateur (Lukat *et al.*, 1992).

Les régulateurs pourraient aussi catalyser leur hydrolyse, avec des temps de demi-vie allant de quelques secondes (ex. phospho-CheY; Hess *et al.*, 1988) à plusieurs heures (ex. phospho-OmpR; Igo *et al.*, 1989). Cette capacité de déphosphorylation intrinsèque au régulateur pourrait être activée par un effet allostérique médié par le transmetteur, qui n'aurait alors aucune activité phosphatase propre.

### 3.Mécanismes de communication:

Bien que les attributs des transmetteurs et des receveurs soient maintenant à peu près connus, on ne comprend pas encore leur activité au niveau moléculaire.

Il y a 4 opérations de communication distinctes (Fig.18):

- a. détection du stimulus par le domaine senseur;
- b. contrôle du senseur sur l'activité du transmetteur;
- c. communication transmetteur-receveur;
- d. contrôle du receveur sur l'activité de l'effecteur.

Aucun de ces événements n'est encore bien compris (Parkinson, 1995).

L'étape fonctionnelle critique dans la transduction du signal est l'interaction entre le senseur et le régulateur de réponse (Parkinson et Kofoid, 1992).

Les senseurs modulent le taux de phosphorylation des régulateurs de deux manières:

- l'activité autokinase contrôle la disponibilité en phosphohistidine, qui est le meilleur phosphodonneur pour les régulateurs (c'est un fait inhérent aux protéines kinases, indépendant des régulateurs);
- le deuxième mécanisme implique une association réversible entre le régulateur non phosphorylé et le transmetteur activé, via des interactions spécifiques, suivie du transfert du phosphate catalysé par le régulateur. La phosphorylation du régulateur a pour conséquence la production d'une réponse adaptative interne qui se maintient jusqu'à ce que le régulateur perde son phosphate; ce qui peut se faire soit spontanément (par une activité autophosphatase intrinsèque), soit grâce à un effet allostérique médié par le transmetteur.

La précision globale de la transduction du signal est déterminée par l'interaction d'"arrimage" entre le transmetteur et le régulateur. Dans la majorité des cas, l'affinité entre transmetteur et receveur appariés est considérablement plus élevée que dans le cas de paires hétérologues

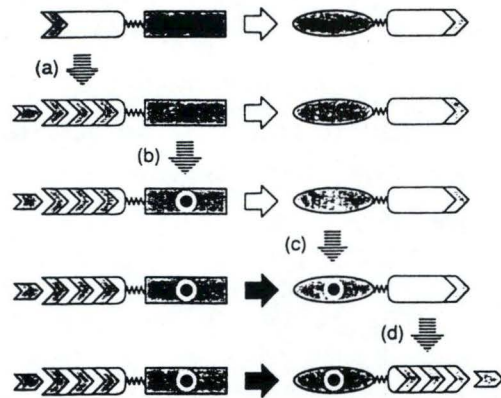


Figure 18: Représentation des quatre opérations de communication entre les différents modules des systèmes à deux composantes; a-d. voir texte (Parkinson, 1995)

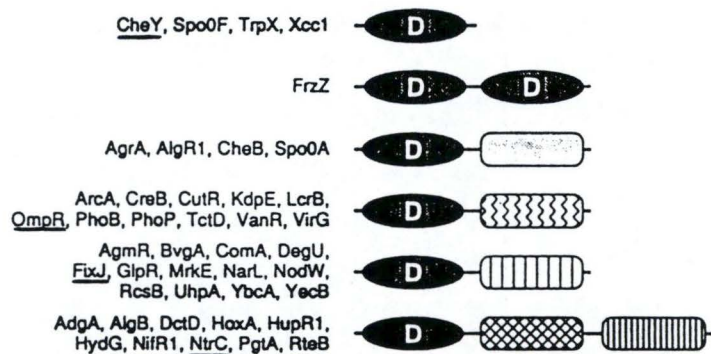


Figure 19: Représentation de la classification des régulateurs en 5 familles en fonction de la similarité de leurs domaines supplémentaires. Les protéines principales composant ces familles sont indiquées (Parkinson et Kofoid, 1992)



(crosstalk ou petits donneurs de phosphate); ce qui autorise un transfert du groupement phosphate au moins cent fois plus rapide.

Si le régulateur distingue son senseur attitré des autres donneurs de phosphate (Lukat *et al.*, 1992), on peut penser que la structure locale du transmetteur au niveau de l'histidine contient l'information nécessaire pour la reconnaissance du substrat. Cependant, il se peut aussi que les contacts spécifiques se trouvent hors des sites catalytiques. En effet, on suppose que la région variable entourant le bloc H chez les senseurs est impliquée dans la reconnaissance du régulateur. Par contre aucun motif caractéristique n'a été identifié sur les régulateurs.

Si un mécanisme de reconnaissance spécifique existe, il devrait impliquer des contacts stéréospécifiques entre des surfaces exposées des régulateurs et des transmetteurs. Des candidats, pour les régulateurs, sont les hélices  $\alpha$  externes, pour lesquelles un changement conformationnel médié par la phosphorylation pourrait déplacer les sites de contacts afin de faciliter les étapes d'arrimage ou de décrochage (Parkinson et Kofoid, 1992).

## B. Conservation structurale et fonctionnelle des régulateurs.

### 1. Historique et liens de parenté entre les différentes familles.

La première indication de similarité parmi les régulateurs de réponse est apparue en 1985, quand il a été montré une parenté entre la séquence de CheY et les protéines régulatrices d'autres réponses cellulaires, comme la sporulation ou la synthèse de protéines membranaires (Stock *et al.*, 1985). Viennent ensuite la description de la nature à "deux composantes" de ces systèmes (Nixon *et al.*, 1986), la cristallisation de CheY (Volz *et al.*, 1986), puis la résolution de sa structure tridimensionnelle chez *S. typhimurium* (Stock *et al.*, 1989) et chez *E. coli* (Volz et Matsumura, 1991). Ce n'est que plus tard, en 1993, qu'une analyse compréhensible de la relation structure-fonction des régulateurs a vu le jour (Volz, 1993). Ce chapitre se base essentiellement sur cette analyse.

La plupart des régulateurs sont des protéines multidomaines, divergeant beaucoup au niveau de la longueur de leur séquence et au niveau de l'organisation de leurs domaines; mais tous sont homologues au niveau de leur domaine receveur amino-terminal.

Ils ont été classés en plusieurs familles (Volz, 1993) sur base de la similarité de leurs domaines supplémentaires (Fig.19):

- la famille de CheY regroupe les régulateurs composés seulement du domaine receveur, contrairement aux régulateurs des autres familles qui sont multidomains. On y retrouve CheY et SpoOF, et trois autres membres ayant dix à vingt acides aminés supplémentaires à leur extrémité carboxy-terminale, dont la fonction est inconnue;

- la famille NtrC regroupe les protéines homologues à NtrC. Elles ont environ 460 acides aminés et possèdent deux domaines après le domaine receveur. Le domaine central interagit avec le facteur  $\sigma^{54}$  de la RNA-polymérase pour former des complexes de transcription ouverts. Le dernier domaine de NtrC est homologue aux protéines FIS (Koch *et al.*, 1988; Johnson *et al.*, 1988) qui sont des protéines régulatrices H-T-H (Hélice-coude-hélice) dont la structure 3D est connue (Kostrewa *et al.*, 1991);

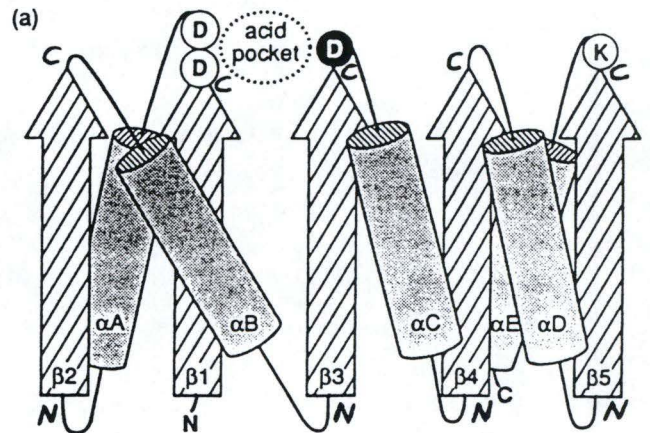
NtrC (Nitrogen regulatory protein ou NRI) est régulé par NtrB suite à une variation de la disponibilité en azote. NtrC active la transcription des gènes et opérons appartenant au régulon Ntr, impliqué dans l'assimilation de l'azote, par la polymérase contenant le facteur  $\sigma^{54}$ .

- la famille OmpR est composée de PhoP, PhoB, VirG et d'autres homologues d'OmpR, d'environ 230 acides aminés de long et ayant un domaine de liaison à l'ADN après le domaine receveur;

OmpR est une protéine cytoplasmique régulant la transcription des porines de membrane externe, OmpC et OmpF, en réponse à l'information reçue de son senseur EnvZ suite à une variation de l'osmolarité (Hall et Silhavy, 1981).

- enfin, la famille FixJ rassemble les protéines comprenant un domaine de type FixJ à leur extrémité carboxy-terminale et qui ont une longueur totale d'environ 220 acides aminés. Cette famille est assez mal connue, mais on leur suppose aussi une capacité de liaison à l'ADN, car FixJ ressemble à la région des facteurs  $\sigma$  impliqués dans la reconnaissance de promoteurs;

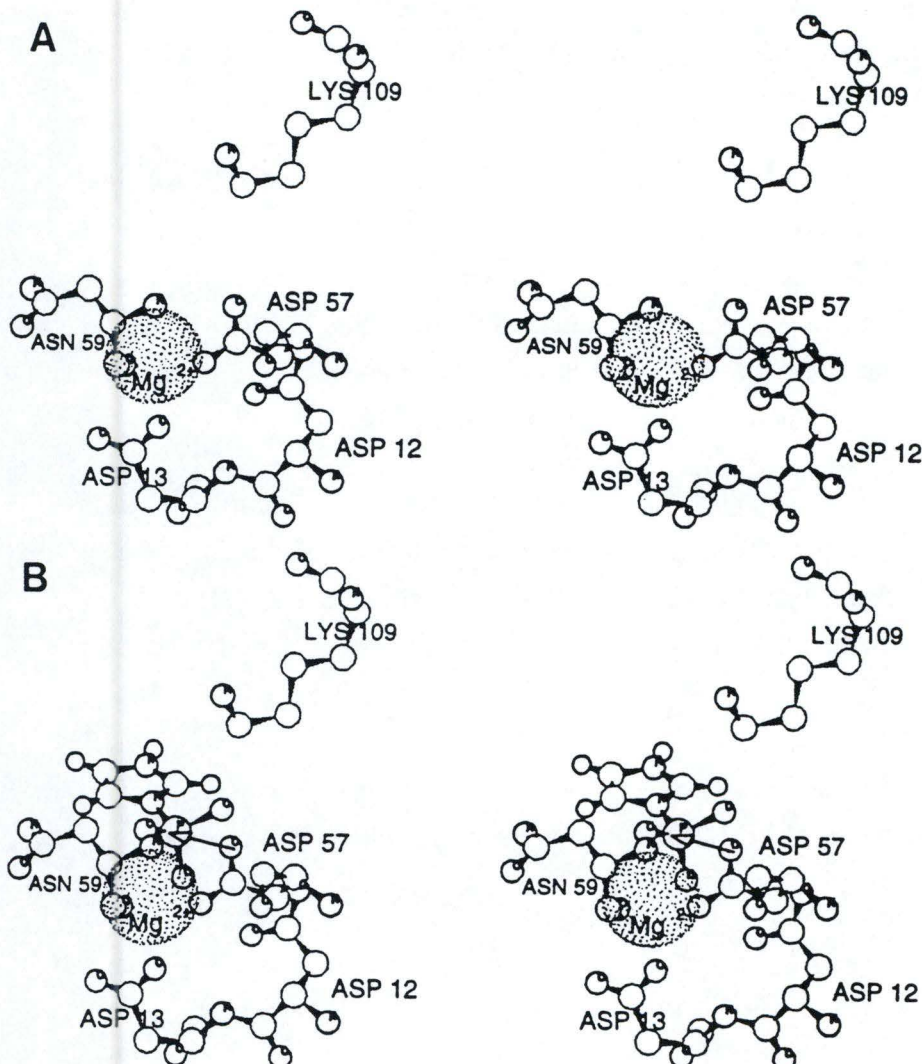
FixJ, sous le contrôle de FixL en réponse à l'anaérobiose chez *E.coli*, régule la transcription des gènes *nifA* et *fixK*, impliqués dans la fixation de l'azote (de Philip *et al.*, 1990).



**Figure 20:**

A, Structure tertiaire de CheY avec un ion  $Mg^{++}$  positionné dans le site actif (Stock et al., 1995)

B, Topologie de CheY; le cercle noir représente l'aspartate servant de site de phosphorylation, les cercles blancs représentent les autres résidus importants du site catalytique (Parkinson et Kofoid, 1992).



**Figure 21:**

A, Vue stereo du site actif de CheY liant montrant la poche acide composée de D11, D12 et de D57 liant le  $Mg^{++}$ . Les résidus N59 et K109 y sont aussi représentés;

B, Vue stereo du site actif de CheY comme montré en A, avec un groupement phosphoimidazole positionné pour donner son phosphate au D57 (Stock et al., 1995)

- on cite parfois une cinquième famille, "miscellaneous" (ou "divers"), qui regroupe les séquences dont les domaines supplémentaires ont des fonctions disparates.

Excepté les protéines les plus homologues (seules les séquences partageant moins de 60% d'identité ont été retenues), le pourcentage moyen d'identité est de 23% après alignement des domaines régulateurs; la valeur la plus faible étant de 6% entre CheY et VirA (Volz, 1993). Cela peut paraître inhabituel de considérer la superfamille de CheY comme conservée avec ce pourcentage moyen d'identité, mais on sait que les familles de protéines  $\beta/\alpha$  contiennent une large proportion de sites structurellement conservés qui restent reconnaissables même s'ils sont très distants d'un point de vue évolutif.

Il est généralement accepté que les domaines receveurs de tous les régulateurs partagent une structure commune. On pense aussi que la plupart des régulateurs utilisent le même mécanisme biochimique d'activation. Il est donc plausible de considérer qu'il y a un lien de parenté spécialement conservé entre les séquences de cette superfamille.

## 2. Structure moléculaire de CheY:

CheY est le prototype des régulateurs, car on possède un modèle 3D pour sa forme apo (forme ne liant pas le  $Mg^{++}$ ), sa forme liant le  $Mg^{++}$  et pour plusieurs formes mutantes.

### a) *Fold* général et éléments de structure secondaire:

CheY est une petite protéine globulaire de 128 acides aminés, repliée en structure  $(\beta/\alpha)_5$ ; les brins  $\beta$  sont arrangés en feuillet parallèle avec la topologie  $\beta_1\beta_2\beta_3\beta_4\beta_5$ . Elle contient 40% d'hélices  $\alpha$ , 20% de brins  $\beta$ , 25% de coudes et les 15% restants constituent des boucles de tailles variées.

Les brins  $\beta_1$ ,  $\beta_3$  et  $\beta_4$  forment un noyau hydrophobe, les brins  $\beta_2$  et  $\beta_5$  étant accessibles au solvant. Les 5 hélices  $\alpha$  sont amphiphiles et sont rassemblées en deux groupes de part et d'autre du feuillet  $\beta$  central,  $\alpha_A$  et  $\alpha_E$  d'un côté et  $\alpha_B, \alpha_C, \alpha_D$  de l'autre (fig. 20).

Residue	Role
F8	Hydrophobic core sequence of internal $\beta_1$ strand
L9	Hydrophobic core sequence of internal $\beta_1$ strand
V10	Hydrophobic core sequence of internal $\beta_1$ strand
V11	Hydrophobic core sequence of internal $\beta_1$ strand
D12	Mg <sup>2+</sup> binding
D13	Mg <sup>2+</sup> binding and catalysis
M17	Hydrophobic cover of $\alpha_1$ - $\alpha_5$ -interface
R18	Hydrophobic cover of $\alpha_1$ - $\beta_1$ -interface
V21	Hydrophobic core
L25	Hydrophobic core
G29	$\alpha_4$ -conformation; C-cap of $\alpha_1$
F30	Hydrophobic cluster with F8, F53, and F124
V33	Sole hydrophobic residue of exposed $\beta_2$ -strand; buried
L46	Hydrophobic core
G52	Aspartate by consensus; start of $\beta_3$
F53	Hydrophobic cluster with F8, F30, and F124
V54	Hydrophobic core sequence of internal $\beta_3$ -strand
I55	Hydrophobic core sequence of internal $\beta_3$ -strand
S56	Hydrophobic core sequence of internal $\beta_3$ -strand
D57	Invariant; required for phosphorylation-induced activation
W58	Hydrophobic closure to $\alpha_3$ - $\beta_4$ - $\alpha_4$ -interface
M60	Hydrophobic plug for crevice between $\beta_1$ , $\alpha_2$ , $\beta_3$ , and $\alpha_3$
P61	Rigid group in $\gamma$ -turn loop
N62	Positive $\phi$ -conformation; required for $\gamma$ -turn
D64	N-cap of $\alpha_3$
G65	Internal start of $\alpha_3$ ; severe steric restrictions on $\beta$ -substituent
L66	Internal residue of $\alpha_3$
L68	Internal residue of $\alpha_3$ ; hydrophobic core
L69	Internal residue of $\alpha_3$ ; hydrophobic core
L72	Internal residue of $\alpha_3$ ; hydrophobic core
P82	Start of $\beta_4$
V83	Hydrophobic core sequence of internal $\beta_4$ -strand
L84	Hydrophobic core sequence of internal $\beta_4$ -strand
M85	Hydrophobic core sequence of internal $\beta_4$ -strand
V86	Hydrophobic core sequence of internal $\beta_4$ -strand
T87	Invariant hydroxyl; proton donor to active site
A88	Reduced bulk; required for access to D57 active site
I96	Conserved solvent accessible hydrophobic residue
G102	$\alpha_1$ -conformation; C-cap of $\alpha_4$
A103	Reciprocally covariant with position 106
Y106	Conserved aromatic; reciprocally covariant with 103
V107	Sole hydrophobic residue of exposed $\beta_5$ -strand; buried
K109	Invariant; required for activation
P110	<i>cis</i> peptide following K109; rigidity to $\beta_5$ - $\alpha_5$ -loop
F111	Outer shell of hydrophobic core
L116	Internal residue of $\alpha_5$ ; hydrophobic core
L120	Internal residue of $\alpha_5$ ; hydrophobic core
F124	Hydrophobic cluster with F8, F30, and F53

**Figure 22:** Rôle structural des résidus conservés de la protéine CheY dans la superfamille des régulateurs (Volz, 1995).

### b) Structure du site actif:

La région de phosphorylation de CheY se situe à l'extrémité carboxy-terminale du feuillet  $\beta$ . Les résidus hautement conservés D12, D13, D57, T87 et K109 de CheY, sont configurés en une structure tridimensionnelle qui forme le site actif: la poche acide (fig. 21).

D'un côté du site actif, D12 et D13 sont groupés sur la même face du coude entre  $\beta_1$  et  $\alpha_A$ , et projettent leur chaîne latérale vers le site actif. Ils sont peu accessibles au solvant et interviennent dans la coordination du  $Mg^{++}$ , qui est essentiel pour les réactions de phosphorylation et de déphosphorylation (Stock *et al.*, 1993). L'acide aminé à la position 12 peut parfois être un glutamate plutôt qu'un aspartate. Quelques régulateurs ont aussi un résidu acide en position 14.

D57, porté par le brin central  $\beta_3$ , est le résidu acide phosphorylé dans le processus d'activation de CheY (Sanders *et al.*, 1989). Des mutants de divers régulateurs (CheY, CheB, NtrC), où cet acide aminé a été substitué, sont incapables d'être phosphorylés et n'ont aucune activité de transmission du signal.

Le brin adjacent,  $\beta_4$ , porte le résidu T87 dans une position telle que le groupe  $\gamma$ -hydroxyl participe au réseau de liaison hydrogène des molécules du solvant dans le site actif.

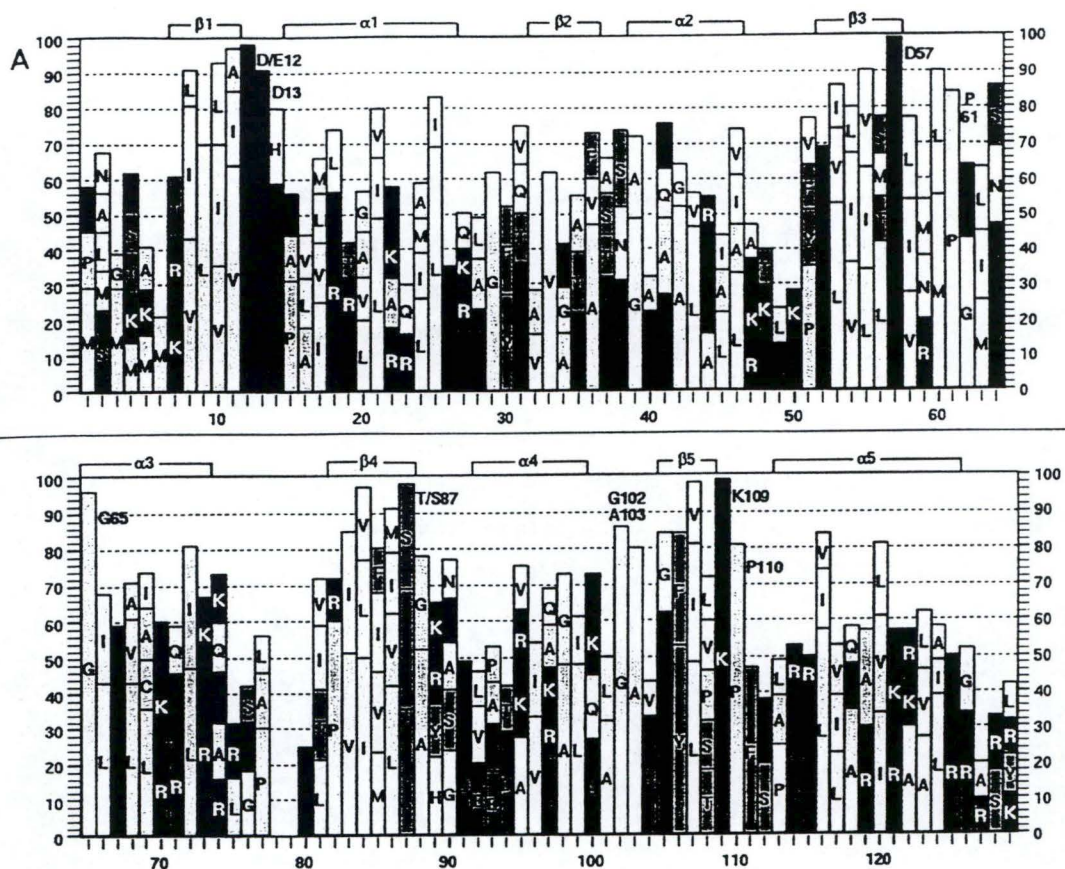
Enfin, K109, à l'extrémité du brin  $\beta$  le plus externe,  $\beta_5$ , a sa chaîne latérale en configuration étendue dans le site actif, formant une liaison hydrogène avec le groupement carboxyl de D57. La paire D57-K109 est enfouie dans le site actif, montrant une surface accessible négligeable. Une substitution de la lysine par une arginine résulte en une perte totale de fonction, bien que la protéine soit encore phosphorylable.

### 3. Résidus conservés:

Outre les résidus du site actif, une série d'autres résidus impliqués dans la formation des structures secondaires et tertiaires sont relativement bien conservés parmi les membres de la superfamille (fig. 22).

#### a) conservation du noyau hydrophobe et de la structure tertiaire.

La représentation d'un alignement multiple des séquences de 79 homologues de CheY est montrée dans l'histogramme de la figure 23.



**Figure 23:** Préférences en acides aminés pour le domaine receveur de la superfamille des régulateurs. Le pourcentage d'occurrence de chaque acide aminé est repris dans l'histogramme en suivant la numérotation de CheY. Cet histogramme est basé sur l'alignement des 79 séquences de régulateurs obtenu par Volz (1993). Les valeurs de moins de 10% ne sont pas représentées pour plus de clarté. Les éléments structuraux de CheY sont montrés au dessus de la figure.

La caractéristique la plus évidente des domaines "CheY-like" est la nature hautement hydrophobe des trois brins  $\beta$  centraux (résidus 7-11, 52-57 et 82-87). Le noyau hydrophobe inclut aussi huit sites hautement conservés des hélices  $\alpha$  avoisinantes (résidus 21 et 25 de  $\alpha_A$ , 46 de  $\alpha_B$ , 68, 69 et 72 de  $\alpha_C$  et 116 et 120 de  $\alpha_E$ ) et 26 résidus des éléments superficiels de la protéine provenant des hélices, des coudes s'interposant entre les brins et les hélices, et des brins  $\beta$  les plus externes. Trois résidus hydrophobes très conservés, les résidus 58, 96 et 99, sont étonnamment très accessibles au solvant dans la molécule CheY. Ces positions pourraient être impliquées dans des contacts interprotéines ou des contacts interdomaines dans le cas de régulateurs multidomaines.

Les cinq hélices diffèrent entre elles au niveau de leur degré de variabilité de séquence, sans doute pour des raisons structurales. Par exemple, l'hélice  $\alpha_C$ , qui est bordée par  $\alpha_B$  et par  $\alpha_D$ , est la plus conservée; alors que la plus grande variabilité est trouvée dans l'hélice  $\alpha_B$ , qui est l'hélice la plus exposée. La variabilité élevée de l'hélice  $\alpha_E$  est probablement due au manque d'exigences structurales pour le lien flexible des domaines supplémentaires. La conservation plus élevée de l'hélice  $\alpha_E$  à l'intérieur d'une même famille supporte cette hypothèse.

En général, les sections de grande variabilité au niveau des alignements de séquences primaires coïncident avec les régions de CheY les plus distantes du site actif, ce qui pourrait signifier que ces régions sont capables de supporter des changements structuraux sans occasionner de conséquences fonctionnelles nuisibles.

#### b) conservation de la structure secondaire.

Tous les autres résidus hautement conservés coïncident avec des caractéristiques structurales importantes de CheY.

Le meilleur exemple est la boucle " $\gamma$ -turn loop", adjacente à la région de phosphorylation de la molécule, qui est une structure compliquée, rigide et exposée au solvant.

Les *gamma-turns* sont définis par un pont hydrogène entre le groupement carbonyle d'un acide aminé et le groupement amine de l'acide aminé deux résidus en amont dans la séquence.

L'apparente "invariabilité" de cette boucle de surface est la conséquence d'une combinaison de cinq éléments structuraux:



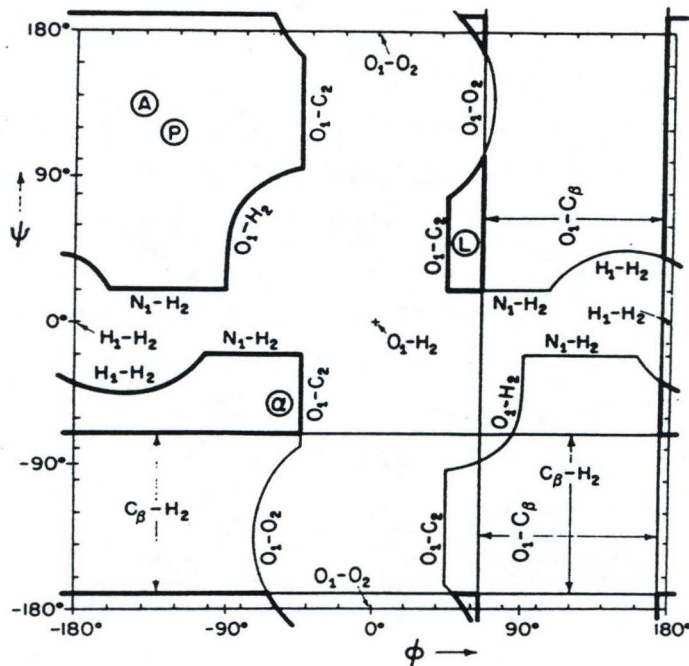


Figure 24: Diagramme illustrant les effets stériques produisant le graphe de Ramachandran. Les axes du graphe sont respectivement les angles dièdres  $\phi$  et  $\psi$ . Des traits gras limitent les zones tolérées et interdites. Les quatre régions tolérées correspondent aux angles dièdres des brins  $\beta$  parallèles (P) et antiparallèles (A), des hélices  $\alpha$  droites et des hélices  $3_10$  ( $\alpha$ ), et des hélices gauches (L).

- la présence exclusive d'un résidu hydrophobe enfoui à la position 60 (M+L+I= 98%).

- une proline conservée (84%) à la position 61;

- le centre de la boucle à la position 62 (G+D+E+N+K+R= 87%), imposant une valeur positive pour l'angle  $\phi$ ;

L'angle  $\phi$  est l'angle de rotation autour de la liaison  $C_{\alpha}$ --NH, par rapport à l'angle  $\psi$  qui est l'angle de rotation autour de la liaison  $C_{\alpha}$ --C= O. Ces angles sont restreints à une série de valeurs énergiquement plus favorables (cfr. graphe de Ramachandran, fig.24). Une valeur d'angle  $\phi$  positive est très rare (hélices  $\alpha$  gauches)

- le résidu de "capping" amino-terminal de l'hélice  $\alpha_C$  à la position 64;

Une hélice  $\alpha$  peut être considérée comme un macro-dipôle, car les ponts hydrogènes qui la maintiennent dans sa conformation pointent tous dans la même direction. Un résidu de "capping" est un résidu qui neutralise les charges partielles ainsi formées aux extrémités de l'hélice (Serrano et Fersht, 1989).

- et le premier résidu de l'hélice  $\alpha_C$ , une glycine conservée (96%) à la position 65.

Le degré élevé d'accessibilité au solvant et la proximité du site actif de phosphorylation suggèrent que cette conservation pourrait être importante pour la reconnaissance intermoléculaire. Cette boucle pourrait servir de "marque de fabrique" retenue par tous les régulateurs pour la reconnaissance par les kinases, alors que les résidus des sites variables proches contribueraient à la spécificité des différents systèmes.

La seconde caractéristique structurale peu commune est la présence d'une région conservée de type "VIb turn" incluant les résidus KPF (109-111). Ce coude se distingue par un lien peptidique *cis* ( $\omega=0$ ) entre K109 et P110.

$\omega$  est l'angle de torsion autour du lien peptidique. La nature du lien peptidique (liaison amide partiellement double, plane et rigide) implique que l'état le plus stable est la position *trans* ( $\omega=180^\circ$ ).

La lysine en position 109 est absolument conservée (99%), alors que le pourcentage d'identité pour la proline en position 110 est de 81%. La conservation de ce coude peut être indirectement reliée à sa fonction, qui peut être d'être un lien rigide entre  $\beta_5$  et  $\alpha_E$ . On peut poser l'hypothèse que si lors de la phosphorylation, les atomes du squelette de la lysine sont

repositionnés, l'effet peut se transmettre plus directement aux régions avoisinantes de la molécule par un lien rigide.

Il y a encore beaucoup de résidus hautement conservés, ayant un rôle particulier au niveau des structures secondaires de CheY; par exemple la présence de résidus impliqués dans le "capping" d'une hélice  $\alpha$  ou la présence de résidus chargés ou d'une proline à l'extrémité aminotermine d'un brin  $\beta$  interne.

#### 4. Contrôle de l'activité de l'effecteur par le domaine receveur.

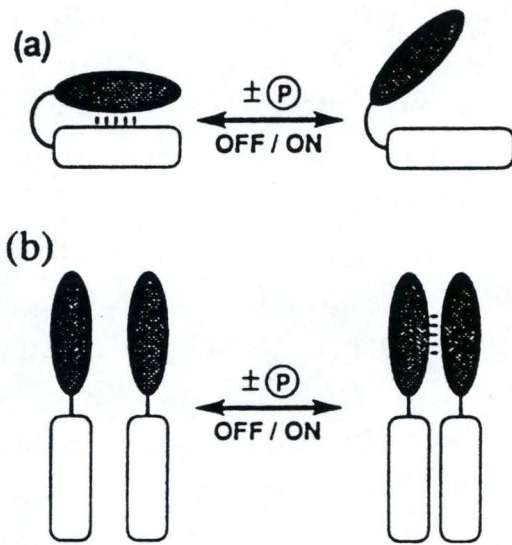
En général, la phosphorylation conduit à l'activation de la fonction de l'effecteur (bien que la forme non phosphorylée de certains régulateurs ait parfois un rôle fonctionnel): la phosphorylation d'OmpR augmente sa capacité à se lier à des séquences régulatrices spécifiques impliquées dans l'activation de la transcription (Forst *et al.*, 1989) et celle de CheB active son activité méthylestérase (Lupas et Stock, 1989).

Des mutations réduisant la phosphorylation bloquent habituellement la réponse adaptative induite par le stimulus. Ces mutations concernent des résidus de la poche acide (ex: CheY, Bourret *et al.*, 1990; 1993)

Cependant, la phosphorylation en elle-même n'est probablement pas requise pour l'activité de l'effecteur. Il y a plusieurs exemples de protéines régulatrices constitutivement actives qui sont très faiblement phosphorylées, ou même pas du tout (au moins *in vitro*).

Ex: CheY pour lequel l'aspartate 13 a été substitué par une lysine (Bourret *et al.*, 1990) engendre une réponse motrice alors qu'elle n'est pas phosphorylée. Cette réponse n'est pas altérée par la délétion de CheA (son senseur), ni par l'introduction d'une deuxième mutation, D57A ou D57E. Ces mutations sont d'un grand intérêt, car on a trouvé des régulateurs multidomaines ayant cette substitution naturellement, comme FrzG de *Myxococcus xanthus*, un homologue de CheB (McCleary *et al.*, 1990) et un régulateur transcriptionnel F1bD de *Caulobacter crescentus* appartenant à la famille de NtrC. Ces régulateurs sont capables d'activer la réponse adaptative en absence de phosphorylation détectable.

Puisque des mutations peuvent mimer l'effet activateur de la phosphorylation des régulateurs, la phosphorylation peut simplement être un mécanisme induisant des changements conformationnels. Un de ces mécanismes pourrait être le déplacement de la chaîne latérale de la lysine 109, changement conformationnel propagé au reste de la molécule par le lien rigide qui la suit.



**Figure 25:** Représentation des deux mécanismes généraux proposés pour le contrôle de l'activité du domaine effecteur par le domaine receveur. a. contrôle direct; b; contrôle indirect.

Bien qu'on ne connaisse pas en détail les mécanismes aboutissant au contrôle de l'activité de l'effecteur, il semble que les domaines receveurs aient développé des stratégies aussi bien positives que négatives pour contrôler leur domaine effecteur associé.

Dans certains cas, les receveurs inhibent l'activité de l'effecteur. En effet, il a été montré que des délétions génétiques ou des clivages protéolytiques du domaine amino-terminal activent le domaine carboxy-terminal. Ceci est valable pour l'activité estérase de CheB (Simms *et al.*, 1985), et pour l'activation transcriptionnelle par FixJ (Kahn et Ditta, 1991), PhoB et SpoOA (Grimsley *et al.*, 1994).

Ex: Le clivage de SpoOA par la trypsine entre les résidus 142 et 143 engendre la séparation des deux domaines. Le domaine amino-terminal retient la capacité de se faire phosphoryler. Le domaine carboxy-terminal est une protéine active liant l'ADN, retenant la spécificité de reconnaissance des boîtes OA des promoteurs *abrB* de la protéine intacte. Il stimule 5 fois plus la transcription *in vitro* à partir des promoteurs *spoIIG* que la protéine intacte non phosphorylée mais environ à moitié autant que SpoOA phosphorylé (Grimsley *et al.*, 1994).

A partir de ces résultats, il est vraisemblable que le domaine amino-terminal, sous sa forme non phosphorylée, inhibe l'activité du domaine effecteur et que la phosphorylation agisse pour relever cet effet inhibiteur.

Cependant, pour d'autres régulateurs, le domaine effecteur est sous contrôle positif du domaine receveur. Par exemple, le fait d'enlever le domaine amino-terminal de OmpR ne stimule pas la capacité de liaison à l'ADN de son domaine effecteur; ce qui signifierait que le domaine receveur est nécessaire pour activer la réponse adaptative (Tsung *et al.*, 1989).

Deux classes de mécanismes de contrôle peuvent donc être mis en oeuvre (fig. 25):

- le contrôle peut impliquer des interactions directes entre le receveur et l'effecteur associé (contrôle négatif), qui sont des contacts spécifiques modulés par l'état de phosphorylation. Ceci requiert que les receveurs coévoluent avec leur partenaires effecteurs.
- alternativement, le contrôle peut ne pas demander de contacts directs entre le receveur et le reste de la protéine. Un mécanisme possible (contrôle positif) pourrait être le changement de l'état de dimérisation (ou d'oligomérisation) du receveur induit par la phosphorylation (Kofoid et

Parkinson, 1988), ce qui altérerait la fonction de leurs effecteurs associés en les rapprochant les uns des autres. Un autre mécanisme (contrôle négatif) serait qu'un simple mouvement de l'hélice  $\alpha_E$  bascule le domaine carboxy-terminal dans une orientation différente par rapport au domaine receveur (Stock *et al.*, 1995), dans le cas où il y aurait interférence stérique entre les deux domaines empêchant l'effecteur d'interagir avec ses cibles moléculaires, ADN ou protéines.

espèce	biovar	morphologie des colonies	hôte préférentiel
<i>B. abortus</i>	1, 2, 3, 4, 5, 6, 9	S*	bovins 
<i>B. melitensis</i>	1, 2, 3	S	ovins-caprins 
<i>B. suis</i>	1	S	porcs 
	2	S	porcs-lièvres 
	3	S	porcs 
	4	S	rennes 
	5	S	rongeurs sauvages
<i>B. neotomae</i>	-	S	néotomes**
<i>B. ovis</i>	-	R	ovins 
<i>B. canis</i>	-	R	chiens 

Figure 26: Les *Brucella* et leurs hôtes préférentiels (Verger, 1993).

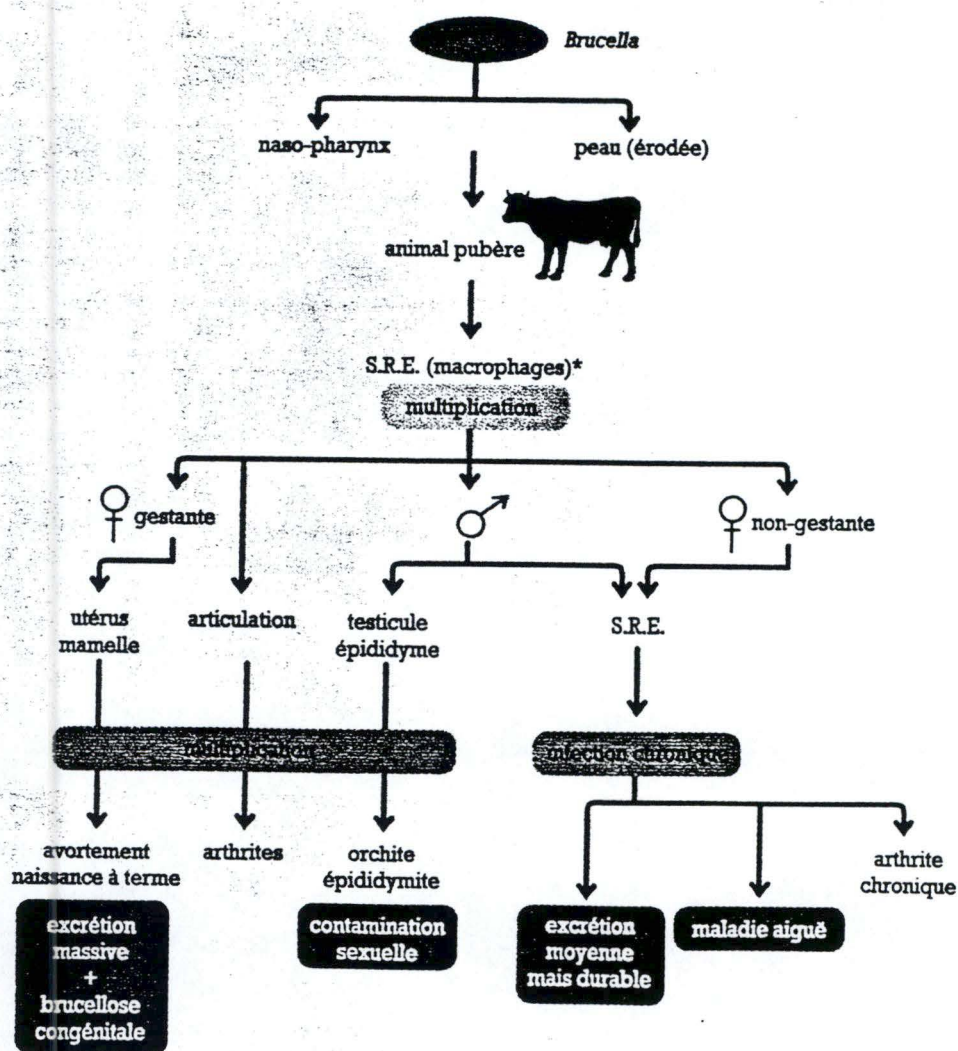


Figure 27: Evolution de la Brucellose chez un individu (S.R.E. = système réticulo-endothélial), (Verger, 1993)

### III. Brucella:

Les brucelles, identifiées par Bruce en 1887, sont de petits coccobacilles de 0,5 à 0,7  $\mu\text{m}$  de large et de 0,6 à 1,5  $\mu\text{m}$  de long, appartenant à la classe des  $\alpha 2$ -Protéobactériaceae. Elles sont immobiles, asporulées, Gram-négatives et aérobies strictes (Alton *et al.*, 1988).

Le genre *Brucella* est divisé en 6 espèces (*B. abortus*, *B. canis*, *B. melitensis*, *B. neotomae*, *B. ovis* et *B. suis*) qui présentent plus de 87% de similarité au niveau de leur ADN et un contenu en G+C de 59%. Ce sont des parasites intracellulaires facultatifs ayant une propension marquée à se développer dans les macrophages et responsables d'un groupe de maladies de répartition mondiale, les brucelloses, communes à l'homme (fièvre de Malte), et à quelques espèces animales telles que les ruminants, les équidés et les porcins (Fig. 26 et 27).

Les symptômes les plus généraux de cette maladie chez les animaux sont l'arthrite, l'avortement chez les femelles et la stérilité chez les mâles. Chez l'homme, on constate des fièvres intermittentes en phase aiguë et des atteintes ostéo-articulaires en phase chronique (Verger, 1993).

En raison de leur impact mondial au niveau économique et sanitaire, la mise en place de programmes de lutte contre les brucelloses dans les populations animales est nécessaire. Outre l'abattage des animaux réagissant aux épreuves du diagnostic sérologique, le recours à la vaccination est possible pour prévenir cette zoonose (Verger, 1993). Malheureusement, les souches vaccinales utilisées génèrent une réponse humorale très proche de celle provoquée par une souche infectieuse. Il est donc impossible de distinguer sans ambiguïté les animaux vaccinés de ceux qui sont contaminés. De plus, ces vaccins n'assurent la protection que de 60 à 75% des animaux. Enfin, ils restent pathogènes pour l'homme.

Une meilleure compréhension de la pathogénie de cette infection (mécanismes d'adaptation et facteurs de virulence) permettrait d'initier une recherche pouvant conduire à la production de souches vaccinales mieux caractérisées et ne présentant pas les inconvénients cités ci-dessus.

Une des voies d'investigation qui se propose à nous est l'étude de systèmes à deux composantes impliqués dans la régulation des mécanismes d'adaptation et de virulence de *Brucella* spp. Cette hypothèse est soutenue par la mise en évidence d'au moins un système chez les bactéries du genre *Brucella* (Halling, 1994).



## Objectifs du travail

Notre travail s'inscrit dans le cadre de l'étude du modèle bactérien *Brucella* sp. A l'heure actuelle, les mécanismes d'adaptation et de virulence de ces bactéries sont encore largement méconnus et de nombreuses pistes restent à explorer, entre autre l'étude de systèmes de régulation et plus particulièrement les systèmes de transduction du signal à deux composantes. Ces systèmes sont retrouvés des archaebactéries aux eucaryotes et interviennent dans plus de 50 voies de transduction du signal.

L'objectif majeur de ce travail est la caractérisation d'une paire senseur-régulateur appartenant au système de transduction du signal à deux composantes chez *Brucella* sp.

Ce travail comporte deux étapes.

La première étape consiste à cloner et séquencer un opéron encodant un couple senseur-régulateur chez *Brucella* sp. Nous avons choisi de cloner un homologue du système de référence qu'est le système OmpR-EnvZ chez *E. coli*.

Dans un deuxième temps, nous caractériserons le(s) partenaire(s) cloné(s) par une approche théorique. Celle-ci consistera à prédire la structure d'un de ces partenaires et élaborer un modèle tri-dimensionnel. Le modèle 3-D du partenaire cloné devra pouvoir être utilisé comme outil dans l'étude des relations structure/fonction de ces protéines: caractérisation des résidus impliqués dans les fonctions catalytiques, ainsi que des régions impliquées dans l'interaction entre les deux partenaires.

# XL1 blue

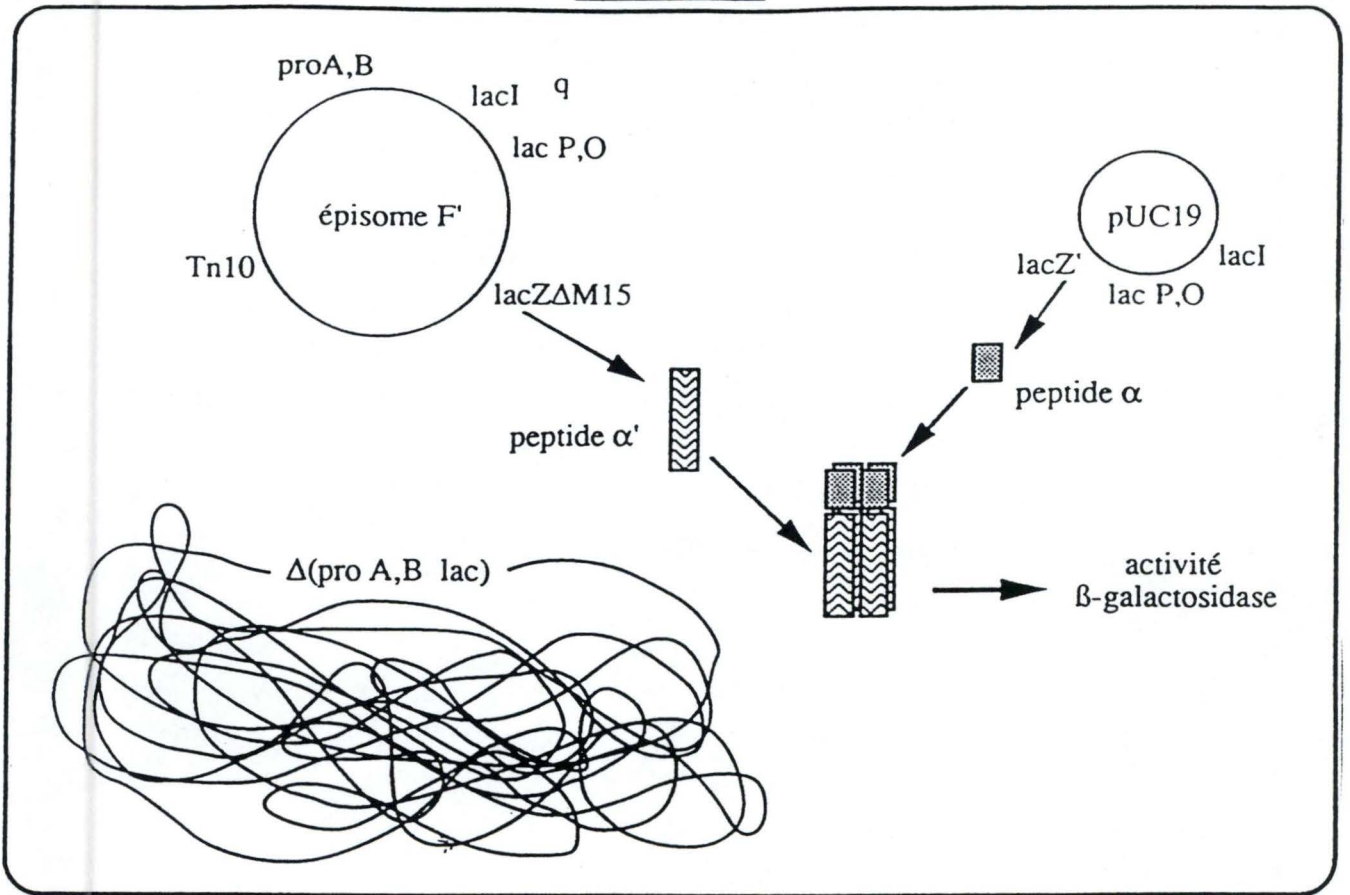


Figure 28: Système de complémentation intragénique permettant le test blanc-bleu dans la souche XL1 blue avec un plasmide pUC19.

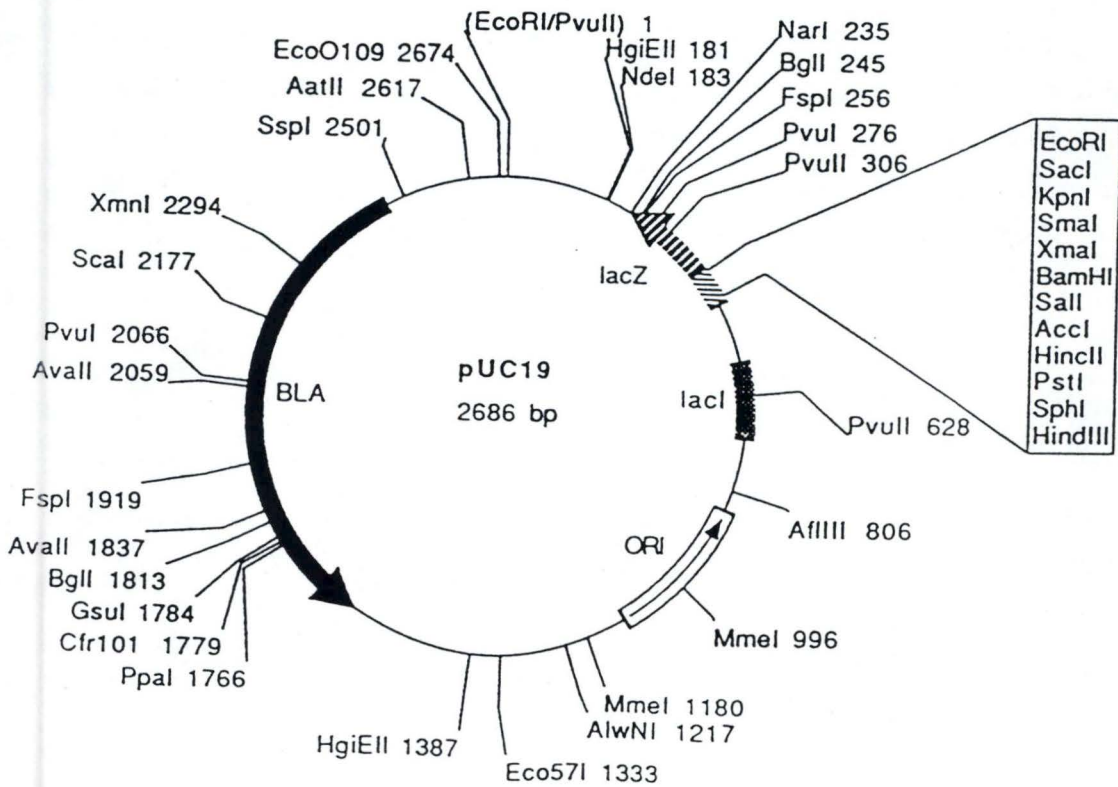


Figure 29: Carte de restriction du plasmide pUC19.

# Matériel et Méthodes

## I. Techniques de génie génétique:

### A. MATERIEL

#### 1. Souches et plasmides.

##### \* *E. coli* XL1-blue (Fig. 28)

Cette souche est utilisée pour les différentes étapes de clonage.

Son génotype est recA1, endA1, gyr A96, thi-1, hsd R-17, Sup E44, rel A1,  $\Delta$  lac, (F', lac IqZ $\Delta$ M15, Tn10).

- recA1 réduit la recombinaison au millième de son taux normal, minimisant ainsi la recombinaison entre ADN endogène et exogène;
- endA1 améliore le rendement et la qualité des préparations d'ADN plasmidique à petite échelle en diminuant le taux d'endonucléases;
- $\Delta$  lac, délétion de l'opéron lactose qui la rend incapable de métaboliser le lactose.

L'épisome F' (facteur de conjugaison bactérienne) porte les caractères suivant:

- Tn10 contient un marqueur de résistance à la tétracycline (Tet<sup>R</sup>) permettant de sélectionner les bactéries contenant l'épisome F';
- lac Iq renforce la production du répresseur de l'opéron lactose;
- et Z $\Delta$ M15, fragment du gène lacZ qui, par complémentation intragénique avec un plasmide portant un autre fragment du gène lacZ (lacZ'), donne une coloration bleue à la colonie en présence d'un chromogène, le X-gal. Si, au niveau du plasmide, le fragment lacZ' est disrupté (plasmide recombinant), la colonie sera blanche en présence de X-gal.

##### \* pUC19 (Fig.29)

Le vecteur utilisé pour la construction de la sous-banque de *B. melitensis* 16M est le plasmide pUC19. IL comporte 2686 paires de bases et porte une origine de répllication pour *E. coli* (col E1 ORI) et *BLA*, un gène de résistance à l'ampiciline ( $\beta$ -lactamase) permettant la sélection des colonies possédant le vecteur.

Le plasmide pUC19 contient :

- lac I, un fragment non opérationnel du gène régulateur lacI

- lac Z', contenant le promoteur lac, l'opérateur et la partie N-terminale du gène de la  $\beta$ -galactosidase. Cette portion du gène encode un peptide  $\alpha$  inactif. Lorsqu'un tel vecteur transforme une bactérie XL1-blue, la complémentation intragénique entre le peptide  $\alpha$  (encodé par le vecteur) et le produit du gène lacZ $\Delta$ M15 (porté par F') permet d'obtenir une  $\beta$ -galactosidase fonctionnelle. L'activité  $\beta$ -galactosidasique est mise en évidence par un test coloré. En conditions inductrices, le X-gal ajouté au milieu de culture est clivé par la  $\beta$ -galactosidase, libérant un chromophore colorant les colonies en bleu.

A l'intérieur de l'extrémité N-terminale du gène lacZ' a été introduite une batterie de sites de clonage (site multiple de clonage ou MCS) où sont localisés 10 sites de restriction uniques. Cette insertion n'altère pas la capacité de complémentation du peptide  $\alpha$ .

Cependant, lors du clonage d'un fragment d'ADN dans un site de restriction de cette batterie (créant un plasmide recombinant), le gène lacZ' est disrupté entraînant l'absence de production du peptide  $\alpha$ : la complémentation n'ayant pas lieu, les colonies restent blanches. Le test de coloration permet donc de discriminer les colonies possédant le plasmide natif et celles possédant un plasmide recombinant.

## 2. Solutions et Tampons.

### a) Solutions:

\* Solutions d'antibiotiques (à utiliser à 1/1000<sup>ème</sup> de volume):

- Ampicilline 10 % dans de l'eau distillée, stérilisée par filtration (filtre Millipore) 0,22  $\mu$ m
- Tétracycline 12,5 mg/ml dans un mélange 50 % eau distillée-50 % éthanol

\* Solution d'IPTG (Isopropyl- $\beta$ -D-thio-galactopyranoside):

- 100 mM IPTG dans de l'eau pure
- Filtrer sur membrane 0,2  $\mu$
- Conserver à -20°C

\* Solution d'X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -galactoside):

- 2% X-gal dans du diméthylformamide
- Filtrer et conserver à -20°C

$\lambda$  Hind III

taille (bp)	ng
23 130	238
9 416	97
6 557	63
4 361	45
2 322	24
2 027	21
564	5,8

$\phi$ x174DNA/HincII DIGEST

taille (bp)	ng
1057	196
770	143
612	114
495	92
392	73
345	
341	189
335	
297	109
291	
210	39
162	30
79	15

Figure 30: Etalons de taille pour gel d'agarose.

\* Solution de RNase (Macherey - Nagel):

- Incuber 1 mg de RNase par ml d'eau pure 60' à 37°C
- A utiliser à 1/10<sup>ème</sup> de volume

\* Solution de bromure d'éthidium: 2mg/ml.

\* Marqueurs de taille pour gel d'agarose (fig. 30)

- λ/ Hind III (Gibco BRL) 100 µl
- tampon de marqueur de taille 700 µl
- Bromophenol bleu 6x 200 µl
- . Incuber 10' à 65°C et aliquoter par 10 µl
- . Conserver à - 20°C
- φx174DNA/HincII DIGEST ou Marqueur V (Eurogentec)

\* Solution de coloration et d'alourdisseur:

Sucrose	50%
EDTA	100 mM
SDS	1%
Bleu de bromophénol	0.1%

\* SSC 25X pH 7.0

- NaCl 3,75M (21,9g/100ml)
- Citrate de sodium 375 mM (11,0g/100ml)
- Ajuster le pH à 7,0 (avec quelques gouttes de NaOH 10M).

\* Solution de phénol/chloroforme:

- phénol saturé au TEN buffer (10mM Tris.HCl pH 8,0, 1mM EDTA, 10 mM NaCl). Ajouter un volume de phénol à un volume de TEN. Mélanger et centrifuger, éliminer la phase supérieure et recommencer jusqu'à ce que le pH du mélange soit supérieur à 7,8.
- ajouter à ce mélange un même volume de chloroforme (UCB) + alcool isoamylique (UCB) (dans une proportion 24/1).

**b) Tampons:**

\* Tampon de marqueur de taille pour gel d'agarose

Tris	10 mM
NaCl	5 mM
EDTA	0,1 mM

\* P1 + RNase :

Tris/HCl	50 mM pH 8,0
EDTA	10 mM

Ajouter 100 µg de RNaseA par ml . Stocker à 4°C.

\* P2 :

NaOH	200 mM
SDS	1 %

Conserver à température ambiante.

\* P3 :

KAc	2,6 M pH 5,2
-----	--------------

Conserver à température ambiante.

\* Disodium Phosphate 0,5M pH 7.2

Na <sub>2</sub> HPO <sub>4</sub>	0,5 M (13,4g/100ml)
H <sub>3</sub> PO <sub>4</sub>	85% (0,4ml/100ml)

\* EDTA 0,2M pH 8,0 7,4g/100ml

\* PBS 10X

Na <sub>2</sub> HPO <sub>4</sub>	0,58M (82,3g)
NaH <sub>2</sub> HPO <sub>4</sub> •H <sub>2</sub> O	0,17M (23,5g)
NaCl	0,68M (40,0g)

- ajouter de l'eau déionisée pour arriver à un volume final de 1l.

\* HSB (High-salt elution buffer)

Tris-HCl	50mM, pH8,0
NaCl	1M
EDTA	10mM, pH8,0

\* TAE 50X :

Tris HCl	242 g
Acide acétique	57,1 g
EDTA 0,5 M (pH 8,0)	100 ml

Dissoudre dans un litre d'eau distillée

### 3. Milieux de culture:

#### \* Milieu Luria Bertani (LB)

-Tryptone	1 g/100 ml.
-NaCl	0,5 g/100 ml.
-Yeast extract	0,5 g/100 ml.

Le milieu est autoclavé (20 minutes à 120°C).

#### \* Milieu LB solide

De même composition que le LB liquide, ce milieu contient 2% d'Agar. On additionne au milieu autoclavé et refroidi à 55°C des antibiotiques ou autres additifs mentionnés et coulés dans des boîtes de Pétri.

#### \* SOC ( milieu utilisé pour les électroporations):

Yeast extract	0,5%
Tryptone	2%
KCl	2.5mM
NaCl	10 mM
MgCl <sub>2</sub>	10 mM
MgSO <sub>4</sub>	10 mM
glucose	20 mM

## B. METHODES:

### 1. Techniques relatives aux cellules:

#### Transformation par électroporation

Cette technique utilise un bref choc électrique pour permettre l'entrée d'ADN étranger dans une cellule bactérienne. La survie et la transformation bactérienne dépendent de l'intensité et de la durée du choc électrique.

#### *Préparation de cellules électrocompétentes*

- A partir d'une colonie de la souche *E. coli* XL1-blue, obtenue sur milieu solide (LB Agar + Tétracycline), effectuer une préculture ON à 37°C dans 5 ml de milieu (LB+ tétra) liquide
- Transférer la préculture dans un Erlen de 1 L contenant 400 ml du même milieu



- Incuber à 37°C jusqu'à l'obtention d'une D.O. à 550 nm de 0,5 et placer 20' sur glace
- Centrifuger 15' à 5000 rpm et à 4°C
- Eliminer le surnageant et ajouter au culot 200 ml d'eau froide et stérile
- Centrifuger 15' à 5000 rpm et à 4°C
- Eliminer le surnageant et ajouter au culot 100 ml d'eau froide et stérile
- Centrifuger 15' à 5000 rpm et à 4°C
- Eliminer le surnageant et ajouter au culot 4 ml de glycérol 10% froid et stérile
- Centrifuger 15' à 5000 rpm et à 4°C
- Eliminer le surnageant et ajouter au culot 600 µl de glycérol 10 % froid et stérile
- Aliquoter par 300 µl et stocker à - 80°C

### *Electroporation:*

- Dégeler les cellules électrocompétentes sur glace
- Dans un Eppendorf stérile froid, mélanger 100 µl de cellules avec 10 ng d'ADN ( de 1 à 10 µl maximum) et laisser au moins 1' sur glace
- Placer le "Gene Pulser" à 25 mF et 2500 volts et le "Gene Controller" à 200 Ω
- Transférer, sans faire de bulles, le mélange d'ADN et de cellules dans une cuvette à électroporation de 0,2 cm de largeur
- Placer cette cuvette entre les 2 électrodes de l'électroporateur et mettre sous tension
- Enlever la cuvette et y ajouter immédiatement 1 ml de milieu SOC ou de LB
- Transférer le mélange dans un Eppendorf stérile et incuber 1 heure à 37°C
- Etaler sur boîte LB Agar (+ Tétracycline + Ampicilline)
- Incuber ON à 37°C

## 2. Techniques relatives à l'ADN.

### a) *MINI*Préparation d'ADN plasmidique:

Cette technique permet de purifier les plasmides produits par *E.coli*. L'opération est conçue pour isoler l'ADN plasmidique de l'ADN génomique et des protéines bactériennes.

Une miniprep comprend 5 étapes: la lyse alcaline des bactéries, l'élimination des protéines et des débris cellulaires, la précipitation de l'ADN plasmidique, le lavage de l'ADN plasmidique et enfin sa resuspension.

- Ensemencer 5 ml de milieu liquide (LB+antibiotiques) avec une colonie isolée.

- Incuber la culture liquide une nuit à 37°C sous forte agitation.
- Centrifuger 1.5 ml de la culture dans un eppendorf (pendant 5 minutes à 12 500 rpm) et éliminer le milieu de culture.
- Resuspendre ce culot dans 300 µl de P1+ RNAaseA, ajouter 300 µl de P2 et mélanger doucement. Incuber 5 minutes à température ambiante.
- Ajouter 300 µl de P3 et mélanger doucement, incuber 10 minutes sur glace.
- Centrifuger 15 minutes à 4°C (à 12500 rpm) et récupérer le surnageant.
- Précipiter l'ADN en ajoutant 0.72 volume d'isopropanol, homogénéiser et centrifuger 15 minutes à température ambiante (à 12500 rpm). Eliminer le surnageant.
- Laver l'ADN, c'est-à-dire ajouter 300 µl d'éthanol 70% (-20°C) et centrifuger 5 minutes à 12500 rpm à 4°C. Eliminer prudemment l'éthanol.
- Sécher sous vide quelques minutes et resuspendre dans 20 à 50 µl d'H<sub>2</sub>O distillée stérile (suivant le nombre de copies du plasmide récupéré).
- Stocker à -20°C.

#### b) Extraction au phénol:

Afin d'extraire les protéines d'une solution d'ADN, on utilise le phénol, solvant non-miscible à l'eau. Les protéines précipitent à l'interface entre la phase organique et la phase aqueuse. Cette opération s'effectue entre deux restrictions ou entre une restriction et une ligation.

- Porter le volume de la solution à 200 µl avec de l'eau pure.
- Ajouter 200 µl de phénol saturé en TEN.
- Agiter au vortex pendant 30 secondes.
- Centrifuger 5' à la microfuge.
- Prélever avec précaution la phase supérieure, en évitant de reprendre l'interface phénol-eau.

#### c) Précipitation de l'ADN.

Il est souvent nécessaire de concentrer une solution d'ADN, par exemple après une extraction au phénol. A cet effet, on précipite l'ADN pour le resuspendre dans un volume inférieur.

- Ajouter à la solution 1/10ème de volume d'acétate de sodium 3M, pH 4,8 et 2,5 volumes d'éthanol absolu
- Mélanger par retournement.
- Précipiter l'ADN 30' à -80°C

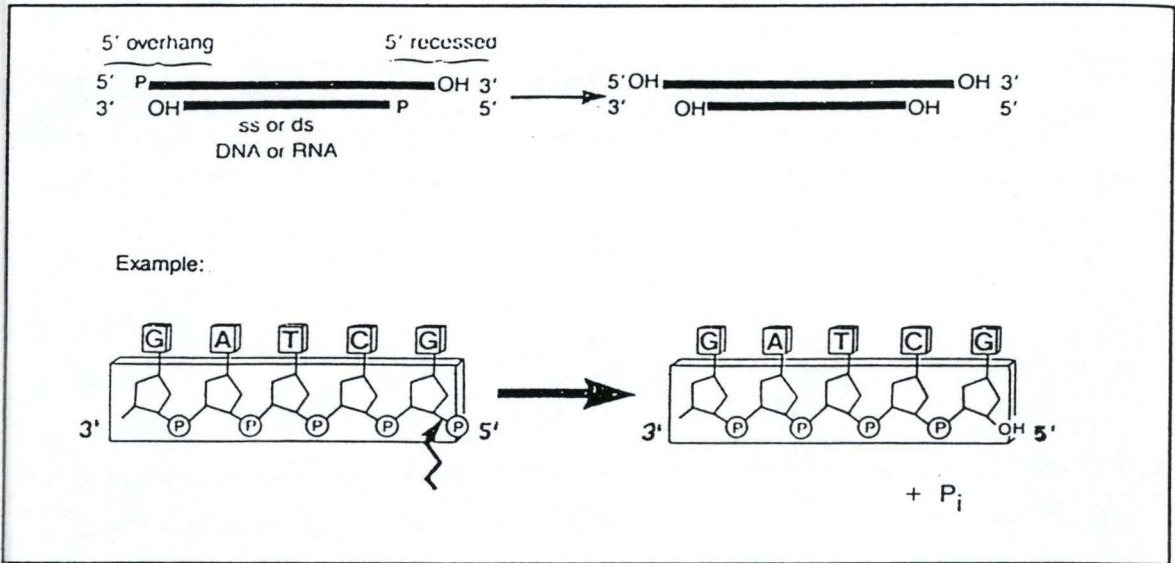


Figure 31: Représentation de l'action de la phosphatase alcaline d'intestin de bovin.

- Centrifuger 15' à 4°C et 13000 g
- Eliminer le surnageant et ajouter 1 ml d'éthanol 70% (désalage)
- Centrifuger 5' à 13000g
- Eliminer le surnageant
- Sécher le culot
- Resuspendre l'ADN précipité dans le volume désiré d'eau pure.

#### d) Restriction d'ADN purifié:

Le clivage de l'ADN bicaténaire à des sites précis se fait par des endonucléases de restriction qui reconnaissent de courtes séquences palindromiques de 4 à 6 paires de bases. Les enzymes de restriction et leurs tampons sont fournis par Boehringer Mannheim.

- Placer dans un eppendorf :
  - . x µl d'ADN à restreindre en suspension dans de l'eau distillée
  - . Tampon de restriction 10x: 1/10ème du volume final (livré avec l'enzyme de restriction)
  - . Enzyme de restriction: entre 2 et 10 unités d'enzyme par µg d'ADN (1 unité correspond à la quantité d'enzyme nécessaire pour restreindre complètement 1 µg d'ADN du phage λ en 1 heure).
- Amener au volume final avec de l'eau pure
- Incuber 1 heure à la température optimale pour l'enzyme (souvent 37°C)
- La réaction peut aussi être stoppée (enzyme inactivée) soit par chauffage (15 à 20 minutes à 65°C) si l'enzyme est thermosensible, soit par une extraction phénol/chloroforme.

#### e) Déphosphorylation

Après la restriction d'un plasmide, les extrémités 5' de l'ADN peuvent être déphosphorylées afin d'augmenter l'efficacité des clonages futurs en empêchant le plasmide de se refermer sur lui même. Cette réaction de déphosphorylation est effectuée par la CIP (Calf Intestine Phosphatase, Boehringer Mannheim). Elle catalyse l'hydrolyse des résidus 5'-phosphate de l'ADN ou de l'ARN. Le produit déphosphorylé possède ainsi une extrémité 5'-hydroxyl (fig. 31)

- mélanger dans un Eppendorf:
  - . x µl de plasmide à déphosphoryler.
  - . 0.5 U de CIP/ 3-5µg de vecteur restreint.
  - . 5 µl de tampon de CIP 10X.
  - . porter à 50 µl avec de l'eau stérile .

- incuber 1H à 37°C
- inactiver la réaction 15 minutes à 75°C
- précipiter à l'éthanol.

#### f) Ligation:

Pour lier de manière covalente deux fragments d'ADN, on utilise l'enzyme ADN ligase du phage T4, qui catalyse la formation de liens phosphodiester entre l'hydroxyle 3' d'une extrémité d'ADN et le phosphate 5' d'une extrémité proche. L'enzyme pour être active nécessite la présence d'ATP et d'ions  $Mg^{++}$ . La réaction nécessite un ADN purifié des protéines et des sels qui pourraient le contaminer. Une extraction au phénol est donc nécessaire entre une restriction et une ligation.

- Mélanger dans un tube Eppendorf :
  - . les fragments d'ADN à liguer, le nombre d'insert étant un peu plus élevé que le nombre de vecteur (au moins 200ng).
  - . 1/10 ème du volume final de tampon ligase 10x livré avec l'enzyme
  - . le nombre d'unités de ligase nécessaire (dépendant de la quantité d'ADN)
- Compléter avec de l'eau distillée stérile.
- Incuber pendant 2 heures à température ambiante.
- Conserver le mélange de ligation à -20°C.

#### g) Electrophorèse en gel d'agarose:

Cette technique permet de séparer des fragments d'ADN en fonction de leur taille. Le pourcentage d'agarose à utiliser sera donc inversement proportionnel aux tailles des fragments à séparer. Le pourcentage habituel est de 0,8%.

La taille des fragments est estimée en faisant co-migrer des fragments d'ADN de taille connue. Nous avons utilisé comme marqueurs de taille: l'ADN du phage  $\lambda$  coupé par *Hind* III (GibcoBRL) ou le marqueur V (Eurogentec).

L'utilisation du bromure d'éthidium, une molécule fluorescente liant l'ADN bicaténaire en s'intercalant entre les bases de la double hélice, permet la visualisation des fragments sous illumination UV.

La fluorescence d'un fragment d'ADN est proportionnelle à sa taille, mais aussi à la quantité d'ADN présente; ainsi, en connaissant la quantité d'ADN du marqueur de taille, on peut estimer la quantité d'ADN du fragment par comparaison des intensités de fluorescence.

#### *Préparation de l'échantillon*

- Ajouter 2  $\mu$ l de solution de coloration et d'alourdisseur à 10  $\mu$ l d'ADN

### *Préparation du gel*

- Dissoudre l'agarose (de 0,8 à 2% selon la séparation désirée) dans le TAE 1x et chauffer la solution jusqu'à transparence.
- Placer les peignes dans la cuve
- Couler l'agarose dans la cuve quand la solution a refroidi à environ 50°C, ajouter 2 µl de bromure d'éthidium (2mg/ml) par 50 ml de gel
- Laisser le gel se solidifier au moins 20'
- Retirer les peignes
- Placer le gel dans la cuve et le recouvrir de TAE 1x

### *Migration*

- Déposer les échantillons préparés dans les puits
- Migrer avec une différence de potentiel de 50 - 100 volts pendant 1 à 2 heures.
- Visualiser sur table UV

### h) Récupération d'ADN sur gel d'agarose par papier:

Les fragments d'ADN sont séparés par électrophorèse en gel d'agarose de concentration appropriée. Deux fentes sont ensuite incisées de part et d'autre des fragments d'ADN d'intérêt et une bande de membrane DEAE-cellulose est insérée dans ces fentes. L'électrophorèse est poursuivie jusqu'à ce que tous les fragments d'intérêt soient récoltés sur la membrane inférieure; la membrane au-dessus des fragments d'intérêt servant d'écran d'arrêt pour les fragments de taille supérieure. La membrane inférieure est ensuite ôtée du gel et lavée. Enfin l'ADN est élué de la membrane. Le protocole qui suit est basé sur la procédure originale de Grivitz *et al.* (1980), qui a utilisé une membrane de dialyse pour attraper l'ADN.

#### *Préparation des papiers*

- Couper les WHATMAN DE81 ion exchange (pour cela: mettre des gants et utiliser des ciseaux stérilisés à la flamme).
- Laisser quelques heures dans du NaCl 2,5M.
- Laver plusieurs fois dans de l'H<sub>2</sub>O.
- Stocker dans de l'EDTA 1mM à 4°C.

#### *Récupération*

- Digérer une quantité d'ADN contenant au moins 100ng de fragments d'intérêt. Séparer les fragments sur un gel d'agarose de concentration appropriée, puis localiser la bande des fragments d'intérêt sous lampe UV.
- Couper devant puis derrière la bande à récupérer et introduire les papiers.
- Tirer le gel de 30' à 1h et vérifier les papiers sous UV avant de les enlever.
- Laver les papiers dans de l'eau pure froide.

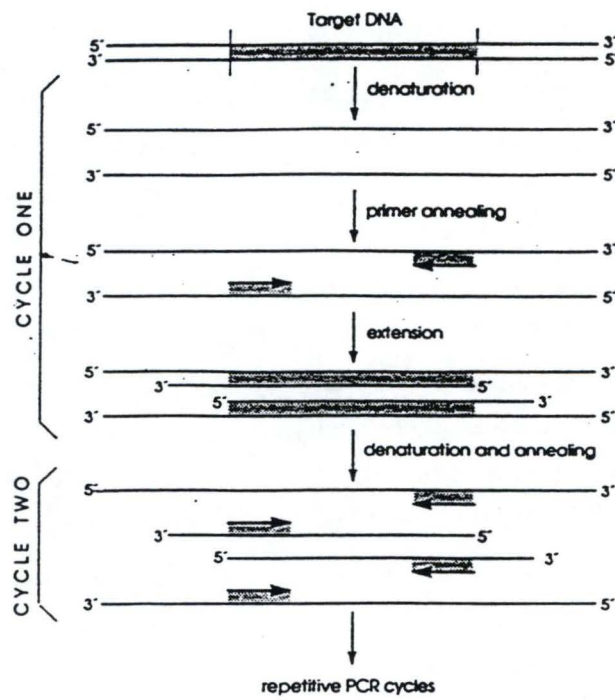


Figure 32: Schéma de l'amplification de gène par PCR (d'après Gerhardt, 1994)

- Sécher les papiers entre deux buvards (pas de trop pour ne pas fixer l'ADN irréversiblement).
- Les mettre dans un Eppendorf et ajouter 350 µl de HSB.
- Secouer et vortexer jusqu'à l'obtention d'une pâte.
- Incuber 2h à 37°C en secouant de temps en temps.
- Trouer le fond de l'Eppendorf, le placer sur un second et centrifuger 5'.
- Ajouter 700 µl de N-Butanol pour extraire le bromure d'éthidium et centrifuger 5'.
- Eliminer le surnageant.
- Ajouter 700 µl d'éthanol 100%.
- Précipiter 10' à -80°C, centrifuger 20' à 4°C et éliminer le surnageant à la pipette pasteur.
- Laver avec 300 µl d'éthanol 90%.
- Centrifuger 20' à 4°C
- Sécher sous-vide, puis resuspendre dans un volume adéquat d'H<sub>2</sub>O (en général 20 µl)

#### i) Amplification PCR:

La PCR permet l'amplification de fragments d'ADN (fig. 32).

L'ADN doit être dénaturé. Des amorces oligonucléotidiques d'environ 20 bases, complémentaires de séquence se trouvant de chaque côté de la région-cible, s'hybrident chacune à un des deux brins. La longueur de la séquence-cible est déterminée par la distance entre les 2 amorces.

Une ADN polymérase thermostable est alors utilisée pour synthétiser à partir de l'extrémité 3'-OH de chaque amorce un brin d'ADN. Un cycle consiste en 3 étapes: dénaturation, hybridation des amorces et élongation. En répétant ce cycle, on augmente de façon exponentielle le nombre de copies de la cible.

La température d'hybridation utilisée est inférieure à la température de melting  $T_m$  des amorces. Ce  $T_m$  se calcule de la façon suivante:  $T_m = 4(G+C) + 2(A+T)$ . Les amorces devront contenir environ le même pourcentage de bases G-C que l'ADN cible. Il est également recommandé de façon à limiter les amplifications non spécifiques que l'amorce contienne un excès de G/C à l'extrémité 5' et au centre et que les deux ou trois dernières bases de l'extrémité 3' soient A/T

#### *Méthode*

-Mélange de réaction réalisé dans un microtube:

- environ 250ng d'ADN génomique ou plasmidique portant la région à amplifier
- 5 à 10 µl de tampon d'ADN polymérase (DynaZyme)10X



- 2  $\mu$ l de dNTP 10 mM
- 1 à 10  $\mu$ l de primer 1 (10 pmole/ $\mu$ l)
- 1 à 10  $\mu$ l de primer 2 (10 pmole/ $\mu$ l)
- 1 à 2  $\mu$ l de Dynazyme (2 U/ $\mu$ l)
- Porter à 50 ou 100  $\mu$ l avec de l'eau distillée stérile
- Déposer 80  $\mu$ l d'huile minérale ultra-pure au-dessus du mélange de réaction afin d'empêcher toute évaporation.
- Programmer l'appareil à PCR comme suit:
  - Dénaturation de 5 minutes à 95°C.
  - 36 cycles d'amplification présentés comme suit:
    - 1 min. à 95°C pour la dénaturation
    - 1 min. à 55°C pour l'hybridation
    - 1 min. à 72°C pour la polymérisation
  - Elongation finale de 10 minutes à 72°C.
- Eliminer l'huile minérale.
- Déposer 10  $\mu$ l sur gel d'agarose et conserver le reste à -20°C.

#### *Purification des fragments obtenus par PCR*

Quand la PCR est une PCR préparative (et une non analytique), il est souvent nécessaire de purifier les séquences amplifiées à partir du mélange réactionnel des contaminants (primers de départ, primers-dimers, protéines, sels,...). Cette purification a été menée en utilisant le kit Wizard PCR Preps (Promega).

- Transférer la phase aqueuse dans un eppendorf propre et stérile, car la présence de l'huile minérale peut entraîner une diminution du rendement de la purification.
- Ajouter 100  $\mu$ l de "Direct Purification Buffer". Vortexer.
- Ajouter 1 ml de résine (PCR Preps DNA purification resin) et vortexer 3 fois 20 secondes.
- Préparer une minicolonne "Wizard" pour chaque produit PCR. Retirer le piston d'une seringue de 3ml et y attacher la minicolonne.
- Injecter le mélange ADN/tampon/résine dans la minicolonne à l'aide du piston (très lentement).
- Détacher la seringue de la minicolonne et retirer le piston. Réajuster le corps de la seringue à la minicolonne et injecter 2 ml de la solution de lavage (isopropanol 80%).
- Transférer la minicolonne sur un eppendorf. Centrifuger le tout 20 secondes à 12000 rpm pour éliminer la solution de lavage et sécher la résine.
- Placer la minicolonne sur un nouvel eppendorf et ajouter 50  $\mu$ l d'eau distillée.
- Centrifuger 20 secondes à 12000 rpm pour éluer l'ADN.

- Stocker l'ADN purifié à -20°C.

#### Direct purification buffer

KCl	50mM
Tris HCl	10mM, pH 8,8
MgCl <sub>2</sub>	1,5mM
Triton X-100	0,1%

#### j) Marquage d'une sonde à la fluorescéine

Le Random Primer Fluorescein DNA Labeling Kit est utilisé pour générer une sonde fluorescéinylée à partir d'un fragment d'ADN. Le fragment de Klenow de l'ADN polymérase I catalyse *in vitro* l'addition de nucléotides fluorescéinylés à partir de primers aléatoires (nonamers) fluorescéinylés. Les éléments du kit (stockés à -20°C) sont placés dans de la glace durant la manipulation.

- Préparer le tampon de réaction en mélangeant 1µl de dUTP-fluorescéine avec 11,5µl de Nucleotide buffer 5x dans un eppendorf stérile.
- Dans un eppendorf, mélanger les réactifs suivants:
  - . 100 ng d'ADN template (1 à 29 µl)
  - . 10 µl de random 9-mer primers
  - . de l'eau stérile pour arriver à un volume de 39 µl
- Chauffer à 95°C pendant 5' puis mettre immédiatement sur glace.
- Ajouter les réactifs suivant dans le tube de réaction:
  - . 10 µl de tampon de réaction
  - . 1 µl de Klenow (5U)
- Mélanger et incuber à 37°C pendant 30 à 60 minutes.
- Arrêter la réaction en ajoutant 2µl de Stop Solution
- Conserver à 4°C à l'obscurité.
- Les sondes fluorescéinylées doivent être dénaturées avant d'être additionées à la solution d'hybridation.

#### *Vérification du marquage de la sonde*

L'efficacité de la sonde est un facteur important pour le succès de l'hybridation et de la détection qui suivent. Le marquage est vérifié de la manière suivante:

- Réaliser 6 dilutions croissantes de la sonde et d'un ADN prébiotynilé (de 10<sup>-1</sup> à 10<sup>-6</sup>).
- Déposer en spot 1 µl de chaque dilution sur une membrane.
- Sécher la membrane et fixer l'ADN sous UV.

- Afin de visualiser les spots, suivre le protocole de l'étape de détection du Southern blot (point k.) Le marquage de la sonde est considéré comme efficace lorsque les spots des dilutions  $10^{-3}$  à  $10^{-6}$  sont visibles à la révélation.

*Réactifs livrés avec le kit*

\* 5X Nucleotide Buffer

- Solution aqueuse
- dATP, dCTP, dGTP 100 $\mu$ M
- dTTP 20 $\mu$ M

\* Random 9-mer Primers

- Tris HCl 10mM, pH 7,5
- EDTA 1mM
- random 9-mer oligonucleotides

\* DNA Polymerase I Kleenow Fragment (Exo-) 5U/ $\mu$ l

- Solution au glycerol

\* Fluorescein-12-dUTP 1mM

- Solution aqueuse

\* Stop Solution

- EDTA 0,5M, pH8,0

k) Southern blotting:

Le Southern blot est une méthode d'analyse de l'ADN imaginée par Southern 1975, pour visualiser les gènes ou toute séquence d'ADN génomique, par hybridation d'une sonde spécifique marquée avec des fragments d'ADN, préalablement séparés par électrophorèse, dénaturés et transférés sur une membrane.

*Restriction des plasmides issus de miniprep*

Les plasmides de la sous-banque purifiés par miniprep sont restreints par deux enzymes: *Eco* RI et *Bam* HI, qui sont de part et d'autre du site de clivage *Sma* I, site d'insertion du fragment génomique de *Brucella* dans le pUC19.

On utilise comme témoin le plasmide pBmo1 portant l'insert de 389 bp amplifié sur le génome de *Brucella* à partir des oligos dégénérés OmpRlike et CheYlike.

Méthode

- 1  $\mu$ l d'ADN plasmidique obtenu par minipréparation.
- 2  $\mu$ l de mélange d'enzyme:
  - . 1/20 tampon de dilution II
  - . 1/4 EcoRI
  - . 1/4 BamHI
  - . 9/20 H<sub>2</sub>O

- 2 µl de tampon B de restriction 10x (Boehringer).
- Compléter avec de l'eau distillée stérile pour obtenir 20 µl au total.
- Incuber 1 heure à 37°C
- Vérifier que la restriction est complète par électrophorèse en gel d'agarose d'1/10 du mélange de restriction.

#### *Electrophorèse en gel d'agarose*

La séparation des fragments d'ADN plasmidique restreint par *BamHI*-*EcoRI* se fait par migration dans un gel d'agarose de 1% . La coloration par le bromure d'éthidium permet, sans interférer avec la suite des opérations de contrôler la qualité de l'ADN, de la migration et de la digestion enzymatique (3 paramètres importants pour le résultat final).

Il est important de signaler que le marqueur de taille utilisé n'était pas un marqueur fluorescéinylé.

#### *Transfert*

Après avoir séparé les fragments d'ADN sur gel d'agarose, on transfère les acides nucléiques sur une membrane de nylon, car l'hybridation de la sonde est impossible au niveau du gel d'agarose. Le principe du transfert de l'ADN du gel vers la membrane repose sur le phénomène de capillarité d'où son nom de blotting.

-Après l'électrophorèse, placer le gel d'agarose sous UV (302 nm) pendant 3 à 5 minutes. Pendant ce temps, le gel peut être photographié. Les UV clivent l'ADN et facilitent ainsi le transfert.

-Immerger le gel dans la solution de dénaturation pendant 30 minutes sous faible agitation. L'ADN est ainsi fragmenté et dénaturé sous forme de simple brin, ce qui est indispensable pour l'hybridation ultérieure.

-Immerger le gel dans la solution de neutralisation pendant 15 minutes.

-Répéter le bain de neutralisation.

-Découper un morceau de membrane Immobilon-S de taille un peu supérieure à celle du gel (porter des gants sans talc pour manipuler la membrane pour éviter de tacher la membrane) et un morceau de papier Whatman.

-Immerger la membrane dans du SSC 2X.

-Par un système de transfert sous vide, l'ADN est transféré sur la membrane en présence de SSC 10X pendant 1 heure. Suivre le schéma ci-contre utilisant le système de transfert MilliBlot-V Vacuum .

-Après le transfert, sécher correctement la membrane (jusqu'à ce que les bords de la membrane s'incurvent) entre deux feuilles de papier Whatman 3MM (le séchage peut se faire rapidement dans un four à une température comprise entre 42°C et 80°C ou à température ambiante pendant une nuit).

### Light Emission Mechanism of CDP-Star™

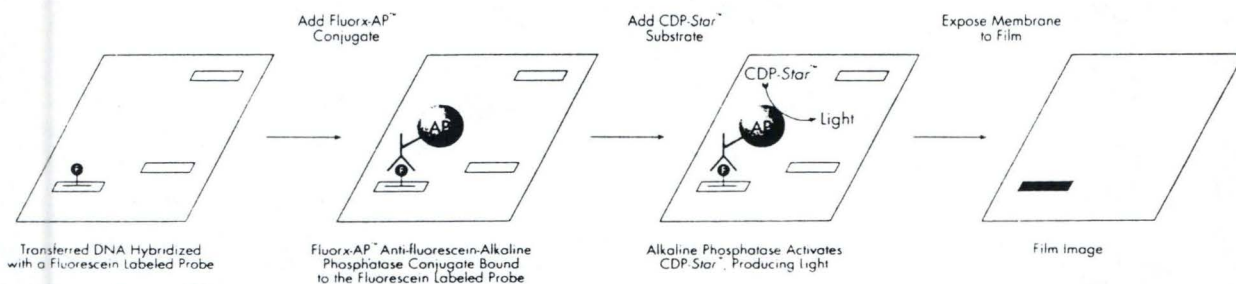
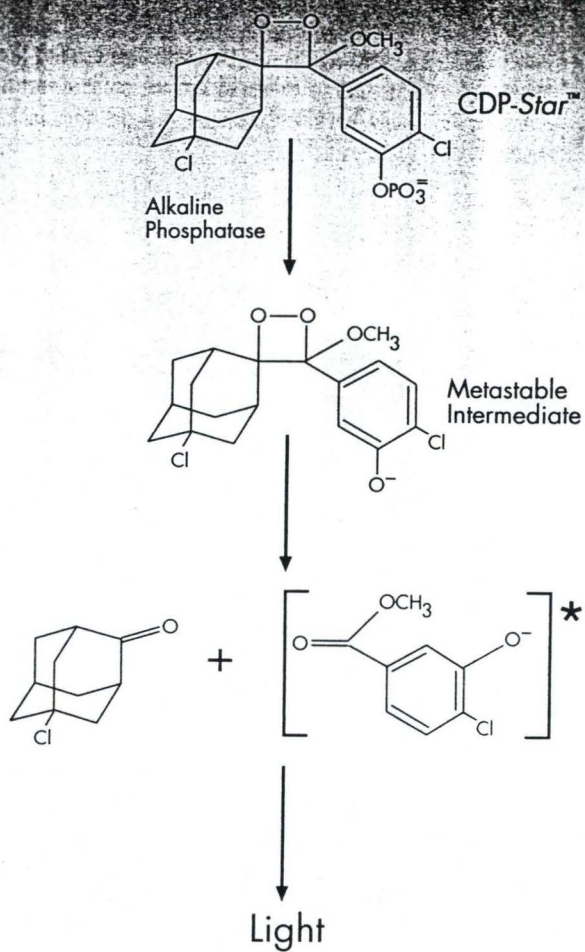


Figure 33: A, Mécanisme d'émission de lumière du CDP-Star™.  
 B, Système de détection de nucléotides avec le Southern Star™.

-Immobiliser l'ADN sur la membrane, en l'exposant aux UV pendant 5 secondes.

#### *(Pré-)hybridation*

Après avoir fixé l'ADN transféré sur la membrane et marqué la sonde, il faut hybrider la sonde aux fragments d'ADN de la membrane. Mais avant, une étape préalable est nécessaire: la préhybridation. L'ADN transféré n'a saturé qu'une faible proportion des sites de fixation potentiels de la membrane. Si la sonde fluorescéinylée était apportée avant cette étape, elle irait directement se fixer sur ces sites plutôt que d'aller s'hybrider sur les séquences complémentaires de l'ADN transféré. La préhybridation a pour but de saturer tous ces sites non spécifiques par de l'ADN hétérologue, dans notre cas de l'ADN de sperme de saumon.

- Préchauffer les solutions de préhybridation et d'hybridation à 64°C
- Bouillir le sperme de saumon pendant 5 minutes.
- Fixer l'ADN sur la membrane en passant celle-ci 5 secondes sous UV.
- Mouiller la membrane dans du disodium phosphate 0,25M pH 7,2.
- Placer la membrane dans un hybritube de 20 ml (Gibco BRL) en l'enroulant autour du plongeur (la face de la membrane portant l'ADN vers le plongeur).
- Préhybridation: ajouter 267µl de sperme de saumon à 16 ml solution de préhybridation et verser dans l'hybritube.
- Laisser incubé 1 heure à 64°C
- Bouillir la sonde 5 minutes
- Hybridation: ajouter la sonde bouillie à la solution de préhybridation, puis mélanger en inversant l'hybritube quelques fois.
- Laisser incubé une nuit à 64°C.

#### *Détection*

La détection avec le kit "Southern-Star™ Chemoluminescent Systems" comprend le blocage, une incubation avec une phosphatase alcaline conjuguée à un anticorps anti-fluorescéine (le "Fluorx-AP conjugate") et une réaction finale de chémoluminescence. Le CDP-Star™ (Tropix) est transformé en un produit responsable de l'émission locale de lumière, qui peut impressionner un film RX, voir figure 33.

Le protocole suivi comprend des temps, des stringences et des températures de lavage qui peuvent être adaptés pour une meilleure élimination du bruit de fond.

- Laver la membrane:
  - . 2 x 5' à température ambiante avec du SSC 2x, 1% SDS
  - . 2 x 5' à 64°C avec du SSC 0,1x, 1% SDS
  - . 2 x 5' à température ambiante avec du SSC 1x
- Laver la membrane 2 x 5' dans du "Blocking Buffer" (0.5 ml/cm<sup>2</sup>).

- Incuber 25' dans du "Blocking Buffer" (1 ml/cm<sup>2</sup>).
- Diluer le "Fluorx-AP conjugate" 1:5,000 dans du "Blocking Buffer". Utiliser 2,0 µl de "Fluorx-AP conjugate" dans 10ml de "Blocking Buffer" par 100 cm<sup>2</sup> (0.1 ml/cm<sup>2</sup>).
- Incuber la membrane 25' à température ambiante sous agitation constante dans la solution de conjugué.
- Laver :
  - . 5' dans du "Blocking Buffer" (0.5 ml/cm<sup>2</sup>).
  - . 3 x 5' dans du "Wash Buffer" (1 ml/cm<sup>2</sup>).
  - . 2 x 2' dans de l'"Assay Buffer" (0.5 ml/cm<sup>2</sup>).
- Eliminer le liquide en posant le coin de la membrane sur un papier Whatman, puis la poser à plat.
- Déposer une mince couche de CDP-*Star* sur la membrane (3 ml/ 100cm<sup>2</sup>) et incubé 5'.
- Eliminer l'excès de CDP-*Star* et placer la membrane dans un feuillet de développement.
- Enlever les bulles et les faux-plis.
- Eposer la membrane au film de 30" à plusieurs minutes.
- Immerger le film 5' dans le révélateur.
- Rincer le film à l'eau distillée et l'immerger 8' dans le fixateur .
- Rincer ensuite à l'eau de ville.

#### *Réactifs livrés avec le kit*

\* CDP-*Star*™ Ready-to-use substrate solution 0,25mM

Conserver à 4°C

\* I-Block™ Reagent

- Caséine hautement purifiée fournie sous forme de poudre sèche et protégée de la contamination par la phosphatase alcaline.

- Conserver à 4°C

\* Fluorx-AP

- Anticorps anti-fluorescéine conjugué avec une phosphatase alcaline.

- Diluer 1:5,000

- conserver à 4°C

\* 10X Assay Buffer

Tris HCl 200mM, pH 9,8

MgCl<sub>2</sub> 10mM

- Diluer 1:10 avec de l'eau MilliQ™ ou équivalent.

- Conserver à 4°C.

## *Solutions*

### \* Solution de dénaturation

NaOH 500 mM

NaCl 1,5 M

- Amener le volume à 1 L.

### \* Solution de neutralisation

Tris 1 M

NaCl 1,5 M

- Amener à pH 7,5.

- Amener le volume à 1L.

### \* Solution de préhybridation

EDTA 1mM (0,25ml EDTA 0,2M)

SDS 7% (17,5ml de SDS 20%)

Disodium Phosphate, pH7,2 0,25M (25ml de disodium phosphate 0,5M, pH7,2)

- Ajouter de l'eau déionisée pour arriver à un volume final de 50ml

- Si le SDS précipite, mettre la solution dans un bain à 50°C pour le redissoudre.

### \* Blocking buffer

I-Block Reagent 0,2% (0,6g)

PBS 1X (30 ml PBS 10X)

SDS 0,5% (7,5 ml SDS 20%)

- Ajouter 30 ml de PBS 10X à 200 ml d'eau déionisée.

- Ajouter l'I-Block Reagent et passer au micro-onde environ 80" (35"/100ml) ou chauffer à 70°C 5', sans faire bouillir.

- Refroidir avant utilisation.

### \* Wash buffer

PBS 1X (50 ml PBS 10X)

SDS 0,5% (12,5 ml SDS 20%)

- Ajouter de l'eau déionisée pour atteindre un volume de 500 ml.



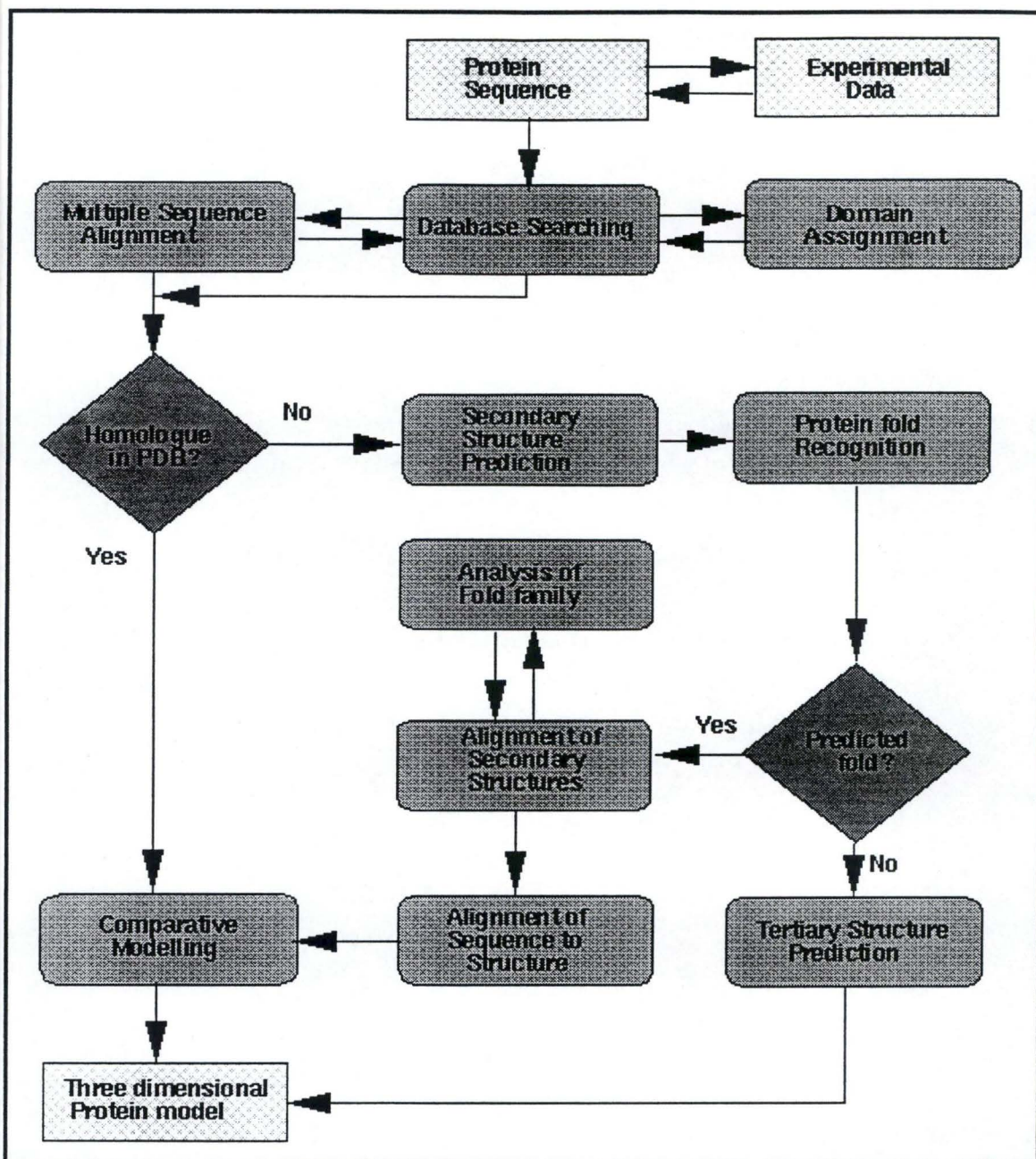


Tableau 1: Tableau schématique récapitulant les étapes généralement suivies lors de prédiction de structure de protéines (Russel, 1996)

## **II. Modélisation de structures tridimensionnelles de protéines**

### **A. Matériel**

Le matériel informatique utilisé est composé d'une station de travail graphique Silicon Graphics, modèle Indigo2, dotée d'un processeur cadencé à 150 MHz, disposant de 64Mb de mémoire vive et sous système d'exploitation Unix (version Irix 5.3), ainsi que d'une station VAX 6220 sous un système d'exploitation VMS v.5-5.2.

Les logiciels utilisés sous Unix sont les suivants :

- Alignement de séquences protéiques: Matchbox.
- Visualisation et manipulation de structures tridimensionnelles de protéines: Insight<sup>®</sup>II (version 95.0 ) (Biosym/MSI, San Diego).
- Calculs de dynamique et de mécanique moléculaires: Discover<sup>®</sup> (versions 2.9.7) (biosym/MSI, San Diego)

### **B. Méthodes de prédiction de la structure tridimensionnelle de protéines.**

La connaissance de l'architecture tridimensionnelle d'une protéine représente un outil de première importance pour la compréhension de propriétés protéiques telles que la spécificité d'interaction entre sous-unités protéiques, l'activité enzymatique, la régulation ou la stabilité.

Les méthodes expérimentales de détermination de la structure 3D de protéines sont, à ce jour, la diffraction aux rayons X ou la résonance magnétique nucléaire (RMN). Mais elles présentent certains inconvénients: la diffraction aux RX demande l'obtention préalable de cristaux protéiques (cette méthode est donc lente et coûteuse); la RMN quant à elle ne donne pas toujours de résultats, et ceux-ci sont parfois difficilement interprétables. Le nombre de structures protéiques déterminées par ces méthodes est faible par rapport au nombre de séquences connues, nombre qui ne fait qu'augmenter vu la tendance actuelle à séquencer les génomes entiers. Cependant, il est possible de prédire la structure 3D de certaines protéines à partir de leur séquence en acides aminés. Le nombre de méthodes élaborées à cette fin a beaucoup augmenté ces dernières années, et leur puissance s'améliore de jour en jour.

Le tableau 1 montre schématiquement les étapes, détaillées ci-dessous, généralement suivies lors d'une prédiction de structure.

BLAST <a href="http://www.ncbi.nlm.nih.gov/BLAST/">http://www.ncbi.nlm.nih.gov/BLAST/</a>	
blastp	Compare une séquence protéique avec les séquences contenues dans les banques de séquences protéiques;
blastn	Compare une séquence nucléique avec les séquences contenues dans les banques de séquences nucléiques.
blastx	Compare les séquences protéiques déduites dans les 6 phases d'une séquence nucléique avec les séquences contenues dans les banques de séquences protéiques.
tblastn	Compare une séquence protéique avec les séquences contenues dans les banques de séquences traduites dans les 6 phases.
tblastx	Compare les séquences déduites dans les 6 phases avec les séquences traduites dans les 6 phases.

**Tableau 2:** Description des usages de cinq programmes du serveur BLAST (Altschul *et al.*, 1990)

SwissProt	<a href="http://expasy.hcuge.ch/sprot/sprot-top.html">http://expasy.hcuge.ch/sprot/sprot-top.html</a>	Banque de séquences protéiques	Department of Medical Biochemistry (University of Geneva) and EMBL Outstation.
GenBank	<a href="http://www2.ncbi.nlm.nih.gov/genbank/query_form.html">http://www2.ncbi.nlm.nih.gov/genbank/query_form.html</a>	Banque de séquences nucléiques.	
PDB	<a href="http://www.pdb.bnl.gov/">http://www.pdb.bnl.gov/</a>	Banque de structures cristallographiques de protéines.	Brookhaven National Laboratory (New-York)

**Tableau 3:** Banques de séquences utilisées par BLAST ou FASTA.

## 1. Recherche dans les banques de données.

La première étape dans l'analyse d'une nouvelle séquence, qu'on appellera séquence-cible, est la comparaison avec d'autres séquences contenues dans des banques de données, afin de trouver des séquences significativement similaires.

La séquence cible (ou protéine cible) est la séquence pour laquelle on recherche une information sur la structure 3D; par rapport à la séquence de référence (ou *template*) dont les caractéristiques de structure secondaire ou tertiaire, connues, vont nous servir de base pour l'élaboration du modèle structural de la séquence cible.

Ces recherches peuvent être exécutées facilement grâce à de nombreux programmes accessibles sur le réseau Internet. Par exemple: BLAST (tableau 2; Altschul *et al.*, 1990) et FASTA (Pearson *et al.*, 1988)

Il est essentiel de consulter différentes banques de séquences (tableau 3), comme SwissProt, PIR ou GENBANK, afin d'obtenir le plus possible de séquences similaires, pour un alignement de séquences plus fiable. Il est important de ne pas négliger la comparaison de la séquence cible à une banque de séquences pour lesquelles des informations sur la structure tridimensionnelle sont disponibles (ex: PDB, "Protein Data Bank" Brookhaven National Laboratory- Cambridge USA). En effet, le fait que la séquence cible soit homologue ou non à une protéine de structure connue n'apparaît pas clairement à travers la recherche dans de larges banques de séquences (Russel, 1996).

## 2. Alignement de séquences.

L'alignement de séquences apporte des données sur les relations fonctionnelles et structurales entre les résidus des différentes protéines alignées.

Les alignements multiples peuvent fournir:

- une information quant à la présence de régions structurellement conservées entre la séquence cible et une protéine de structure connue;

Nous pouvons parler de régions structurellement conservées, car des séquences homologues ont une grande chance d'adopter la même structure (voir point 4)

- la localisation d'un résidu supposé être impliqué dans une fonction protéique, comme les résidus du site catalytique;

- une information sur des résidus supposés être enfouis dans le noyau protéique ou exposés au solvant; les régions hydrophobes / hydrophiles conservées seront plus probablement à l'intérieur de la protéine / exposées au solvant.

Notons que l'information obtenue avec des alignement multiples est plus fiable qu'avec un alignement pairé pour des applications telles que la modélisation par homologie et la prédiction de structure secondaire.

Certains programmes sont spécifiquement adaptés à l'alignement de deux séquences, comme Bestfit (Smith et Waterman, 1981) ou Gap (Needleman et Wunch, 1970), d'autres à l'alignement de plusieurs séquences, comme Matchbox (Depiereux et Feytmans, 1992) ou Clustalw (Thompson *et al.*, 1994).

### 3. Attribution de domaines.

Si la séquence cible fait plus de 500 acides aminés, il est très probable qu'elle soit composée de domaines de fonctions distinctes (Russel, 1996). Il est alors préférable de fragmenter la protéine et de considérer chaque domaine séparément.

La localisation des domaines peut être effectuée de différentes manières:

- Si la similarité à d'autres séquences complètes ne se retrouve que sur une portion de la séquence cible, c'est la meilleure indication de la présence de domaines structuraux distincts.

- Des régions de faible complexité, non-globulaires, séparent souvent les domaines dans les protéines (Wooton, 1994). De longues extensions de résidus répétés (P, Q, S ou T) indiquent généralement des séquences de liaison et sont habituellement de bons endroits pour fragmenter la protéine en domaines. Il existe des méthodes automatiques de segmentation de protéines (par exemple, grâce à l'algorithme SEG; Wooton, 1994)

### 4. Modélisation par homologie.

Deux séquences similaires ont une forte probabilité d'adopter une structure tertiaire identique: deux séquences de plus de 70 acides aminés partageant au moins 25% d'identité ont une très forte probabilité d'adopter la même conformation générale (Doolittle *et al.*, 1981).

Donc si la protéine cible montre une homologie significative (>25% d'identité ou >30% de similarité) avec une protéine de structure tridimensionnelle connue, un modèle de la structure tertiaire de la séquence cible pourra être obtenu.

Les étapes de base de la modélisation par homologie sont (Vinals, 1996):

- la détermination de régions structurellement conservées (SCR) sur base d'alignements multiples de séquences et éventuellement de structures;

- l'établissement des correspondances entre la séquence cible et la (ou les) structure(s) similaire(s) connue(s) à partir de l'alignement multiple;

- la construction d'une partie du squelette sur base des SCRs d'une part, des VR (régions variables) d'autre part;
- le positionnement des chaînes latérales;
- l'optimisation par minimisation d'énergie ou dynamique moléculaire;
- et l'évaluation. Dans notre cas, nous avons utilisé un programme de vérification de structure utilisant un nouveau potentiel de force moyenne à l'échelle atomique, développé au laboratoire (Melo et Feytmans, soumis)

L'optimisation du modèle est l'affinage de l'ensemble de la structure de façon à éliminer les invraisemblances introduites lors de l'assemblage des différentes "pièces détachées", liaisons covalentes inadéquates, contacts stériques invraisemblables, conformations locales défavorables, ... Pour cela, il faut caractériser plus ou moins précisément les différents types d'énergies associées aux atomes et à leurs interactions. Une fonction énergie doit donc être décrite, puis utilisée pour tenter de calculer la conformation native de la protéine. Celle-ci est la conformation correspondant au minimum absolu de la fonction énergie libre multidimensionnelle, somme du potentiel intramoléculaire, de l'entropie et de l'énergie libre de solvation.

Le principe de la minimisation d'énergie potentielle est de faire évoluer le système vers la conformation minimale la plus proche de la conformation initiale, sans garantir que ce minimum est le minimum global de la fonction énergie. Si ce minimum global est séparé de l'état initial par une barrière énergétique, un algorithme de minimisation d'énergie est incapable de l'atteindre.

La dynamique moléculaire permet au système de passer des barrières énergétiques grâce à l'énergie cinétique additionnelle, et d'atteindre ainsi un minimum d'énergie éventuellement plus proche du minimum global de la fonction.

Un exemple de modélisation par homologie est la modélisation de la D-lactate deshydrogenase (D-LDH) à partir de la formate deshydrogenase (FDH) (Vinals *et al.*, 1995).

## 5. Modélisation de protéines non-homologues à une structure 3D connue.

### a) Prédiction de structure secondaire:

Sans homologie de structure connue à partir de laquelle élaborer un modèle 3D, il est cependant possible de prédire la structure secondaire de la protéine cible. Le but de la prédiction de structure secondaire est de fournir la localisation des hélices  $\alpha$  et des brins  $\beta$  à l'intérieur d'une protéine (ou famille de protéines).

SERVEUR	Sequence(s)	URL	Références
PHD	+	<a href="http://www.embl-heidelberg.de/predictprotein/predictprotein.html">http://www.embl-heidelberg.de/predictprotein/predictprotein.html</a>	Rost et Sander, 1993
SSPred	+	<a href="http://www.embl-heidelberg.de/sspred/sspred_info.html">http://www.embl-heidelberg.de/sspred/sspred_info.html</a>	Mehta, 1995
PSSP-BCM	1	<a href="http://dot.imgen.bcm.tmc.edu:9331/psspprediction/pssp.html">http://dot.imgen.bcm.tmc.edu:9331/psspprediction/pssp.html</a>	Smith <i>et al.</i> , 1996
SOPM	1	<a href="http://www.ibcp.fr/serv_pred.html">http://www.ibcp.fr/serv_pred.html</a>	Geourjon et Deleage, 1994

**Tableau 4:** Serveurs de prédiction de structure disponibles sur le réseau. + indique que l'algorithme nécessite un alignement multiple pour effectuer la prédiction. 1 indique qu'une seule séquence est suffisante.

On prédit les structures secondaires depuis presque un quart de siècle. Les premières méthodes souffraient d'un manque de données: il y avait relativement peu de structures 3D connues à partir desquelles dériver les paramètres. De plus, les prédictions étaient effectuées sur des séquences uniques plutôt que sur des familles de séquences homologues. Les méthodes initiales les plus connues sont probablement celle de Chou et Fasman (1974), celle de Garnier, Osgruthorpe et Robson (GOR; 1978) et celle de Lim (1974). Ces auteurs annoncent d'assez hauts pourcentage de succès (70-80%), définis par le nombre de résidus correctement prédits faisant partie soit d'une hélice  $\alpha$ , soit d'un feuillet  $\beta$ , soit d'un coude, divisé par le nombre total de résidus dans la protéine. Mais dans les 19 protéines que Chou et Fasman, par exemple, avaient prises pour évaluer leur méthode, 15 étaient incluses dans l'ensemble de protéines à partir desquelles ils avaient estimé les paramètres. Après examen plus soigneux (un test avec 62 séquences), les méthodes montrent seulement entre 56 et 60% d'exactitude (Kabsch et Sander, 1983).

La disponibilité de grandes familles de séquences homologues a révolutionné la prédiction de structure secondaire. Les méthodes traditionnelles, quand elles sont appliquées à une famille de protéines plutôt qu'à une séquence unique, se révèlent bien plus conformes à l'identification d'éléments de structure secondaire du noyau protéique, allant jusqu'à des précisions supérieures à 70% (Levin *et al.*, 1993). L'augmentation de pourcentage semble faible, mais ces prédictions sont réellement plus puissantes puisqu'elles sont susceptibles de prédire le noyau protéique plus précisément. Une autre amélioration a été l'introduction de nouvelles technologies informatiques comme les réseaux neuronaux. De nombreuses méthodes de prédiction de structure secondaire (tableau 4) sont disponibles sur le réseau comme PHD (Rost et Sander, 1993), SSPRED (Mehta, 1995), PSSP (BCM; Smith *et al.* 1996),...

En pratique, il est recommandé d'utiliser le plus d'approches possibles et de combiner en un alignement leurs résultats avec les idées que l'on a de la conservation des résidus pour en retirer une prédiction "consensus" beaucoup plus précise et plus fiable.

### b) Reconnaissance de *fold*:

Même sans homologue de séquence dont la structure 3D est connue, il est possible de trouver un *fold* approprié pour la protéine cible parmi les structures 3D connues. En effet, il a été reconnu depuis longtemps que les protéines adoptent souvent des folds similaires, malgré l'absence de similarités de séquence ou de fonction significatives, et que la nature est apparemment restreinte à un nombre



limité de repliements protéiques. Pour beaucoup de protéines (~70%), il y aurait une structure appropriée dans les banques de données à partir de laquelle on pourrait construire un modèle (Orengo, 1994). Malheureusement, le manque de similarité de séquence fait que la similarité de structure n'est détectée que lorsque la structure 3D de la séquence cible est déterminée. Les méthodes de reconnaissance de *fold*s protéiques tentent de détecter des similarités entre structures 3D qui ne sont accompagnées par aucune similarité de séquence significative. Il y a beaucoup d'approches différentes, mais le thème unificateur est d'essayer de trouver des *fold*s qui sont compatibles avec une séquence particulière. Contrairement à la seule comparaison de séquences, ces méthodes profitent de l'information supplémentaire rendue disponible par l'information de la structure 3D. En effet, le problème est pris à l'envers: plutôt que de prédire comment une protéine se replie, elles prédisent comment un *fold* connu convient à la séquence protéique.

Les méthodes de reconnaissance de *fold* peuvent être classées en deux catégories:

- **Threading:** (potential-based threading )

Pratiquement, ces méthodes sont basées sur des algorithmes qui "enfilent" (threading) la séquence d'une protéine sur la structure connue d'une autre, et qui cherchent un alignement correspondant à un modèle structural énergétiquement favorable. Comme elles sont basées sur un calcul d'énergie (potentiel empirique) plutôt que sur un calcul de distance à un niveau évolutif, les méthodes de threading élargissent la prédiction de structure à une plus large classe de séquences que dans le cas de modélisation comparative. On y retrouve des méthodes comme ProFIT (Sippl, 1993) ou THREADER (Miller et al., 1996).

- **Pseudo-threading:** (prediction-based threading)

Ces méthodes n'utilisent pas de potentiels empiriques. Elles détectent, puis 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, comme par exemple Topits (Rost, 1995).

Malgré les résultats initiaux prometteurs, les méthodes de reconnaissance ne sont pas toujours précises. En général, on observe au mieux ~50% d'exactitude: on a une chance sur deux d'avoir le *fold* correct placé au sommet de la liste. Mais bien que beaucoup de méthodes échouent à détecter le *fold* correct comme meilleur score, celui-ci est souvent trouvé dans les dix premiers scores.

Ensuite, le *fold* une fois prédit, il convient d'analyser les caractéristiques structurales des membres de sa famille, en analysant des banques de données telles CATH (protein structural classification) ou SCOP (Structural Classification Of Proteins; Murzin *et al.*, 1995). Les structures secondaires du noyau protéique,

comme celles comprises dans les tonneaux  $\beta$ , devraient être présentes dans le *fold*. Si les structures secondaires prédites ne peuvent être appariées avec ce qu'on pense être le noyau du *fold* protéique, alors la prédiction peut être fausse! (N'oublions pas que la prédiction de structure secondaire peut aussi contenir des erreurs!)

Enfin, la dernière étape est l'alignement de la séquence à la structure tertiaire.

Aucune méthode spécifique n'a à ce jour été développée pour aligner séquence et structure, bien qu'une méthode logique pourrait inclure:

- s'assurer que les résidus prédits comme enfouis ou exposés au solvant s'alignent à ceux connus être enfouis ou exposés dans la structure *template*

On peut prédire l'accessibilité des résidus manuellement, ou en utilisant des serveurs automatisés comme PHD (Rost et Sander, 1993).

- s'assurer que les patterns de liaison hydrogène ne sont pas disruptés dans les structures composées de feuillets  $\beta$ . Ceci peut se faire grâce à un pseudo-potential mesurant l'interaction résidu-résidu et les contacts entre brins  $\beta$  pour n'importe quelle séquence protéique (Hubbard et al., 1996)

- essayer de conserver les propriétés des résidus des structures connues (taille, polarité, hydrophobicité) le mieux possible pour les structures inconnues.

### c) Prédiction *ab initio*:

La prédiction *ab initio* est le calcul d'une conformation d'énergie minimale à partir de potentiels physicochimiques sans aucune donnée structurale de départ. Elle implique toutes les méthodes qui ne se servent à aucun moment de l'information contenue dans une banque de donnée quelconque.

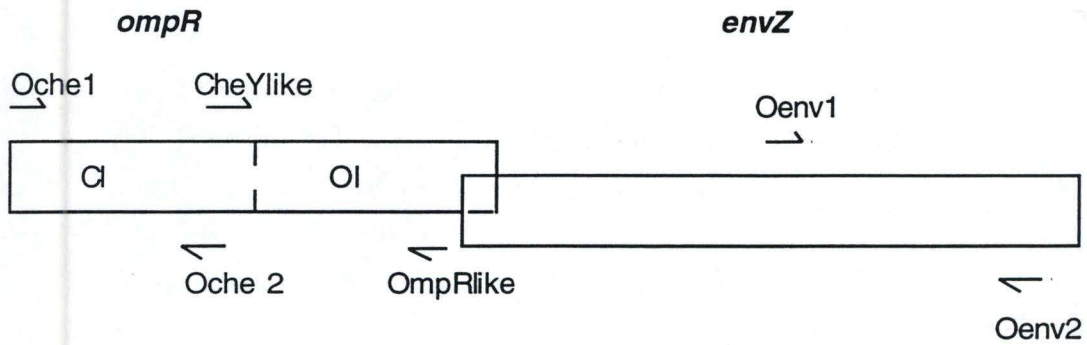
Actuellement, la prédiction *ab initio* n'est encore qu'un rêve inaccessible, car bien qu'il y ait eu beaucoup de progrès dans cette direction ces dernières années, aucune méthode n'a encore montré de résultats convaincants pour des molécules plus importantes que des peptides de quelques acides aminés. On retrouve dans cette catégorie des méthodes comme celle de Brasseur (1995), qui utilise une équation semi-empirique décrivant l'énergie interne de la molécule (énergies de torsion, van der Waals et électrostatiques) ainsi que l'énergie libre de solvation pour calculer *ab initio* la conformation d'un petit polypeptide. D'autres exemples sont la dynamique moléculaire ou le Monte Carlo (Holm et Sander, 1992) qui sont des méthodes essayant de contourner le problème des minima multiples.

## 6. Conclusion.

Toutes les méthodes citées ci-dessus ne sont pas encore à la pointe de la perfection et le domaine de la prédiction de structure 3D de protéines est encore loin d'être totalement exploré. Beaucoup d'améliorations peuvent encore être tentées en reconsidérant les paramètres, ce qui pourra se faire quand le nombre de structures et de folds connus aura augmenté, ou grâce aux alignements multiples qui pourraient être adaptés pour calculer des potentiels de contacts plus spécifiques.

Cependant, il est possible actuellement d'obtenir des modèles assez précis de protéines, qui ont déjà servi dans plusieurs domaines: conception d'inhibiteurs de la protéase de HIV, de médicaments anti-cancéreux (Balaji et al. 1985), d'antihypertenseurs grâce au modèle de la rénine humaine (Hutchins et Greer, 1991),... Ces modèles permettent aussi de planifier des expériences de mutagenèse dirigées.

Opéron *ompB*



**Figure 34:** Représentation schématique de l'opéron *ompB* et des paires d'oligonucléotides dégénérées Oche1/2, CheYlike/OmpRlike et Oenv1/2. CI (CheYlike) est le domaine amino-terminal receveur et OI (OmpRlike) est le domaine carboxy-terminal effecteur de OmpR.

# Résultats et Discussion

## Partie 1: Clonage d'une paire senseur-régulateur chez *B. abortus*.

Le premier objectif de ce mémoire était de cloner, chez *Brucella* spp. un opéron encodant un couple senseur-régulateur, équivalent à l'opéron *ompB* d'*E. coli* encodant le régulateur OmpR et le senseur EnvZ. Ce système OmpR-EnvZ a été choisi à priori, puisqu'étant la référence chez *E. coli*.

Pour ce faire, trois méthodes ont été menées en parallèle:

- des amplifications par PCR sur *B. melitensis* de fragments de gènes homologues à *ompB*,
- la création et le criblage d'une sous-banque génomique de *B. melitensis* 16M,
- le criblage d'une banque génomique de *B. abortus*.

### I. Amplification PCR de fragments de gènes homologues à *ompB*:

Différentes amorces PCR ont été déterminées dans trois domaines de l'opéron *ompB* (encodant les domaines receveur d'OmpR, effecteur d'OmpR et transmetteur d'EnvZ) ayant pour but l'obtention de sondes spécifiques de *B. abortus*. La position de chaque amorce sur l'opéron *ompB* d'*E. coli* est représentée sur la figure 34.

#### A. Tentative d'amplification d'un domaine homologue au domaine effecteur de OmpR.

La première paire d'amorces PCR sur le domaine effecteur d'OmpR a été déterminée afin de tester si réellement un gène similaire à l'OmpR d'*E. coli* était présent chez *Brucella* spp.

Une série de séquences similaires de la protéine OmpR d'*E. coli* a été obtenue suite à la recherche dans les banques de données par le serveur BLAST (Altschul *et al.*, 1990), correspondant à l'OmpR et à d'autres régulateurs de différentes espèces d'organismes, de mitochondries et de

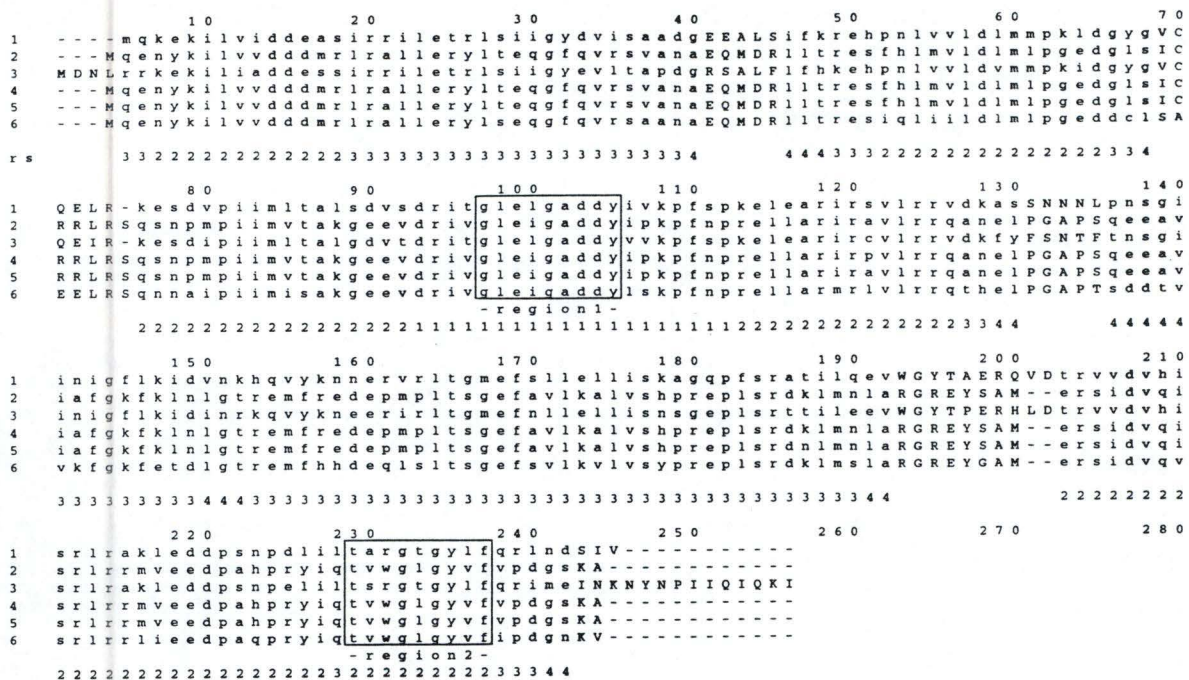


Figure 35: Aligement multiple des proteines OmpR de 1) *Porphyridium aeruginum*, 2) *Escherichia coli*, 3) *Cyanidium caldarium*, 4) *Salmonella typhimurium*, 5) *Salmonella typhi*, and 6) *Xenorhabdus nematophilus* d'après le serveur Matchbox. rs= score de confiance; Les régions choisies pour la détermination des oligonucléotides sont encadrées.

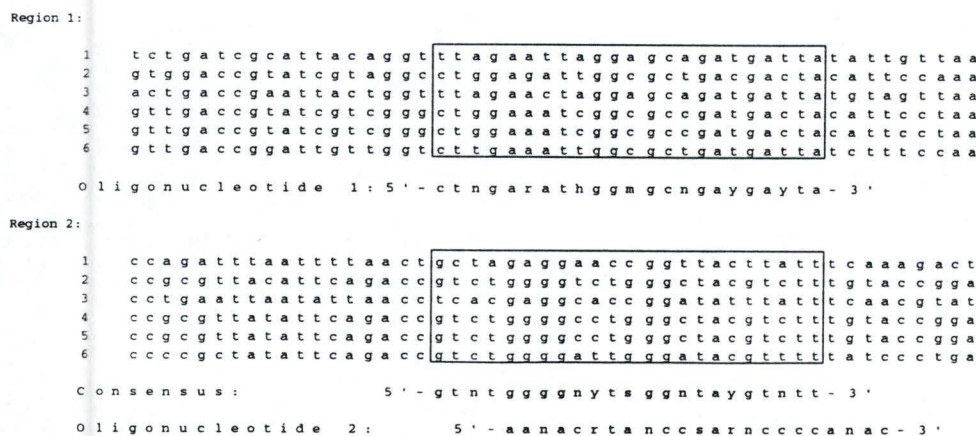


Figure 36: Aligement nucléotidique des régions déterminées (Figure 35). Les bases utilisées pour la détermination des oligonucléotides sont encadrées. La numérotation des séquences est identique à celle de la Figure 35. Les abréviations utilisées dans cette figure ainsi que dans les figures 38 et 39 sont: n=a,c,g,t; r=a,g; y=c,t; m=a,c; k=t,g; s=c,g; w=a,t; h=a,t,c; b=t,c,g; d=a,t,g; v=a,c,g

chloroplastes. Seules les séquences des OmpRs (*E. coli*, *Salmonella typhimurium*, *Salmonella typhi*, *Xenorhabdus nematophilus*, *Cyanidium caldarium* et *Porphyridium aerugineum*) ont été retenues et alignées grâce au logiciel Matchbox (les paramètres étant:  $w=7$ ,  $cutoff=400$ , matrice DFK-phys et balayage maximal).

Le résultat de cet alignement est montré à la figure 35.

Cet alignement a permis de déterminer les zones de la séquence d'OmpR où les résidus sont généralement bien conservés. Parmi ces zones, deux régions de 9 résidus ont été retenues selon les critères suivants:

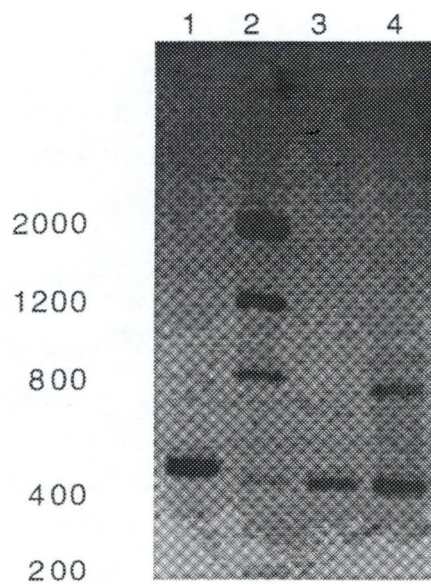
- le score de confiance obtenu par Matchbox pour ces régions doit être le plus bas possible (confiance élevée),
- elles doivent avoir un contenu en leucine, sérine et arginine le plus faible possible. En effet, ces trois résidus présentent le plus haut niveau de dégénérescence des codons (6 codons pour un acide aminé),
- elles doivent si possible être localisées dans le domaine effecteur, plus spécifique d'OmpR.

Une seule région se conforme à ces critères: celle qui se situe à l'extrémité carboxy-terminale du domaine effecteur. Pour la deuxième amorce, aucune zone du domaine effecteur ne répondant à nos critères, celle-ci fut choisie à la fin du domaine receveur et fut nommée "CheYlike". Par opposition, l'amorce du domaine effecteur fut appelée "OmpRlike".

Ensuite, les séquences nucléotidiques des amorces ont du être déterminées sur base de l'alignement des séquences nucléotidiques des deux régions retenues dans les OmpR's en tenant compte de la dégénérescence du code génétique. Cet alignement est illustré à la figure 36.

Une amplification PCR sur l'ADN génomique de *S. typhimurium* a servi de contrôle positif (bande attendue: 418 bp) pour l'amplification sur l'ADN génomique de *B. melitensis* 16M. Le résultat de la PCR, après migration sur gel d'agarose 2%, est une amplification d'un fragment d'environ 400 bp chez *B. melitensis* et chez *B. abortus*, et du fragment de 418 bp attendu chez *S. typhimurium* (Fig. 37)

L'étape suivante a été de vérifier si le fragment amplifié pour le contrôle positif est bien l'OmpR de *S. typhimurium* et si les produits d'amplification PCR de *Brucella* spp. ont bien une séquence homologue aux séquences connues des OmpR's. Pour cela, les fragments PCR ont



**Figure 37:** Amplification PCR sur l'ADN génomique de *S. Typhimurium* (1) *B. melitensis* (3) et *B. abortus* (4) à partir des oligonucléotides dégénérés CheYlike/OmpRlike. Le marqueur utilisé est un ladder de 100 à 2000 bp.



été purifiés par extraction au phénol et précipités à l'éthanol avant d'être insérés dans la matrice pGEM-T (Promega). Les plasmides obtenus ont été nommés pBmo1 (*B. melitensis* OmpR) et pSto1 (*S. typhimurium* OmpR).

Le séquençage de ces plasmides a été réalisé à partir des amorces "Forward" et "Reverse" s'hybridant de part et d'autre du site multiple de clonage du plasmide pGEM-T. La séquence de l'insert de pSto1 correspond entièrement à la portion correspondante du gène OmpR de *S. typhimurium*. La séquence de l'insert de pBmo1 a été soumise à une recherche de similarité dans les banques de séquences par BLAST. Les protéines homologues présentant le plus haut degré de similarité avec la séquence déduite de l'insert de pBmo1 sont des régulateurs de transcription appartenant à la famille des systèmes de transduction du signal à deux composantes. Parmi les protéines les plus similaires, on retrouve l'OmpR de *S. typhimurium* et de *E. coli*.

Enfin, après amplification PCR et purification, l'insert du plasmide pBmo1 a été marqué à la fluorescéine par "Random priming labelling". C'est cette sonde qui a été utilisée lors des Southern blots et des criblages de (sous-) banques de *B. abortus* et de *B. melitensis*.

## B. Tentatives d'amplification de domaines homologues aux domaines receveur de OmpR et transmetteur de EnvZ.

Les oligonucléotides dégénérés suivants ont été définis selon les mêmes critères que les oligonucléotides "OmpRlike" et "CheYlike".

Le second alignement multiple a porté sur les protéines CheY de 6 espèces (*Listeria monocytogenes*, *Halobacterium salinarium*, *E. coli*, *Rhodobacter sphaeroïdes* et *S. typhimurium*) et sur les domaines N-terminaux d'OmpR's de 7 espèces (*Thermotoga maritima*, *Campylobacter jejuni*, *E. coli*, *Porphyridium aerugineum*, *Cyanidium caldarium*, *S. typhi*, *S. typhimurium* et *X. nematophilus*), afin de déterminer une paire d'amorces amplifiant le domaine receveur d'OmpR. Cet alignement ne contient que les protéines CheY et OmpR de façon à limiter les possibilités d'amplification par ces amorces, vu que tous les domaines receveurs des régulateurs de systèmes à deux composantes sont similaires et que plus de 50 systèmes à deux composantes sont suspectés chez *E. coli* (Swanson *et al.*, 1994). Les

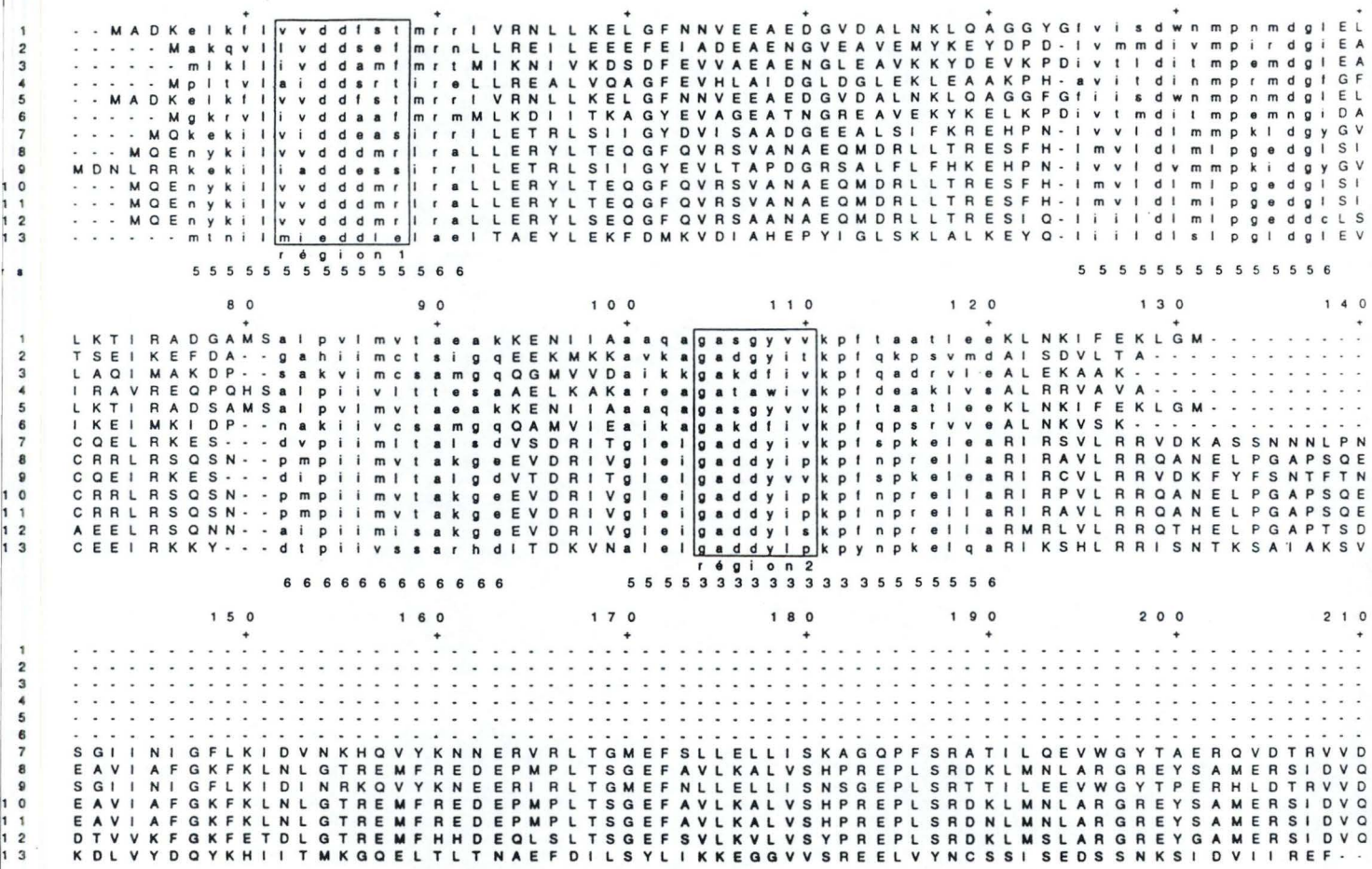


Figure 38a: Aligement multiple des protéines CheY de 1) *E. coli*, 2) *H. salinarium*, 3) *L. monocytogenes*, 4) *R. sphaeroides*, 5) *S. typhimurium* et OmpR de 6) *T. maritima*, 7) *P. aeruginum*, 8) *E. coli*, 9) *C. caldarium*, 10) *S. typhimurium*, 11) *S. typhi*, 12) *X. nematophilus* et 13) *C. jejuni* d'après le serveur Matchbox. rs= score de confiance. Les régions choisies pour la détermination des oligonucléotides sont encadrées.

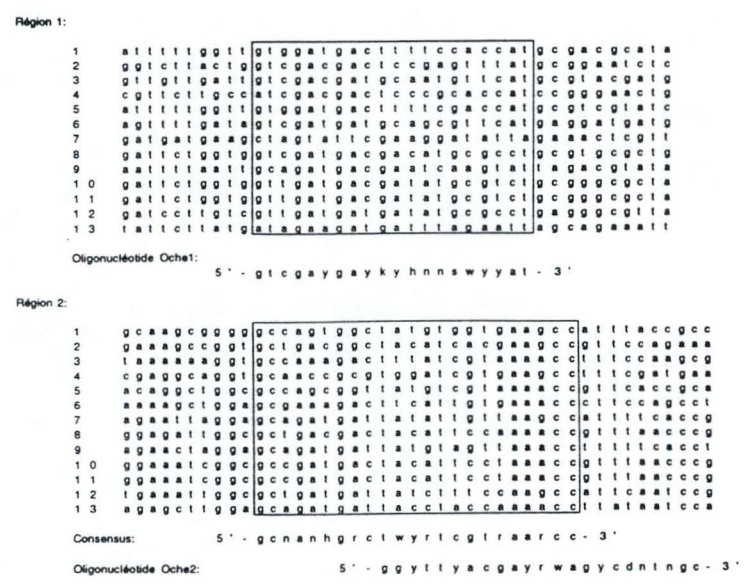


Figure 38b: Aligement nucléotidique des régions déterminées (Figure 38a). La légende est identique à celle de la figure 38a et les abbréviations à celles de la figure 36.



Figure 39a: Alignement multiple des protéines EnvZ de 1) *E. coli*, 2) *S. typhimurium*, 3) *X. nematophilus* d'après le serveur Matchbox rs= score de confiance. Les régions choisies pour la détermination des oligonucléotides sont entourées.

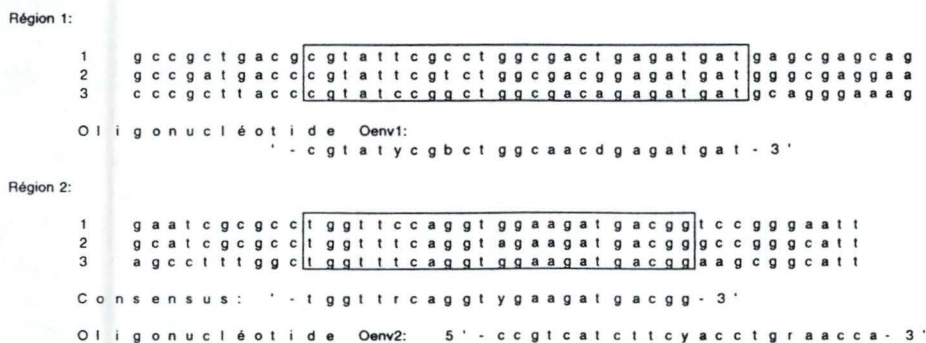


Figure 39b: Alignement nucléotidique des régions déterminées (Figure 39a). La légende est identique à celle de la figure 39a et les abréviations à celles de la figure 36.

oligonucléotides définis à partir de cet alignement ont été appelés Oche 1 et Oche 2. Les alignements protéiques et nucléotidiques sont montrés à la figure 38.

Afin de restreindre encore la probabilité d'amplifications multiples, une troisième paire d'oligonucléotides dégénérés a été déterminée sur base d'un alignement multiple des protéines EnvZ de *E. coli*, *S. typhimurium* et *X. nematophilus*. Ce dernier alignement de séquences (fig. 39) nous a permis de définir les oligonucléotides Oenv 1 et Oenv 2.

Des amplifications PCR ont été testées en utilisant ces 2 paires d'amorces, dans les conditions décrites au point I.B.2.i. du Matériel et Méthodes, avec de l'ADN génomique de *S. typhimurium* comme contrôle positif. Aucune amplification significative n'a été obtenue sur l'ADN génomique de *B. melitensis* à partir de ces amorces, ni sur l'ADN génomique de *S. typhimurium* à partir des oligonucléotides Oche1 et Oche2, même en variant les conditions d'amplification (quantité d'ADN de départ, nature de la polymérase, volume de réaction ainsi que les temps de dénaturation, d'hybridation des amorces et d'élongation).

## II. Identification par Southern blot d'un fragment de restriction de *B. melitensis* 16M susceptible de contenir l'opéron homologue à *ompB* en entier.

L'étape préliminaire à la création d'une (sous-)banque génomique est de restreindre le génome afin de cloner les fragments de la taille désirée dans le vecteur. Pour cela, l'ADN génomique de *B. melitensis* a été restreint par différentes enzymes (*Bam* HI, *Eco* RV, *Hind* III, *Pvu* I et *Sph* I) et la sélection de l'enzyme la plus adéquate s'est faite par Southern blot avec l'insert du plasmide pBmoI comme sonde. Cette analyse (fig. 40) montre une hybridation de la sonde sur une bande unique pour l'ADN restreint par les enzymes *Eco* RV (bande d'environ 4 Kb), *Hind* III (bande d'environ 1,5 Kb) ou *Bam* HI (bande d'environ 20 Kb). L'enzyme *Sph* I clive l'ADN génomique de *B. melitensis* dans l'opéron *ompB*. L'hybridation de la sonde sur les fragments *Sph* I montrent plusieurs bandes. Cette enzyme est donc à proscrire dans notre cas, puisqu'il ne nous serait pas possible d'obtenir l'opéron homologue à *ompB* en entier.

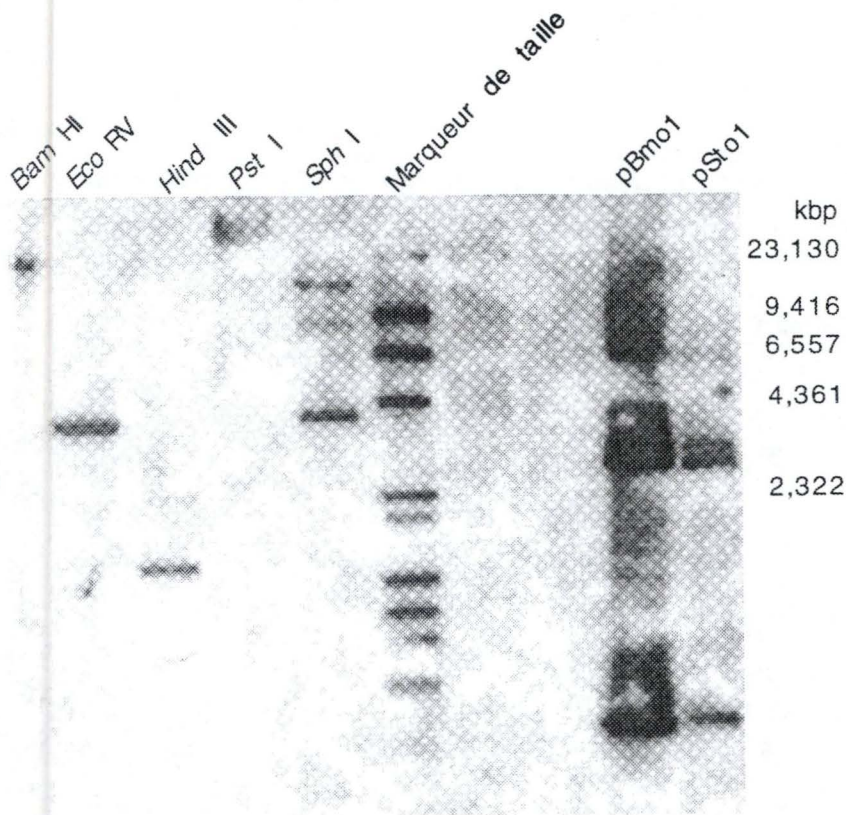


Figure 40: Southern blot de l'insert de pBmo1 sur l'ADN génomique de *B. melitensis* 16M restreint par les enzymes *Bam* HI, *Eco* RV, *Hind* III, *Pst* I et *Sph* I. Les deux pistes de droite contiennent les plasmides de contrôle pBmo1 et pSto1.

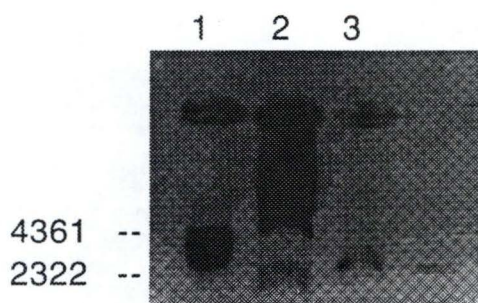


Figure 41: Gel d'agarose ayant servi à la quantification:

1. des fragments d'ADN génomique de *B. melitensis* 16M restreint *Eco* RV de taille comprise entre +/- 3000 et 4300 bp après récupération par papiers.
3. du plasmide pUC19 restreint par *Sma* I, déphosphorylé, reliqué sur lui même et récupéré par papiers.

La piste 2 correspond au marqueur de taille  $\lambda$  DNA/*Hind* III.

L'ORF de l'opéron *ompB* faisant environ 2070 paires de bases chez *E. coli*, c'est le fragment *Eco RV* qui est le plus susceptible de contenir l'ORF entière de l'opéron homologue chez *B. melitensis*. Une sous-banque contenant des fragments *Eco RV* de taille proche de 4 Kb a donc été réalisée.

### III. Criblage d'une sous-banque génomique de *B. melitensis* 16M:

#### A. Création de la sous-banque

##### 1. Récupération des fragments aux alentours de 4 Kb sur gel d'agarose.

De l'ADN génomique de *B. melitensis* 16M a été restreint par *Eco RV*, 1 heure à 37°C, dans un volume total de 200 µl, dont 5 µl ont été mis sur gel d'agarose 1% pour vérifier la qualité de la restriction:

30 µl d'ADN génomique (plus de 30 µg)

20 µl de tampon B

140 µl H<sub>2</sub>O

10 µl *Eco RV*

La séparation des fragments génomiques des 195 µl restant du mélange de restriction s'est faite sur gel d'agarose 1% pendant 4 heures. Les fragments d'intérêt ont été récupérés par la technique des papiers.

Afin de récupérer les fragments d'une taille proche de 4 Kb, le gel a été incisé à deux endroits:

- au niveau de la bande du fragment de 4361 bp du marqueur de poids moléculaire  $\lambda$ , les papiers à ce niveau bloquant la migration des fragments de taille supérieure;

- un peu en dessous, environ à 3000 bp, les papiers à ce niveau retenant les fragments d'intérêt de taille comprise entre +/- 3000 et 4300 bp.

Après migration sur gel d'agarose, la quantité d'ADN génomique récupérée a été estimée à 350 ng (fig. 41).

## 2. Préparation du plasmide.

Le plasmide utilisé pour la création de la sous-banque est le pUC19. Celui-ci a été restreint avant d'y insérer les fragments d'ADN génomiques de *B. abortus* récupérés par papiers. Le pUC19 ne contenant pas de site de clivage *Eco* RV, la restriction s'est faite en *Sma* I, générant un plasmide linéaire à bouts francs.

La restriction a été réalisée pendant 1 heure à température ambiante dans un volume de 10  $\mu$ l:

4  $\mu$ l de pUC19 ( plus de 5  $\mu$ g)

1  $\mu$ l de tampon A

1  $\mu$ l *Sma* I

4  $\mu$ l d'H<sub>2</sub>O

La qualité de la restriction a été vérifiée par migration sur gel d'agarose.

Le plasmide restreint a ensuite été déphosphorylé, extrait au phénol et précipité à l'éthanol, en vue de la ligation. Pour tester l'efficacité de la déphosphorylation, le plasmide déphosphorylé a d'abord été incubé en présence de ligase, 1 heure à température ambiante. Seuls les plasmides non déphosphorylés peuvent se religuer sur eux-même. Les plasmides linéarisés (déphosphorylés) ont été séparés des plasmides reliqués sur eux-même par migration sur gel d'agarose (fig.41) et récupération par papiers. La quantité de plasmides récupérés a été estimée après migration sur gel d'agarose à 208ng.

## 3. Ligation.

Environ 200ng de plasmides restreints et récupérés ont été mis en présence d'une quantité un peu supérieure de fragments génomiques *Eco* RV (300 ng). Le rapport entre les quantités de fragments génomiques et plasmidiques dans les échantillons est équivalent à celui entre les molarités de ces fragments génomiques et plasmidiques, puisque les fragments génomiques et plasmides ont environ la même taille.

La ligation a été effectuée avec ces quantités d'ADN dans un volume final de 35  $\mu$ l, à température ambiante.

#### 4. Transformation.

Une première transformation-test de bactéries XL1-blue s'est faite avec +/- 70ng du mélange de ligation. Après étalement sur milieu LB-tétracycline-ampiciline, nous avons obtenu 17 colonies blanches pour environ 200 colonies bleues.

Le reste du mélange de ligation a été engagé dans trois transformations par électroporation et nous avons obtenu une centaine de colonies blanches pour environ dix fois plus de colonies bleues.

Ensuite, des *minipréparations* ont été effectuées sur les colonies blanches et quelques colonies bleues qui ont servi de contrôle négatif.

#### B. Criblage de la sous-banque par Southern blot.

Le criblage de notre sous banque "*EcoRV*" a été réalisé par Southern blot en utilisant comme sonde l'insert du pBmo1 marqué à la fluorescéine.

Les plasmides de la sous-banque purifiés par minipréparation ont été restreints simultanément par deux enzymes: *Eco RI* et *Bam HI*, qui clivent de part et d'autre du site de clivage *Sma I*, site d'insertion du fragment d'ADN génomique de *Brucella* dans le pUC19.

Le plasmide pBmo1 a été utilisé comme témoin positif. Ce plasmide a été restreint par les enzymes *Xba I* et *Pst I* clivant de part et d'autre du site d'insertion de l'insert, pour libérer celui-ci.

Après vérification de la qualité de la restriction, la séparation des fragments plasmidiques s'est faite par migration lente sur gel d'agarose 1%. Comme le marqueur de taille utilisé n'était pas marqué à la fluorescéine, la position de chacune des bandes du marqueur a été mesurée lors de l'exposition du gel sous UV, afin de pouvoir estimer la taille des fragments éventuellement visibles après impression du film RX.

Parmi les 83 colonies blanches testées, l'analyse des films RX après impression ne nous a pas permis d'identifier de bandes positives autres que celle de l'insert du pBmoI (contrôle positif), correspondant à des clones positifs. D'autres clones auraient encore pu être analysés, mais cette démarche a été interrompue à ce niveau vu que la troisième technique utilisée en parallèle a révélé deux clones positifs.



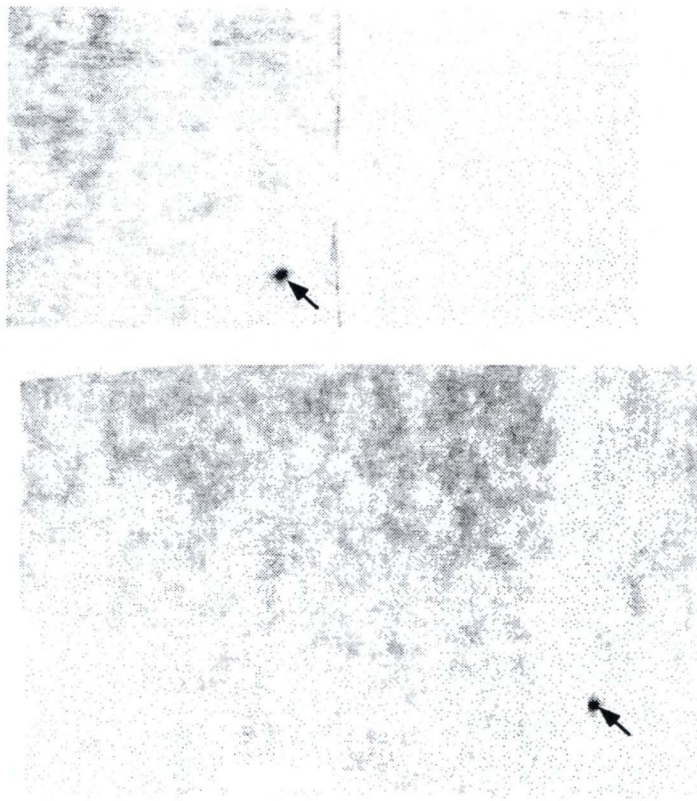


Figure 42: Criblage de la banque génomique de *B. melitensis* 16M en phage  $\lambda$ gt11 par l'insert du plasmide pBmo1.  
Les deux phages positifs sont indiqués par une flèche.

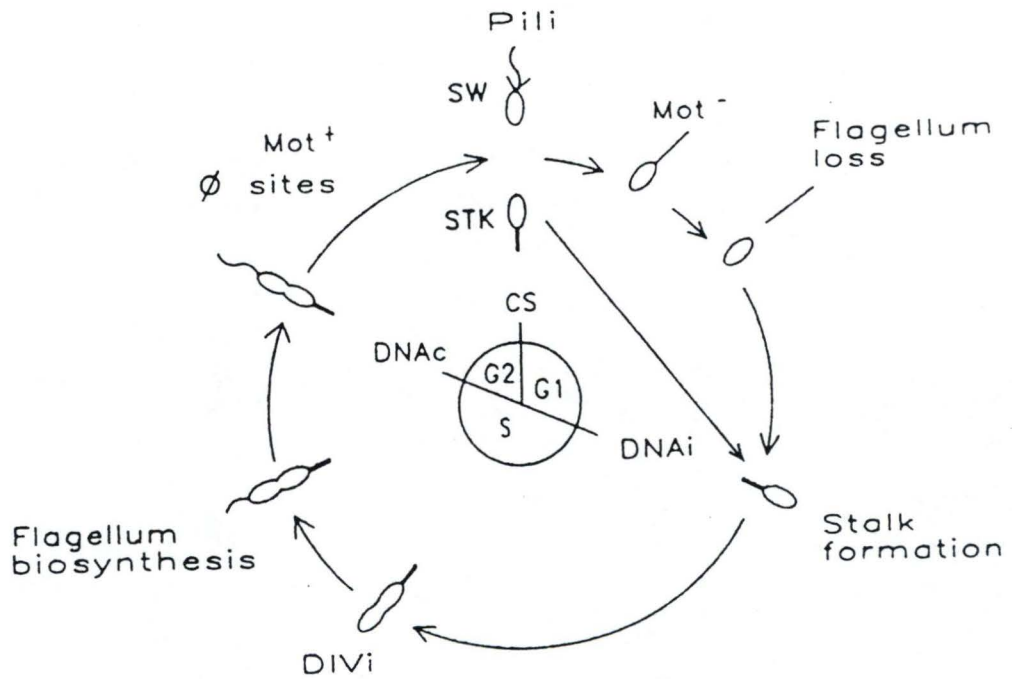
#### IV. Criblage de la banque génomique de *B. abortus*:

La troisième méthode menée en parallèle au laboratoire a été le criblage, avec l'insert du plasmide pBmo1 comme sonde, de la banque de *B. abortus* établie dans le phage  $\lambda$ gt11 (Tibor *et al.*, 1994). Lors du criblage, deux plages de lyse positives ont été identifiées (fig. 42), à partir desquelles des phages ont été isolés. Ces phages ont été soumis à plusieurs cycles de purification (prélèvement, infection, criblage) jusqu'à ce que toutes les plages de lyses criblées soient reconnues par la sonde.

L'ADN de ces deux phages purifiés a été préparé après lyse en culture liquide et précipitation au polyéthylène glycol. L'ADN du premier phage a été restreint par l'enzyme *Eco* RI et l'insert a été cloné dans le vecteur pBluescript KS- (Stratagène) linéarisé par la même enzyme. Le plasmide résultant a été appelé pOlb1. Une réaction PCR à partir des amorces Forward et Reverse  $\lambda$ gt11 a été réalisée sur l'ADN du premier phage. L'insert amplifié d'une taille approximative de 1,6Kb a été purifié et séquencé à partir des amorces Forward et Reverse  $\lambda$ gt11 et de l'oligonucléotide OmpRlike, en utilisant un séquenceur automatique ABI PRISM 377. L'insert du deuxième phage n'a pas été séquencé car il possède le même profil de restriction que le premier.

Des séquences obtenues à partir de ces amorces, une séquence de 1515 bp a été reconstituée, contenant une seule phase ouverte de lecture potentielle d'une taille de 696 bp qui encoderait une protéine de 232 acides aminés. La séquence nucléotidique et la séquence protéique déduite de cette ORF ont chacune été soumises à une procédure de recherche par similarité dans les banques de séquences par BLAST. Toutes les séquences homologues répertoriées sont des régulateurs de réponse appartenant aux systèmes à deux composantes. La séquence la plus homologue est une protéine régulatrice de la bactérie *Caulobacter crescentus*, CtrA, partageant 80,5% d'identité sur la longueur totale de la séquence protéique.

Il est à noter que lors de la recherche de similarité dans les banques de séquences pour l'insert du plasmide pBmo1 (389 bp), la séquence de la protéine CtrA n'était pas encore disponible. La même recherche effectuée actuellement nous renseigne CtrA comme la protéine la plus homologue. CtrA ferait partie de la sous-famille OmpR des régulateurs de réponse (Quon *et al.*, 1996)



**Figure 43:** Cycle cellulaire de *Caulobacter crescentus* et séquence des événements développementaux. Deux cellules-filles distinctes sont produites lors de la division: une cellule nageuse (SW) et une cellule pédonculée (STK). Les événements développementaux incluent l'assemblage des pili (Pili), la perte de motilité (Mot<sup>-</sup>), l'initiation de la division (DIVi), l'activation des récepteurs polaires pour les bactériophages  $\phi$ LC72 ( $\phi$ sites), le gain de motilité (Mot<sup>+</sup>) et la séparation cellulaire (CS); (Newton et Otha, 1990).

## V. Discussion

Nous avons cloné chez *B. abortus* un homologue du gène *ctrA* de *C. crescentus*.

Arrivés à ce point des résultats, un éclaircissement à propos de *C. crescentus* et de son cycle cellulaire complexe est indispensable.

*C. crescentus* est une bactérie Gram-négative aquatique et nécrophage, vivant dans les étangs et les ruisseaux, et appartenant à la classe des  $\alpha$ 2-Protéobactériaceae comme *Brucella* spp. Elle est caractérisée par un cycle cellulaire inhabituel. Chaque division cellulaire donne naissance à deux types cellulaires différents qui, bien qu'ils héritent de génomes identiques, expriment des programmes génétiques différents (Degnen et Newton, 1972; fig.43):

- une cellule pédonculée immobile possédant une excroissance de la membrane cellulaire à un de ses pôles;
- une cellule nageuse sensiblement plus petite, portant un flagelle et des pili à sa base.

La cellule pédonculée commence immédiatement à répliquer son chromosome et se prépare pour la division suivante. A mi-chemin de la réplication chromosomique, la différenciation débute au pôle opposé du pédoncule, aboutissant à la synthèse d'un flagelle, de pili et de récepteurs membranaires reconnus par des phages à ADN. Juste avant la division, le flagelle commence à tourner et la cellule nageuse peut se détacher de la cellule pédonculée et se déplacer.

La cellule nageuse est immature et doit subir un processus de maturation avant de répliquer son chromosome et de se diviser. Après une période de mobilité, le flagelle est perdu et une tige est synthétisée à la place. Cette nouvelle cellule pédonculée est maintenant capable de se diviser.

Donc, le nouveau pôle formé subit deux évènements de différenciation:

- premièrement, en tant que partie de la cellule prédivisionnelle, il assemble un flagelle, des pili, et des récepteurs pour les bactériophages;
- ensuite, en tant que pôle différencié de la cellule nageuse, il doit perdre le premier "set" d'organelles et le remplacer par un pédoncule.

Ces transitions se produisent invariablement durant le cycle cellulaire et sont sujettes à des contrôles ponctuels dépendants du cycle cellulaire (Huguenel et Newton, 1982), ce qui suggère qu'un système régulateur

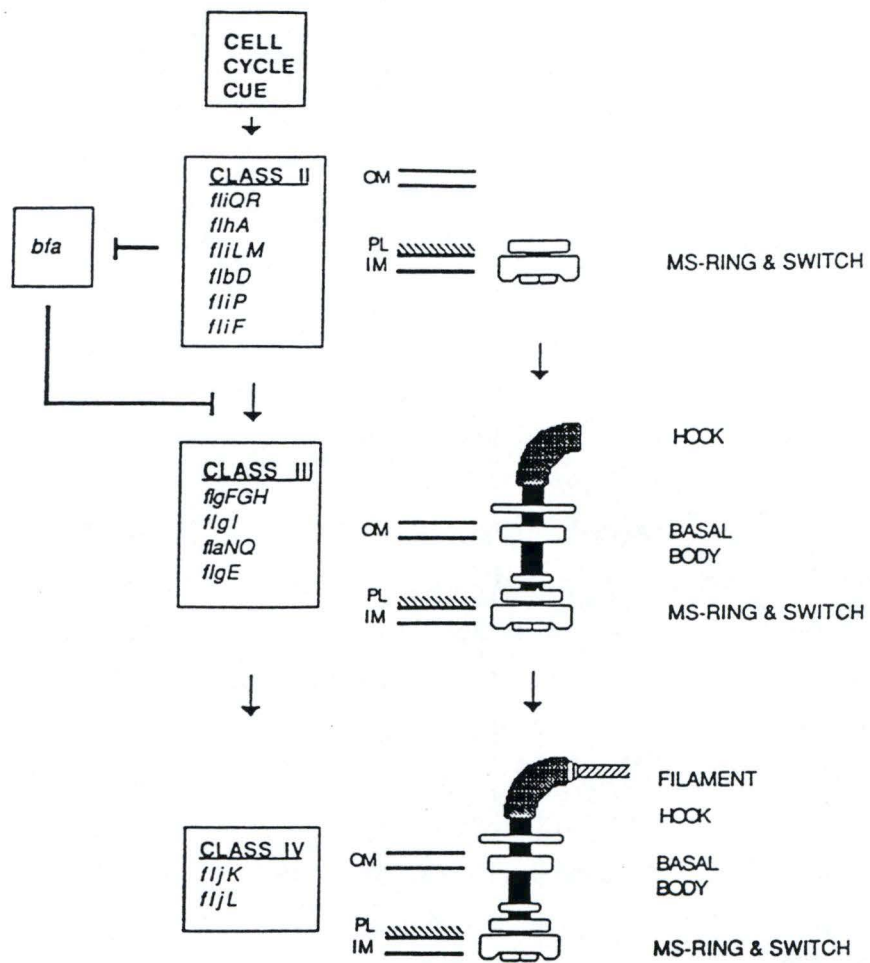


Figure 44: Hierarchie de la régulation flagellaire (Gober et Marques, 1995).

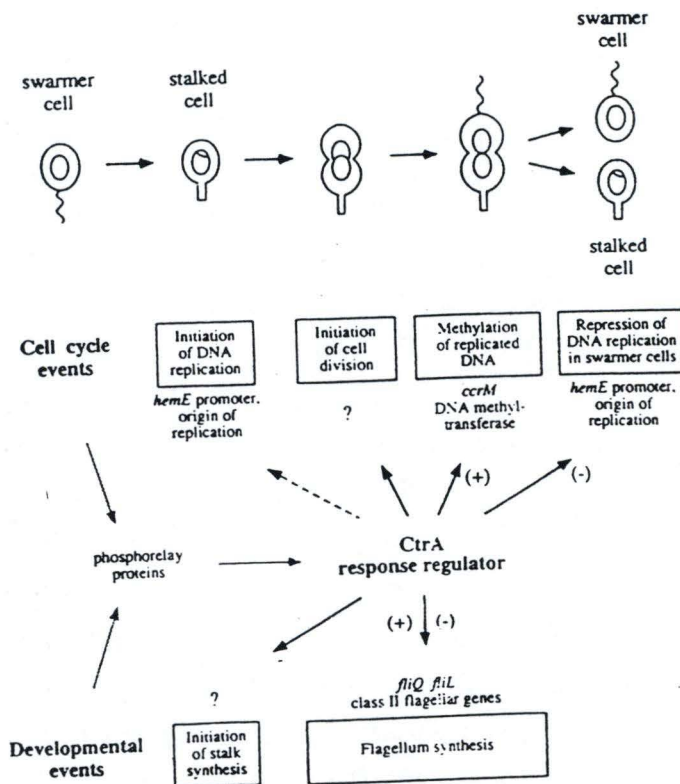


Figure 45: Rôles de CtrA dans la régulation du cycle cellulaire et du développement chez *Caulobacter crescentus*.

unique puisse contrôler aussi bien les développements morphologiques que le cycle cellulaire (Quon *et al.*, 1996).

La biogenèse du flagelle est l'aspect le plus étudié dans le développement morphologique de *Caulobacter* (Gober et Marques, 1995). Plus de 40 gènes impliqués dans ce processus ont été identifiés et la transcription de beaucoup d'entre eux est contrôlée temporellement. Ces gènes sont organisés d'une manière hiérarchique et sont répartis en quatre classes (de I à IV): il faut que les gènes d'une classe supérieure soient exprimés pour que les gènes de la classe inférieure le soient à leur tour (fig.44). Les gènes de classe II, qui sont presque au sommet de la hiérarchie, sont exprimés tôt dans le cycle cellulaire et encodent les composants les plus internes du corpuscule basal du flagelle ("MS ring") et les protéines du "moteur flagellaire" ("switch complex"). L'expression de tous les gènes de classe II est requise pour que l'expression des gènes de classes III et IV puisse avoir lieu. Les gènes de classe III encodent les composants externes du corpuscule basal et du coude, et leur expression est requise pour l'expression des gènes de classe IV, dont les produits sont les flagellines *fljK* et *fljL*.

Plusieurs équipes ont essayé de rechercher ce qui déterminait la succession temporelle des événements développementaux du cycle cellulaire, entre autre l'équipe de Lucille Shapiro (Quon *et al.*, 1996) qui a isolé une mutation dans le gène *ctrA* (cell cycle transcriptional activator) affectant la régulation négative du promoteur de classe II, *fliQ*.

Le gène *ctrA* encode un régulateur de réponse essentiel, agissant directement au niveau des promoteurs de plusieurs gènes impliqués dans le développement du flagelle ainsi que dans de nombreux événements liés au cycle cellulaire et à la méthylation de l'ADN (fig.45). CtrA contient un domaine carboxy-terminal liant l'ADN et reconnaissant des séquences nucléiques présentes dans des promoteurs de gènes flagellaires de classe II. CtrA contient tous les acides aminés conservés qui définissent les membres de la superfamille des régulateurs, dont le site de phosphorylation, l'aspartate 51, et d'autres résidus de la poche acide. CtrA appartiendrait à la sous-famille d'OmpR, car il est similaire à OmpR sur toute sa longueur: 29% d'identité au total se partageant en 33% d'identité pour le domaine amino-terminal et 24% pour le domaine carboxy-terminal. La mutation isolée par Shapiro et ses collègues (Quon *et al.* 1996) est une substitution de la thréonine 170 en isoleucine. Ce résidu est conservé soit comme une thréonine, soit comme une sérine

dans la majorité des membres de la sous-famille d'OmpR (Quon *et al.*, 1996).

D'autres membres du "système à deux composantes" intervenant dans la régulation du cycle cellulaire ou de la biosynthèse du flagelle ont été isolés, comme les senseurs PleC et DivJ, et les régulateurs DivK et PleD; mais l'organisation de la voie de transduction du signal est encore ignorée.

Notre objectif premier de cloner un homologue de l'opéron *ompB* de *E. coli* est pratiquement atteint. Nous avons réussi à cloner un homologue de OmpR d'*E. coli*, faisant partie des systèmes à deux composantes, bien qu'aucun senseur ne lui soit associé en opéron.

De plus, nous avons mis en évidence la présence chez *B. abortus* d'un régulateur similaire au régulateur transcriptionnel contrôlant le cycle cellulaire et la synthèse flagellaire chez *C. crescentus*.

## Partie 2: Prédiction de la structure tridimensionnelle du régulateur cloné

L'approche que nous avons suivie est résumée au tableau 1, dans le Matériel et Méthodes.

### I. Recherche d'homologues dans les banques de séquences.

Après séquençage du clone positif, une première recherche superficielle dans les banques de séquences avait déjà pu nous renseigner sur l'identification de notre séquence, que nous avons naturellement appelée BaCtrA (CtrA de *B. abortus*), et sur son appartenance à un système de transduction du signal à deux composantes. Une recherche plus approfondie a été menée, via le serveur BLAST dans une banque de séquences non redondantes incluant les banques SwissProt, Genbank, PIR et SPupdate afin d'obtenir le plus possible de séquences similaires en vue de l'alignement multiple.

Parmi les 230995 séquences présentes dans la banque, plus de cent montrent une similarité significative avec BaCtrA. Toutes ces séquences homologues sont des régulateurs faisant partie de systèmes à deux composantes. Parmi celles-ci une première sélection arbitraire a été faite, en ne gardant qu'une septantaine séquences dont le score BLAST était supérieur à 100.

Lorsqu'une recherche BLAST est menée dans la banque PDB seule, la séquence la plus similaire est celle de CheY, pour laquelle la structure 3-D a été déterminée sous plusieurs formes (forme liant le  $Mg^{++}$ , forme apo, formes mutantes). Cependant, CheY se situe parmi les protéines dont le score BLAST est inférieur à 100, sans doute vu le fait qu'elle n'est homologue à BaCtrA que sur la moitié de la séquence de BaCtrA (rappelons que CheY ne possède que le domaine receveur amino-terminal, alors que BaCtrA possède en plus un domaine effecteur carboxy-terminal).



```

      10      20      30      40      50      60      70
      +      +      +      +      +      +      +
1  -----mrvllieddsaiAQSIEMLKSESFNVYTTDLGEEGIDLGLKYDYD--iilldlnlpmmsgye
2  -----mrvllieddsatAQTIELMLKSEGFNVYTTDLGEEGVDLGLKIYDYD--lilldlnlpmmsgid
3  -----mkilliveddtllLQGLILAAQTIEGYACDVTVTARMABQSLQLEAGHYS--lvvldlglpdedglh
4  -----mkillvaedepktGIYLLQGLREAGFNVDVVTGTDVAVDQALNEAYD--lilldvmpglgdgwe
5  -----mrvlvvednallRHHLKVQIQDAGHQVDDAEDAKEYDYLLNEHIPD--iaivdglpdedgls
6  -----MmrvlvvednallRHHLKVQLQDSGHQVDAEDAKEYDYLLNEHLPD--iaivdglpdedgls
7  -----mrilvveddvnlnRQLADTLKEAGYVVDQAFDGEEGHFLGDTEPYD--ailldglpemdgvT
8  -----mkillvedddvlIKVLTKNLTTHHYIVDVVKDGEWGTYGSTFEYD--liildimlpnldgis
9  MIDMDEKlriIlceddenlGMLLREYLQAKGYSAELYPDGEAGFKAFLLKNKYD--lcvfdvmpkkgdft
10 MIDMDEKlriIlceddenlGMLLREYLQAKGYSAELYPDGEAGFKAFLLKNKYD--lcvfdvmpkkgdft
11 -----MikillveddglSNSVDFLDDFADVMQVFDGEEGLYEAESGVYD---lilldmlpeknqfg
12 -----mkillveddetdGMAIKKVLVSEKYVVDWVTDGSQAWDYLENQWTEYTLaiavdwllpglsgle
13 -----mrlllaednrelAHWLEKALVQNGFAVDCVFDGLAADHLLHSEMYA--lavldinmpgmdgle
14 -----mrillieddnliGNGLQIGLTKLGFVDFWFTDGTGMAALTSAPYD--avvldltlpkldgle

      111111111234                                4444444444444444

      80      90      100      110      120      130      140
      +      +      +      +      +      +      +
1  vLRTLRLSKVktpililsgmaGIEDKVRglgfgaddymtkpfhkdeliarihavrrrskGHAQSVITITGD
2  vLRTLRLVAKIntpimilsgssEIDTKVktfaggaddymtkpfhkdemiarihavrrrskGHAQSVIKITGD
3  fLARIRQKkytlpvliltardTLTDKIagldvgaddylvkpfaleelharirallrrhnNQGESEILVGN
4  vLRRFRTAGQYtvpvllftardGVDDRVkglelgaddylvkpfalsellarvrtllrrgsSLQVQTSGLIG
5  lIRRWRSSNDVslpilvltareSWQDKVevlsagaddylvkpfhieevmarmqalmrnsGLASQVVISLPP
6  lIRRWRSSDVslpvlvltareGWQDKVevlssgaddylvkpfhieevmarmqalmrnsGLASQVINTPP
7  vLEKWRGAGRgvpvliltardRWSDKVagidagaddylvkpfhvevlariralirraaGHSSEIICGP
8  lCKELREKGItpvlimtakeSLDDKGhgfelgaddylvkpfleelkmriqallkrsgKFNENLTYGN
9  lAQEVRAANAeipiifltaktLKEDILegfkiagaddytikpfsmeeltfrieailrrvrGKKNKESNIYK
10 lAQEVRAANAeipiifltaktLKEDILegfkiagaddytikpfsmeeltfyieailrrvrGKKNKESNIYK
11 vLKELREKGItpvlimtakeSLDDKGhgfelgaddylvkpfleelkmriqallkrsgKFNENLTYGN
12 lCQKRLTQGNslpvlmltalgEPENRVegldagaddylvkpfvmaellarlralqrspQFQPQLTLGN
13 vVQRLRKRQQtlpvllltarsAVADRVkglngvaddylvkpfleeldarlrallrrsaGQVHEVQQLGE
14 vLQWRSSNHQdvpvlliltardTLDERVkgllqsgaddylvkpfalaevaarlqalirrryGYHHSVIEQAG

4          3333333333          411111111111111111113333444444444444

      150      160      170      180      190      200      210
      +      +      +      +      +      +      +
1  LVVNLDAKTVEVAGQRVH-----ltgkeyqmlELLSLRKGTTLTKEMFLNHLVGGMDEPELKIIDVF
2  IVVNLDAKTVEVNGNRVH-----ltgkeyqmlELLSLRKGTTLTKEMFLNHLVGGMDEPELKIIDVF
3  LTLNMRGRQVWVGGEELI-----ltpkeyallSRLMLKAGSPVHREILYNDIYNWDNPESTNTLEVH
4  DLQVDLLKRRATRGKRIE-----ltakefallELLMRRQGEVLSKSLIASQVWDMNFDSDTINVEVA
5  FQVDLSRRELSINDEVIK-----ltafeytimETLIRNNGKVSKDSLMLQLYPDAELRESHTIDVL
6  FQVDLSRRELSVNEEVIK-----ltafeytimETLIRNNGKVSKDSLMLQLYPDAELRESHTIDVL
7  VRLDTKSSKATVNGTTLK-----ltsheyrlLAYLMHMHGVEVSRTELVEHMYDQDFDRDSNTIEVF
8  LLLNPNSTCEVTYNGCPLN-----littmeydlleLLLRNCQHVSSEELLDKLWSEDFPSEATVRSH
9  IGKFTFDQKQILAIAGDKQTK----littkesellGLLCAHANEILQRDFALKTIWIDDNYFNARSMDVY
10 IGKFTFDQKQILAIAGDKQTK----littkesellGLLCAHANEILQRDFALKTIWIDDNYFNARSMDVY
11 IVVNLSINTVIVVEDTPE-----llgkefdllVYFLQONQNVILPKTQIFDRLWGFSDTITISVVEVY
12 FSLDPSNNLSVITISEPLNLERQEIALtvrefqifQYLMQNPERIISGSKIRQQLWDLDEEPMNSVAAQ
13 LIFHDEGYFLLQGGPLA-----ltpregallTVLMYRRTRPVSRQQLFQVFLNDEVSPESIELEY
14 VKLDQNRQSVWLNQNPIS-----ltsreykllELFMLNKDRVLSRSSIEEKLSSWDEEISSGALDVH

      4444444444

      220      230      240      250      260      270      280
      +      +      +      +      +      +      +
1  ICKLRKKLDAVSGNQSYietvwgrgyvlREPDAEMRESA-----
2  ICKLRKKLAASAHGKHHietvwgrgyvlRDPNEQVNAA-----
3  IHNLRDKVGKAR-----irtvrgfgymLVANEEN-----
4  IRRLRAKIDDDFEVKL-lhtcrmgmyleAQDEG-----
5  MGRLRKKIQAQYPQEV-ittvrgqgyfELR-----
6  MGRLRKKIQAQYPHDV-ittvrgqgyfELR-----
7  VGRRLRKKMGVDL-----ietvrglgyriQAPKHAN-----
8  VRRLRHLKVAAGAPHDFiatmhgrgyylKAPSTEEVNNLSVTPENNShSAAVIVKSRETESDRPQHLI...
9  ITKLRHLKDDDSIE--iinihgkyklITPEPES-----
10 ITKLRHLKDDDSIE--iinihgkyklITPEPES-----
11 VSKVRKKLKGTTFAEN-lqtlrsvgyllKDVQ-----
12 MRLIRRKLAAQNCPCP-iktvpqgyrftLSP-----
13 ITHRLRKLQGSQDVR---ittlrglgyvlERGDEVG-----
14 IYNLRQKLGKQF-----irtvhggyalGQVEK-----

      444444444444

```

Figure 46: Regions prédites comme conservées par Matchbox pour les 14 séquences du groupe 1A.

1	232	ctra	2	231	caulo	3	222	basr	4	227	copr	5	223	phop
6	224	phops	7	223	popp	8	632	rcac	9	236	rpry	10	236	rpryb
11	224	strep	12	234	sycy2	13	224	tstd	14	221	ygix			

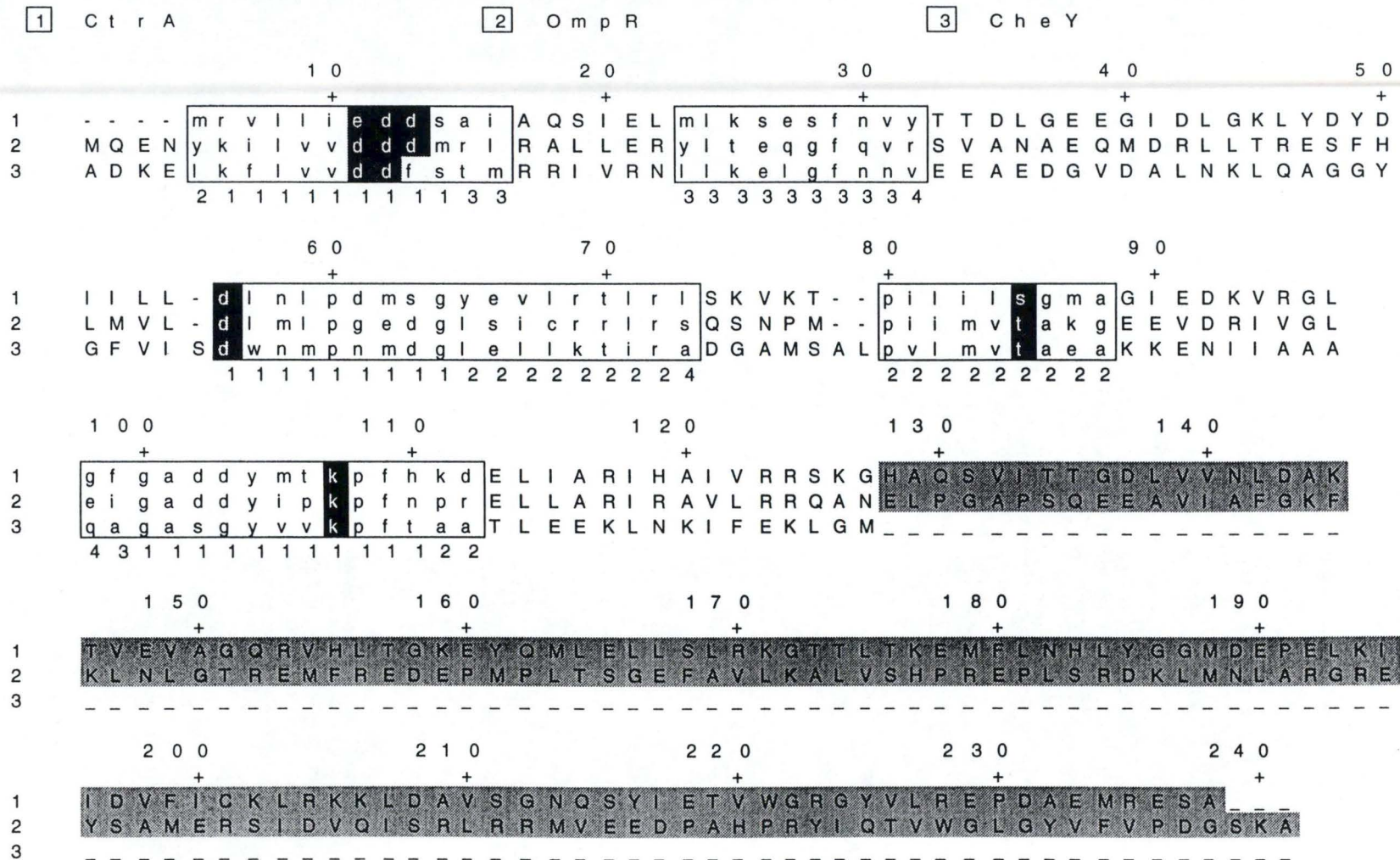


Figure 47: Aligement Matchbox pour les séquences de BaCtrA, OmpR et CheY.  
 Les résidus de BaCtrA et d'OmpR impliqués dans le site catalytique sont ceux dont le fond est gris foncé.  
 Les domaines effecteurs sont montrés par un fond gris clair.  
 Les boxes conservées retrouvées par Matchbox sont en minuscules et sont encadrées.

## II. Alignement de séquences et délimitation des domaines de la protéine.

Les séquences retenues dans les banques de données ont été soumises à une analyse factorielle par l'algorithme Matchbox et classées en quatre groupes en fonction de leur similarité.

Le groupe contenant la séquence BaCtrA a été appelé Groupe1. Les 50 séquences de ce groupe ont ensuite été alignées par Matchbox et une seule boîte conservée parmi toutes les séquences a été mise en évidence par Matchbox. A l'intérieur de ce groupe 1, différents sous-groupes de similarité peuvent être définis. Le sous-groupe 1A comprend BaCtrA ainsi que 13 autres séquences. Lorsque ces 14 séquences sont alignées, 6 boîtes conservées sont décelées par Matchbox (fig.46). OmpR d'*E. coli* est retrouvé dans l'alignement des 50 séquences.

Un troisième alignement incluant les protéines BaCtrA, CheY (*E. coli*) et OmpR (*E. coli*) a aussi été effectué par Matchbox. Dans ce cas, 5 boîtes conservées ont été trouvées, mais elles ne semblent pas correspondre à des structures secondaires telles que les hélices  $\alpha$  et les feuillets  $\beta$  connus pour CheY.

Cependant, le dernier alignement effectué nous a permis de localiser les résidus impliqués dans l'activité catalytique (fig.47 et p'13):

- le glutamate 7 (E7) et l'aspartate 8 (D8) correspondants aux D12 et D13 de CheY. Un troisième résidu acide en position 9 (D9) est présent dans une majorité de régulateurs dont BaCtrA, mais inexistant à la position correspondante chez CheY;

- le site de phosphorylation en position 51 (D51) correspondant au D57 de CheY;

- la sérine 79 (S79) correspondant à la T87 de CheY;

- enfin, la lysine 101 (K101) correspondant à la K109 de CheY;

L'alignement de BaCtrA et de CheY nous a aussi permis de séparer CtrA en deux domaines (fig. 47). Le domaine amino-terminal compte 121 acides aminés. Il présente 29,4% d'identité et 60,5% de similarité avec CheY. Il s'agit donc du domaine "receveur" de la protéine ou domaine "CheYlike". Le domaine carboxy-terminal, composé des 111 résidus restants, est homologue au domaine effecteur des régulateurs. Il est aussi appelé domaine "OmpRlike".

Dans la suite du travail de prédiction, nous allons considérer séparément les deux domaines, comme s'il s'agissait de protéines

différentes (notées BaCtrA N-term et BaCtrA C-term). La prédiction de structure séparée des différents domaines d'une protéine est préférable lorsqu'une modélisation par homologie de la protéine entière ne peut être réalisée (Russel, 1996).

### III. Modélisation du domaine amino-terminal de BaCtrA.

Nous savons qu'il existe une protéine homologue à BaCtrA N-term présente dans les banques de séquences de structures connues (PDB): CheY (code PDB: 3chy). Cependant, bien que BaCtrA N-term et CheY partagent 29,4% d'identité au niveau de leur séquence, aucune zone conservée au niveau des structures secondaires n'ont été retrouvées par Matchbox. Nous disposons donc de peu de régions structurellement conservées (SCRs). Or ce sont ces régions qui servent de base à la modélisation par homologie. Des recherches complémentaires ont donc été menées afin d'obtenir plus d'information sur les structures secondaires et tertiaires de BaCtrA, avant de tenter une modélisation par homologie.

#### A. Recherches complémentaires sur la conformation du domaine amino-terminal de BaCtrA:

L'information complémentaire quant à la conformation adoptée par BaCtrA N-term provient de trois types de recherches:

- une prédiction de structure secondaire par PHD;
- une reconnaissance de folds par comparaison de structures secondaires et de patterns d'accessibilité (via TOPITS)
- une reconnaissance de folds par Threading.

##### 1. Prédiction de structure secondaire.

Les structures secondaires des domaines receveurs de BaCtrA et de OmpR ont été prédites par PHD. PHD est un algorithme qui prédit les structures secondaires à partir d'alignements multiples. Pour obtenir une prédiction relativement précise, il est conseillé par les auteurs d'utiliser au minimum 5 séquences dans l'alignement, pour lesquelles nous retrouvons des degrés de similarité différents.

	envoi	% hélices $\alpha$	% feuillet $\beta$	% loop	Classe	Topologie
PHD BaCtrA	Groupe 1	40	24	36	$\alpha/\beta$	$\beta\alpha\beta\alpha\beta\beta\alpha\alpha$
	Groupe 1A	42	22	36	$\alpha/\beta$	$\beta\alpha\beta\alpha\beta\alpha\alpha$
PHD OmpR	Groupe 1	33	18	49	mixed	$\beta\alpha\beta\alpha\beta\beta\alpha$
CheY	-	40	20	40	$\alpha/\beta$	$\beta\alpha\beta\alpha\beta\alpha\beta\alpha$

Tableau 5: Prédiction de structures secondaires obtenues par PHD pour les séquences de BaCtrA N-term et OmpR N-term.

Les caractéristiques topologiques d'après la structure de CheY sont aussi indiquées. Les groupes 1 et 1A correspondent respectivement aux 50 et aux 14 séquences ayant été sélectionnées pour les alignements Matchbox.

BaCtrA N-term.			
Code PDB	Famille	Score	Classe
3chy	CheY- like	3.36	$\alpha/\beta$
1pii	Tryptophan biosynthesis enzymes	2.12	$\alpha/\beta$
2dri	L-arabinose binding protein-like	2.1	$\alpha/\beta$
1pfk	Phosphofructokinase	2.06	$\alpha/\beta$
2liv	L-arabinose binding protein-like	2.04	$\alpha/\beta$

Tableau 6:

Caractéristiques des structures ayant obtenu les meilleurs scores pour la reconnaissance de folds par Topits au niveau du domaine BaCtrA N-term.

La première série de séquences soumises à une prédiction par PHD a été le groupe 1A (14 séquences homologues) pour la prédiction de structure de BaCtrA N-term. L'alignement obtenu en même temps que la prédiction correspond pratiquement à l'alignement Matchbox pour ces mêmes 14 séquences (Fig. 46). Les pourcentages d'identité entre les séquences alignées et la séquence cible s'échelonnent entre 28 et 37%.

Une deuxième série de séquences a été soumise, et elle correspond aux 50 séquences du groupe 1, avec des degrés de similarité compris entre 39 et 51% (26 et 37% d'identité) pour la prédiction de structure de BaCtrA N-term, ainsi que pour celle de OmpR N-term.

Les résultats de ces prédictions sont rassemblés dans le tableau 5.

Si nous comparons les topologies obtenues pour les deux domaines receveurs BaCtrA N-term et OmpR N-term avec celle connue de CheY, nous pouvons voir qu'elles sont quasiment identiques à l'exception des extrémités carboxy-terminales des domaines. Ceci pourrait s'expliquer par le fait qu'un deuxième domaine, le domaine effecteur succède au domaine receveur chez BaCtrA et OmpR, alors que le régulateur CheY ne possède qu'un seul domaine.

## 2. Reconnaissance de folds par comparaison de structures secondaires et de patterns d'accessibilité.

Une reconnaissance de fold par comparaison de structures secondaires et de patterns d'accessibilité a été effectuée par Topits. Après avoir prédit la structure secondaire de la séquence cible, cet algorithme détecte, puis aligne des motifs similaires de structure secondaire et d'accessibilité entre cette séquence de structure inconnue et un fold connu. Des scores de "probabilité" sont calculés pour chaque fold détecté, puis sont classés par ordre décroissant. Un premier score supérieur à 3 est correct dans 50% des cas (s'il est supérieur à 3,5 le fold est correct dans 60% des cas).

La topologie prédite par Topits est  $\beta\alpha\beta\alpha\beta\alpha\alpha$  et le fold détecté en première position est CheY avec un score de 3,36 (tableau 6).

BaCtrA N-term.			
Code PDB	Famille	Score	Classe
3chy	CheY-like	-8.83	$\alpha\beta$
2blt-b	$\beta$ -Lactamase / $\delta$ -ala carboxypeptidase	-7.31	$\alpha\beta$
1351	Calmodulin-like	-6.86	all $\alpha$
1dyn-b	Pleckstrin-homology domain	-6.68	all $\beta$
1mjc	Cold shock DNA-binding domain	-6.56	all $\beta$

Tableau 7:

Caractéristiques des 5 structures ayant obtenus les scores les plus élevés pour la reconnaissance de folds par PROFIT pour le domaine BaCtrA N-term.

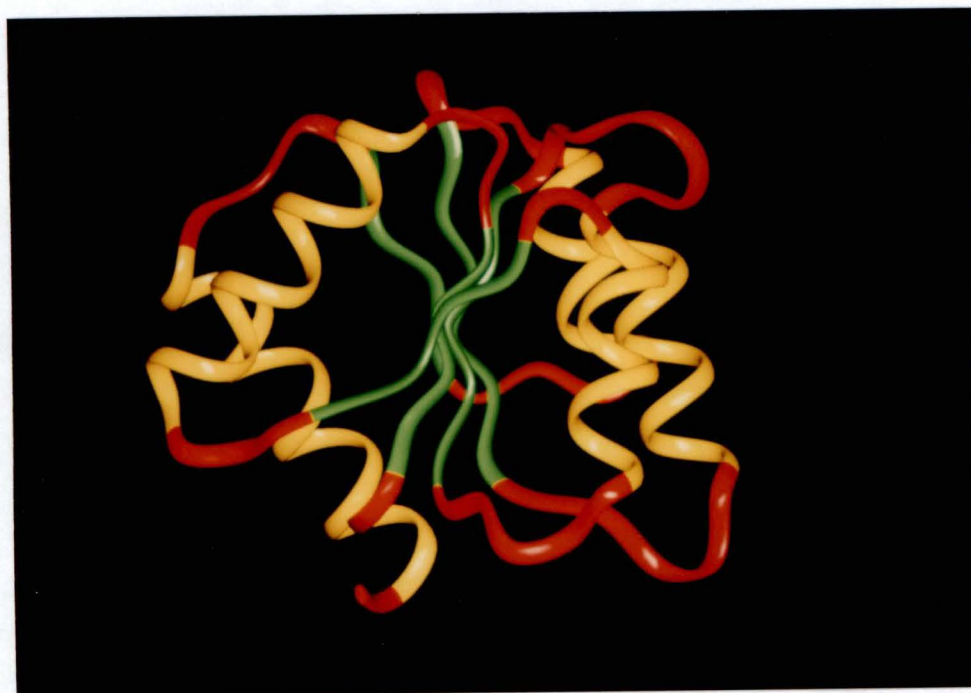
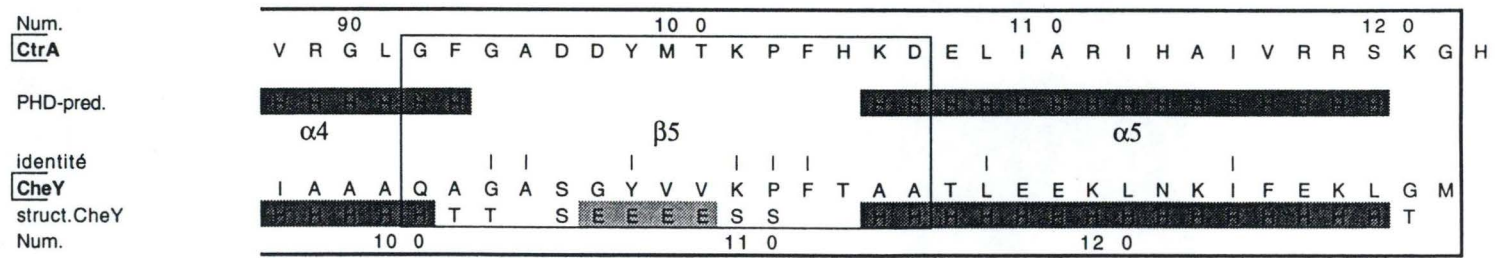
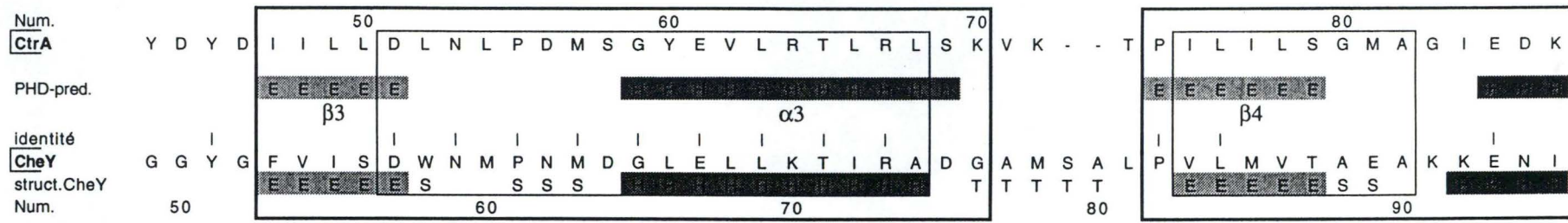
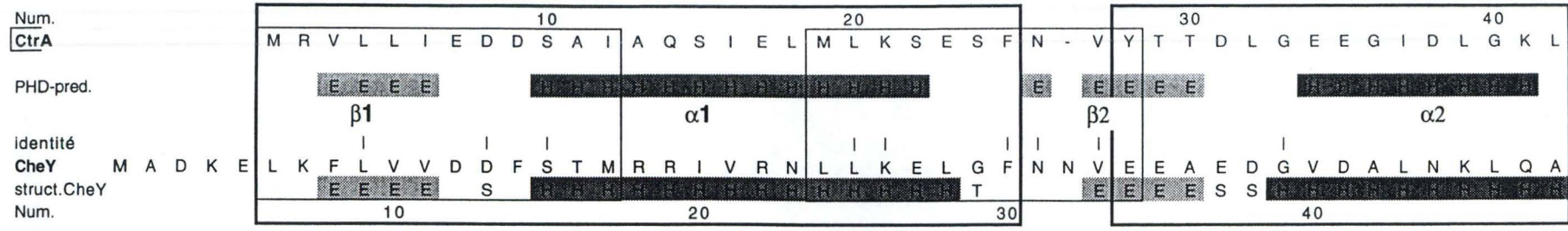


Figure 49: Modèle proposé pour le domaine amino-terminal de BaCtrA sur base de l'homologie avec CheY. Les hélices  $\alpha$  sont représentées en jaune, le feuillet  $\beta$  central en vert et les loops en rouge.



■ = Hélice  $\alpha$   
 ■ = Brin  $\beta$

**Figure 48:** Délimitation des SCRs sur base des régions prédites par Matchbox comme étant structurellement conservées et sur base de l'alignement produit par ProFIT entre BaCtrA N-term et CheY. Les boxes délimités par des traits fins sont les boxes conservés retrouvés par Matchbox, ceux dont les traits sont gras représentent les régions que nous avons considérées comme SCRs.



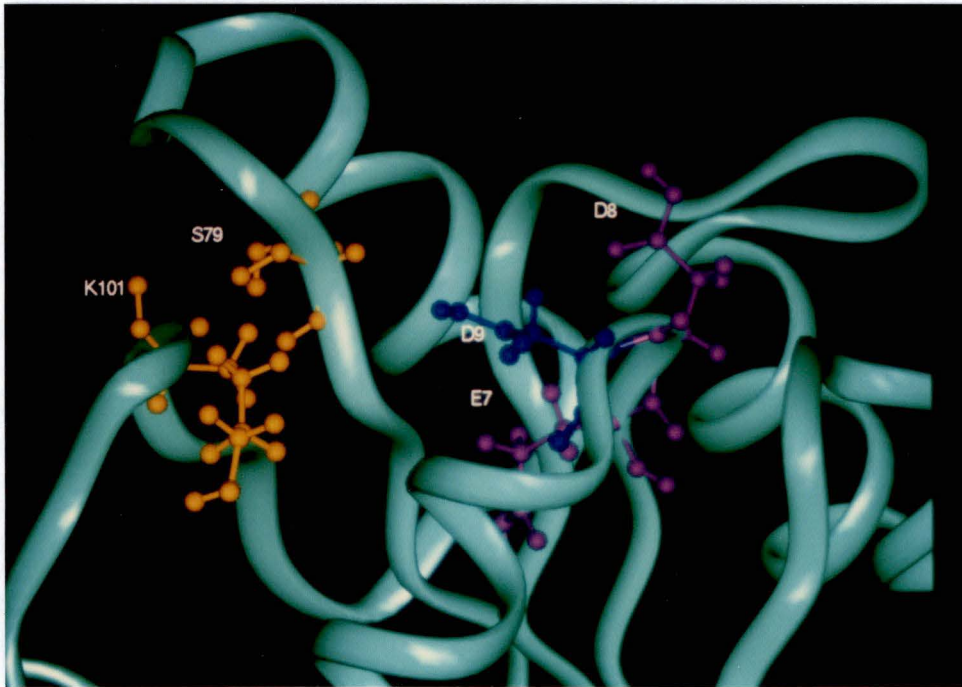


Figure 50: Site catalytique. Le site de phosphorylation (D57) est représenté en rouge et les résidus impliqués dans la liaison au  $Mg^{++}$  (E7, D8 et D9) sont en mauve, Les résidus S79 et K101 sont représentés en jaune.

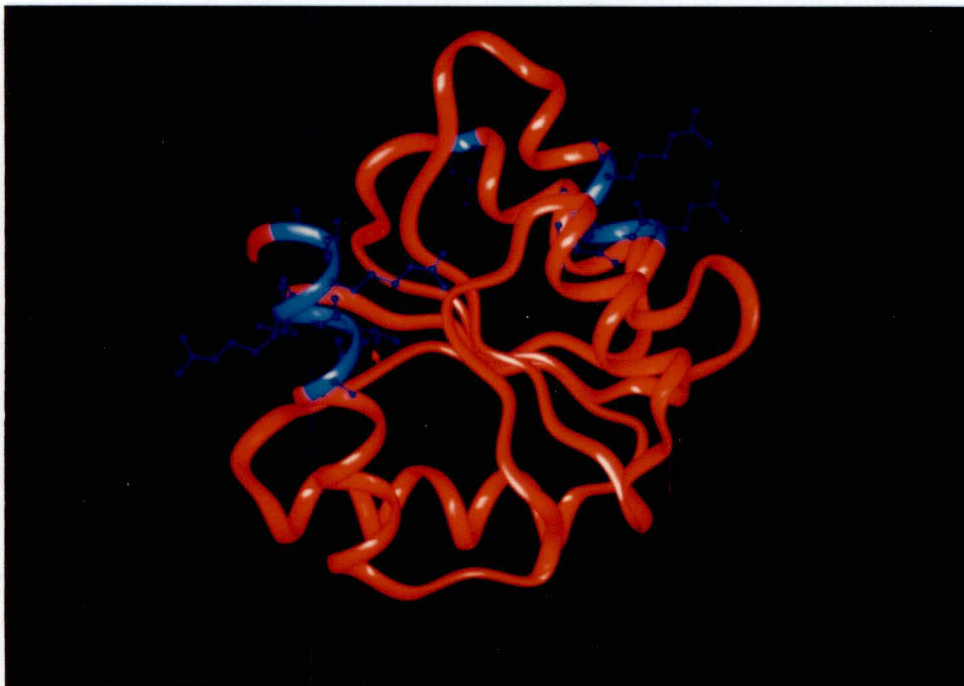


Figure 51: Régions détectées comme énergétiquement défavorables par un potentiel de force moyenne à l'échelle atomique.

### 3. Reconnaissance de folds par Threading.

Enfin, une reconnaissance de fold par Threading a été effectuée (via ProFIT) sur la séquence de BaCtrA N-term.

ProFIT calcule un score pour chaque structure testée qui correspond à une estimation de l'énergie qu'aurait la protéine cible si elle adoptait le fold testé. Ce score est modulé par la longueur des séquences, les pourcentages d'identité et de similarité, la présence de gaps entre la séquence du fold testé et la séquence de la protéine cible,... Il doit être le plus négatif possible.

L'information apportée par les 5 meilleurs scores obtenus pour la séquence de BaCtrA N-term est résumée dans le tableau 7.

De nouveau, CheY est classé comme meilleur score (-8,83).

En conclusion, ProFIT et Topits retrouvent tous les deux en majorité des structures de classe  $\alpha/\beta$ . Parmi ces structures, le meilleur score obtenu est celui de CheY. De plus, PHD prédit une topologie semblable à celle de CheY pour le domaine amino-terminal de CtrA et de OmpR.

Ces résultats, combinés à ceux de Matchbox, confirment que CheY peut être utilisé comme structure *template* pour la modélisation du domaine amino-terminal de CtrA et que d'autres SCRs peuvent être déterminées sur bases d'autres critères que la similarité de séquence obtenue par le logiciel Matchox.

### B. Modélisation par homologie du domaine "CheYlike"

La première étape de la modélisation par homologie est la détermination des régions structurellement conservées (SCRs). Dans notre cas, les SCRs sont soit:

- les régions prédites par Matchbox comme étant structurellement conservées;
- les régions où ProFIT aligne BaCtrA N-term à la structure de CheY. Dans ces zones alignées, les structures secondaires prédites par PHD pour le domaine BaCtrA N-term correspondent en général à des structures secondaires de CheY.

Ces SCRs recouvrent presque la totalité des séquences de CheY et de BaCtrA, à part le loop entre  $\alpha_B$  et  $\beta_3$ , le loop entre  $\alpha_C$  et  $\beta_4$  et le "gap" dans le loop entre  $\alpha_A$  et  $\beta_2$  (fig. 48).

Les coordonnées cartésiennes des atomes du squelette de CheY inclus dans les SCRs sont attribuées aux atomes du squelette de la cible BaCtrA. Les coordonnées de la chaîne principale des résidus des deux loops de CheY ne faisant pas partie des SCRs ont quand même été assignées aux résidus conservés de BaCtrA. Cependant, lors de l'optimisation du modèle, la chaîne principale de ces loops ne sera pas fixée et sa conformation pourra être modifiée par la minimisation d'énergie.

L'étape suivante consiste à placer sur la cible la chaîne latérale des résidus, en utilisant une librairie de rotamères. Dans cette librairie, on trouve les quelques conformations fréquemment observées dans les structures cristallographiques pour les chaînes latérales de tous les résidus. Chaque rotamère donne lieu à un calcul d'énergie potentielle, et le rotamère de plus basse énergie est attribué au nouvel acide aminé, pour autant que cette énergie soit inférieure à celle du conformère par défaut.

A l'issue de cette étape, le modèle est appelé non affiné. Il faut procéder à des minimisations d'énergie afin d'obtenir une structure réaliste (optimisation). Deux minimisations successives ont été effectuées, lors desquelles nous avons fixé la chaîne principale de tous les résidus (leur position ne sera donc pas modifiée par cette minimisation), à l'exception des résidus entourant les délétions dans BaCtrA par rapport à CheY, et appliqué une contrainte sur les chaînes latérales de ces résidus.

Le modèle obtenu après minimisation (en utilisant l'algorithme Steepest descent) est montré sur la figure 49. La figure 50 représente le site catalytique. Le modèle a ensuite été soumis à une évaluation. Celle-ci vérifie les structures en utilisant un potentiel de force moyenne à l'échelle atomique développé au laboratoire (Melo et Feytmans, soumis). Trois zones ont été détectées comme énergétiquement défavorables (figure 51), comprenant au total 15 résidus. Ces trois zones sont situées en surface dans les hélices  $\alpha_B$  (L42),  $\alpha_C$  (E61-R64) et  $\alpha_E$  (A114-S119). Si on examine le modèle à ces endroits, on peut remarquer que de nombreux groupements  $\text{CH}_2$  sont exposés au solvant. Ces zones correspondent peut-être à des zones de contact entre sous-unités ou domaines. Le reste de la protéine semble énergétiquement réaliste. La méthode d'évaluation montre que le niveau de confiance du modèle est bon (12,40% d'erreur).

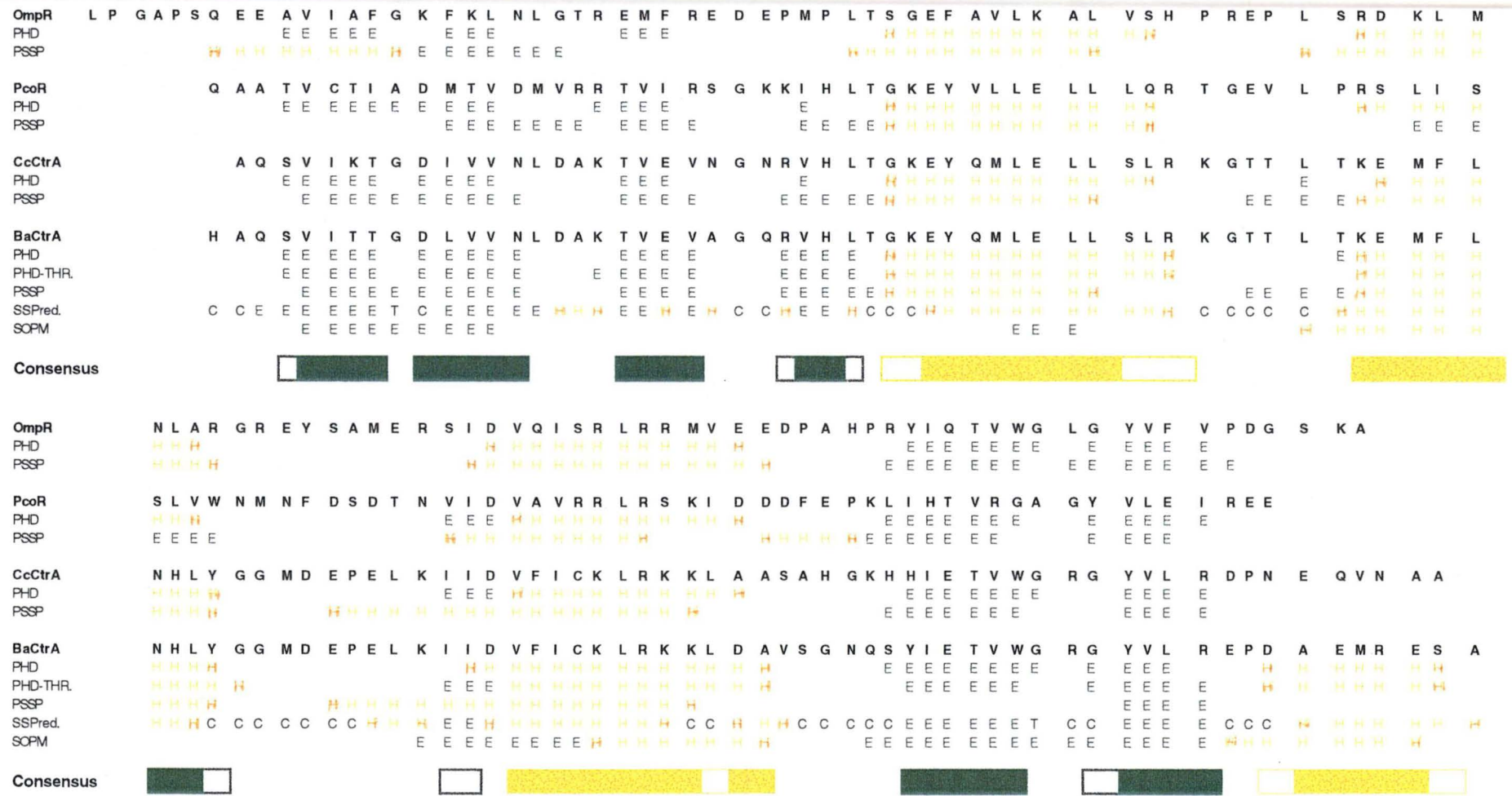


Figure 52: Prédiction de structure secondaire pour les domaine C-term des quatre séquences cibles alignées OmpR, PcoR, CcCtrA et BaCtrA. Les hélices a sont représentées en jaune et les brins b en vert. Le consensus défini comme décrit dans le texte est représenté par des boxes pleins; des régions à incertitudes étant aussi mentionnées (boxes creux).

## IV. Tentative de modélisation du domaine carboxy-terminal "OmpRlike"

Comme aucun homologue de séquence du domaine BaCtrA C-term n'a été trouvé dans la banque PDB, des prédictions de structures secondaires et des reconnaissances de *folds* ont été effectuées, de la même façon que pour le domaine BaCtrA N-term. Ces recherches ont porté sur les séquences de BaCtrA C-term et de CcCtrA C-term, de OmpR C-term et de PcoR C-term (PcoR est la seconde séquence la plus homologue à BaCtrA relevée dans les banques de séquences).

### 1. Prédiction de structure secondaire.

Les prédictions de structures secondaires pour les quatre séquences citées ci-dessus ont été accomplies par différents serveurs: PHD, SSPred., PSSP, SOPM.

Les deux premières méthodes de prédiction de structures secondaires nécessitent la soumission d'une liste de plusieurs séquences homologues qu'elle commencent par aligner; les deux autres prédisent la structure directement à partir de la séquence cible. SOPM (Geourjon et Deleage, 1994) utilise différents algorithmes de prédiction de structure secondaire et définit un consensus à partir des différentes prédictions obtenues.

Les résultats obtenus par ces quatre serveurs pour les quatre protéines sont récapitulés dans l'alignement de la figure 52. Ces résultats nous ont permis de définir arbitrairement une topologie "consensus": un résidu est considéré faisant partie d'une structure secondaire ( $\alpha$  ou  $\beta$ ) s'il est prédit par 4 méthodes sur 5 pour BaCtrA C-term ou par 8 méthodes sur 11 pour les prédictions sur les quatre protéines. La topologie "consensus" prédite par les différentes méthodes pour les quatre protéines est  $\beta\beta\beta\beta\alpha\alpha\beta\beta(\alpha)$ .

Suite à ces prédictions de structures secondaires effectuées sur les séquences d'OmpR, de BaCtrA, de CcCtrA et de PcoR, la topologie "consensus" définie nous a paru très semblable à celle du Cro ( $\beta\alpha\alpha\beta\beta$ ). La structure tridimensionnelle du dimère de Cro complexé à un fragment d'ADN a été résolue à 2.35 Å (Mondragon et al., 1989). Une recherche de similarité par Matchox a été effectuée afin de vérifier s'il existe des régions conservées entre les deux séquences, régions dont



le rôle serait la liaison à l'ADN et la régulation transcriptionnelle. L'alignement a été effectué sur les séquences de BaCtrA C-term, OmpR C-term et Cro. Trois boîtes ont été mises en évidence par Matchbox, qui se superposent quasiment à la totalité de la séquence de Cro (fig. 53). Rappelons que Blast ne retrouve pas de séquence significativement similaire dans PDB, alors que la séquence de Cro se trouve bien dans cette banque. Les hélices 1 et 3 prédites par PHD pour BaCtrA se superposent aux hélices 1 et 3 de Cro dans l'alignement Matchbox. L'hélice 2 pose problème, car un gap y est introduit par l'alignement. Il est cependant raisonnable de penser que l'hélice 2 de Cro correspond à l'hélice 2 prédite pour BaCtrA. Les trois premiers brins  $\beta$  de BaCtrA n'ont pas d'équivalents chez Cro.

Les folds de BaCtrA et de Cro sont donc probablement semblables: trois hélices  $\alpha$ , dont la troisième lie l'ADN et au moins 3 brins  $\beta$  arrangés en feuillet anti-parallèle, avec le dernier brin impliqué dans la dimérisation. Chez BaCtrA, trois brins supplémentaires sont présents dans ce feuillet.

## 2. Reconnaissance de folds par comparaison de structures secondaires et de patterns d'accessibilité.

Comme pour le domaine BaCtrA N-term, la reconnaissance de fold peut être effectuée par Topits. La topologie prédite est  $\beta\beta\beta\beta\alpha\alpha\beta\beta\alpha$  et est reprise dans l'alignement des différentes prédictions de structure secondaire de la figure 52. Les scores obtenus pour les structures détectées sont inférieurs à 2 (tableau 8). Il n'y a donc qu'environ 33% de chances qu'un de ces folds soit le bon.

## 3. Reconnaissance de folds par threading.

Au niveau de la reconnaissance de folds, différentes méthodes de threading ont été testées sur les quatre séquences cibles: ProFIT, THREADER et 123D. Comme il existe plusieurs protéines liant l'ADN dans les banques de structures des différentes méthodes de threading, nous espérons associer BaCtrA à un de ces folds connus qui permet la liaison à l'ADN, vu que cette fonction est également attribuée à BaCtrA. Les structures retrouvées sont généralement des classes "all  $\alpha$ " ou " $\alpha/\beta$ ", mais des structures des classes "all  $\beta$ " et " $\alpha+\beta$ " sont assez

Protéine cible	Code PDB	Famille	Classe	Score
<b>Cro</b>	1cse-i	Subtilases	$\alpha/\beta$	-6.82
<b>CcCtrA</b>	1bnh	Ribonuclease inhibitor	$\alpha/\beta$	-6.91
<b>BaCtrA</b>	1ppi*	$\alpha$ -amylase	$\alpha/\beta$	-7.22
<b>PcoR</b>	1mjc	Cold shock DNA binding domain	all $\beta$	-6.63
<b>OmpR</b>	1enh	Homeodomain-like	all $\alpha$	-8.26

Tableau 9: Structures ayant obtenu le meilleur score par ProFIT pour le domaine C-term des 5 protéines cibles (Cro, CcCtrA, BaCtrA, PcoR et OmpR). La structure notée \* a été reprise dans l'alignement de la figure 51.

Code PDB	Famille	Classe	Cro	CcCtrA	BaCtrA	PcoR	OmpR
2rsl-c*	$\gamma$ - $\delta$ resolvase	$\alpha/\beta$	4	3	9	10	
1bnh*	Ribonuclease inhibitor	$\alpha/\beta$	3	1	5		
1cdl-b	Calmodulin-like	all $\alpha$	10	5	2		
5p21*	G protein	$\alpha/\beta$		10	3	6	
1pdn-c*	Homeodomain-like	all $\alpha$			8	3	2
1nci-a	Cadherin	all $\beta$		2	4		
1pne	Profilin (actin-binding protein)	$\alpha+\beta$	8		6		
1lmb3	phage repressors	all $\alpha$				5	9
1frd	2Fe-2sS ferredoxin	$\alpha+\beta$				9	8

Tableau 10: Structures ayant été classées dans les dix premiers scores par ProFIT pour au moins deux protéines cibles (Cro, CcCtrA, BaCtrA, PcoR et OmpR). Les structures notées \* sont celles qui ont été reprises dans l'alignement de la figure 51.



ctra

H A Q S V I T T G D L V V N L D A K T V E V A G Q R V H L T G K E Y Q M L E L L S L R K G T T L T K E M F L N H L Y G G M

Topits

EEEEEEEEEEEEEE EEEE EEE HHHHHHHHHHHHHHHH E HHHHHHHHHHHH

Profit:

1ppi

EEEEEE SEEEETTSSBS BTT BBGGGTBBT TTTB GGG B TTT

5p21

HHHHHHH SEEEEEEETT HHHHHT HHHHHHHHHHHH TT S EEEEE T

1bnh

HHHHHHHHHHHHHHHHT S EEE TTS HHHHHHHHHHHHHH EEE HHHHHHHHHHHH

1pdn-c

HHHHHHHHHHHHHT T HHHHHHHHHT HHHHHHHHHHHHHH SS S

2rsl-c

HHHHH TT EEEEESSGGGT S SHHHHHHHHHHHHHHHT T EEEET

1enh(ompr)

HHHHHHHHHHHHHHH SS HHHHHHHHHHHHT

123D:

All\*

EEEEEE EEEE HHHHHHHH EE HHHHHHHHHHHHHH HHHHHHHHHHHH

dom2\*

EEE EE HHHHHHHH HHHHHHHHHHHHHH HHHHHHHHHHHH

CONSENSUS



ctra

D E P E L K I I D V F I C K L R K K L D A V S G N Q S Y I E T V W G R G Y V L R E P D A E M R E S A

Topits

HHHHHHHHHHHHHHHHH EEEEEEEE EEEE HHHHHHHH

Profit:

1ppi

SSS SB T T HHHHHHHSBGGGE EEB TTS HHHHHHHH

5p21

T S S HHHHHHHHHHHHHT EEE TT TT HHHHHHHHHHHH

1bnh

T TT EEE HHHHHHHHHHHH SS EEE TTS HHHHHHHHHHHH

1pdn-c

SS S THHHHHHHHTTT HHHHHHHHHHHHTSS S STT HHH

2rsl-c

TT B SHHHHHHHHHHHHHHHH

1enh(ompr)

HHHHHHHHHHHHHHH

123D:

All\*

HHHHHHHHHHHHHHHHHHHHH HHHHHHHH EE HHHHHHHH

dom2\*

HHHHHHHHHHHHHHHHHHHHH HHHHHHHH HHHHHHHH

CONSENSUS



Figure 54: Alignement des structures secondaires des folds reconnus par threading et pseudo-threading avec la séquence de BaCtA C-term, et définition d'un consensus (boxes pleins) comme décrit dans le texte; les boxes creux représentent les régions "à incertitudes".

\* dom2- Prédiction 123D effectuée seulement à partir de la séquence de BaCtA C-term.  
All = Prédiction 123D effectuée à partir de la séquence entière de BaCtA.

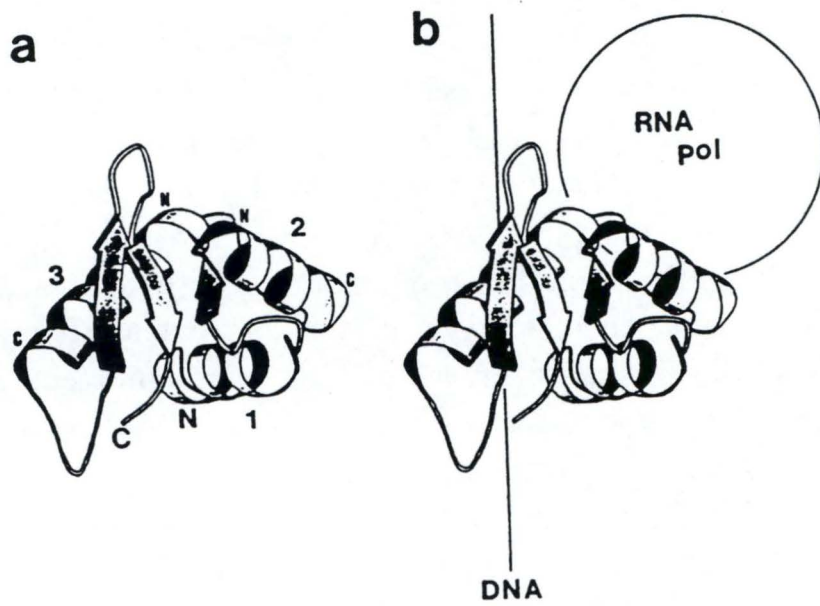
fréquentes aussi, même parmi les dix premiers scores. Aucune méthode ne nous a fourni de structures pouvant être adoptées en entier par notre séquence, à l'exception des deux premières structures retrouvées par ProFIT pour OmpR (les homéodomains 1enh avec un score de -8,26 et 1pdn-c avec un score de -7,55, qui sont des protéines régulatrices liant l'ADN). Les meilleurs scores pour les quatre autres séquences ne sont pas inférieurs à -7,3 et ne correspondent à aucun des folds connus liant l'ADN (tableau 9).

Le tableau 10 montre le rang des structures que ProFIT classe dans les dix premiers scores pour au moins deux séquences. Ce tableau indique aussi les folds reconnus pour la séquence du répresseur Cro. Puisque la structure de Cro n'existe pas dans les banques de ProFIT, THREADER, 123D ni dans celle de Topits, elle ne peut être reconnue comme ayant un fold énergétiquement favorable pour la séquence de BaCtrA. Nous avons donc testé avec Cro la reconnaissance de folds par ProFIT et 123D de la même façon que pour les 4 séquences citées ci-dessus. Nous pouvons ainsi comparer les résultats obtenus pour les cinq séquences et essayer de trouver un fold commun reconnu être énergétiquement favorable. Dans les structures classées parmi les dix premiers scores retrouvés pour la séquence de Cro, quatre sont aussi retrouvées pour BaCtrA et trois le sont pour CcCtrA; alors que OmpR n'a qu'une seule structure retrouvée en commun avec BaCtrA. Ceci pourrait signifier que CtrA a une structure plus semblable à celle de Cro qu'à celle de OmpR.

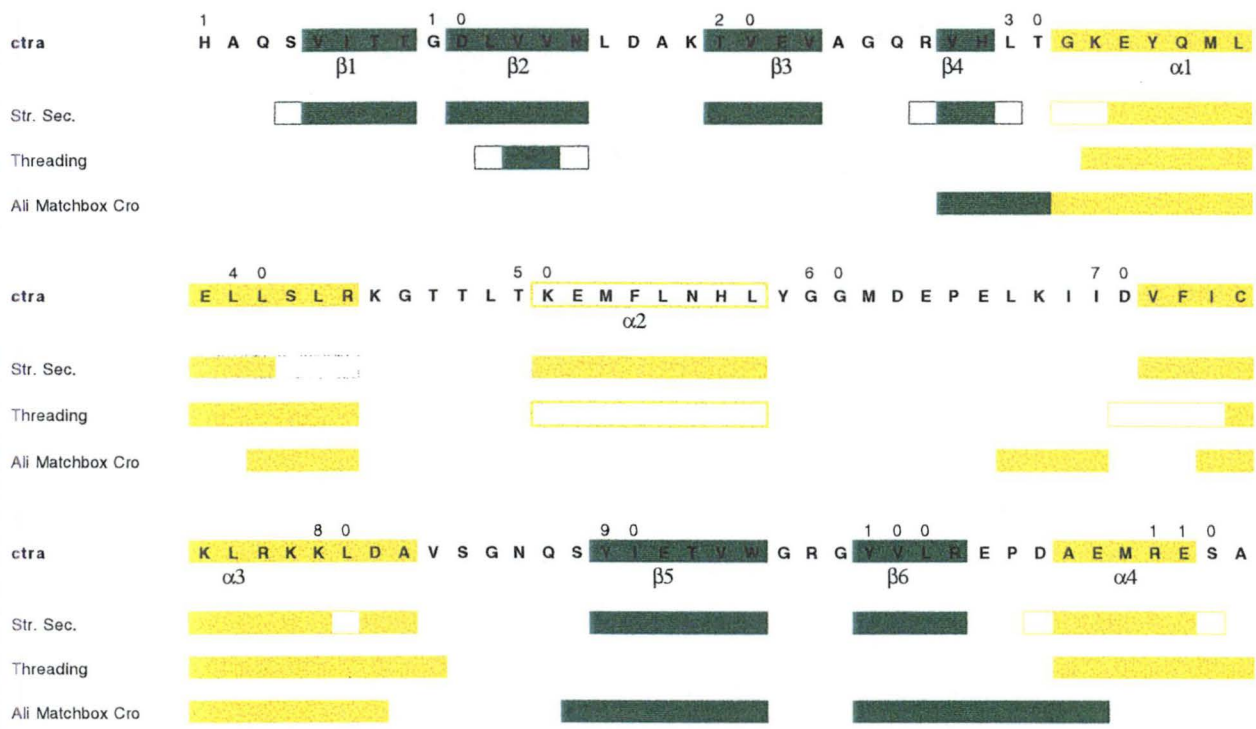
La figure 54 montre l'alignement de la structure classée en première position pour la séquence de BaCtrA C-term et de quatre structures relativement bien classées pour au moins trois séquences cibles avec la séquence BaCtrA C-term. Sur ce tableau sont aussi repris la prédiction de structure secondaire de Topits et les consensus définis par 123D sur la séquence entière et sur la séquence C-term de BaCtrA. Le consensus qui y est défini ne nous apporte rien de particulier, excepté que la première hélice  $\alpha$  prédite par PHD est retrouvée quasiment pour toutes les structures alignées.

#### 4. Consensus

Les résultats des différentes approches testées sont récapitulés dans la figure 55, afin de définir la topologie consensus générale et d'affiner la position des éléments de structure secondaire prédits. La topologie



**Figure 56:** Structure (a) et mode de liaison à l'ADN (b) proposés pour le domaine de liaison à l'ADN de PhoB (Suzuki et Makino, 1995)



**Figure 55a:** Détermination du consensus de structure secondaire de BaCtrA C-term à partir des méthodes de prédiction de structures secondaires et des méthodes de reconnaissance de folds. Les hélices  $\alpha$  et les feuillets  $\beta$  sont représentés en jaune et vert respectivement.

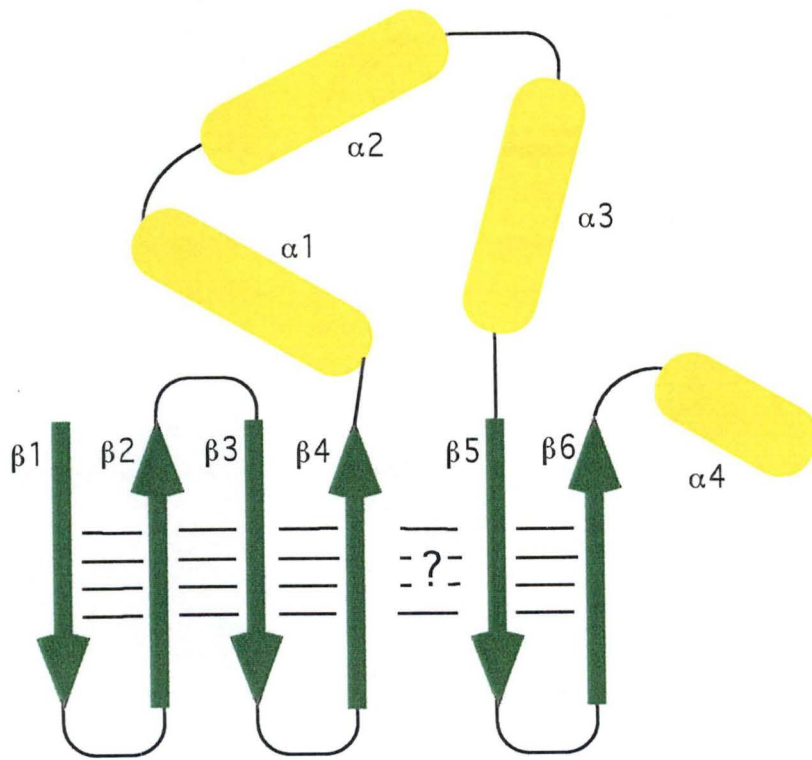


Figure 55b: Modèle topologique de BaCtrA C-term. Les hélices  $\alpha$  et les feuillets  $\beta$  sont représentés en jaune et vert respectivement.

consensus a été déterminée comme étant  $\beta\beta\beta\beta\alpha\alpha\beta\beta\alpha$ . Les trois premiers brins  $\beta$  prédits par PHD ont quand même été repris dans le consensus même s'ils ne sont pas retrouvés par d'autres méthodes car les prédictions à ce niveau sont relativement fiables et aucune des structures de protéines liant l'ADN présentes dans les banques utilisées n'ont de brins  $\beta$  à cet endroit (ce fold n'aurait donc jamais pu être détecté).

## V. Discussion

Un modèle de la structure 3D de CtrA N-term a été établi sur base de l'homologie avec CheY. Les résidus impliqués dans le site catalytique ont été déterminés par correspondance avec CheY. Le modèle affiné par minimisation d'énergie possède un haut niveau de confiance, comme le montre l'évaluation réalisée grâce à un potentiel d'énergie extrêmement précis (niveau atomique). Il peut donc servir à concevoir des expériences de mutagenèse dirigée.

Pour CtrA C-term, nous avons prédit un modèle topologique à partir des données obtenues par prédiction des structures secondaires et par threading. Ce modèle est appuyé par la similarité de séquence et fonctionnelle entre BaCtrA C-term et Cro. Ceci nous a permis de déterminer quels éléments de structures secondaires pourraient être impliqués dans la fonction de régulation de la transcription.

Nous attribuerions le rôle de liaison à l'ADN aux trois hélices  $\alpha$  centrales, plus particulièrement à l'hélice 3 qui s'insérerait dans le sillon majeur de l'hélice d'ADN, comme le fait la troisième hélice  $\alpha$  de Cro. L'hélice 2 jouerait probablement un rôle d'interaction avec la polymérase, rôle partagé avec l'hélice 3.

Les quatre premiers brins  $\beta$  formeraient un feuillet anti-parallèle, qui pourrait être complété par les deux derniers brins qui suivent l'hélice 3. L'alternative serait que ces deux brins forment un autre plan anti-parallèle avec les deux brins correspondant d'un autre monomère BaCtrA, lors de la dimérisation comme dans le cas de Cro. Dans ce cas, la présence d'une hélice  $\alpha$  à l'extrémité carboxy-terminale du domaine pourrait gêner la dimérisation.

Ces résultats sont comparables à ceux obtenus par Suzuki et Makino (1995), qui ont proposé un mode de liaison à l'ADN et d'interaction avec la polymérase pour les régulateurs OmpR et PhoB, par comparaison avec l'histone H5 (fig. 56). Celle-ci contient 3 hélices  $\alpha$ , un petit brin  $\beta$  (à l'extrémité amino-terminale de l'hélice 2) et une épingle

$\beta$  carboxy-terminale composée de 2 brins  $\beta$  formant un feuillet anti-parallèle avec les 2 brins du deuxième monomère. Ce modèle a été obtenu par la seule comparaison de séquence entre l'histone H5, OmpR, PhoB et quelques autres régulateurs de la sous-famille OmpR. Toutes ces séquences ayant montré une similarité, les auteurs en ont déduit une similarité de structure.

Le modèle élaboré pour BaCtrA C-term est bien sûr beaucoup moins précis que celui élaboré pour BaCtrA N-term. Cependant, il permet de poser une hypothèse d'explication structurale à la mutation T170I, qui rend CcCtrA thermosensible. Dans BaCtrA, le résidu 170 est également une thréonine, située du côté N-terminal de l'hélice 1 d'après notre modèle. La thréonine 170 est probablement impliquée dans la stabilisation de cette hélice par "capping" (p'15; Serrano et Fersht, 1989). La mutation T170I provoquerait donc une déstabilisation de l'hélice 1, qui entrainerait l'inactivation de l'enzyme à 37°C.

Bien qu'il soit admis que la prédiction de structure séparée des différents domaines d'une protéine est préférable lorsqu'une modélisation par homologie de la protéine entière ne peut être réalisée (Russel, 1996), le fait de séparer ces domaines introduit des inexactitudes à certains endroits du modèle de chacun de ces domaines. C'est ce qui semble être le cas pour 3 régions du modèle de BaCtrA N-term. Actuellement, aucune méthode théorique ne permet d'assembler les différents domaines. Cependant, nous avons un indice quant à leur position respective: le domaine C-term recouvrirait les zones estimées énergétiquement défavorables par la méthode d'évaluation du modèle de BaCtrA N-term. Il reste à présent à compléter le modèle de BaCtrA C-term par les nouvelles méthodes qui ne tarderont pas à être disponibles (prédiction des interactions entre brins  $\beta$ , par exemple; Hubbard et al., 1996) et surtout à mettre au point des expériences de mutagenèse dirigées qui devraient confirmer nos hypothèses structurales.

## Conclusion et Perspectives

Les deux objectifs de ce travail ont été atteints. Nous avons réussi à cloner chez *B. abortus* une protéine régulatrice, membre des systèmes à deux composantes, homologue de OmpR d'*E. coli* et identique à 80,5% à CtrA de *C. crescentus* (CcCtrA), qui sont tous deux des régulateurs transcriptionnels.

Au vu de cette homologie, l'hypothèse d'une conservation de la fonction de CcCtrA est posée pour le régulateur cloné chez *B. abortus*. Etant donné que *B. abortus* ne présente pas les particularités morphologiques liées au développement qui caractérisent *C. crescentus*, nous sommes conduits à supposer que la régulation exercée par BaCtrA concerne probablement d'autres fonctions, comme la régulation d'enzymes impliquées dans la topologie de l'ADN, dans sa réparation,... Ces fonctions pourraient être en rapport avec les mécanismes de virulence développés par les bactéries du genre *Brucella*.

Ce régulateur, appelé BaCtrA (CtrA de *B. abortus*), a été analysé par une approche théorique, en tentant d'élaborer un modèle tridimensionnel de la protéine.

A cette fin, nous avons considéré les deux domaines de BaCtrA séparément: le domaine amino-terminal, receveur, homologue à CheY ("prototype" des régulateurs) qui a été appelé BaCtrA N-term et le domaine carboxy-terminal effecteur, appelé BaCtrA C-term.

Un modèle tridimensionnel de BaCtrA N-term a été élaboré sur base de l'homologie avec CheY, puis évalué grâce à un potentiel d'énergie extrêmement précis (niveau atomique), mis au point au laboratoire. Cette évaluation confirme que ce modèle a un haut niveau de confiance: seulement 12% de résidus sont détectés comme ayant une conformation énergétiquement défavorable. Les résidus impliqués dans l'activité catalytique ont été déterminés par correspondance avec CheY, et sont les premiers candidats à pouvoir être testés lors d'expériences de mutagenèse dirigée, principalement l'aspartate 51 qui est le site de phosphorylation prédit.

D'un point de vue méthodologique, nous avons utilisé une approche originale pour la modélisation de BaCtrA N-term, qui combine des méthodes de modélisation par homologie et des méthodes de prédiction de structures par threading. Cette approche nous a permis de compléter



l'alignement de séquence entre BaCtrA et CheY, tout en contournant les problèmes propres au threading.

Un modèle topologique beaucoup moins précis de BaCtrA C-term a été prédit à partir des données obtenues par prédiction de structure secondaire et par threading. Ce modèle est appuyé par la similarité topologique et fonctionnelle entre BaCtrA C-term et Cro. La fonction de liaison à l'ADN a été attribuée à l'hélice  $\alpha_3$ , qui partagerait aussi avec l'hélice  $\alpha_2$  le rôle d'interaction avec la polymérase. Les brins  $\beta$  seraient tous arrangés en un même feuillet anti-parallèle ou, alternativement, les quatre premiers brins et les deux derniers constituant deux feuillets distincts, les deux derniers brins seraient en interaction avec les deux brins correspondant d'un autre monomère au sein d'un dimère.

Ce modèle permet de définir des résidus à tester en mutagenèse dirigée.

La première mutation à effectuer est la mutation T170I qui rend la protéine thermosensible chez *C. crescentus*. Cette mutation déstabiliserait l'hélice 1 du domaine en substituant un résidu impliqué dans le "capping" de cette hélice. D'autres mutations semblables de résidus probablement impliqués dans le capping des hélices  $\alpha$  peuvent être testées, par exemple une substitution des résidus T151 (N-capping de l'hélice 1), K165 (C-capping de l'hélice 1), D191 (N-capping de l'hélice 3) et D225 (C-capping de l'hélice à l'extrémité carboxy-terminale du domaine). Ces mutations pourraient conduire à l'obtention de mutants thermosensibles, outils importants de la caractérisation future de BaCtrA.

Deuxièmement, nous pourrions tester des résidus correspondant à ceux d'OmpR d'*E. coli* dont la mutation a montré une perte de fonction, par exemple les résidus M182, P185 et L187 correspondant aux résidus E193, A196 et E198 pour qui la mutation empêche l'interaction avec la polymérase, ou les résidus V192 et S209 correspondant aux résidus V203 et R220 pour qui une mutation empêche la liaison à l'ADN.

Au terme de ce travail, le modèle structural d'un régulateur transcriptionnel de *B. abortus* a donc été élaboré et peut servir d'outil d'interprétation et de prédiction pour des expériences futures. Cet outil reste perfectible, d'une part par l'apport de données expérimentales, d'autre part par l'application de nouvelles méthodes théoriques en cours de développement.

# Bibliographie

- \* **Aiba, H., F.Nakasai, S.Mizushima, and T.Mizuno.** 1989. Evidence for physiological importance of the phosphotransfer between the two regulatory components, EnvZ and OmpR, in osmoregulation of *Escherichia coli*. *J.Biol.Chem.* **264**: 14090-14094.
- \* **Alton, G.G., L.M.Jones, R.D.Angus, and J.M.Verger.** 1988. Techniques for the brucellosis laboratory. INRA, Paris.
- \* **Altschul, S.T., W.Gish, W.Miller, E.W.Myer and D.J.Lipman.** 1990. Basic local alignment search tool. *J.Mol.Biol.* **215** : 403-410
- \* **Balaji, V.N., J.Scott Dixon, D.H.Smith, R.Venkataraman, and K.C.Murdock.** 1985. Design of anticancer drugs using modeling techniques. In: B.Venkataraman, R.J.Feldmann, ed. "Macromolecular structure and specificity: computer-assisted modeling and applications". New York: The New York Academy of Sciences.; vol **439**.
- \* **Bork, P., C.Ouzounis, and C.Sander.** 1993. From genome sequences to protein function. *Curr.Opin.Struct.Biol.* **4** : 393-403.
- \* **Bourret, R.B., J.F.Hess, and M.I.Simon.** 1990. Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY. *Proc.Natl.Acad.Sci.USA.* **87** : 41-45.
- \* **Bourret, R.B., S.K.Drake, S.A.Chervitz, M.I.Simon, and J.J.Falke.** 1993. Activation of the phosphosignaling protein CheY. Analysis of activated mutants by <sup>19</sup>F NMR and protein engineering. *J.Biol.Chem.* **268** : 13089-13096.
- \* **Brasseur, R.** 1995. Simulating the folding of small proteins by use of the local minimum energy and the free solvation energy yields native-like structures. *J.Mol.graph.* **13**: 1-11.
- \* **Bruix, M., J.Pascual, J.Santoro, J.Prieto, L.Serrano, and M.Rico.** 1993. <sup>1</sup>H- and <sup>15</sup>N-NMR assignment and solution structure of the chemotactic *Escherichia coli* CheY protein. *Eur.J.Biochem.* **215** : 573-585.
- \* **Chang, C., and E.M.Meyerowitz.** 1994. Eukaryotes have "two-component" signal transducers. *12th Forum in Microbiology*, 481-486.
- \* **Chou,P.Y. and G.D.Fasman.** 1974. Prediction of protein conformation. *Biochemistry.* **13** : 211-222.
- \* **de Philip, P., J.Batut, and P.Boistard.** 1990. *Rhizobium meliloti* FixL is an oxygen sensor and regulates *R.meliloti nifA* and *fixK* genes differently in *Escherichia coli*. *J.Bacteriol.* **172**: 4255-4262.
- \* **Degnen, S.T., and A.Newton.** 1972. Dependence of cell division on the completion of chromosome replication in *Caulobacter crescentus*. *J.Bacteriol.* **110** : 852-856.
- \* **Depiereux, E. and E.Feytmans.** 1992. MATCH-BOX. A fundamentally new algorithm for simultaneous alignment of several protein sequences. *Comput.Appl.Biosci.* **8** : 501-509.
- \* **Doolittle, R.F.** 1981. Similar amino-acid sequences: chance or common ancestry. *Science.* **214** : 1449-159.
- \* **Forst, S., J.Delgado and M.Inouye.** 1989. Phosphorylation of OmpR by the osmosensor EnvZ modulates expression

of the OmpF and OmpC genes in *Escherichia coli*.  
*Proc.Natl.Acad.Sci.USA.* **86** : 6052-6056.

\* **Garnier, J., D.J.Osguthorpe and B.Robson.** 1978.

Analysis of the accuracy and implications of simple methods for predicting secondary structure of globular proteins.  
*J.Mol.Biol.* **120** : 97-120.

\* **Geourjon, C., and G. Deleage.** 1994. SOPM : a self optimised prediction method for protein.  
*Protein Engng.* **7**, 157-164

\* **Gober, J.W., and M.V.Marques.** 1995. Regulation of cellular differentiation in *Caulobacter crescentus*.  
*Microbiol.Rev.* **59** : 31-47.

\* **Grimsley, J.K. et al.** 1994. Subunit composition and domain structure of the SpoOA sporulation transcription factor of *Bacillus subtilis*.  
*J.Biol.Chem.* **269** : 16972-82.

\* **Grivitz, S.C., S.Bacchetti, A.J.Rainbow, and F.L.Graham.** 1980. A rapid and efficient procedure for the purification of DNA from agarose gels. *Anal.Biochem.* **106**: 492-496.

\* **Halling S.** 1994. 47th Annual Meeting Brucellosis Research Conf.

\* **Hall, M.N., and T.J. Silhavy.**1989. Genetic analysis of the OmpB locus in *Escherichia coli* K-12.  
*J.Mol.Biol.* **151**:1 -15.

\* **Hess, J.F., K.Oosawa, N.Kaplan, and M.I.Simon.** 1988. Phosphorylation of three proteins in signaling pathway of bacterial chemotaxis.  
*Cell* **53**: 79-87.

\* **Holm, L., and C.Sander.** 1992. Fast and simple Monte-Carlo algorithm for side chain optimization in proteins- Application to the model building by homology.  
*Prot.Struct.Funct.Genet.* **14**: 213-223.

\* **Hubbard, T., Park, J., Lham, A., Leplae, R. & Tramontano, A.** (1996) Protein structure prediction: playing the fold.  
*Trends in Biochemistry Sciences*, **21**:279-291.

\* **Huguenel, E.D., and A.Newton.** 1982. Localization of surface structures during procaryotic differentiation: role of cell division in *Caulobacter crescentus*.  
*Differentiation* **21** : 71-78

\* **Hutchins, C., and J.Greer.** 1991. Comparative modeling of proteins in the design of novel renin inhibitors.  
*Crit.Rev.Biochem.Mol.Biol.* **26**: 77-127.

\* **Igo, M.M., A.J.Ninfa, J.B.Stock, and T.J.Silhavy.** 1989. Phosphorylation and dephosphorylation of a bacterial transcriptional activator by transmembrane receptor.  
*Genes dev.* **3**: 1725-1734.

\* **Johnson, R.C, C.A.Ball, D.Pfeffer, and M.I.Simon Jr.** 1988. Isolation of the gene encoding the Hin recombinational enhancer binding protein.  
*Proc.Natl.Acad.Sci.USA.* **85**: 3484-3488.

\* **Kabsch, W., and C.Sander.** 1983. How good are predictions of protein secondary structure ?  
*FEBS.Lett.* **155**: 179-182.

\* **Kahn, D., and G.Ditta.** 1991. Modular structure of FixJ : homology of the transcriptional activator domain with the -35 binding domain of sigma factors.  
*Mol.Microbiol.* **5** : 987-997.

\* **Koch, C., J.Vandekerckhove, and R.Kahmann.** 1988. *Escherichia coli* host factor for site-specific DNA inversion: cloning and characterization of the *fis* gene.  
*Proc.Natl.Acad.Sci.USA.* **85** : 4981-4985.

\* **Kofoid, E.C., and J.S.Parkinson.** 1988. Transmitter and receiver modules in bacterial signaling proteins.  
*Proc.Natl.Acad.Sci.USA* **85**: 4981-4985.

- \* **Kostrewa, D., J.Granzin, C.Koch, H.-W.Cho, S.Raghunathan, W.Wolf, J.Labahn, R.Kahmann, and W.Saenger.** 1991. Three-dimensional structure of the *E.coli* DNA binding protein FIS. *Nature* **349** : 178-180.
- \* **Kumar, A., B.Grimes, N.Fujita, K.Makino, R.A.Malloch, R.S.Hayward, and A.Ishihama.** 1994. Role of the sigma 70 subunit of Escherichia coli RNA polymerase in transcription activation. *J.Mol.Biol.* **235** : 405-413.
- \* **Levin, J.M., S.Pascarella, P.Argos, and J.Garnier.** 1993. Quantification of secondary structure prediction improvement using multiple alignments. *Prot.Eng.* **6** : 849-854.
- \* **Lim, V.I.** 1974. Structural principles of the globular organization of proteins chains. A stereochemical theory of globular proteins secondary structure. *J.Mol.Biol.* **88** : 857-872.
- \* **Lukat, G.S., W.R.McCleary, A.M.Stock, and J.B.Stock.** 1992. Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proc.Natl.Acad.Sci.USA* **89** : 718-722.
- \* **Lupas, A., and J.Stock.** 1989. Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis. *J.Biol.Chem.* **264** : 17337-17342.
- \* **McCleary, W.R, M.McBride, and D.Zusman.** 1990. Developmental sensory transduction in *Myxococcus xanthus* involves methylation and demethylation of FrzCD. *J.Bacteriol.* **172** : 4877-4887.
- \* **Mehta, P., J.Meringa and P.Argos.** 1995. A simple and fast approach to prediction of protein secondary structure from multiple aligned sequences with accuracy above 70%. *Prot.Science.* **4** : 2517-2525.
- \* **Miller, R.T., Jones, D.T. and Thornton, J.M.** 1996. Protein fold recognition by sequence threading: tools and assessment techniques. *FASEB Journal*, 10:171-178.
- \* **Moy, F.J., D.F.Lowry, P.Matsurama, F.W.Dahlquist, J.E.Krywko, and P.J.Domaille.** 1994. Assignment, secondary structure, global fold, and dynamics of chemotaxis Y protein using three- and four-dimensional heteronuclear (<sup>13</sup>C, <sup>15</sup>N) NMR spectroscopy. *Biochemistry.* **33** : 10731-10742.
- \* **Mondragon, A., Wolberger, C. & Harrison, S.C.** 1989. Structure of phage 434 Cro protein at 2.35 angstroms resolution. *J. Mol. Biol.* 205:179-189
- \* **Murzin, A.G., S.E.Brenner, T.Hubbard and C.Chothia.** 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J.Mol.Biol.* **247** : 536-540.
- \* **Needleman, S.B., and C.D.Wunch.** 1970. A general method applicable to the search for similarities in the amino acid sequences of two proteins. *J.Mol.Biol.* **48** : 443-453.
- \* **Neidhardt, F.C., J.L.Ingraham, and M.Schaechter.** 1990. Physiology of the bacterial cell - A molecular approach. Sinauer Associates
- \* **Newton, A., and N.Ohta.** 1990. Regulation of the cell division cycle and differentiation in bacteria. *Annu.Rev.Microbiol.* **44** : 689-719.
- \* **Ninfa, A.J., and B.Magasanik.** 1986. Covalent modification of the *glnG* product, NRI, by the *glnL* product NRIL, regulates the transcription of the *glnALG* operon in *Escherichia coli*. *Proc.Natl.Acad.Sci.USA* **83** : 5909-5913.

- \* **Ninfa, A.J., and R.L.Bennett.** 1991. Identification of the site of autophosphorylation of the bacterial protein kinase/phosphatase NRII. *J.Biol.Chem.* **266** : 6888-6893
- \* **Ninfa, E.G., M.R.Atkinson, E.S Kamberov, and A.J.Ninfa.** 1993. Mechanism of autophosphorylation of *Escherichia coli* nitrogen regulator II (NRII or NtrB): trans-phosphorylation between subunits. *J.Bacteriol.* **175** : 7024-7032.
- \* **Nixon, B.T., C.W.Ronson, and F.M. Ausubel.** 1986. Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *Proc.Natl.Acad.Sci.USA* **83** : 7850-7854.
- \* **Orengo, C.** 1994. Classification of protein folds. *Curr.Op.Struct.Biol.* **4** : 429-440.
- \* **Orengo, C.A., T.P.Flores, W.R.Taylor, J.M.Thornton.** 1993. Identification and classification of protein fold families. *Protein Engng.*, **6**: 485-500.
- \* **Otha, N., and A.Newton.**1996. Signal transduction in the cell cycle regulation of *Caulobacter* differentiation. *Trends in Microbiology.* **4** : 326-332.
- \* **Otha, N., T.Lane, E.G.Ninfa, J.M.Sommer, and A.Newton** .1992. A histidine protein kinase homologue required for regulation of bacterial cell division and differentiation. *Proc.Natl.Acad.Sci.USA* **89** : 10297-10301.
- \* **Parkinson, J.S.** 1993. Signal transduction schemes in bacteria. *Cell* **73** : 857-871.
- \* **Parkinson, J.S.** 1995. Genetic approaches for signaling pathway and proteins, p'9-23. In J.A.Hoch and T.J.Silhavy (ed), *Two-component signal transduction* . American Society for Microbiology, Washington.
- \* **Parkinson, J.S., and E.C.Kofoid.** 1992. Communication modules in bacterial signaling proteins. *Annu.Rev.Genet.* **26** : 71-112.
- \* **Pearson, W.R and D.J.Lipman.** 1988. *Proc.Natl.Acad.Sci.USA.* **85** : 2444-2448.
- \* **Pratt, L.A., and T.J.Silhavy.** 1995. Porin regulon of *Escherichia coli*, p'105-127. In J.A.Hoch and T.J.Silhavy (ed), *Two-component signal transduction* . American Society for Microbiology, Washington.
- \* **Quon, K.C., G.T.Marczynski and L.Shapiro.** 1996. Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84** : 83-93.
- \* **Rawn, J.D.** 1990. *Traité de Biochimie.* De Boeck Université.
- \* **Rost, B.** 1995. TOPITS: Threading one-dimensional predictions into three-dimensional structures. *The third international conference on Intelligent Systems for Molecular Biology ( ISMB)*, 314-321.
- \* **Rost, B. and C.Sander.** 1993. Prediction of protein secondary structure at better than 70% accuracy. *J.Mol.Biol.* **232** : 584-599.
- \* **Russel, R.B.** 1996. A guide to Structure Prediction. <http://bonsai.lif.icnet.uk/people/rob/CCP1/IBBS/index.html>
- \* **Salyers, A.A., and D.D.Whitt.** 1994. Bacterial pathogenesis : a molecular approach. ASM Press.
- \* **Sanders, D.A., B.L.Gillece-Castro, A.L.Burlingame, and D.E.Koshland, Jr.** 1992. Phosphorylation site of NtrC, a protein

phosphatase whose covalent intermediate activates transcription. *J.Bacteriol.* **174**: 5117-5122.

\* **Sanders, D.A., B.L.Gillece-Castro, A.M.Stock, A.L.Burlingame, and D.E.Koshland, Jr.** 1989. Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. *J.Biol.chem.* **264**: 21770-21778.

\* **Serrano, L., and A.R.Fersht.** 1989. Capping and  $\alpha$ -helix stability. *Nature*, **342**: 296-299.

\* **Simms, S.A., Keane M.G., and J.Stock.** 1985. Multiple forms of the CheB methyltransferase in bacterial chemosensing. *J.Biol.Chem.* **260** : 10161-10168.

**Sippl, M.J.** 1993. Boltzmann's Principle, Knowledge Based Mean Fields and Protein Folding. An Approach to the Computational Determination of Protein Structures. *J.Computer Aided Mol.Design*, **7**, 473-501

\* **Smith, R.F., B.A.Wiese, M.K.Wojzynski, D.B.Davison, and K.C.Worley .** 1996. BCM Search Launcher - An Integrated Interface to Molecular Biology Data Base. Search and Analysis Services Available on the World Wide Web *Genome Research* **6**: 454-462.

\* **Smith, T.F., and M.S.Waterman.** 1981. Comparison of biosequences. *Adv.Appl.Math.* **2**: 482-489.

\* **Sommer, J.M. and A.Newton.** 1988. Sequential regulation of developmental events during polar morphogenesis in *Caulobacter crescentus* : assembly of pili on swarmer cells requires cell separation. *J.Bacteriol.* **170** : 409-415.

\* **Sommer, J.M. and A.Newton.** 1989. Turning off flagellum rotation requires the pleiotropic gene *pleD* : *pleA*, *pleC*, and *pleD* define two morphogenic pathways in *Caulobacter crescentus*. *J.Bacteriol.* **171** : 392-401.

\* **Sommer, J.M. and A.Newton.** 1991. Pseudoreversion analysis indicates a direct role of cell division genes in polar morphogenesis and differentiation in *Caulobacter crescentus*. *Genetics* **129** : 623-630.

\* **Stock, A.M., D.E.Koshland, Jr, and J.B.Stock.** 1985. Homologies between the *Salmonella typhimurium* CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. *Proc.Natl.Acad.Sci.USA* **82** : 7989-7993.

\* **Stock, A.M., E.Martinez-Hackert, B.F.Rasmusen, A.H.West, J.B.Stock, D.Ringe, and G.A.Petsko.** 1993. Structure of the Mg<sup>++</sup>-bound form of the CheY protein and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry* **32** : 13375-13380.

\* **Stock, J.B., A.J.Ninfa, and A.M.Stock.** 1989. Protein phosphorylation and regulation of adaptative response in bacteria . *Microbiol.Rev.* **53** : 450-490.

\* **Stock, J.B., A.M.Stock, and J.M.Mottonen.** 1990. Signal transduction in bacteria. *Nature (London)* **344** : 395-400.

\* **Stock, J.B., M.G.Surette M.Levit and P.Park.** 1995. Two-component signal transduction systems: structure-function relationships and mechanisms of catalysis, p'25-51. In J.A.Hoch and T.J.Silhavy (ed), *Two-component signal transduction* . American Society for Microbiology, Washington.

\* **Suzuki, M., and K.Makino.** 1995. The DNA-binding domains of transcription factors, PhoB and OmpR, adopt the same folding as the histone H5 and bind to RNA polymerase using the end of two  $\alpha$ -helices. *Proc.Japan Acad.*, **71**: 132-137.

\* **Swanson, R.V., L.A.Alex, and M.I.Simon.** 1994. Histidine and

aspartate phosphorylation: two-component systems and the limits of homology. *TIBS* **19**: 485-490.

\* **Swanson, R.V., R.B.Bourret, and M.I.Simon.** 1993. Intermolecular complementation of the kinase activity of CheA. *Mol.Microbiol.* **8** : 435-441.

\* **Thompson, D.J., D.G.Higgins and T.J.Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Res.* **22** : 4673-4680.

\* **Tibor, A., V.Weynants, P.Denoel, B.Lichtfousse, X.DeBolle, J.N.Limet, and J-J.Letesson.** 1994. Molecular cloning, nucleotide sequence, and occurrence of a 16.5-kilodalton outer membrane protein of *Brucella abortus* with similarity to PAL lipoproteins. *Infim.*, **62**: 3633-3639.

\* **Tsung, K., R.E.Brissette, and M.Inouye.** 1989. Identification of the DNA-binding domain of the ompR protein required for transcriptional activation of the OmpF and OmpC genes of *Escherichia coli* by in vivo DNA-footprinting. *J.Biol.Chem.* **264** : 10104-10109.

\* **Verger, J.M.** 1993. Le point sur la brucellose : dépistage et prophylaxie. *Le point vétérinaire* **152** : 1-37.

\* **Vinals, C., X.DeBolle, E.Depiereux, and E.Feytmans.** 1995. Knowledge-based modeling of the D-Lactate Deshydrogenase three-dimensional structure. *Proteins* . **23** : 307-318.

\* **Vinals,C.** 1996. Modélisation de structures tridimensionnelles de protéines: application à l'étude de la stéréospécificité des lactate déshydrogénases. Thèse de doctorat. Facultés Universitaires Notre-Dame de la Paix.

\* **Volz, K.** 1993. Structural conservation in CheY superfamily. *Biochemistry* **33** : 11741-11753.

\* **Volz, K.** 1995. Structural and functional conservation in response regulators, p'53-63. In J.A.Hoch and T.J.Silhavy (ed), *Two-component signal transduction* . American Society for Microbiology, Washington.

\* **Volz, K., and P.Matsumura.** 1991. Crystal structure of *Escherichia coli* CheY refined at 1.7-Å resolution. *J.Biol.Chem.* **266** : 15511-15519.

\* **Volz, K., J.Beman, and P.Matsumura.** 1986. Crystallization and preliminary characterization of CheY, a chemotaxis control protein from *Escherichia coli*. *J.Biol.Chem.* **266** : 15511-15519

\* **Wang, S.P., P.L.Sharma, P.V.Schoenlein and B.Ely.** 1993. A histidine protein kinase is involved in polar organelle development in *Caulobacter crescentus*. *Proc.Natl.Acad.Sci.USA.* **90** : 690-694.

\* **Wootton, J.C.** 1994. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput.Chem.* **18** : 269-285.

\* **Wootton, J.C., and M.H.Drummond.** 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.* **2** : 535-543.

\* **Yang, Y., and M.Inouye.** 1991. Intermolecular complementation between two defective mutant signal-transduction receptors of *Escherichia coli*. *Proc.Natl.Acad.Sci.USA.* **88** : 11057-11061.

\* **Zvelebil, M.J., J.G.Barton, W.R.Taylor and M.J.Sternberg.** 1987. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J.Mol.Biol.* **195**: 957-961.